University of Glasgow

Fiani, Cristina (2026) Safeguarding children from harassment in social VR: integrating child, parent and expert perspectives into the design of automated embodied moderators. PhD thesis.

https://theses.gla.ac.uk/85863/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

Safeguarding Children from Harassment in Social VR: Integrating Child, Parent and Expert Perspectives into the Design of Automated Embodied Moderators

Cristina Fiani

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

SCHOOL OF COMPUTING SCIENCE
COLLEGE OF SCIENCE AND ENGINEERING
Social AI CDT



DECEMBER 2025

Abstract

Children grow up in a digital world with constant online connection, changing the way they socialise, interact and communicate. Social Virtual Reality (VR) has emerged as one of the technologies that attracts children and teenagers, despite age restrictions typically 13+. In contrast to 2D digital media, social VR leverages VR headsets' unique affordances that allow embodiment, presence and immersion. Users can now interact via an embodied avatar synchronously in 3D-360-degree high-fidelity social virtual environments, providing rich visual, auditory, haptic and phantom touch sensations, which allow for mimicking face-to-face interactions. However, with these new forms of socialising come challenges, including new forms of harassment, not only verbal, but also physical (e.g., avatars invading personal space) and environmental (e.g., displaying inappropriate content). Children and adults now share these unsupervised social VR spaces, while still being anonymous, which has led to numerous harassment events, including group harassment. First, a gap remains in understanding child and parental involvement in this space. Second, child-friendly solutions are lacking, and the effectiveness of the existing safety features and their usage by children are not evaluated. Third, there is a noticeable absence of involvement from key stakeholders such as experts, parents, and children themselves in research studies. To address these gaps, in Part I of this thesis, through three empirical studies (a mixed-methods questionnaire with 79 parents and 70 non-parent social-VR aware adult users, interviews with 16 experts and a user study with 40 children 8-16 years old) the research aims to 1) explore and better understand issues and concerns associated with child use of social VR, 2) inform the design of child safety features in social VR. The findings reveal concerns regarding the freedom children have in social VR, the increased harassment situations and the current lack of safety and supervision tools. Then, the research underscores the critical insight that existing safety solutions place responsibility on children and the necessity of creating child-friendly, real-time safety tools such as quick-access, discreet and engaging features to effectively safeguard children against one-to-one and group harassment in social VR. With AI moderation becoming promising and leveraging embodiment, in Part II, I propose a novel approach to safeguard children in these spaces using AEMs, AI-driven embodied safety agents to respond to harassment in social VR in real-time, designed with parents, children and experts. I developed "Big Buddy", a Wizard-of-Oz AEM prototype and through five studies with children aged 8-16 years old, parents and experts (user studies with 43 children and 17 parents, 34 children, 34 parent-child dyads respectively, workshops with 13 children 8-16 years old, 5 parents, 3 grandparents and interviews with 16 experts), the research aims to 1) explore both experiential and conceptual aspects of AEMs from stakeholder perspectives and 2) identify the requirements for AEMs to effectively engage both children and their parents, while ensuring children remain empowered and supported in the process. Considering the lifecycle of a moderation incident (before, during and after a child is harassed in social VR), and informed by the findings, I present how I envision an AEM to work in practice, highlighting considerations around the role of AEMs, child users and parental settings and how parents and children can be involved. Finally, I propose design guidelines to develop child-friendly social VR safety features. This thesis addresses critical gaps and establishes a solid foundation for child safety in social VR.

Contents

Abstract ii

Ethics & Resources xv

Declaration xvi

Additional Ouputs & Impact xviii

Acknowledgements xx

Abbreviations xxii

I Introduction 1

1 Contextualisation 2
- 1.1 Motivation . . . 2
- 1.2 Terminology . . . 3
- 1.3 Scope . . . 4
  - 1.3.1 Type of Harassment in social VR . . . 4
  - 1.3.2 Relevant Stakeholders . . . 4
  - 1.3.3 Child Social VR Users . . . 5
  - 1.3.4 Incidents and Interventions Time-frame . . . 7
- 1.4 Research Questions . . . 7
  - 1.4.1 The Problem and Moderation Space (RQ1, RQ2) . . . 7
  - 1.4.2 Safeguarding Children from Harassment in Social VR using Automated Embodied Moderators (RQ3, RQ4) . . . 8
- 1.5 Thesis Statement . . . 8
- 1.6 Thesis Outline & Walkthrough . . . 9

2 Literature Review 15
- 2.1 Introduction . . . 15
- 2.2 Child Safety in 2D Digital Media and Lessons from the Online World . 15
  - 2.2.1 Prevalence of the Online World and Its Impact on Children . . 15
  - 2.2.2 Relevant Aspects of Children's Social-Emotional Perceptions and Influences on Prosocial Behaviours . . . 16
  - 2.2.3 Bullying and Harassment: Offline and Online . . . 18

2.2.4 Parental Role, Responsibility and Parental Mediation of Children's Interaction with Digital Media . . . 21
2.2.5 Relevant Topics from Criminology and Law . . . 22
2.3 Child Safety in Social Virtual Reality . . . 23
2.3.1 Social Virtual Reality . . . 23
2.3.2 Prevalence of Children in Social VR . . . 26
2.3.3 Harassment in Social VR . . . 27
2.3.4 Existing Safety Features in Social VR and their Limitations . . . 29
2.4 Research Methods for Child-Computer Interaction . . . 31
2.4.1 The Importance of Integrating Diverse Perspectives in the Research of Child Safety Features in Social VR . . . 31
2.4.2 Conducting CCI Research with Children . . . 32
2.4.3 Designing and Conducting VR Studies with Children on Sensitive Topics such as Harassment . . . 33
2.5 Gaps in Prior Work . . . 35
2.5.1 Involving Different Perspectives to Investigate Child Safety in Social VR . . . 36
2.5.2 Child-Centric Solution and Evaluation Methodology . . . 36
2.5.3 Novel Solutions: Designing for Safety AI and the Affordances of VR . . . 37

II The Problem and Moderation Space 38

3 Issues and Concerns Associated with Child Use of Social Virtual Reality 39
3.1 Chapter Introduction . . . 39
3.1.1 Motivation . . . 40
3.1.2 Overview of Study #1 . . . 40
3.2 Study #1 - Through the Eyes of Social VR-Aware Parents & Adult Users: Issues and Concerns Associated with Child Use of Social Virtual Reality . . . 41
3.2.1 Related Work . . . 41
3.2.2 Methodology . . . 43
3.2.3 Results . . . 47
3.2.4 Discussion . . . 58
3.3 Chapter Conclusion . . . 61
3.3.1 Contribution to Research Question 1 . . . 62

4 Informing The Design of Child Safety Features in Social Virtual Reality 63
4.1 Chapter Introduction . . . 63
4.1.1 Motivation . . . 64
4.1.2 Overview of Studies . . . 64
4.2 Study #2 - Expert-Driven Design Considerations for Real-Time Safety Features in Social VR to Safeguard Children Against Individual and Group Harassment . . . 66

4.2.1 Motivation . . . 66
4.2.2 Related Work . . . 66
4.2.3 Methodology . . . 68
4.2.4 Results . . . 74
4.2.5 Discussion . . . 93
4.2.6 Conclusion . . . 97
4.3 Study #3 - Children's Perceptions of Blocking Strategies for Group Harassment in Social VR . . . 97
4.3.1 Motivation . . . 97
4.3.2 Related Work . . . 98
4.3.3 Methodology . . . 99
4.3.4 Results . . . 106
4.3.5 Discussion . . . 117
4.3.6 Conclusion . . . 119
4.4 Chapter Conclusion . . . 119
4.4.1 Contribution to Research Question 2 . . . 120

5 Summary & Reflection of Part II 121
5.1 Introduction . . . 121
5.2 Converging and Diverging Insights from Four Stakeholder Groups . . . 122
5.2.1 Converging Insights . . . 122
5.2.2 Diverging Insights . . . 124
5.2.3 Integrating Stakeholder Views . . . 125
5.3 Limitations and Future Work . . . 125
5.4 Implications Informing the Next Part of the Thesis . . . 126
5.4.1 Leveraging AI-Moderation . . . 127
5.4.2 Leveraging Embodiment . . . 127
5.4.3 Proposed Solution . . . 128

III Safeguarding Children from Harassment in Social VR using Automated Embodied Moderators 129

6 Exploring Stakeholder Perspectives on AEMs 130
6.1 Chapter Introduction . . . 130
6.1.1 Motivation . . . 131
6.1.2 Overview of Studies . . . 131
6.2 Study #4 - Big Buddy: Exploring Child Reactions and Parental Perceptions towards a Simulated Automated Embodied Moderator for Social VR . . . 133
6.2.1 Motivation . . . 133
6.2.2 Related Work . . . 133
6.2.3 Methodology . . . 134
6.2.4 Results . . . 143
6.2.5 Discussion . . . 155
6.2.6 Conclusion . . . 159

6.3 Studies #5 and #6 - Collecting Multi-stakeholder Perspectives on Automated Embodied Moderators for Child Safety in Social VR 160
6.3.1 Motivation 160
6.3.2 Related Work 160
6.3.3 Methodology - Collecting Multi-stakeholder Perspectives on Automated Embodied Moderators for Child Safety in Social VR . . 162
6.3.4 Methodology - Study #5 One-to-one Interviews with Experts . 162
6.3.5 Methodology - Study #6: Workshops with Guardians and Children 164
6.3.6 Limitations and Future Work 168
6.3.7 Results of Studies #5 and #6 169
6.3.8 Discussion of Studies #5 and #6 190
6.3.9 Conclusion 194
6.4 Chapter Conclusion 194
6.4.1 Contribution to Research Question 3 195

7 Deriving Requirements for AEMs from Stakeholder Experiences 196
7.1 Chapter Introduction 196
7.1.1 Motivation 196
7.1.2 Overview of Studies 197
7.2 Study #7 - Exploring Levels of Automation in Social VR: From Manual to Automated Interventions in Group Harassment Targeting a Child . . 198
7.2.1 Motivation 198
7.2.2 Related Work 199
7.2.3 Methodology 200
7.2.4 Results 206
7.2.5 Discussion 218
7.2.6 Conclusion 219
7.3 Study #8 - Context-driven Parental Involvement for Children in Social VR via AEM-app: Balancing Safety and Privacy 219
7.3.1 Motivation 219
7.3.2 Related Work 220
7.3.3 Methodology 222
7.3.4 Results - Quantitative 232
7.3.5 Results - Qualitative 238
7.3.6 Discussion 259
7.3.7 Conclusion 261
7.4 Chapter Conclusion 262
7.4.1 Contribution to Research Question 4 262

8 Summary & Reflection of Part III 263
8.1 Design Guidelines of Automated Embodied Moderators to Safeguard Children in Social VR 263
8.1.1 A Triadic Human-in-the-Loop AEM Framework to Safeguard Children in Social VR Pre/During/Post-Event 263
8.1.2 Journeys of Families Navigating AEM 269
8.2 Limitations and Future Work 271

IV Conclusion **273**

9 Contributions & Final Reflection **274**
9.1 Contributions Overview . . . 274
9.1.1 Empirical Contributions . . . 274
9.1.2 Methodological Contributions . . . 275
9.1.3 Conceptual Contributions . . . 275
9.2 Design Guidelines Derived from Findings to Safeguard Children in Social VR . . . 275
9.2.1 Preventive, Proactive and Reactive Harassment Interventions . 275
9.2.2 Safety Tools . . . 277
9.2.3 Child Participation and Agency . . . 278
9.2.4 Human Oversight and Involvement . . . 280
9.2.5 Child Comfort and Privacy . . . 282
9.2.6 AEM Design in practice . . . 283
9.3 Methodological Approaches to Evaluating Safety Tools for Children in Social VR . . . 283
9.3.1 From Third-Party to First-Person Engagement . . . 283
9.3.2 First-Person Perspective in Social VR: Designed Disruptions for Ethical Insight . . . 284
9.4 Research Questions . . . 285
9.4.1 Research Question 1 . . . 285
9.4.2 Research Question 2 . . . 286
9.4.3 Research Question 3 . . . 287
9.4.4 Research Question 4 . . . 289
9.5 Future Work . . . 290
9.6 Conclusion . . . 291

Appendices **293**

A Study #1 **294**
A.1 Intervention Categories Presented to Participants . . . 294

B Study #3 **296**
B.1 Quantitative Summary Tables . . . 296
B.2 Non-Significant Statistical Results . . . 297

C Study #7 **298**
C.1 Quantitative Summary Tables . . . 298
C.2 Non-Significant Statistical Results . . . 299

D Study #8 **300**
D.1 Parents Results . . . 300
D.2 Children Results . . . 301

References **304**

List of Tables

1.1 Table summarising studies and outputs. 14

3.1 Questionnaire Demographics Table Summary. (N=149, 79 Parents, 70 Non-Parents). The questionnaire was mainly released on forums about VR/social VR; there may be self-selection bias, in which more engaged parents and non-parent adults in VR/social VR may have been more likely to participate in this study. However, this enabled me to get responses from users familiar with VR/social VR. 46
3.2 Four-Quadrant Matrix presenting themes from social VR adults' and children's experiences and perceptions from non-parent adults' and parents' perspectives. 50

4.1 Ten Safety Features in Social VR Platforms, five common and five distinct ones across VRChat [204] and Rec Room [187] (May 2023 - list is non-exhaustive for time and evaluation constraints). 69
4.2 Demographics of experts who participated in one-to-one interviews (Part 1). Experts had a Mean age of 43.3 years (SD=10.5). 72
4.3 Demographics of experts who participated in one-to-one interviews: child bullying knowledge, child cyberbullying knowledge, Social VR awareness, Social VR experience, owner of VR headset, VR experience. Experts had a mean of Social VR Awareness of 4 (SD=1.5). 73
4.4 Solutions Proposed by Experts for Scenario 1 (Section 4.2.3.3): One-to-one harassment towards a child. (Half the participants had to propose a solution for scenario 1, while the other half had to propose a solution for scenario 2.) 81
4.5 Solutions Proposed by Experts for Scenario 2 (Section 4.2.3.3): Group harassment towards a child. (Half the participants had to propose a solution for scenario 1, while the other half had to propose a solution for scenario 2.) 82
4.6 Participants' Online Gaming, VR, and social VR Habits (33 families, N=40 children) 105
4.7 Qualitative Findings: Categories, Themes, Sub-Themes with Counts, gathered from child interviews at the end of the user study. 107
4.8 Weighted Ranking of Tools Based on Favourite, Second Favourite, and Least Favourite Scores. The ranking is based on a scoring system where favourite votes are worth 3 points, second favourite votes are worth 2 points, and least favourite votes are worth 1 point, with the total weighted score being the sum of these points. 115

4.9 Weighted Ranking of Views Based on Favourite and Least Favourite Scores. The ranking is based on a scoring system where favourite votes are worth 2 points and least favourite votes are worth 1 point, with the total weighted score being the sum of these points. 115
4.10 Two-way ART ANOVAs results 1) with the independent variables as Group Size and Tools (C1,C2,C3) 2) with the independent variables as Group Size and IP (C1 VS C4). While *[PointBlock]* enjoyment mean is higher (4.8, SD=1.2) compared to *[MassBlock]* (4.5, SD=1.2) and *[SelectBlock]* (4.5, SD=1.2), the post-hoc test did not reveal significant differences between individual pairs. (Small effect: ≈0.01, Medium effect: ≈0.06, Large effect: >0.14). 117

6.1 Table summarising demographics of parents, parents who were interviewed and children. (N_parents = 31, N_parents(interviewed) = 17, N_children=43). MEAN [SD]. 142
6.2 Summary of children's qualitative explanations for their AEM scores. Percentages indicate the proportion of responses in negative, neutral, or positive categories for each scale. 153
6.3 Summary of parents' qualitative explanations for their AEM scores. Percentages indicate the proportion of responses in negative, neutral, or positive categories for each scale. 154
6.4 Workshops Schedule Breakdown for Guardian and Child Participants. . . . 165
6.5 Table Summarising participant groups' attendance at workshops. Whilst the presence of guardians varied, there was always a responsible adult accompanying children (the librarian was present in all workshops). P(Participant Number)(P parent or G grandparent)-(Workshop#) (e.g., P3P-2) with corresponding child as P(Participant Number letter)(C)-(Workshop#), a letter is added if more than one child per guardian came (e.g., P3aC-2, P3bC-2). 165
6.6 Demographics of parents and grandparents who participated in the workshops. Parents had a mean age of 45.5 (SD=8.4). Grandparents had a mean age of 63.5 (SD=2.8). 167
6.7 VR ownership, experience, and attitudes of the guardians who participated in the workshops. Guardians had a mean Social VR awareness of 1.9 (SD=1.4), and a mean attitude towards social VR of 2.6 (SD=1.1). 167
6.8 Demographics of Children who participated in the workshops. Children had a mean age of 11.5 (SD=2.7). These were reported by the guardians. . 167

7.1 Participants' Online Gaming, VR, and social VR Habits (26 families, N=34 children) 205
7.2 Qualitative Findings - Categories, Themes, Sub-Themes with Counts, gathered from child interviews at the end of the user study. 207
7.3 Rankings and favourite children's interventions with score breakdowns and total weighted scores. The ranking is based on a scoring system where favourite votes are worth 4 points, second favourites 3 points, third favourites 2 points and least favourites 1 point, with the total weighted score being the sum of these points. Ranked as top 1 is (C1) [manual], followed by (C3) [AEM gives options], (C4) [AEM fully automated], and last, (C2) [Call AEM]. 215

7.4 One-way ART ANOVAs results 1) with the independent variable as the Interventions (C1, C2, C3, C4) 2) with the independent variable as Time (Before VS After bullies blocked). (Small effect: <0.01, Medium effect: ≈0.06, Large effect: >0.14) . . . 217
7.5 Child Participant Demographics and VR/Social VR Experience and Familiarity (N=34). (Responses given by parents) . . . 229
7.6 Parent Participant Demographics and VR/Social VR Experience and Familiarity (N=32) . . . 229
7.7 One-way ART ANOVA and Post-hoc Results Across Conditions for All Measures from Parent Participants. I set the significance level at 0.05, and interpreted effect sizes using partial eta squared η_p^2 with small effects (< 0.01), medium effects (≈ 0.06) and large effects (≈ 0.14 or higher) [286, 287]. 232
7.8 One-way ART ANOVA and Post-hoc Test Results Across Likert Scales from Children Participants with significance and effect sizes. I set the significance level at 0.05, and interpreted effect sizes using partial eta squared η_p^2 with small effects (< 0.01), medium effects (≈ 0.06) and large effects (≈ 0.14 or higher) [286, 287]. . . . 235
7.9 **Availability Level 1 Parents' Weighted Rankings** - Rankings were converted to coefficients, with 4 points assigned to the most preferred option, 3 to the second, 2 to the third, and 1 to the least preferred. Cell colour intensity corresponds to the weight, with stronger colours indicating higher preference. . . . 237
7.10 **Availability Level 2 Parents' Weighted Rankings** - Rankings were converted to coefficients, with 4 points assigned to the most preferred option, 3 to the second, 2 to the third, and 1 to the least preferred. Cell colour intensity corresponds to the weight, with stronger colours indicating higher preference. . . . 237
7.11 **Availability Level 3 Parents' Weighted Rankings** - Rankings were converted to coefficients, with 4 points assigned to the most preferred option, 3 to the second, 2 to the third, and 1 to the least preferred. Cell colour intensity corresponds to the weight, with stronger colours indicating higher preference. . . . 237
7.12 Children's weighted rankings of conditions based on their most and least preferred choices. Rankings were converted into weighted scores: 4 points for the most preferred, 3 for the second, 2 for the third, and 1 for the least preferred option. Cell colour intensity corresponds to the weight, with stronger colours indicating higher preference. . . . 238

8.1 Examples of AEM Embodiment Archetypes based on Clusters from Study #4 . . . 266

B.1 Summary statistics by condition and likert scale Study #3. . . . 296
B.2 Non-Significant Results from Two-Way ART ANOVAs for Group Size and Tools/Intervention Perspective (IP) Study #3. . . . 297

C.1 Summary Data for Emotional Perceptions . . . 298
C.2 Summary Data for Usability Perceptions . . . 299

C.3 Non-Significant Results from One-Way ART ANOVAs for Conditions Study #7. 299

D.1 One-Way ART ANOVA results Parents. 300
D.2 ART ANOVA and Post-hoc Results Across Conditions for All Measures from Parent Participants 300
D.3 One-Way ART ANOVA results for Children with effect sizes and significance. 301
D.4 ART ANOVA and Post-hoc Test Results Across Likert Scales from Children Participants 302

List of Figures

1.1 Diagram showing directions for the thesis, the scope including the intervention space, perspectives and types of harassment in social VR. I consider moderation as prevention & intervention, and I aim to focus on the real-time intervention space. 5

1.2 Diagram of child Social VR user ecosystem stakeholders. The inner circles represent the closest stakeholders to the social VR child user, while the outer circles indicate those more distant. 6

3.1 Questionnaire Blocks and Question Types 44

3.2 Likert Scales showing percentages of children using (a) VR in general or (b) Social VR in the sample. Among children under 13 years, 24% use social VR daily or weekly and 24% use VR in general daily or weekly. 43% use social VR at least every two weeks. 48

3.3 Violin and boxplots of inputted ages corresponding to the minimum appropriate ages in social VR for three different supervision states (a, b, c), by parent and non-parent adult groups, divided into familiar and non-familiar subsets. 48

3.4 Interventions presented to the participants and the selected age range. The violin boxplots summarise age ranges selected for four categories of interventions (pre-planned, knowledge, real-time actions and after-the-fact). Each intervention was selected by n participants. The mean of the minimum ages given for all interventions is 10 years old, and the mean of the maximum ages is 15 years old. 57

4.1 Counts of top 3 expert-selected tools for two harassment scenarios and the least useful tool perceived by experts: addressing one-on-one harassment and group harassment targeting a child. 75

4.2 **(a)** Three harassers (small group), one throwing objects at the victim, another showing a drawing of the child, and the last saying "Loser", with mocking laughs, **(b)** Nine harassers (large group) replicating the three types of harassment. 101

4.3 As participants look at the virtual watch on their hand, **(a)** a green diamond appears with the button corresponding to the condition's blocking tool: **(b)** *SelectBlock* (profile pictures of users to block each individually), **(c)** *MassBlock* (block all users, unblock list appears) or **(d)** *PointBlock* (Magic Wand to point towards avatar and block). (Researcher-Developed using Unity.) 101

4.4 User's perspective in the large group harassment scenario (Ground Level IP VS Above IP). The left column shows the scene in a profile view, where the white camera represents the user's point of view. The right column shows what the user sees through their eyes in the large group harassment scenario. 102
4.5 Likert Scale Responses 116

6.1 Experiment chronology and VR game conditions. 136
6.2 Example of the VR game scenario occurring in round 2 with (C4) [BB, Reset+Notify Parents+Exclusion] from the user's eyes. **(a)** Tower Building Game. **(b)** Provocative situation. **(c)** Big Buddy intervenes: points reset to 0, parents notified and exclusion. **(d)** Competitor appears punished. **(e)** Big Buddy Close-Up. 137
6.3 Self-Assessment Manakin (SAM) Likert scales. Re-adapted the third scale from 'not in control-in control' [65] to 'scared-safe'. 138
6.4 Attributes Scales For Preferred Embodied Moderator in social VR for children and parents. 139
6.5 Violin-boxplots of three SAM Likert scales' scores for each condition. **(a)** Children felt significantly sadder in the round with (C1) [No BB] compared to the round with (C4) [BB, Reset+Notify Parents+Exclusion]. **(b)** No significant effect. **(c)** Children felt significantly safer in (C2) [BB, Reset Points] than in (C1) [No BB]. Significant pairwise comparisons are labelled. 144
6.6 Violin and boxplots of the three 4-point scales' scores for Big Buddy conditions. **(a)** No significant effect. **(b)** Children felt that Big Buddy could see them significantly more in (C4) [BB, Reset+Notify Parents+Exclusion] than (C2) [BB, Reset Points]. **(c)** Children felt that the punishment in (C4) [BB, Reset+Notify Parents+Exclusion] was significantly fairer than in (C2) [BB, Reset Points]. Significant pairwise comparisons are labelled. . 146
6.7 Clusters for an AEM attributes scores given by children and parents. The first principal component (dim1) represents the direction of greatest variation in the data, and the second principal component (dim2) represents the direction of the second greatest variation that is orthogonal to the first component. Labels inside clusters correspond to participants' IDs. 152
6.8 Violin-Boxplots of children's and parents' scores given for the attributes (social/physical) of the scales for a preferred AEM. 153
6.9 Findings structure across all three stakeholder groups (experts, guardians and children). 169
6.10 A selection of children's post-its transferred to a digital Miro board, regarding physical appearance, when and where their moderator would appear. . 183
6.11 PART 1: Examples of children's storyboarding designs with their created automated AEM in a scenario of harassment. 184
6.12 PART 2: Examples of children's storyboarding designs with their created AEM in a scenario of harassment. 185
6.13 Designs from workshop 0 combining children's moderators' ideas into one final design (cat, snake, assassin). 185
6.14 Summary Map of Findings regarding Automated Moderation with corresponding perspectives (E for Experts, G for Guardians and C for Children). 191

6.15 Summary Map of Findings regarding Embodiment of AEMs with corresponding perspectives (E for Experts, G for Guardians and C for Children). 191

7.1 Diagram plotting the four conditions of Study on two axes: on the y-axis, the initiation (who starts the intervention) and on the x-axis, the decision/action (who decides and acts). The diagram highlights the varying levels of child and Automated Moderator (AEM) involvement across conditions. 201
7.2 The four conditions presented. **a) C1:** Participant uses their watch to click on the button and uses the *[PointBlock]* tool from study #3 to block harassers' avatars. **b) C2:** Participant uses their virtual watch to click on the button that calls the Automated Moderator (AEM). The AEM then blocks the bullies. **c) C3:** The AEM initiates the intervention, asking the child if they want the AEM to block the bullies, and gives the child three options: yes, no or wand. Child clicks on the chosen button. **d) C4:** The AEM initiates and blocks the bullies automatically. 202
7.3 The Automated Moderator in the virtual scene. 203
7.4 Emotional Responses Likert Scales showing results from child participants for each condition. Responses were provided for both before and after the bullies were blocked. (Colourblindness note: Legend labels are in the order of blocks from left to right on the plots.) 216
7.5 Usability Likert Scales showing results from child participants for each condition.(Colourblindness note: Legend labels are in the order of blocks from left to right on the plots.) 216
7.6 Visuals with screenshots of **parent** perspectives (mobile App view) for each condition. 224
7.7 Visuals with screenshots of **child** perspectives (in VR view) for each condition. Note: the text in bubbles is spoken out loud in the VR simulation. 224
7.8 Three hypothetical availability scenarios: 1) Fully Available, 2) Partially Available and 3) Minimally Available. 227
7.9 Parent Likert Scales (ranging from Strongly Agree to Disagree) as percentages for each condition. 234
7.10 Children Likert scales results as percentages for each of the four conditions. 235

8.1 A Triadic Human-in-the-Loop AEM Framework to Safeguard Children in Social VR 264
8.2 Diagram representing the three actors (Child in social VR, parent and AEM) in a triadic model. 264
8.3 Schema representing two journeys as examples to see how the system may work with different preferences from the child and the parent. 269

Ethics & Resources

All studies conducted in the course of the PhD and presented in this thesis have ethical approval by the College of Science & Engineering (CoSE) and I had the Scottish Protecting Vulnerable Groups (PVG) scheme [1] to conduct studies with children. Each study involving children has a "Selection and Participation of Children" section. The associated ethics application reference numbers are:

- Chapter 3: Study #1 - 300210173
- Chapter 4: Study #2 - 300220184, Study #3 - 300220209
- Chapter 6: Study #4 - 300210315, Study #5 - 300220184, Study #6 - 300220177
- Chapter 7: Study #7 - 300220209, Study #8 - 300240054

Note that studies #1, #3, #4, #5 & #6 and #7 have been published and studies #2 and #8 are under review. In this thesis, they closely reflect the final publications. I served as the lead author and primary researcher, while co-authors contributed to the qualitative analysis (as detailed in footnotes and in each subsection of the Methods describing the qualitative analysis).

OSF pre-registrations were completed for study #3 (osf.io/m7yd6), study #4 (osf.io/4jnue) and study #7 (osf.io/pgre5).

Unity/iOS projects I developed for studies #3, #4, #7 and #8 are available on GitHub by demand.

You can view the **video figures** illustrating both the prototypes and the studies' setups I developed via the provided links: study #4 (`https://youtu.be/eii8oak28X0`), #3 and #7 (`https://youtu.be/YuPFg-7Q5V4`) and #8 (`https://youtu.be/l-DSKKmJQA4`).

Declaration

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work.

The work presented in this thesis has contributed to research publications in leading conferences (e.g., CHI, CSCW, IDC) where I am the first author.

- **Study #1 in Chapter 3** has been published. *(Full Paper CSCW 2024 [2])*. Cristina Fiani, Pejman Saeghe, Mark McGill, and Mohamed Khamis. 2024. Exploring the Perspectives of Social VR-Aware Non-Parent Adults and Parents on Children's Use of Social Virtual Reality. Proc. ACM Hum.-Comput. Interact. 8, CSCW1, Article 54 (April 2024), 25 pages. https://doi.org/10.1145/3652867
- **Study #2 in Chapter 4** is under review for the International Journal of Child-Computer Interaction (IJCCI) *(Pre-print)*. Cristina Fiani, Joseph O'Hagan, Mark McGill and Mohamed Khamis. 2024. Expert-Driven Design Considerations for Real-Time Safety Features in Social VR to Safeguard Children Against Individual and Group Harassment. http://dx.doi.org/10.2139/ssrn.4839984
- **Study #3 in Chapter 4 and study #7 in Chapter 7** have been published in the same paper. *(Full paper IDC 2025 [3])*. Cristina Fiani, Robin Bretin, Mark McGill, and Mohamed Khamis. 2025. "Abracadabra, Block the Bullies!": Child Perceptions of Manual, Semi-Automated and Automated Interventions against Group Harassment in Social VR. In Proceedings of the 24th Interaction Design and Children (IDC '25). Association for Computing Machinery, New York, NY, USA, 314–339. https://doi.org/10.1145/3713043.3728850
- **Study #4 in Chapter 6** has been published.
  - *(Full paper IDC 2023 [4])* Cristina Fiani, Robin Bretin, Mark McGill, and Mohamed Khamis. 2023. Big Buddy: Exploring Child Reactions and Parental Perceptions towards a Simulated Embodied Moderating System for Social Virtual Reality. In Proceedings of the 22nd Annual ACM Interaction Design and Children Conference (IDC '23). Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3585088.3589374
  - *(LBW CHI 2023 [5])* Cristina Fiani, Robin Bretin, Mark McGill, and Mohamed Khamis. 2023. Big Buddy: A Simulated Embodied Moderating System to Mitigate Children's Reaction to Provocative Situations within Social Virtual Reality. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA'23). Association for Computing Machinery, New York, NY, USA, Article 53, 1–7. https://doi.org/10.1145/3544549.3585840
- **Studies #5 and #6 in Chapter 6** have been published. *(Full paper CHI 2024 [6])*. Cristina Fiani, Robin Bretin, Shaun Alexander Macdonald, Mohamed Khamis, and Mark McGill. 2024. "Pikachu would electrocute people who are misbehaving": Expert, Guardian and Child Perspectives on Automated Embodied

Moderators for Safeguarding Children in Social Virtual Reality. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 113, 1–23. https://doi.org/10.1145/3613904.3642144

- **Study #8 in Chapter 7** is under review for the journal of Transactions on Computer-Human Interaction (TOCHI). Cristina Fiani, Robin Bretin, Shaun Alexander Macdonald, Joseph O'Hagan, Mark McGill and Mohamed Khamis.

Additional Ouputs & Impact

In addition to the work included in this thesis, I led and collaborated on additional relevant projects ([7], [8], [9]):

- Cristina Fiani and Stacy Marsella. 2022. Investigating the Non-verbal Behavior Features of Bullying for the Development of an Automatic Recognition System in Social Virtual Reality. In Proceedings of the 2022 International Conference on Advanced Visual Interfaces (AVI '22). Association for Computing Machinery, New York, NY, USA, Article 67, 1–3. https://doi.org/10.1145/3531073.3534492
- Cristina Fiani, Mark McGill, and Mohamed Khamis. 2023. Ensuring Child Safety in Social VR: Navigating Different Perspectives and Merging Viewpoints. In CHI '23 Workshop: Combating Toxicity, Harassment, and Abuse in Online Social Spaces, April 23–28, 2023, Hamburg, Germany. ACM, New York, NY, USA. https://www.mkhamis.com/data/papers/fiani2023chiworkshop.pdf
- Maheshya Weerasinghe, Shaun Macdonald, Cristina Fiani, Joseph O'Hagan, Mathieu Chollet, Mark McGill, and Mohamed Khamis. 2025. Beyond mute and block: adoption and effectiveness of safety tools in social VR, from ubiquitous harassment to social sculpting. IEEE Transactions on Visualization and Computer Graphics 31, 5 (2025), 3275–3284. https://doi.org/10.1109/TVCG.2025.3549860

In addition to academic dissemination, I have established partnerships with schools and organisations for this project. Notably, *Giggleswick School* and *RespectMe* (Scotland's anti-bullying charity https://www.respectme.org.uk/) collaborated on the project, supporting my research with families. I also created a parental guidance video addressing VR-related bullying for RespectMe https://www.youtube.com/watch?v=LvosLTCTL3E&feature=youtu.be. *Glasgow Libraries* provided a platform to run co-design workshops with families.

The project has contributed to high-profile consultations and reports:

- Ofcom Consultation: Protecting children from harms online (August 2024). https://www.rephrain.ac.uk/wp-content/uploads/REPHRAIN-Ofcom-Consultation-Response-August-2024.pdf
- IEEE White Paper - The metaverse and its impact on human rights, the rule of law and democracy (May 2024). https://rm.coe.int/the-metaverse-and-its-impact-on-human-rights-the-rule-of-law-and-democ/1680b178b0
- REPHRAIN White Paper: the Metaverse and Web 3.0 (May 2023). https://bpb-eu-w2.wpmucdn.com/blogs.bristol.ac.uk/dist/1/670/files/2023/05/REPHRAIN-White-Paper-Metaverse-and-Web-3.pdf

Public engagement events I attended include:

- Innovate UK Immersive Tech Awards 2025 - London, UK (June 2025). Winner in the UX & UI category for prototypes I developed. https://iuk.immersivetechnetwork.org/news/winners-celebrated-at-the-buzzing-innovate-uk-immersive-tech-awards-2025/

- Data Science and AI Festival - Advanced Research Centre, Glasgow (27/10/2025).
- RIG Launch Event (XR) - Advanced Research Centre, Glasgow (28/03/2025).
- Festival of Privacy (REPHRAIN) - Bristol (26/03/2024).
- SPRITE+ Conference - Belfast (2023). Presentation speaker and Best Poster Award.
- Guest Talk at Stanford Autonomous Agents Lab - Virtual (25/08/2023).
- ARCadia - Glasgow (24/09/2022).

To support the research and my attendance at conferences, I successfully secured grants and funding:

- CHI 2024 Student Volunteer
- Gary Marsden Award ($1,850) for IDC 2023
- SICSA Research Scholar (£500) for CHI 2023
- SPRITE+ Belfast 2023 Funding
- IDC 2023 Student Volunteer

I conducted an internship as a VR Safety Intern at Sony Interactive Entertainment, PlayStation (Sept-Dec 2023). However, the work completed during the internship is not included in this thesis.

Acknowledgements

Selfishly. I would like to thank myself first. Doing a PhD does not show the importance of intelligence, but the importance of discipline and consistency. Everyone told me it's a "marathon" not a "sprint". I think it's multiple "sprints". Sometimes you have to move fast and sometimes you must take a step back, re-question everything, and take a deep breath before sprinting again. I would also say that my training journey (boxing, Hyrox, running, dancing) has helped me in this PhD journey and vice versa. If anyone is reading this and starting a PhD, having another field where those skills of resilience need practice can be extremely helpful. And also, when one crumbles, at least you can rely on the other one!

Importantly, I want to thank **my supervisors, Mohamed Khamis and Mark McGill**. Thank you for believing in me more than I believed in myself during this project, and thank you for allowing me to explore my ideas. You gave me the perfect balance of freedom and feedback, which allowed me to grow and led me to incredible opportunities. I really could not have asked for better supervisors.

I can't thank enough all the coffee shops that have welcomed me, where I have spent hours and hours, working on my PhD, probably stealing a lot of their potential customers, but at least they had a very very regular one! Special mention to **Songbird Speciality Coffee** in Partick. They've opened when I was in one of my toughest times, and in this, sometimes (oftentimes), isolated journey, going there became my most productive and safest place to go to everyday.

I, of course, want to thank all **my family**, Merci! Special thanks to **my mother and father**. I think they probably would say that the stress they had to go through these past years is as if they did a PhD themselves. I can't count the number of times I called them, saying I give up, I can't do this, I'm not good enough, someone else deserves the spot, but they've always had my back and never gave up on me. Shout out to **my sister too, Natalia**. At times, I told myself "What would Natalia do in this situation?" and the answer was, she'd probably just laugh about it and move on. I tried to do that, but it's one of the hardest things to do. I want to thank my grandma, **Tata**, my number one fan! I tried to call her every Sunday afternoon. She's been admiring my work and inspiring me to push myself and stay grounded. I would also like to give a special shout-out to **Elena**, my cousin, one of the strongest people I know, who has also inspired me since the start of my undergraduate degree. And, last but not least, a special mention to **my cat, Leo**.

I am very thankful for **my colleagues from SIRIUS and MIG groups: Robin, Joseph, Shaun, Ammar, Melvin, Omar, Yasmine, Habiba, Florian, Tom, Diego and more**. I could not have done this PhD without them. Not only have they helped me with my work, participated in my research, but they have also supported me. We went on great trips for conferences, and I now have amazing memories with them in places like Hawaii, Hamburg, Belfast and Frascati.

Many thanks to my **Social AI CDT friends: Juliane, Dominik and Nicole**. They've seen me at my worst and were still there for me. I am grateful to the Social AI CDT directors for giving me this opportunity and the great staff, special shout-out to Jared and Saskia.

I also want to thank the **friends** I had and the new ones I made over these past years: **Corentin, Jeanne, Sofia, Ruchita, Urvi, Coline, Maria, Flor, Aryana, Lucie, Lena**.

Huge thanks to all the **participants** for taking the time to take part in my research: experts, children, parents (special thanks to Laura), the experts, school teachers (at Giggleswick School), the **Miniversity** team, the charity **RespectMe**, **Claire Quigley** for her assistance in coordinating the workshops at **Mitchell Library and Bellcraig Community Centre in Glasgow**!

Thank you, **William Lewington**, for the opportunity to gain valuable experience at SIE and shout out to **Steven Faye**.

This work was supported by the **UKRI Centre for Doctoral Training in Socially Intelligent Artificial Agents** (grant number EP/S02266X/1), partly by **REPHRAIN**: The National Research Centre on Privacy, Harm Reduction and Adversarial Influence Online under UKRI (grant number EP/V011189/1) and partly sponsored by a **2020 Meta Research Award on Responsible Innovation**.

Abbreviations

- `AEM` - Automated Embodied Moderator
- `AI` - Artificial Intelligence
- `ART` - Aligned Rank Transform
- `BB` - Big Buddy, a name given to the automated embodied moderator in the user studies' prototypes
- `C#` - Condition number
- `CCI` - Child-Computer Interaction
- `Child#` - Child participant number
- `CoSE` - College of Science & Engineering
- `IET` - Institution of Engineering and Technology
- `IP` - Intervention Perspective
- `NLP` - Natural Language Processing
- `NPC` - Non-Player Character
- `OSA` - Online Safety Act
- `P#` - Adult participant number (May refer to adult, expert or parent participants in specific studies)
- `Parent#` - Parent participant number in specific studies
- `SchoolParent#` - Parent from school participant number in specific studies
- `SD` - Standard Deviation
- `SIP` - Social Information Processing
- `TMSC` - Transactional Model of Stress and Coping
- `TOSS` - Teen Online Safety Strategies
- `UK` - United Kingdom
- `UN` - United Nations
- `UNCRC` - United Nations Convention on the Rights of the Child
- `UNICEF` - United Nations Children's Fund
- `USA` - United States of America
- `VR` - Virtual Reality

PART I
Introduction

Chapter 1

Contextualisation

1.1 Motivation

Imagine stepping into a 3D virtual world where you can socialise with people from all around the world as if you were in the same room. This is what social Virtual Reality (VR) aspires to be. You put on a VR headset, choose your social VR platform, and as you enter the virtual world, you can select and personalise the avatar that embodies your "real-world" body, either to represent yourself similar to the offline world or by completely staying anonymous, choosing a different gender/body or non-human avatar and recreating your "self" [10, 11]. Movements and gestures are captured and mirrored in this virtual world. You can see, hear, and even have the illusion of touch [12]. There are events, games, clubs, and you can create your own virtual worlds [13, 14]. The boundaries and physics of the real world can be completely redefined.

This attracted attention, especially during the COVID-19 pandemic, as people who stayed at home used it to escape their own rooms and step into a new world, interacting with others, which mimicked face-to-face interactions [15, 16]. A few years later, VR headsets, such as Meta Quest [17], are now on the Christmas lists of children [18]. In social VR, social norms are not explicitly defined; however, communities have emerging norms, often with volunteers for moderation, for example [19]. There is little regulation and law enforcement, yet, minors now step into these virtual worlds [20, 21, 22, 23]. Two of the earliest and most popular social VR platforms serve as the primary reference examples in this thesis when developing prototypes or identifying common safety tools and features. Both platforms rank in the top 50 of most-played VR games [24]. When narrowed to social VR, they are among the top 10: *VRChat* [25] (released in 2017, with a total of 15 million players as of 2025 [26]) and *Rec Room* [27] (released in 2017, with a total of 3 million players as of 2025 [28]). While these can also be accessed on other devices, such as PCs, my focus is on their use through VR headsets.

Children can now socialise at any time, with anyone, in immersive environments, without leaving the perceived safety of their home. However, while social VR offers new opportunities, it also brings new risks. In social VR, children can face not only verbal harassment like in online social media or multiplayer games [29, 30], but they can also face new forms of harassment unique to social VR: physical harms [31] (e.g., another avatar can invade their personal space and be sexually abused, cases of "Meta-rape" where avatars are already being reported to the police [32, 33]) or environmental harms (displaying graphic content or allowing children in virtual clubs [34]). The impact of such harms is anticipated to be different in social VR compared to 2D digital

media due to the unique affordances VR offers: immersion, presence and embodiment, which increase the "feeling of being there" [35], while still allowing for anonymity. The effect of harassment is likely to be even more amplified for children as they find it more difficult to distinguish the real world from the virtual world, and are more easily socially influenced in VR compared to 2D digital media [36, 37, 38].

And while children access these social virtual worlds, it is challenging for parents to act as bystanders and intervene when their children put on head-worn devices that completely occlude reality, and do not support bystander awareness and supervision by default [39]. Published by the Institution of Engineering and Technology (IET) at the start of this PhD, a survey with 1,000 responses from parents of children aged 5-10 years old in April 2022, in the United Kingdom (UK), revealed that of parents whose children already interact with VR, over a quarter (26%) admitted they did not know what their child was accessing in this new virtual world [18]. This underlines the need for robust safeguarding to be introduced. Additionally, existing safety features such as blocking, reporting or personal space bubbles place the responsibility on the children to shield themselves, but they might not have the maturity to do so. And while some platforms employ human moderators, such as Rec Room and Horizon World [19], they cannot safeguard children 24/7, as a study has shown that only 24% of incidents in social VR were addressed by human moderators [19]. In this regard, automated moderation has emerged as a potential solution that could enhance the effectiveness of moderating social VR environments [40]. However, we lack an understanding of how they should be integrated into the social VR environment from children's and parents' perspectives.

Therefore, this thesis aims to explore the intervention space, including existing safety tools, as well as new possible tools that can effectively safeguard children in social VR.

1.2 Terminology

- In this thesis, the terms "children" and "minors" refer to individuals under the age of 18. When "teenager" is mentioned, it refers to minors aged 13 and above. The studies involved children aged 8 to 16 years old (see section 1.3).
- "Parent" term is used throughout to refer to biological, adoptive and legal guardians unless otherwise specified. "Guardians" may be used to gather both parents and grandparents.
- "Experts" are individuals with at least 5 years of experience (research and/or industry) in child development/educational psychology, psychiatry and social media/social VR online safety.
- "Metaverse" refers to a broad concept of a fully immersive and interactive virtual world [41]. The term "social VR" represents one part of this concept, focusing specifically on social interactions between remote VR users in virtual environments.
- A user causing social disruptions to child users is referred to as "harasser". The affected children were referred to as "victims". However, when involving child participants, the terms "bully" or "bullied child" were also used for simplicity.

- The concepts of "cyberbullying" and "online harassment" are still developing as technology evolves and are subjective section 2.2.3.1. I use "harassment in social VR" to describe offensive behaviour in social VR identified with three categories in recent research [31]: verbal, environmental, and physical. Although I acknowledge the nuanced difference between "intentional" and "unintentional" harassment, especially as highlighted by experts in study #2, my focus remains on the child's victim experience, irrespective of intent.
- I consider "agency" as "beliefs that children can affect an outcome; to have self-efficacy and be effective as agents" [42].

1.3 Scope

The overarching aim of this research is to *envision new safeguards for children in social VR*. To achieve this, I had to thoroughly understand the extent of the problem and the nuances of the moderation landscape. This includes: 1) defining harassment and types of harassment in social VR towards children (see section 2.2.3.1, section 2.3.3), 2) focusing on the intervention rather than the detection (with the assumption that automated detection of harassment will arise through advances in AI section 9.5), 3) identifying the targeted stakeholders in this research, and 4) determining the time-frame and evaluation measures (i.e., short-term or long-term).

I lay out the scope and planned direction of the thesis in figure 1.1.

1.3.1 Type of Harassment in social VR

Based on the classification of types of harassment in social VR [31], as described in section 2.3.3, I focus on physical harassment (e.g., unwanted touching, standing too close – medium: avatar movement) and environmental (e.g., displaying sexual or violent content, throwing objects – medium: shared screens, virtual objects) towards children. I also examine the power dynamic, i.e., one-to-one harassment or group harassment. Although previous research has identified possibilities for "unintentional" harassment, such as in social VR settings where adults may be unaware that the other user is a minor [23], this thesis does not foreground the distinction between "intentional" and "unintentional" harassment. Instead, I remain focused on the victim's experience regardless of intent. Nevertheless, when participants raise this distinction, I consider it and treat it as a point of discussion (e.g., in section 4.2).

1.3.2 Relevant Stakeholders

When considering the ecosystem of a child social VR user (see figure 1.2), it includes a range of stakeholders both within and outside social VR platforms. Internally, this involves other users such as children, adults, and bystanders, as well as platform developers, designers, and professionals within the VR industry. Externally, it encompasses legal guardians or parents, family members such as siblings and grandparents, educators, teachers, classmates, and broader institutions, including schools, regulatory bodies, and policymakers. Additionally, it includes subject matter experts in areas such as children and technology, online safety, and child development.

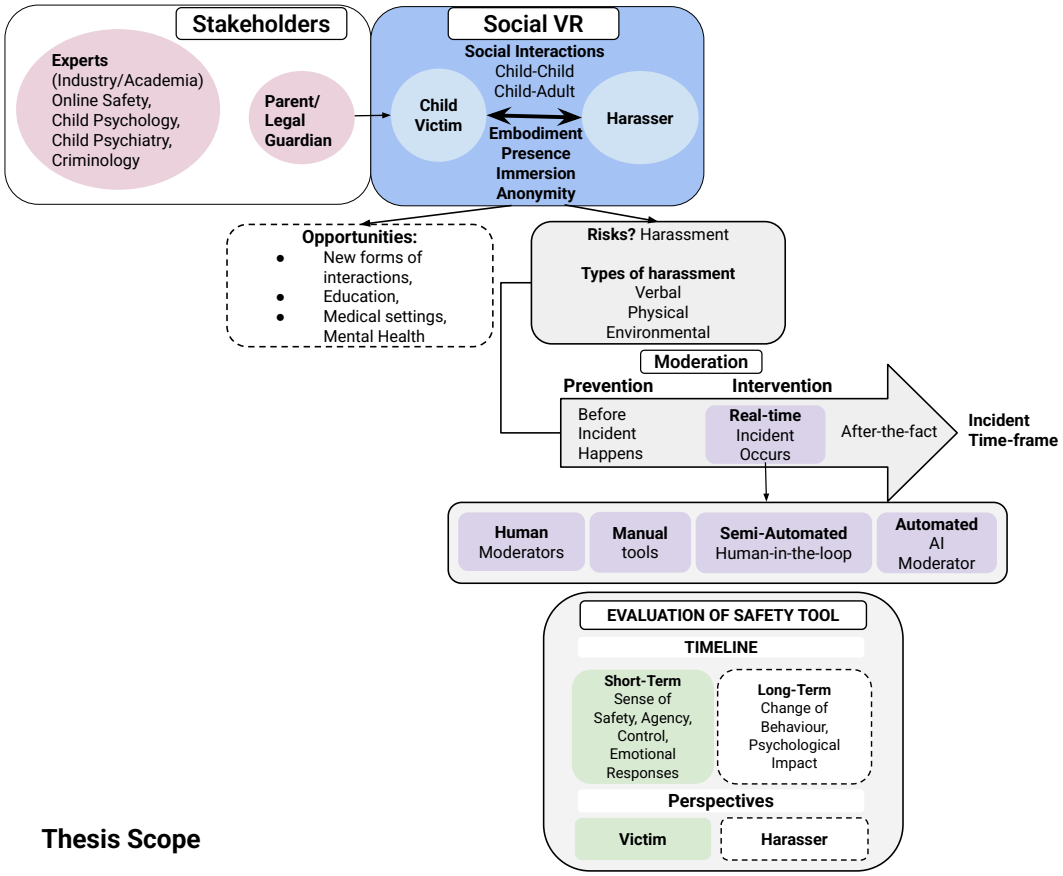


Figure 1.1: Diagram showing directions for the thesis, the scope including the intervention space, perspectives and types of harassment in social VR. I consider moderation as prevention & intervention, and I aim to focus on the real-time intervention space.

Collaboration and integration of various perspectives is crucial to combat social VR child harassment and develop strategies that are easy for children to grasp and use, while also safeguarding children and reassuring guardians. In this thesis, I focus on exploring the perspectives of children, parents, and experts. The importance of these perspectives is detailed in section 2.4.1. I also refer to designers and developers mainly when discussing design guidelines based on the findings.

1.3.3 Child Social VR Users

I focus on harassment towards children aged 8 to 16 years old. I chose this age range for my PhD research based on both practical and developmental considerations. In the UK, primary school-age years start at 5-6 years old and are a stage for children that focuses on learning and developing competence [43]. A preliminary pilot test I conducted with a 6-year-old, using a first social VR prototype I had built, revealed that the task demands were too challenging at that age, prompting me to raise the lower age limit to 8. I chose to adjust the starting age to 8 rather than 7 as a precaution, to maintain consistency with prior VR research with children aged 8–13 [44], which informed my methodological design. The first 8-year-old participants in the study demonstrated increased attention, focus, and comprehension of the task. The upper

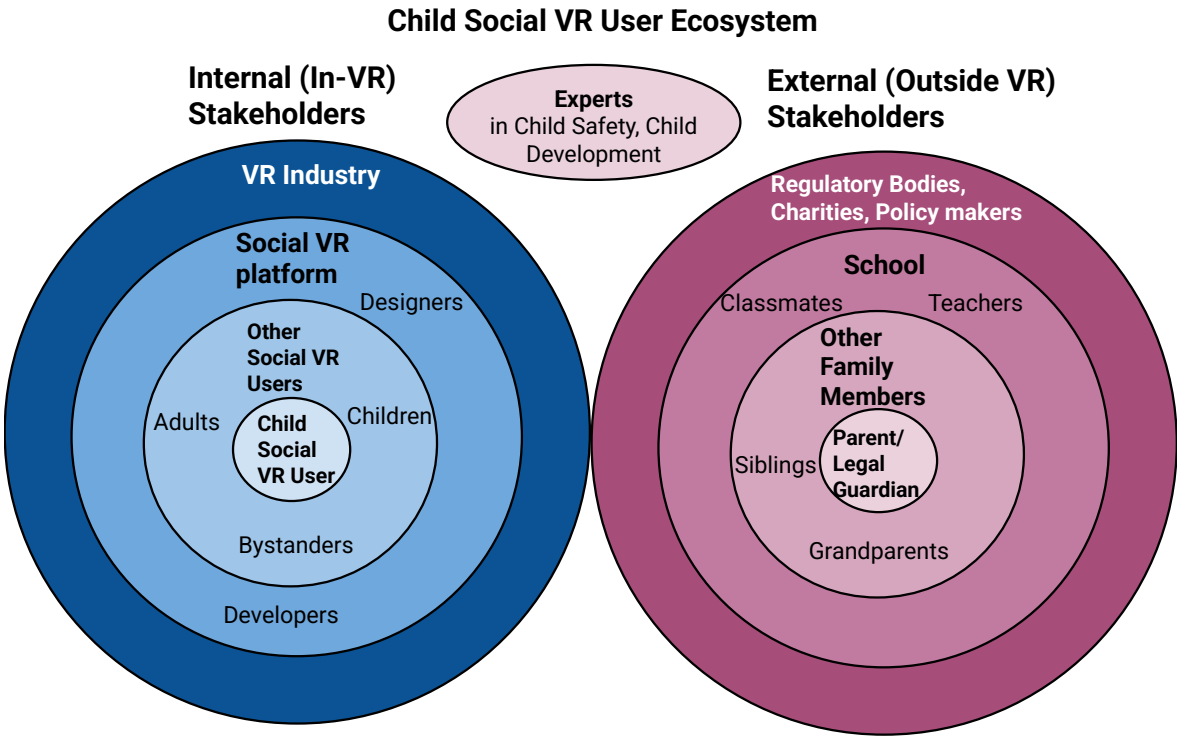


Figure 1.2: Diagram of child Social VR user ecosystem stakeholders. The inner circles represent the closest stakeholders to the social VR child user, while the outer circles indicate those more distant.

age boundary was set at 16 primarily because the first user study was conducted in a UK secondary school, which typically includes students in Years 7 to 11 (ages 11 to 16). While it would have been possible to include participants up to 18 years old, limiting the age range to 16 ensured consistency with the school setting, streamlined the recruitment process and ethics, which require separate forms for participants under 16. This age range also captures a key developmental period in which self-regulation, decision-making, psychological maturity, social behaviours such as peer influence and bullying are especially salient and still maturing [45, 46, 47, 48]. The broad age range (8-16) is also intended to support recruitment efforts. Overall, 164 children participated across all studies. The pooled mean age is 11.41 years (SD = 2.29), with 15 participants aged 8 and six participants aged 16, in total.

The user studies are conducted with children as players in social VR worlds, measuring their reactions in disruptive situations as "pure" victims [49], described as a child who experiences victimisation but does not reciprocate or engage in bullying behaviour. I chose not to focus on the harassers' perspectives for ethical reasons, as doing so would require longitudinal studies with in-game testing or the recruitment of individuals with a history of bullying others, for example. This approach would necessitate specialised expertise in criminology and psychology to ensure accuracy and ethical integrity.

While the primary focus of the user studies in this thesis is on the victim's perspective, I also gather insights into the harasser's perspective through qualitative data collected with experts and parents. For example, how they believe the interventions might impact the harasser.

1.3.4 Incidents and Interventions Time-frame

As moderation includes both prevention and intervention, in this thesis, I focus on intervening when the incident happens, in real-time, which can consist of *immediate* actions such as blocking but also *short-time-delayed* interventions from human moderation (also described as "Boundaries Settings" and "Quick Reactions" in prior work [50]). This enabled me to conduct short-term user studies and foundational prototypes with stakeholder perspectives that have not yet been researched. However, when participants raise prevention or after-the-fact responses, I consider these contributions and address them in the discussions.

The PhD research partly follows a research-through-design methodology, a process that involves designing prototypes as a means to generate knowledge [51, 52].

1.4 Research Questions

In this thesis, I set out to explore a series of research questions involving children, parents, and experts, in two parts: the first part is aimed at better understanding the context, and the second part investigates a solution I propose to safeguard children from harassment in social VR: **Automated Embodied Moderators (AEMs)**.

1.4.1 The Problem and Moderation Space (RQ1, RQ2)

At the start of the PhD in 2021, there was limited knowledge about how children use social VR and what safeguarding practices are appropriate for them. Therefore, it was essential to first understand the context by exploring the perspectives of parent and adult social VR users on children's use of social VR.

RQ1

What are the issues and concerns associated with children's use of social VR?

Based on the adult social VR user and parent perspectives and the identified issues derived from RQ1, including child harassment, I proceeded to explore existing and potential safeguarding approaches from the viewpoints of experts and children:

RQ2

What design considerations can help shape safety features to address child harassment in social VR?

1.4.2 Safeguarding Children from Harassment in Social VR using Automated Embodied Moderators (RQ3, RQ4)

Drawing from findings and the literature review, it is evident that age limits are frequently disregarded, current safety tools have limitations, placing the responsibility on the child user, and there is a lack of parental awareness and guidance. There is an important need for child-friendly solutions, improved supervision enabled by the system and awareness from parents.

To address these challenges, I propose a concrete and novel approach that incorporates new safeguarding practices shown to be effective in social media, including automated moderation and the unique affordances of VR (embodiment) to safeguard children from harassment in social VR: AEMs.

RQ3

How are Automated Embodied Moderators (AEMs) perceived as a solution to safeguard children from harassment in social VR, and what factors are considered important to the prospective usage of AEMs?

Finally, based on the findings related to RQ3, I derive specific requirements for the AEM in terms of its levels of automation and guardian involvement while considering the child's participation.

RQ4

Informed by RQ3, what are the requirements of the Automated Embodied Moderators to safeguard children in social VR while still empowering them?

1.5 Thesis Statement

This thesis demonstrates that the existing moderation solutions in place for safeguarding children from harassment in social VR are inadequate, highlighting the need for supervision and child-friendly tools (RQ1, RQ2). It proposes a novel approach to safeguarding these spaces using Automated Embodied Moderators (AEMs), designed in collaboration with parents, children and experts (RQ3, RQ4). This thesis demonstrates how AEMs can be used to enhance the perceived sense of safety and positive emotional valence without compromising the child's sense of agency. This thesis also contributes researcher-developed prototypes in simulated disruptive situations (individual and group harassment) in social virtual environments. Finally, this thesis specifies how AEMs can enact AI-driven moderation decisions and how parents can be involved while balancing child safety, privacy and agency in social VR.

1.6 Thesis Outline & Walkthrough

The thesis will be organised in the following way (see table summary in table 1.1):

Part I - Introduction - Chapter 1 & 2 After introducing the thesis context and scope in this *Chapter 1*, *Chapter 2: Literature Review* aims to provide the reader with a comprehensive overview of the research landscape and to present the concepts needed to justify the chosen approaches. The literature review is organised into three sections, beginning with an examination of 2D online and offline bullying, harassment and cyberbullying among children, mitigation strategies, relevant topics in criminology and law regulations, as well as parental mediation. Then, I present the unique challenges of social VR, mitigation efforts and limitations. The chapter concludes with methodological considerations for conducting research with children and the gaps addressed in this PhD project.

Part II - The Problem and Moderation Space (RQ1, RQ2) - Chapter 3, 4 & 5 In **Chapter 3** *Issues and Concerns Associated with Child Use of Social Virtual Reality*, I aim to gain an initial understanding of children's participation in social VR and what challenges emerge from adult user and parent perspectives with ***Study #1*** *- Exploring the Perspectives of Social VR-Aware Non-Parent Adults and Parents on Children's Use of Social Virtual Reality*, a mixed-methods questionnaire with 149 participants, including 70 parents, which allowed me to answer RQ1.

Publication 1

[**Full Paper Published CSCW 2024**] Cristina Fiani, Pejman Saeghe, Mark McGill, and Mohamed Khamis.[a] 2024. Exploring the Perspectives of Social VR-Aware Non-Parent Adults and Parents on Children's Use of Social Virtual Reality. Proc. ACM Hum.-Comput. Interact. 8, CSCW1, Article 54 (April 2024), 25 pages. `https://doi.org/10.1145/3652867`

[a]Note: P. Saeghe helped with proofreading and road testing arguments for the qualitative analysis; M. McGill and M. Khamis provided supervision and contributed to the iterative process of forming themes.

Chapter 4 *Informing The Design of Child Safety Features in Social Virtual Reality* dives deeper into design considerations of safety features in social VR from the viewpoints of experts and children to answer RQ2, with two studies conducted. ***Study #2*** *- Expert-Driven Design Considerations for Real-Time Safety Features in Social VR to Safeguard Children against Individual and Group Harassment.* One-on-one interviews were conducted with 16 experts (each with 5+ years of experience in child safety or psychology) to gather their perspectives on existing social VR safety features and to identify additional recommendations for addressing one-to-one and group harassment of children in social VR.

Publication 2

[**Under Review (Journal IJCCI) and Pre-print available**] Cristina Fiani, Joseph O'Hagan, Mark McGill and Mohamed Khamis.[a] Expert-Driven Design Considerations for Real-Time Safety Features in Social VR to Safeguard Children Against Individual and Group Harassment. `http://dx.doi.org/10.2139/ssrn.4839984`

[a]Note: J. O'Hagan was the second coder for the qualitative analysis; M. McGill and M. Khamis provided supervision and contributed to the iterative process of forming themes.

Based on the findings from study #3, I then conducted a user study with children ***Study #3*** - *Real-time Tools in Practice: Exploring Blocking Strategies in Group Harassment Situation towards a Child in Social VR* where 40 children aged 8 to 16 years old tested three different blocking tools I developed and two intervention perspectives, in a built-simulated social VR group harassment.

Publication 3 (Part 1)

[**Full Paper Published at IDC 2025**[a]] Cristina Fiani, Robin Bretin, Mark McGill, and Mohamed Khamis.[b] 2025. "Abracadabra, Block the Bullies!": Child Perceptions of Manual, Semi-Automated and Automated Interventions against Group Harassment in Social VR. In Proceedings of the 24th Interaction Design and Children (IDC '25). Association for Computing Machinery, New York, NY, USA, 314–339. `https://doi.org/10.1145/3713043.3728850`

[a]This publication covers studies #3 and #7. I present them separately in the thesis for better integration with the rest of the thesis.

[b]Note: R. Bretin was the second coder for the qualitative analysis; M. McGill and M. Khamis provided supervision and contributed to the iterative process of forming themes.

Chapter 5 offers a comprehensive summary and reflection of Part II, limitations and future work, illustrating how it leads to Part III.

Part III - Safeguarding Children from Harassment in Social VR using Automated Embodied Moderators (RQ3, RQ4) - Chapters 6, 7 & 8 This part focuses on the proposed novel approach to safeguard children from harassment in social VR: Automated Embodied Moderators (AEMs), explored with children, parents and experts.

In **Chapter 6** *Exploring Child, Guardian, and Expert Perspectives on Automated Embodied Moderators* addresses RQ3 through three studies.

Study #4 – Big Buddy: Exploring Child Reactions and Parental Perceptions towards a Simulated Automated Embodied Moderator for Social VR is a foundational user study with children that explores their perspectives on AEMs in social VR in a simulated disruptive scenario featuring "Big Buddy", a Wizard-of-Oz AEM prototype to safeguard children. The study involved 43 children aged 8 to 16. Additionally, 17 parents were interviewed at the end of the experiment as they observed the VR sessions on a tablet.

Publications 4 & 5

- [**Full Paper IDC 2023**] Cristina Fiani, Robin Bretin, Mark McGill, and Mohamed Khamis.[a] 2023. Big Buddy: Exploring Child Reactions and Parental Perceptions towards a Simulated Embodied Moderating System for Social Virtual Reality. In Proceedings of the 22nd Annual ACM Interaction Design and Children Conference (IDC '23). Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3585088.3589374
- [**LBW CHI 2023**] Fiani, C., Bretin, R., McGill, M., and Khamis, M. [1]*a* Cristina Fiani, Robin Bretin, Mark McGill, and Mohamed Khamis. 2023. Big Buddy: A Simulated Embodied Moderating System to Mitigate Children's Reaction to Provocative Situations within Social Virtual Reality. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA'23). Association for Computing Machinery, New York, NY, USA, Article 53, 1–7. https://doi.org/10.1145/3544549.3585840

[a]Note: R. Bretin was the second coder for the qualitative analysis; M. McGill and M. Khamis provided supervision and contributed to the iterative process of forming themes.

Further research was needed to investigate design considerations for effective AEMs from the additional stakeholders. ***Studies #5 & #6 - Multi-Stakeholder Perspectives on Automated Embodied Moderators for Child Safety in Social VR*** Through interviews with 16 experts in online child safety and psychology, and workshops with 8 guardians (5 parents and 3 grandparents) and 13 children (8-16 years old), the studies aimed to contribute to a comprehensive overview of how AEMs can safeguard children in social VR, exploring perceived concerns, benefits and preferences across the stakeholder groups and gather first-of-their-kind recommendations and reflections around AEM design.

Publication 6

[**Full Paper CHI 2024**] Cristina Fiani, Robin Bretin, Shaun Alexander Macdonald, Mohamed Khamis, and Mark McGill.[a] 2024. "Pikachu would electrocute people who are misbehaving": Expert, Guardian and Child Perspectives on Automated Embodied Moderators for Safeguarding Children in Social Virtual Reality. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 113, 1–23. https://doi.org/10.1145/3613904.3642144

[a]Note: R. Bretin was the second coder for the qualitative analysis of the expert interviews, while S. MacDonald supported the family workshops (specifically with parents) and was the second coder for their qualitative analysis; M. McGill and M. Khamis provided supervision and contributed to the iterative process of forming themes.

Chapter 7 *Deriving Requirements for Automated Embodied Moderators from Child and Parent Experiences* aims to answer RQ4 through two user studies based on Chapter 7's findings.

In ***Study #7 – Exploring Levels of Automation in Social VR: From Manual to Automated Interventions in Group Harassment Targeting a Child***, I contrasted manual blocking with different levels of automation provided by an AEM. I investigated how the initiation and decision of blocking, whether by the child or by an AEM, impact children's perceptions. 34 children aged 8 to 16 years old participated in the experiment.

Publication 3 (Part 2)

[**Full Paper Published at IDC 2025**] Cristina Fiani, Robin Bretin, Mark McGill, and Mohamed Khamis. [a] 2025. "Abracadabra, Block the Bullies!": Child Perceptions of Manual, Semi-Automated and Automated Interventions against Group Harassment in Social VR. In Proceedings of the 24th Interaction Design and Children (IDC '25). Association for Computing Machinery, New York, NY, USA, 314–339. https://doi.org/10.1145/3713043.3728850

[a]Note: R. Bretin was the second coder for the qualitative analysis; M. McGill and M. Khamis provided supervision and contributed to the iterative process of forming themes.

Study #8 – Parenting in Social VR: Designing Levels of Support that Foster Children's Autonomy and Safety, based on findings on *Parental Involvement* Based on findings from studies #4, #5 and #6, this final user study investigated how different levels of parental involvement, mediated through *Big Buddy*, impact both parent and child experiences following simulated Social VR harassment with 34 pairs of parent-child aged 8 to 16 years old. Parents used a custom-built mobile app to receive real-time notifications and provide support at varying levels of engagement.

Publication 8

[**Under Review (Journal TOCHI)**] Cristina Fiani, Robin Bretin, Shaun Alexander Macdonald, Joseph O'Hagan, Mark McGill and Mohamed Khamis.[a]

[a]Note: S. MacDonald was the second coder for the qualitative analysis; J. O'Hagan proofread the paper; M. McGill and M. Khamis provided supervision and contributed to the iterative process of forming themes.

Chapter 8 offers a reflection of Part III, with a design framework of how the AEM would work in practice, limitations and future directions to further investigate and develop AEMs.

Part IV - Chapter 9 describes the main contributions and presents final remarks and conclusion.

Each study will comprise related work, methodology, results, and discussion. The related work and discussion sections will predominantly refer to the overarching chapters, specifically, Chapter 2 (literature review), Chapters 5 and 8 (reflections), and Chapter 9 (guidelines).

RQ	Study	Description	Methods	Outputs
RQ1	Study #1	Social VR-aware parents and adult users) Exploring the Perspectives of Social VR-Aware Non-Parent Adults and Parents on Children's Use of Social Virtual Reality **(N=149, 79 parents)**	Mixed-Methods Questionnaire (Quantitative, Qualitative)	Full Paper published at CSCW 2024
RQ2	Study #2	Expert-Driven Design Considerations for Real-Time Safety Features in Social VR to Safeguard Children against Individual and Group Harassment. **(16 Experts)**	One-to-One Interviews with Experts (Qualitative)	Under Review (Journal IJCCI) and Pre-print available
	Study #3	Real-time Tools in Practice: Exploring Blocking Strategies in Group Harassment Situation towards a Child in Social VR. **(40 children 8-16yo, Mean Age=11.15 (SD=2.33))**.	User study and Short Semi-Structured Interviews (Quantitative, Qualitative)	(Part 1) Full Paper published at IDC 2025
RQ3	Study #4	Example of AEM (Big Buddy) in simulated social VR: Exploring Child Reactions and Parental Perceptions towards a Simulated Automated Embodied Moderator for Social Virtual Reality. **(43 children 8-16yo Mean Age=11.49 (SD=2.09) and 17 parents)**	User study and Short Semi-Structured Interviews (Quantitative, Qualitative)	LBW published at CHI 2023, Full Paper published at IDC 2023
	Studies #5 and #6	(Building on Study #4) Expert, Parent and Child Perspectives on Automated Embodied Moderators for Safeguarding Children in Social Virtual Reality. **(16 experts, 13 children, 5 parents, 3 grandparents)**.	Workshops and Interviews (Qualitative)	Full Paper published at CHI 2024
RQ4	Study #7	Child Perceptions of Manual Semi-Automated and Automated Interventions against Group Harassment in Social VR **(34 children 8-16yo, Mean Age=11.47 (SD=2.35))**	User study and Short Semi-Structured Interviews (Quantitative, Qualitative)	(Part 2) Full paper published at IDC 2025
	Study #8	Parenting in Social VR: Designing Levels of Support that Foster Children's Autonomy and Safety **(34 pairs of parent-child 8-16yo, children Mean Age=11.50 (SD=2.43))**	User study and Short Semi-Structured Interviews (Quantitative, Qualitative)	Submission to TOCHI

Table 1.1: Table summarising studies and outputs.

Chapter 2

Literature Review

2.1 Introduction

This chapter provides an overview of the key research areas relevant to this thesis. It begins by providing context on children's development and behaviour, defines harassment, bullying, and cyberbullying, and examines child safety within 2D digital media, drawing valuable lessons from both online and offline environments (section 2.2). It then defines VR and social VR, exploring child prevalence in social VR, highlighting the unique challenges and limitations of existing safety features (section 2.3). Finally, it reviews research methods in Child-Computer Interaction (CCI), outlining methodological approaches and challenges, as well as trends that inform the design and evaluation of studies with parents and children (section 2.4).

2.2 Child Safety in 2D Digital Media and Lessons from the Online World

2.2.1 Prevalence of the Online World and Its Impact on Children

Children now have access to the internet, social media and multiplayer online games at very young ages, with parental oversight becoming more difficult to maintain [53]. According to the Office of Communications (Ofcom)'s report (within the remit of the UK) on children's interaction with media and the online world (April 2024) [54], nearly one in four children aged 5 to 7 own a smartphone and 38% use social media platforms. Furthermore, roughly one-third (32%) of children aged 8 to 17 report encountering something distressing or unpleasant online in the past year and only 20% of parents with children in this age group say their child told them about seeing something that frightened or upset them during the same period [54]. An earlier Ofcom report [55] showed that 55% of children aged 8 to 17 played online games with someone they knew, 25% with someone they didn't know, and 47% chatted with someone they knew, versus 22% with someone they didn't know outside the game.

2.2.2 Relevant Aspects of Children's Social-Emotional Perceptions and Influences on Prosocial Behaviours

2.2.2.1 Psychological and Developmental Considerations

The United Nations (UN) Convention on the Rights of the Child (UNCRC) considers anyone of age 0-18 years, a child, and over time [56], the UN has formally recognised adolescence as the stage between ages 10 and 19 [57]. However, across countries, the social definitions may vary [58]. While age ranges are useful for marking and shaping generalised laws, policies and programs designed to offer child protection and safeguarding online and offline, biological and psychological development are usually not considered [58]. Erik Erikson's theory [59] outlines eight stages of human development, each shaping a person's social, economic, psychological, and biological aspects throughout their life [43]. In infancy (0–1 year), children learn trust when caregivers meet their needs. During early childhood (2–3 years), they develop autonomy by exploring independence. In the preschool years (3–5), they build initiative through play and decision-making. School-age children (6–11) focus on learning and developing competence and confidence, while in adolescence (12–18), they work on forming a clear sense of personal identity. The role of children's interactions with others from a very young age, whether it is with their caregivers, schools and now even more so with other online users, is also crucial in a child's development. If failed to reach these needs, the theory describes potential consequences such as lack of trust (from infancy), development of shame and doubt (from early childhood), a sense of guilt and lack of initiative (from pre-school years), a feeling of inferiority (from school-age stage) and weak sense of identity, and insecurity (from adolescence) [43].

As children develop, various dimensions of social knowledge are also expected to emerge over time [60]. These include understanding likely outcomes of actions, recognising different social goals, inferring others' intentions, identifying causes of social events, and evaluating the social appropriateness of behaviours [60]. These developments are shaped by experience and adult socialisation, such as learning norms, expectations, and consequences through guidance, modelling, and reinforcement [60].

2.2.2.2 Child Perception Towards Sanctions and Unfairness in Social Disruptive Situations

In the context of child development, as children grow, they begin to socialise and interact with others, which exposes them to potentially being rejected, unfairness and other negative and disruptive social situations, both offline (e.g., at school) and online [44]. Recent studies evaluated children's perception of rewards, sanctions and unfairness for different psychological and educational applications such as reinforcement learning, effective parenting, discipline and social norms for appropriate behaviour [44, 61, 62, 63]. The design and methodologies of these studies are of interest to the experimental design and interventions involving children for this thesis.

Children's Emotional Reaction and Self-Reflection in Provocative Situations Anger and frustration are negative emotions according to Russell's complex [64], which can be elicited in tasks such as toy removal and have been shown to orient children towards desirable goals or objects [62]. The latter study examined the effects of reward and punishment on the anger responses of 40 children aged 5-6 years old. Emotions were self-reported by children using the established Self-Assessment Manikin (SAM) Likert Scale [65], a scale that uses non-verbal pictorial figures to measure arousal, dominance and valence in response to a stimulus (more details about Likert Scales for children in section 2.4.2). The first experiment manipulated a situation in which children encountered obstacles (unfair treatment from the competitor) while pursuing a desirable goal. Each participant would compete with an unfamiliar player. The game was made so that the participant would always lose. Participants were then asked to rate their emotional state using the SAM scale. The experiment showed that anger is associated with attention biases towards rewards rather than punishment. Another study investigated children's (N=32, 8-13 years old) aggressive behaviour, focusing on aggressive Social Information Processing (SIP) [44]. Peer provocation (being excluded from a game and having their game ruined) triggered more anger, hostile intent attributions, revenge goals, and aggressive responses than neutral or instrumental contexts (stealing blocks/balls or sabotaging a peer to win) [44]. Building on the provocative scenarios from this prior work, I developed a prototype featuring provocative simulations for user studies #4 and #7, which was also used as a demonstration in studies #5 and #6.

Perceptions and Influence of Sanctions on Children and Parents Teachers in education have widely used reinforcement learning to maintain discipline in the classroom [61]. Perceptions of punishments and rewards for pupils' behaviour have been investigated, looking at the relative effectiveness of school-initiated rewards and punishments as perceived by children in primary school and parents via a survey (N_children=49, N_parents=64). Regarding punishments, children rated "information being sent home", "teacher explaining what is wrong with their behaviour in front of the class" and "being stopped from going on a school trip" as the top three of the most effective punishments [61].

The prosocial behaviour of children is largely influenced by adult figures, authority and media. A psychological study looked at the effects of being watched, monitored, and agency in parent-child and pupil-teacher relationships [42]. Agency can be defined as "beliefs that children can affect an outcome; to have self-efficacy and be effective as agents" [42]. Children make a distinction in their perception of agency depending on the relationship context. The study showed children perceived the least agency with teachers and the most agency with peers [42]. Indeed, agency should be regarded as a "complex and contradictory process developed over time" rather than a static attitude and can vary based on the interactions, the roles and positions of others [66].

Another study demonstrated that Disney characters can inspire children to help others [67]. Children (N=113, 8-10 years old) were divided into two groups: one was exposed to a Disney clip where the character was helping, and the other group was exposed to a Disney clip without helping behaviour. It was shown that children were more likely to help their friends after watching the animated clip with the helping

character [67]. Being watched can also affect someone's behaviour, called the *audience effect* [68]. When people are observed by others, they tend to have an increased prosocial behaviour, i.e., efforts or intentions designed to benefit the welfare of individuals, groups or society [68, 69]. Children's conception of authority from an individual is shown to depend on their status, the context and the domain of the act, depending on children's age [70]. Younger children's awareness of mothers' and teachers' authority was shown to be greater than that of police [70, 71].

2.2.3 Bullying and Harassment: Offline and Online

2.2.3.1 Definitions

Various terminology exists to refer to offensive behaviours offline and online. The Protection from Harassment Act 1997 [72] considers behaviour to be *harassment* if it causes a person to feel threatened, distressed, humiliated, or fearful of further harm. *Bullying* was initially referred to as repeated, intentional acts of aggression carried out by an individual or group against a target repeatedly over time [73]. A definition was later proposed as: "an aggressive goal-directed behaviour that harms another individual within the context of a power imbalance", considering not only the frequency and intensity of bullying and victimisation, but also the victim perceptions of harm, to capture all bullying behaviours [74].

These terms were then expanded into the online world and are still evolving. *Cyberbullying* is a form of bullying via electronic mediums [75]. However, research pointed out the challenges with respect to applying the presented definition of *bullying* to *cyberbullying*, for measurement and classification, notably, whether and how the power imbalance criteria in the general definition of traditional bullying can be applied to cyberbullying [76, 75, 74]. Online, power imbalance may not appear as physical differences, but rather in terms of social status or via anonymity [74]. *Cyberbullying* also encompasses the terms *online bullying* and *online harassment*. *Online harassment* was recently presented as a wide range of harmful behaviours facilitated by digital platforms and directed at one or more individuals [31]. These behaviours can include verbal attacks such as name-calling or insults (*flaming*), disclosure of personal information (*doxing*), or using another person's identity without permission [77]. When defining and measuring these behaviours, it is also important to consider the difference between objective actions and the victim's subjective perception [74, 31].

The distinction between bullying and cyberbullying lies not only in the opportunity of harm that online anonymity provides, the lack of supervision and expected mediators (e.g., teachers in schools), but also, as devices are used for different purposes, anytime and anywhere, cyberbullying can hound a child from school to their home [78].

Depending on the medium within which harassment occurs, group harassment has been depicted using various terms. On social networks, it is referred to as *networked harassment*, where an individual is "harassed by a group of people networked" through instant messaging or other forms of posting [79]. In the offline world, terms such as *group aggression* (collective physical or psychological harmful behaviour directed towards an individual [80]) and *mobbing* (form of group aggression characterised by repeated actions against the same victim [81]) are often used in relation to children in schools [81, 82], as forms of bullying.

A bullying or cyberbullying incident typically involves at least two individuals, one identified as the *bully* or *harasser* and the other as the *victim*. However, prior research has conceptualised these behaviours along a continuum, recognising that individuals may occupy various roles, including *pure* bully, *pure* victim, *bully-victim* (also referred to as *aggressive* or *provocative* victims [49]) and/or bystander [83]. Unlike pure bullies or pure victims, a bully-victim can have both roles. A study identified a similar group of bully-victim children (*target/aggressor*) involved in internet harassment [84].

The psychosocial challenges that can emerge from cyberbullying victimisation are associated with several negative consequences similar to those seen in offline traditional bullying [85, 86, 87]. These include victims having reduced self-esteem, increased loneliness, levels of depression and anxiety, and they can experience difficulties in their everyday life (e.g., relationship, academic) [85].

2.2.3.2 Evidence-based Intervention for Bullying and Harassment (Offline & Online)

A study has shown that the most effective interventions to tackle offline traditional bullying and victimisation in schools appeared to be disciplinary sanctions by communicating to the bully that the behaviour was unacceptable and reporting the event to other adults [88]. As a second most effective intervention, they showed that teacher-facilitated group discussions would increase the likelihood of peers being defenders to the victim [88]. Teachers, therefore, have a major influence towards student behaviours. Moreover, non-interventions can reinforce the message that victimising others has no negative consequences and bullies would not recognise their behaviours as unacceptable. While most anti-bullying interventions involve teacher supervision and consequences, another study [89] suggests that youth mentoring interventions (e.g., Big Brothers/Big Sisters of America) and support groups have beneficial effects on peer relationships. In a recent study, bullied children in elementary school (N=12) were paired with mentors who visited the school twice a week. It was shown that children paired with a mentor experienced less peer victimisation. Within classrooms, having classmates who showed more peer support and who were older in general than children in other classrooms reduced the risk of being a victim of bullying [90]. Indeed, it was shown that peers act as a group to influence the development of individual children [91].

As described above, cyberbullying can lead to negative psychological consequences (e.g., depression, anxiety, loneliness). Research proposes different strategies to prevent and cope with cyberbullying. While the focus of the thesis is on interventions, I will shortly describe prevention strategies shown to be effective. These include 1) educating about online safety, 2) training bystanders to intervene, and 3) active parental involvement and monitoring of social media use [92]. Another effective preventive solution involves reflective messages to dissuade someone from posting content detected as cyberbullying. Royen and her colleagues [93] found that such reflective messages can reduce the intention to engage in cyberbullying.

Regarding interventions and coping mechanisms, previous research has examined responses to cyberbullying using the *Transactional Model of Stress and Coping (TMSC)* [94] as a theoretical framework to explain how different strategies can result in either positive or negative outcomes. Identified coping styles include problem-focused, emotion-focused, and avoidance-focused approaches [94]. *Problem-focused* strategies aim to directly resolve the issue, such as seeking social support from adults or peers, which is consistently endorsed as positive and effective [95, 96]. *Emotion-focused* coping manages the emotional impact through actions like internalising or externalising emotions [94]. *Avoidant-focused* strategies involve disengaging, either physically or cognitively, from the stressful situation [94]. Moreover, victims often employ both online/technical solutions and traditional (offline) methods [97]. Avoidant online strategies, such as blocking the bully, unfriending on social media, changing account IDs, or deleting profiles [78], were rated as very helpful and effective for stopping the cyberbullying, though fewer children reported using them [97]. Conversely, traditional strategies like talking to a friend or adult were perceived as very helpful, especially for emotional support [97]. Strategies like retaliation, confrontation (especially online), and searching for advice online were generally rated as less helpful [97]. The selection and effectiveness of these strategies are influenced by an individual's appraisal of the threat and their self-efficacy, with problem-focused coping being more likely when a stressor is perceived as controllable [97].

Furthermore, observers' or bystanders' reaction is critical as they may influence those involved and the occurrence of the events [98]. They can change the course of the situation and reduce the negative effects on the victim by confronting bullies and reporting to adults [99].

With the increase of cyberbullying and online harassment on social media and its detrimental impact on mental health, the need for content moderation becomes stronger [100]. The large amount of content posted everyday, however, makes human moderation logistically challenging [101]. In response to the scale of content posted, automated moderation is used. It consists of deploying various artificial intelligence (AI) techniques to filter and process user-generated content, applying preset rules to reject or approve the image or text posted online [102]. Methods such as text classification for harassment-related keywords and natural language processing (NLP) have been shown to detect hate speech with high accuracy [103, 104, 105]. While AI-moderation shows some promise, it has key limitations, including difficulties in parsing ambiguity, identifying false positives [102], and reinforcing biases. Furthermore, there are ethical considerations such as the lack of transparency [106], visibility [107], privacy and user agency [108, 109]. A study with 59 teenagers showed that these interventions may impact children's rights, and emphasised the need to involve children in AI design decisions for platforms [110]. The study further showed that content removal does not solve the actual problem nor support the victim [110].

2.2.4 Parental Role, Responsibility and Parental Mediation of Children's Interaction with Digital Media

2.2.4.1 Approaches to Parental Mediation

Parents typically hold responsibility for their child's well-being and safety. Parents can shape and influence their children's media use and habits [111]. In the digital age, *parental mediation*, defined as "the management of the relationship between children and media" [112], has become an essential part of parenting. Originally conceptualised in the context of television through three approaches (Active, Restrictive and Co-Viewing [113]), the framework has since evolved to address the complexities of online media, which is mobile and ubiquitous. Livingstone and Helsper (2008) [112] expanded the model to include four main categories of mediation in these newer digital contexts: 1) Restrictive mediation, setting rules and regulations, 2) Active mediation, with direct communication with the child about their online usage, 3) Parental Monitoring, such as checking child online interactions, friends lists and downloaded applications and 4) Technical Mediation, such as parental controls, filtering and monitoring software.

Recent approaches to online child safety increasingly emphasise a broad, multi-stakeholder perspective. These include industry measures (e.g., age restrictions), policy and educational initiatives (e.g., school programs), social mediation (e.g., active parental involvement), and technical solutions (e.g., parental controls) [114].

However, research also highlighted growing tensions between parental control and teenagers' need for self-regulation [115, 116]. To address this, a new framework was proposed, *Teen Online Safety Strategies (TOSS)*, which complements parental control with the promotion of teen self-regulation [117]. TOSS includes three pillars: 1) Self-monitoring, encouraging teens to manage their own online behaviours, 2) Impulse control, fostering the ability to delay short-term gratification in favour of long-term benefits, and 3) Risk-Coping, equipping teens to handle online challenges by problem-solving or seeking guidance when needed. A recent study investigated 58 parental control applications through reviews, examining how features align with the TOSS framework [118]. They found a prevalence of restricting and monitoring features and suggested that these can raise ethical concerns, including excessive restrictions, privacy concerns, and a reduction of child autonomy. They also found that applications providing support for parent-child communications, transparency and features with medium granularity (i.e., flexibly setting boundaries) were perceived as more positive by parents and children [118]. While the TOSS framework provides a useful lens for evaluating digital safety strategies in 2D screen-based environments, it has yet to be applied to VR technology.

Two other approaches to parenting regarding digital media and devices were determined: the ethic of *expressive empowerment*, where parents emphasise the individual rights and trust of their child and the ethic of *respectful connectedness*, where parents prioritise the rights of the family and familial goals [119]. Moreover, studies on social media showed that parent-child trust is crucial, and trust and control are not mutually exclusive [120, 114].

2.2.4.2 Challenges in Implementing Parental Mediation

As technology continues to evolve rapidly, many parents face challenges related to unfamiliarity and the need for constantly updated digital skills [117, 121, 122]. A study showed that limited digital skills often lead parents to adopt either restrictive or indulgent approaches to managing their children's online activity [121]. In contrast, those with stronger digital skills are more likely to engage in effective monitoring and mediation [121]. This skills gap has been observed not only in the context of social media, but also in gaming and emerging technologies such as VR. Over 90% of parents feel uncomfortable letting their children explore the metaverse through VR unsupervised [18]. Furthermore, recent research found that many parents feel overwhelmed by the demands of keeping their children safe online and are ill-equipped to develop their children's digital safety skills [123].

Additionally, parental mediation is often viewed as one single approach, and it does not fully take into account how a parent's ability to help depends on their availability [124, 125]. In many families today, both parents work, which means they often have limited time and attention [126, 127, 125]. Distinguishing between levels of parental availability, whether fully, partially available, or minimally available, is important, as it affects both the feasibility and effectiveness of digital intervention tools. Different rules, parental controls used, and safeguarding practices also vary depending on the children's ages, cultural backgrounds and parent-child relationships [120, 121].

2.2.4.3 Parental Concerns of Children's Digital Media Usage

Parents tend to be concerned that digital media amplifies risk, including cyberbullying, children's social isolation, involvement in illegal activities, a threat to academic and physical development and mental health decline [128, 129, 130]. A study questioned French parents (N=149) of school-aged children, regarding their concerns and children's use of digital devices [131]. Parents were mainly concerned that digital devices could harm the child's socio-emotional and cognitive development, and they feared the technology would be at the expense of other activities. They were also concerned about the risk of being exposed to inappropriate content and the risk of addiction [131]. One of the main difficulties parents have is controlling the children's content and screen time [131, 132, 133].

2.2.5 Relevant Topics from Criminology and Law

2.2.5.1 Capable Guardian

In criminology, researchers have identified three essential components that describe the necessary conditions for a criminal event: an offender, a suitable target and the absence of *capable guardianship* [134, 135]. *Capable guardianship* concept was refined as "the presence of a human element which acts, whether intentionally or not, to deter the would-be offender from committing a crime against an available target" [134]. Researchers reported that capable guardians are closely linked to the completion of a

crime [134], as the offenders tend to move toward a place where there is a target without a guardian. A common example of guardianship is active CCTV [135]. More recent research has highlighted the crucial role of guardians in online spaces as a preventive measure against cybercrime [136].

2.2.5.2 Regulations and Policies for Child Online Safety and Child Privacy

This section lays out the current regulations and policies for child online safety. While these regulations are recent and apply to any online spaces, including social VR, they do not specifically tackle the new technology, which, as described in section 2.3, has unique aspects that create new forms of harm and risks.

Internationally, governments are increasingly adopting regulations for online safety [137], notably the European Union's Digital Services Act (DSA) (2022) [138] and the Online Safety Act (OSA) in the UK (2023) [139], now fully implemented as part of the law. Regulations were also enacted across Asia (e.g., Singapore Online Safety (Miscellaneous Amendments) Act 2022 [140]) and Oceania (e.g., Australia Online Safety Act (2021) [141] and Social Media Age Minimum Amendment (2024) [142]). As the thesis project is in the UK, I will only introduce the UK OSA [139]. The primary purpose of OSA is to protect adults and especially children online [143]. Platforms must be both reactive and proactive in addressing the risk of illegal content [143, 137]. If children are likely to access the platform, age assurance measures must be implemented, and some platforms must meet requirements to improve transparency and accountability [143, 137]. However, it has led to debates regarding its potential impact on freedom of expression and user privacy [144, 137]. Policy advisors and academics suggest that international human rights law should guide the regulation of online spaces, as it encompasses protections for privacy and children's rights [137]. The UK has signed up to the UNCRC [145], which sets out a number of children's rights, including the right to [146]: survive, grow, participate and fulfil their potential. Children's rights must be respected and protected both online and offline [147]. Moreover, they should have the same rights as adults over their personal data, including control over its sharing and the requirement to give consent before it is used [146]. The 5Rights framework also reflect international treaties and laws, and combine the following rights for children in the digital world [148]: 1) the right to remove content, 2) the right to know, 3) the right to safety and support, 4) the right to informed and conscious use, and 5) the right to digital litteracy.

2.3 Child Safety in Social Virtual Reality

2.3.1 Social Virtual Reality

The past decade has seen VR reach the consumer market and grow significantly, especially with the pandemic accelerating adoption [16, 149, 15]. In contrast to 2D social media platforms, social VR is a 3D, 360-degree synchronous social experience through the use of VR headsets [23, 22], offering embodied avatar-based interactions in im-

mersive virtual environments, that give the feeling of "being there" [35]. One of the mainstream social VR platforms, VRChat, has an estimated total of 15 million players (according to PlayTracker [150]) and a peak of 127,919 players simultaneously [151] (as of November 22nd 2025).

Before delving into the specifics of social VR, I will first explain what VR is.

2.3.1.1 What is Virtual Reality?

VR is defined as "the sum of the hardware and software systems that seek to perfect an all-inclusive, immersive, sensory illusion of being present in another environment" [152]. The technology encompasses three main features: *Presence*, *Immersion* and *Embodiment*.

Presence is referred to as "the subjective experience of being in one place or environment, even when one is physically situated in another" [153]. It is achieved through *Place Illusion* and *Plausibility Illusion* according to Mel Slater's theory [154]: *Place Illusion* is the experience of feeling present in a location despite knowing that one is not physically there. *Plausibility Illusion* describes the perception that events occurring within the virtual environment are actually happening, shaped by how convincingly the system responds to user actions and environmental changes. When both *Place Illusion* and *Plausibility Illusion* are strong, users are more likely to behave as if the virtual events were real [154].

The term *Immersion* does not have a single agreed-upon definition. Some researchers explain it as a psychological state in which users feel separated from the real world and fully absorbed by the virtual environment, forgetting about the outside world [153]. Others define immersion as a feature of the VR system itself, where some systems are considered more immersive than others depending on factors such as the number of sensors used or the width of the field of view [155]. In this thesis, I consider immersion as the latter.

VR technology also enables a *Sense of Embodiment* (SoE) consisting of a sense of self-location, sense of agency, and sense of body ownership [156], which can enhance and influence social interactions. In social VR, adult users have highlighted that engaging in shared activities in embodied and more physical ways (e.g., virtual hug [157]) led to fostering extended relationships with other players and enhanced connectedness due to verbal and non-verbal communication modalities [158, 10]. Research has also shown the impact of embodiment in social relationships and interactions [159], in particular as one of the four social embodiment effects: "perceiving bodily states in other people actually results in bodily mimicry in oneself", the observer's embodied responses can mimic the perceived embodied stimuli (e.g., when someone smiles at you) [160].

The feeling of presence in VR has been shown to have a significant impact on children psychologically [36, 37]. For example, VR 3D characters have a stronger impact on decision-making than 2D TV characters [36]. Virtual embodiment and socially responsive virtual characters in immersive VR could consequently have a notable impact on the way children experience, and react to, social interventions [37]. Embodied avatars in immersive social VR could leverage an influential appearance and relationship: prior work shows that children's perception of such figures (e.g., Disney characters [67], teachers VS peers [42] as described in section 2.2.2.2) influences their behaviour,

highlighting the need to carefully consider the avatar's representation as it may lead to negative (decreased agency) or positive (inspire good behaviours) impact [67, 42]. Few studies have examined social embodiment in VR among children [37, 36, 161]. For example, a recent study [161] investigated how character type and perceptual realism in VR influence the social-emotional perceptions of 5 to 9 years old children (N=25) by having them interact with three embodied agents representing varying levels of social realism: a human (high realism), a giraffe (moderate realism), and a muppet (low realism). Children responded most positively to the giraffe and interacted with the characters' embodiment and personification similarly, experiencing the virtual bodies as if they had physical presence while still acknowledging they are fictional [161].

2.3.1.2 What is Social Virtual Reality?

Social VR refers to the application of VR where multiple users connect to interact within a shared virtual environment [162].

When using social VR platforms, combined with presence, immersion and embodiment, almost all senses are engaged: sight, hearing, and even touch through what is called *phantom touch* i.e., tactile sensations without physical stimulation [163]. Research has shown that social VR users perceive *phantom touch* positively as it increases their feeling of presence and enables more meaningful interactions with others. Others view it negatively, undesirable and odd as it may lead to new forms of harassment [12].

Because of these VR affordances, three sociotechnical traits of social VR can be identified [11]: 1) social VR enables real-time full-body motion and gestures in high fidelity 3D-360 immersive virtual environments, 2) it provides rich spatial and temporal experiences and evokes sensory responses comparable to in-person interactions and 3) it mediates both verbal and non-verbal communication (via gestures, gaze, facial expressions, proxemics).

Among the top popular VR games, we can find the following social VR games [164]: VRChat [25], Rec Room [27], Gorilla Tag [165] and Horizon Worlds [166]. These platforms are usually free and easily accessible, offering a wide variety of activities, immersive virtual content, and a high-fidelity, realistic social environment [167]. Social VR platforms leverage avatars as the embodied interfaces and proxies through which users interact within VR [167, 168]. Upon entering the social VR platform, users can personalise their avatars and appearance [167]. Because the avatar serves as a direct proxy for the player's physical self, it allows users to engage with the virtual world through a redefined identity, one that is crafted and personalised [11]. Through this process, users can impersonate their avatars with unique behaviours, intentions, and styles, effectively seeing their virtual avatar as a "second self" [11]. Compared to 2D online gaming, in social VR, the connection, where the avatar's movements mirror the user's physical body, via partial or full-body tracking, in real-time, creates a heightened sense of presence and embodiment, enabling the user to experience self-representation [11]. It can also lead to the *Proteus Effect* where individuals conform to their avatar's characteristics [169], influencing their own behaviour and perception [11]. For some, particularly transgender individuals, the embodied experience in social VR is shown to be a means to explore and affirm their gender identity [11].

The design choices made across various social VR applications have a significant impact on communication and interaction [167]. Based on interviews with industry experts, a study categorised what social VR platforms can offer into the following features: 1) embodied locomotion, 2) avatar aesthetics, 3) personal space mechanics, and 4) social mechanics [167]. 1) Users learn to control their avatars through teleportation or by supporting a third-person walking mechanic to minimise motion sickness. 2) Avatar aesthetics include design choices, with some choosing simpler cartoon-like avatar appearances in applications such as Rec Room [27] while others, like VRChat, provide extensive freedom for users to create and import their own fully custom avatars [25]. 3) Social VR applications implement features that use the uniqueness of digital proxemics (i.e., how the space in virtual environments is used and how the presence of others influences our behaviours, interactions, and movements [170]), with the implementation of features such as personal safety bubbles. 4) Social VR integrates social mechanics to enhance interaction. These include friending systems like Rec Room's fist-bump or shake-hand gestures, and VRChat's community gestures like head pats and virtual hugs [167].

Social VR platforms enable users to create their own activities, games, rooms, and entire worlds. In Rec Room, this is done through "Rec Room Studio" [13] and "Maker AI" [171], and in VRChat, players can upload fully custom worlds built with Unity [14]. These creations range from casual hangout lounges, cinema theatres, to interactive games, art galleries, dancing events, classes and even live performance venues and concerts, fostering a constantly evolving ecosystem of user-generated content [21, 172].

Social VR platforms have also been employed for a variety of purposes, from maintaining long-distance relationships [173] and hosting weddings [174], to more concerning uses, such as virtual drinking becoming popular [175] or even users falling asleep with other users while using social VR [172].

2.3.2 Prevalence of Children in Social VR

As VR headsets become more affordable, with an anticipated market growth from $30 billion in 2022 to over $520 billion by 2031 [176], children today are increasingly interacting and socialising in social VR. UN Children's Fund (UNICEF) reported that "tens of millions of children and young people are already active in virtual environments and game spaces" [177]. The Institution of Engineering and Technology (IET) estimates that the next generation will spend the equivalent of around 10 years of their lives in virtual reality [18], and according to the IET study, more than a fifth of 5–10-year-olds (21%) already have a VR headset of their own.

Over the past decade, research on children's use of VR has mainly focused on the effectiveness of the technology in educational or medical settings for children [178, 179, 180, 181, 182, 183]. In particular, VR has been used as an intervention tool for children with autism spectrum disorder [179, 182], and for rehabilitation and therapy in children [178, 183]. With the pandemic, the need for remote instruction tools increased, and VR has enabled immersive learning, shown to be more effective than other traditional tools [180]. The latter study explored the innovative opportunities and challenges associated with using VR for children's remote education via an online survey (N = 311) [180]. They collected data from the parents and legal guardians of children who used VR at

home during the first months of the global pandemic. The study showed that 71.5% of participants' children increased their VR usage. Recently, Maloney and their colleagues explored minors' social VR use, not only presenting opportunities for this disruptive technology but also raising challenges about children's social VR use [23, 22, 21].

While originally designed for adults (e.g., Horizon Worlds was originally 18+ in 2021 [184]), platforms like VRChat, Horizon Worlds and Rec Room [25, 27, 185], now attract many children under 13 [22, 23, 18], even though most social VR platforms recommend a minimum user age of 13 [186, 187], with recent exceptions allowing children aged 10–12 to participate through junior, parent-managed accounts [185, 188]. Similar age restrictions apply to VR headsets, lowered from 13 to 10 by Meta in June 2023 [189, 17]. Children and teenagers are drawn to social VR platforms because of their rich social features, opportunities for play and games, and the natural excitement young people have for new technologies [22, 190].

2.3.3 Harassment in Social VR

The unique VR affordances that mimic real face-to-face interactions while still being anonymous have led to new forms of harm. Drawn from 25 interviews [31], three main types of harassment have been identified in social VR: 1) as found in 2D media, *verbal* (i.e., voice or chat), and unique to social VR, 2) *physical* (e.g., avatars invading personal space or physical attack), and 3) *environmental* (e.g., displaying graphic content in a virtual space). As VR enables full embodiment and immersion, it fosters a sense of presence in the virtual environment [35]. Moreover, due to the anonymity and freedom of these new platforms, there are looser social and ethical norms that can result in increased hostile behaviours from users [20]. Through analysis of the videos and YouTube users' comments, a recent study reveals severe and unexpected safety risks in social VR, emphasising the influence of varying understandings of community norms on reactions to these risks [50]. Among 1000 reviews of Rec Room on the Meta Quest page, 114 were found to report experiencing and/or witnessing harassment during use [191].

The shared use between children and adults has also led to concerning reports of harm not only towards children but also towards adults, including exposure to harassment, sexually inappropriate behaviour, and peer pressure [50, 20, 192, 193]. A growing number of child sexual abuse cases, where children describe sensations of *phantom touch* [194]; referred to as "Meta-Rape", conveying the intense and embodied nature of these experiences [32], in social VR has raised serious concerns in recent years [195].

2.3.3.1 Harassment in Social VR towards Children

Children are vulnerable when experiencing harm in social VR due to the affordances of the technology (presence, immersion, embodiment) [36, 37]. This underscores the importance of understanding children's social VR experiences and perceptions and the need for suitable safety-enhancing tools. Interviews with 30 social VR users revealed widespread observations of children being involved in interactions far beyond their maturity level [22]. The psychological impact of these harms is amplified for children due to their heightened sense of presence in VR. Using a head-mounted device, children aged 6–18 felt greater levels of presence and perceived "realness" in VR compared to adults aged 19–65 [196]. If children perceive VR as more authentic, they could be more susceptible to its impact, potentially amplifying the effects of negative experiences. Another study showed that children displayed increased social responsiveness to VR characters, compared to characters on 2D television [36].

Similarly to the continuum of bullying behaviours described section 2.2.3.1, in social VR too, users can be found to be bully-victims, i.e., social VR users who have experienced both victimisation but also have been accused of being harassers [197]. Social VR users can hold different views on what constitutes harassment versus behaviour that is merely inappropriate or playful, and there is often a mismatch between children's and adults' expectations regarding appropriate social behaviours in social VR environments [22, 23].

2.3.3.2 Group Harassment in Social VR

Group harassment in social VR has emerged as a blend of *networked harassment* [79], *group aggression* [80] and *mobbing* [81], combining the embodied, real-time physical-world interactions with the anonymity and virtual environments characteristic of social media [31, 20, 198, 199, 200, 201]. In prior work, a user shared their experience where "a huge group of strangers surrounded him and shouted swears and slurs when he first entered the space, making him feel quite threatened and unsafe" [20]. Another user mentioned "More times than not, there would be one guy or a group of guys engaging in harassing behaviour because that's how they enjoy their weekend." [200]. The realistic and engaging nature of VR, combined with the potential for group harassment, can amplify the emotional and psychological impact on children. For instance, the phenomenon of "flash mobs" (i.e., large gatherings of people) in VRChat demonstrates how collective actions, like noise bombardment or physically boxing in users, can cause greater disruptions to experiences than individual instances of harassment [201].

2.3.4 Existing Safety Features in Social VR and their Limitations

Unlike traditional social media platforms, the immersive nature of social VR demands a reevaluation and expansion of existing safety features and regulations. The language and current regulations described in section 2.2.5.2 (e.g., Online Safety Act (2023) in the UK) are still too focused on 2D online interactions rather than immersive, 3D experiences. Lessons learned from conventional social media platforms certainly carry weight for verbal harassment, but other types of harassment specific to social VR, i.e., physical and environmental [31], underscore the necessity for innovative safety regulations and features tailored specifically to the space of social VR.

Social VR platforms have put in place Codes of Conduct [202, 203] that include behaviour rules such as "Sexist, racist, discriminatory or harassing language, behaviour, or content is not welcome" and "Players 12 or younger must use a Junior Account" [202]. Ages are generally restricted to 12 or 13 and older for most VR headsets and social VR platforms [204, 205, 27, 206], though some platforms have recently lowered the minimum age to 10 [17]. However, most VR and social VR platforms do not implement mandatory age-verification systems. The UK Online Safety Act (2023) [139] mandates age verification for certain content (specifically pornography and other material deemed harmful to children) but does not require it across all VR or social VR platforms.

Platforms have also incorporated existing safety social media tools like blocking and reporting a user [204, 187, 205], and developed new safety features, fitting into categories of *boundary settings* (e.g. personal space bubbles, shield levels), and *quick reactions* (e.g., reporting, "stop hand" gesture) [50], to protect their users from physical and environmental harassment [204, 187, 205, 50].

While little is known about their effectiveness and usage, a recent study involved researchers observing YouTube videos of social VR users' experiences and identified a list of major in-game real-time safety features across four social VR platforms, including VRChat and Rec Room [50]. They grouped social VR safety features into three categories: boundary settings to manage interactions before an incident (e.g., intimacy proxemics, trust/reputation systems, social spaces, shield levels, and controls for voice and avatars), quick-response tools for use after an incident occurs (e.g., safety gestures, vote-to-kick, safety zones, and incident reports), and agreements that apply at all times (e.g., a code of conduct) [50]. The researchers highlighted that relying only on proxemics is inadequate since different safety threats require varied social cues due to the VR affordances. Current features lack reactive gestures (such as the "talk-to-the-hand gesture" used to block in Rec Room [187]), and adding these could enhance users' ability to respond to threats [50]. Most existing safety tools are reactive, designed to address issues after they occur. This puts pressure on victims to effectively use these tools while under distress, which may be especially challenging for children. The study advocates for a move towards preventive safety mechanisms, incorporating social cues to identify risks early, such as using signal processing data to detect gestures that signal safety concerns [50].

In 2024, a survey of 100 adult social VR users examined the use and effectiveness of those safety tools across 12 popular SVR platforms [9]. Findings showed that effective safety tools predominantly include familiar reactive features such as Mute and Block (61% of respondents had used these features) and Reporting (48% had used this feature), which users widely prefer for their immediacy and ease of use, despite the latter's perceived lack of transparency regarding outcomes [9]. Alongside these, social VR-specific proactive tools like *Social Spaces*, *Intimacy Proxemics*, and *Interaction Shields* enable users to "socially sculpt" their virtual experiences, providing pre-emptive protection and customisation, though these are less frequently adopted due to lower awareness, inconsistent availability across platforms and risk creating fragmented social spaces [9]. Despite the availability of these tools, harassment remains prevalent and is often considered an "inherent quality" of social VR, necessitating more standardised, transparent, and proactive solutions [9].

Safety tools that focus on increasing distance, such as "StayAway" [207], are particularly effective as they reduce the auditory and physical impact of abuse, provide victims with a clearer overview of the harassers' actions, and remove them from the immediate source of harassment. Among other tools evaluated, "StayAway" was the preferred response strategy in social VR due to its high efficiency, lack of restrictions, and freedom of operation, which allows users to leave the room at any time [207].

Besides existing mitigation tools, there are human moderators to moderate and address incidents in real-time in the social virtual environments [31, 20]. Yet, only 24% of incidents observed in a study in social VR were addressed by human moderators, highlighting the lack of scalability for human moderators in a growing VR ecosystem and the need for new moderation tools [19].

Automated/AI moderation can address this scalability issue and has, therefore, been proposed as an alternative safety tool to combat harassment in social VR [40]. A recent study has begun investigating harassment detection using AI, demonstrating that their system effectively uses pose information as contextual data to identify harassment, achieving a detection accuracy of up to 98.26% in experimental evaluations [208]. However, users are concerned about false positives and undermining trust [40].

VR headsets (e.g., Meta Quest) offer casting [209] (i.e., when someone uses a VR headset, casting lets others watch the same view in real time on another device), but using it for parental oversight requires a time-consuming setup, compatible display devices, and constant real-time monitoring, which is impractical for repeated multi-hour play sessions. In a survey [210], parents noted that existing safeguarding ways are not flexible for parental monitoring and control, and while there is a casting feature, consistent monitoring poses challenges. Moreover, parents expressed wanting non-intrusive parental monitoring tools for child VR activities without needing to wear the headset themselves. For example, recorded video clips of child usage that can be reviewed or an additional small screen on the headset or mobile phone for parents to quickly check on their children's current VR activity [210].

Some platforms offer standard parental control features and junior accounts. For instance, Meta Horizon allows parents to supervise their child's activity through a Family Centre dashboard (introduced in March 2022 [211]), where they can approve follower requests, manage privacy settings, monitor screen time, and restrict communication [212]. Rec Room provides linked junior accounts that enable parents to control spend-

ing and disable in-app purchases [213]. Despite these measures, there is a clear lack of consistent and robust parental control tools across platforms. Notably, VRChat [25], one of the most widely used social VR environments [214], offers no parental controls at all, leaving children potentially exposed. Moreover, even when tools are available, their effectiveness heavily relies on active parental engagement, consistent use, and digital literacy, factors that vary significantly between households. Indeed, in a recent study, researchers interviewed 20 parents with children under 18 who use VR at home and found that parents lack awareness and the majority rely on active mediation, such as verbal communication, rather than technical mediation due to their perceived technical constraints [215].

2.4 Research Methods for Child-Computer Interaction

2.4.1 The Importance of Integrating Diverse Perspectives in the Research of Child Safety Features in Social VR

Children are now users of online media at a very early age, so the importance of involving children in online safety research has increased. Giving them a voice and soliciting direct child user feedback allows us to understand children's needs and develop child-centric solutions [216, 217, 218, 219, 220, 221, 222, 110]. The 1989 UN Convention on the Rights of the Child affirms that children have the right to take part in decisions affecting them and to share their opinions [56]. Recent research also increasingly involves both parents and children when developing technologies and online safety solutions, recognising that family-centred design is crucial for supporting parent-child relationships and addressing families' values, needs, and preferences [223, 224]. Parents play a key role in shaping their children's media habits [111], with their own technology use and attitudes directly affecting how children engage with it [111, 225]. However, they can struggle with the unfamiliarity of technology, wanting greater transparency about their children's use of it [115]. Balancing the participation of guardians and children in the research is crucial to develop technological solutions that balance both child and parental requirements [226].

While parents and children offer valuable subjective perspectives, experts contribute their specialised knowledge. They have also played a key role in child online safety research [227], drawing on domain-specific expertise from academia and industry, supported by a strong track record of research and leadership in the field. Professionals bring their expertise and knowledge to offer important insights [228] that parents may not have and influence the acceptance and application of safety measures designed to shield children from harm in virtual settings. A study involving parents, children and industry stakeholders to select applications that are considered "good" or "bad" showed that all stakeholders considered similar topics but with a difference of weights and priorities across stakeholders [228].

Collaboration and integration of those three perspectives from children, guardians, and professional experts are essential for designing effective child safety interventions [228]. Such data and methods "triangulation" approach, i.e., crossing varying stakeholders and methodologies, [229], offers richer insights, helping to combat child harassment in social VR and to create strategies that are both easy for children to understand and use, while also protecting them effectively and considering guardians' requirements.

2.4.2 Conducting CCI Research with Children

Research around emerging technologies like social VR with children brings the need to take into account several methodological and ethical considerations. However, ethics in CCI remains under-defined because of the field's interdisciplinary and innovative nature [230].

Recruiting children for user studies outside school settings is a challenge and requires teacher or parent involvement [231]. Research has explored reasons why families may not participate, showing that these primarily include intrusion and privacy concerns, scheduling and logistical issues, conflicts or mistrust of the researchers [232]. Moreover, they suggest the use of broad networking, using associations, local libraries and mailing lists [231]. Challenges encountered when recruiting children were also due to small age range boundaries, with parents contacting researchers even though their child's age did not fit the target age [231]. Therefore, they suggest a larger age range.

Informed consent by both children and guardians is necessary and should be understandable by children with simple language [233]. They must be able to understand what is expected from them, and it should be voluntary. However, even if the child and the parent have given consent, the researcher must still consider non-verbal reactions and gestures as potentially an opposite view to their participation [234]. Confidentiality, privacy and building trust are also key when involving children in research [234].

In terms of data collection, previous research has examined best practices for conducting ethical studies with children, emphasising the need to account for their well-being and developmental abilities [233]. Building on initial Druin's theory on the role of children in the design of new technology [235], a child can have different roles in the design process, they can be: a user (e.g., children are observed), a tester (e.g., children test prototypes), an informant (e.g., informs the design), a design partner (e.g., children design part of the technology), a co-researcher (e.g., children participate in the analysis of their own practices) or protagonist (e.g., they are encouraged become the main agents of the design process) [236]. Overall, the researcher must consider the psychological factors of conducting research with children. First, it is crucial to engage with children on equal footing and minimise hierarchical dynamics [236]. Second, creating a safe environment is also described as important to promote creative freedom [236].

Within qualitative research, researchers highlight the importance of treating interviews with children as meaningful conversations and ensuring that children are given opportunities to be heard. Co-design and group workshops are also a common way of involving children in research. Drawing activities are used to explore children's

views [237]. However, a limited issue may be that children imitate what other children draw. Design workshops featured in prior work were used to develop mediation tools for 2D social media [238, 239], conversational agents [226, 240], or robots to foster anti-bullying [241]. Workshop activities can include brainstorming via post-its and hand-drawn storyboarding for children [242].

Within quantitative research, various Likert Scales using pictorials have been developed to collect data from children. Specifically, emotions are measured on two dimensions: valence (unpleasantness/pleasantness dimension) and arousal (deactivation/activation dimension), based on Russell's (1980) Circumplex Model [64]. Children can recognise and differentiate emotions from a very early age, as research shows that even infants are capable of distinguishing basic emotional expressions [243, 244, 245]. A well-known pictorial Likert scale, the SAM scale [246], was developed and used to measure emotional valence and arousal. Its non-verbal and low language requirement design has allowed for conducting research with children and the elderly [247]. However, while the SAM scale remains widely used, recent research has identified strategies to improve the comprehension of Likert scales for children. These strategies include incorporating modernised pictorial designs, such as emojis, to aid understanding [248] and applying vibrant and colourful schemes to enhance visual engagement [249]. Research also emphasises the importance of avoiding abstract concepts, such as the neutral midpoint on a scale, which can be difficult for children to interpret [250]. Additionally, limiting the number of scale points can prevent cognitive overload and improve response accuracy [251]. Overall, scales should use concrete, easily understood visual cues that clearly represent varying degrees of emotion or intensity, reducing confusion and improving response accuracy.

2.4.3 Designing and Conducting VR Studies with Children on Sensitive Topics such as Harassment

In this PhD thesis, I address two central ethical questions when conducting user studies: how to safely conduct VR-based user studies with children, and how to ethically design studies involving children on the topic of harassment.

2.4.3.1 Conducting VR-based User Studies with Children

As previously mentioned, most user studies involving children with VR have been conducted in the fields of psychology, education and healthcare, with a focus on short-term applications [178, 179, 180, 182, 183]. There is limited research on the long-term physical, cognitive and psychosocial effects of exposing children to VR [252]. Reported short-term effects of exposure to VR often include cybersickness, visual disturbances and sleep disorders [253]. Moreover, a study found that immersive VR can disrupt the way 6 to 8-year-olds move with their head and upper body [254]. Therefore, despite the limited evidence, researchers should exercise caution when exposing children to VR. Child usage guidelines recommend that VR sessions should not exceed 15 minutes of continuous use [255, 256].

A 2020 VR-based study conducted with 31 autistic children aged 6 to 16 years old described their ethical procedure to mitigate as much as possible the potential risks of exposing children to VR [182], which I followed in my user studies. They provided seating and regular breaks in exposure (2-3 minutes at a time) [182]. The researcher was always present during VR use and checked for any possible negative effects or problems regularly. The application and virtual environment were carefully selected to ensure pleasant and appropriate for children.

Another VR-based study involved 25 children, 5 to 9 years old, and children were excluded if they had a seizure disorder, epilepsy, or any condition that could increase their risk of dizziness, disorientation, or nausea [257]. I have included this check in the pre-survey forms filled in by parents. The researcher also helped adjust the VR headset so that it was comfortable [257]. Indeed, the headsets may feel heavy [258] and therefore, checking with the child is important. Adjustable straps can be useful to provide more comfort, balancing the headset's weight across the head, rather than concentrating it on the face [259, 260].

2.4.3.2 Designing Ethical Studies with Children on the Topic of Harassment

Most studies that research sensitive topics such as bullying and harassment do not usually conduct user studies, but rather questionnaires, interviews, workshops and participatory designs [242, 29, 110, 222, 261]. Some studies have used virtual scenarios, putting the child as observers or decision-makers of situations and simulations [30], which give students space to express their thoughts and feelings about trauma and bullying [262].

Recently, research started involving the use of serious games for prevention and intervention bullying programs with children [263, 264]. Players are placed in different bullying situations, and then they have to decide how their avatar should act [264]. For example, "FearNot!" is a research-based anti-bullying 2D video game designed to help children develop coping skills for bullying situations. The player observes bullying situations from the perspective of a victimised character and is then asked to advise the character on how to respond [265, 266]. Another study developed "Com@Viver" [264], where 221 adolescents observed 2D virtual scenarios of bullying, and had to choose how to respond as bystanders, advise virtual characters, and see the consequences of their actions. This game can help improve adolescents' cognitive empathy and prosocial behaviour in cyberbullying [264].

To develop ecologically valid social VR simulations that assess interventions from children's perspectives, I needed to explore ways to involve children in scenarios that could evoke emotional reactions similar to those in social VR, while maintaining ethical standards. Unlike the serious games using 2D virtual scenarios described above, immersive VR, with its affordances of presence and embodiment, can blur the lines between the real and virtual worlds. This effect heightens ethical considerations when working with children. To navigate this, I employed both third-party and first-party approaches, drawing on prior work to inform the design while ensuring ethical mitigations (described in the ethical applications approved by the CoSE University committee).

First-Party Perspective A first approach is to put the child participant in a first-party perspective, facing a provocative situation used in a research study [44]. The study investigated the use of interactive VR scenarios to assess aggressive SIP in boys with behavioural problems. Prior SIP assessment methods often rely on asking children to reflect on hypothetical scenarios via videos or vignettes [267], which lack ecological validity and may not fully engage emotions. Researchers developed a virtual classroom environment where 32 boys interacted with virtual peers in various social contexts to examine their SIP and compared the results to a standard vignette-based assessment. They assessed six scenarios developed to assess SIP patterns underlying both reactive aggression (triggered by provocation) and proactive aggression (aimed at instrumental gain). The sequence of scenarios included an initial practice scenario (for familiarisation without virtual peers) and a neutral scenario (involving small talk). To assess proactive aggression, two instrumental gain scenarios were used: "Object Acquisition", where the participant could steal a block or ball for additional points, and "Competition", where the participant could win by sabotaging the virtual peer's game. The reactive aggression context included two provocation scenarios: "Object Provocation", such as the virtual peer ruining the participant's game (e.g., bumping over a block tower), and "Social Provocation", where the participant was refused entry into a game by two virtual peers. Results showed that VR scenarios involving peer provocation led to more aggressive responses than neutral and instrumental gain scenarios.

Third-Party Perspective A second approach is to put the child participant in a third-party perspective while still asking them to imagine themselves as a first-party victim. This is similar to the serious games to teach children empathy and prosocial behaviours [264, 266, 265, 263, 268]. For example, with *"Fearnot!"* [268], children are placed in a virtual simulation of a primary school where bullying scenes unfold, and they are asked to give coping advice to the victim. It not only teaches them coping skills and empathy, but also encourages them to become defenders [269]. This approach draws on research showing that empathy, defined as "emotional responses aligned with the emotions of others", emerges in infancy [270, 271], suggesting children can emotionally engage with the victim's perspective even as third parties.

2.5 Gaps in Prior Work

Although there has been significant research on digital media and parenting, and cyberbullying moderation for social media, there is a lack of investigation into child safety in social VR. The existing effective interventions for digital media may not be directly applicable to social VR due to its affordances, full immersion and embodiment. It is therefore important to explore the intervention space, including existing moderation tools, as well as new tools that can effectively safeguard children in social VR.

2.5.1 Involving Different Perspectives to Investigate Child Safety in Social VR

Prior work has shown the importance of integrating various perspectives into investigating safety solutions for children (section 2.4.1). The integration of perspectives of children, parents and experts regarding children's usage of social VR and interventions to protect children in social virtual environments is lacking and can provide essential feedback and requirements to design effective safety solutions for children.

While participatory observations [23] improved the understanding of social VR interactions between adults and minors, **study #1** extends this by directly asking parents and non-parent adults about their feelings about interactions with children in Social VR, actions taken afterwards and the consequences on their children or their own use, as well as asking them to share notable first-hand experiences of encounters with children in social VR, interventions they currently use, and interventions they aspire to have. Diverging from prior studies focusing heavily on children or adult social VR users [23, 272, 193], in **study #2**, I explore safety-enhancing technologies for child users of social VR from expert viewpoints as experts insights can offer objective, evidence-based approaches, leveraging their knowledge for effective safeguards for children (section 2.4.1).

2.5.2 Child-Centric Solution and Evaluation Methodology

As described in section 2.3.4, existing safety features in social VR have limitations, especially for children, and there is a need to develop child-centric solutions as the participation of children in social VR is on the rise. And while more recent studies have explored tools through surveys, observations and interviews [50, 9, 207], first, there is a lack of empirical user studies that examine how children react to intervention tools in harassment situations. Second, there has been no investigation into the role of distancing vantage points (e.g, "StayAway" tool), particularly with children and under controlled experimental conditions. Third, research has overlooked the concept of "group" in the context of harassment in social VR, potentially limiting the effectiveness of interventions. This is potentially due to the challenges and ethical requirements of conducting research with children as described in section 2.4.2.

Studies #3, #4, #7 and #8 aimed to investigate children's perceptions of common or novel solutions in empirical user studies. To evaluate interventions, I either used first-party perspective (based on [44] in a one-to-one provocative situations as it was shown that the game was challenging enough but not too difficult for children aged 8-13, and demonstrated a provocative situation designed to provoke an emotional response) (**study #4**) or third-party perspective (in **studies #3 and #7**) in a simulated group harassment. I based the interventions in study #4 on the school punishments rank [61] described in section 2.2.2.2.

2.5.3 Novel Solutions: Designing for Safety AI and the Affordances of VR

AI holds promise in addressing moderation challenges on social media (section 2.2.3.2). The research also underscores the importance of embodiment in social interactions and VR, particularly for children (section 2.3) and the influence an authority figure can have on children (section 2.2.2.2), highlighting its potential to enhance automated moderation actions with a visible embodied character. Moreover, as criminology capable guardianship theory showed, the sole presence of a guardian can be an effective way of deterring offences (section 2.2.5.1). Given the possibility of implementing automated embodied moderation in social VR, my proposed approach is to introduce an Automated Embodied Moderator (AEM) that can help reduce harassment towards children. This underscores the need for insights into designing and implementing automated moderation systems that leverage embodied experiences in virtual spaces. Such systems could be a promising child-friendly solution that would require further investigation. This solution is explored involving all three stakeholders in workshops (**study #6**), interviews (**study #5**) and empirical user studies (**studies #4 and #7**) (section 2.4.1).

Parents are usually at the core of a child's engagement with technology, but can be unfamiliar and unaware of social VR. They noted that existing safeguarding ways are not flexible for parental monitoring and control, and while there is VR casting, constantly monitoring poses challenges. Moreover, parents expressed a desire for non-intrusive parental monitoring tools for child VR activities without needing to wear the headset themselves. Involving parents in the solutions for parental mediation while also involving the child is crucial to developing family-centred designs (**study #8**). I developed these solutions while considering children's rights (section 2.2.5.2) and parental challenges described in section 2.2.4.2.

PART II
The Problem and Moderation Space

Chapter 3

Issues and Concerns Associated with Child Use of Social Virtual Reality

"With the added obfuscation a headset provides in seeing these age-inappropriate situations, it is very easy for a child to get into them without parental knowledge." - [Study #1 Non-parent Adult Social VR user (P13)]

3.1 Chapter Introduction

At the start of my PhD in 2021, research on children's use of VR was mainly focused on the effectiveness of VR in education, therapy and medical settings [179, 273, 274, 275, 276, 277, 253]. Most VR headset guidelines recommend a minimum user age of 13, lowered to 10 by Meta in June 2023 [189, 17]. Similar age restrictions (13+) apply to social VR platforms [186, 187], although they were initially designed for adults (e.g., Horizon Worlds was originally for 18+ in 2021 [184]). Social VR platforms have increasingly been adopted by teenagers [21], with recent exceptions allowing children aged 10–12 to participate through junior, parent-managed accounts [185, 188]. In 2020, Maloney and his colleagues explored children's social VR use, not only presenting opportunities for this disruptive technology but also raising challenges [23, 22, 21]. Further events and studies have shown the emergence of issues and concerns, notably, harassment in social VR [31, 20, 21]. This chapter lays out issues and concerns associated with child use of social VR, with an initial questionnaire with both parent and non-parent social VR user perspectives.

3.1.1 Motivation

Parents are the legal guardians of children who may have a different view of their own children's social VR usage. Non-parent adults who use social VR are likely to have relevant lived encounters with children, as these platforms are used by both minors and adults [23, 22, 21]. By examining these experiences through the lenses of the main adult stakeholders within the child social VR ecosystem figure 1.2, this chapter aims to gain insights into the dynamics and considerations that shape the child's social VR interactions. This includes encounters with children, as well as experiences of one's children in social VR.

This chapter aims to answer the following research question:

RQ1

What are the issues and concerns associated with children's use of social VR?

Addressing this question paves the way to solutions and guidelines for improving children's safety in social VR, considering the issues and concerns first.

3.1.2 Overview of Study #1

While prior work studied why teenagers use social VR [21] and how children interact with other children and with adults in social VR [23], there was limited understanding of how parents and non-parent adults within the child social VR ecosystem perceived the appropriateness of social VR for different age groups, the issues encountered with children, concerns, and safeguarding practices. In this study, I aimed to investigate the following research questions:

S1-RQ1 How often do children use social VR based on parental estimates, and what ages are perceived as appropriate for its use?

S1-RQ2 What negative experiences and perceptions do parents and non-parent adults have of social VR involving minors?

S1-RQ3 What kind of interventions or oversight do parents and non-parent adults deem necessary to protect children in social VR, and do parents note particular interventions that are more or less appropriate for differing age groups?

Methodology This study presents empirical results of a mixed-methods questionnaire (149 responses) in which parents (N=79) and non-parent adults (N=70) reflected on their experiences with children in social VR. Both perspectives are important as they contribute to a new understanding of the experiences and issues of younger users in social VR. Participants reported on notable experiences they have had in social VR that involved children, the estimated age of the children they encountered in said experiences, and the participant's feelings and actions taken after the experience. Par-

ents shared the estimated frequency of their children's usage of social VR, and current safeguarding practices and parental control tools they use to protect their children. Parents and non-parent adults also reflected on which ages and on what interventions they perceive to be appropriate for social VR.

Key Findings

- According to parents in the sample, 43% of children less than 13 years old use social VR at least every two weeks.
- In the data, most adults were concerned about the lack of supervision and safeguarding tools and the freedom users have in these environments, with mentions of how parents fail to meet age limit recommendations.
- Perceptions of age appropriateness are influenced by parenthood, familiarity with social VR and supervision in social VR. For example, parents familiar with social VR are significantly more likely to let younger children use social VR without supervision compared to non-parents who are unfamiliar with social VR.
- Themes obtained from the thematic coding were similar to those obtained in Maloney's and his colleagues' work [23], including the following: "dealing with harassment and bullying" (in minor-minor interactions), "social distancing from minors" and "adults discussing inappropriate topics for children" (in adult-minor interactions). Findings depict concrete experiences affecting adults or affecting children, to perceptions and concerns towards adults or children.
- Parents seek to apply a breadth of safeguarding practices, restrictions and interventions to protect their children, with little adaptation based on age. This suggests that parents need support in providing interventions and guidance around what interventions are suitable for a given developmental age, lest social development be impacted by overly restrictive measures.

Taken together, findings expand and reveal new insights into the extent to which minors use social VR; how in the sample parenthood, familiarity and supervision influence perceptions of children's use; and how parents would choose to moderate otherwise limit their child's usage of social VR across their childhood (in particular findings 1, 2, 6, 7, 8). Negative experiences with minors in social VR, perceptions and concerns described by parents and non-parent adults confirm results in prior work (findings 3, 4) and add knowledge regarding parental views to the field of social VR (finding 5).

3.2 Study #1 - Through the Eyes of Social VR-Aware Parents & Adult Users: Issues and Concerns Associated with Child Use of Social Virtual Reality

3.2.1 Related Work

This section synthesises related work on the specific topics addressed in this study. (A more comprehensive literature review is presented in chapter 2, with specific relevant sections: section 2.3.2, section 2.3.3 and section 2.2.4).

3.2.1.1 Adult and Minors Interactions in Social VR

As social VR has attracted various age groups, questions about social dynamics are raised. Maloney et al. (2020) [23] used a participatory observation approach to gather 80+ hours of data from mainstream social VR platforms (Rec Room [27], VRChat [25]). This allowed investigating the perspectives and interactions among minors and between minors and adults, observing nuanced interactions, positive and negative. From minor-minor interactions, themes captured were: 1) virtual intimacy and emotional connections via virtually-embodied non-verbal communication, trust and teamwork; 2) social interactions beyond gameplay, where older teenagers seemed to enjoy interacting more than playing with each other compared to younger teenagers; 3) sharing and disclosing in groups; and 4) dealing with harassment and bullying. Looking at adult-to-minor interactions, researchers noted mixed sentiments about each other, with the main themes being: 1) barriers, tensions and frustrations where adults would be 'annoyed' or 'frustrated' by minors' behaviours seeking attention; 2) mutual learning (relationship building, cultural learning); 3) social distancing from minors; and 4) adults discussing content inappropriate for children [23]. While this study improved the understanding of social VR interactions between adults and minors, I extend their work significantly by collecting data directly from users who had a first-hand experience of interaction in social VR with minors, asking users about their feelings, actions taken afterwards and how it may have led to instigating new rules on their children or their own use. I also provide parents' perspectives and new insight into potential interventions for target age groups.

Social VR is a great opportunity to engage and interact with others innovatively, but it also creates problematic situations and encounters, especially with the presence of minors unsupervised, and leads to new forms of abuse, harm and harassment. A study that included 30 semi-structured interviews with adults [22], focused on their social experiences in social VR, without directly asking them about interactions with children or adolescents. The main themes in the data included experiences, tensions and frustrations when encountering minors. For example, the prevalence of immature behaviours such as children screaming, negative social environments where young people are exposed to inappropriate behaviours and children's excessive use of social VR. Adults also expressed safety concerns for younger users, including the exposure of children to negative social environments and the excessive use of social VR, and worries about health and academic performance. In contrast to that work, my questionnaire focused exclusively on experiences in social VR that involve children.

Several recent events and studies have shown the emergence of harassment in social VR [31, 20], as described in section 2.3.3.

I therefore focus on problematic events and negative situations in social VR to better understand the concerns and needs from a larger and diverse sample of adults' and parents' perspectives to create a safer social VR environment for children.

3.2.1.2 Parenting and Technology

Prior work has investigated parents' perspectives, concerns, roles and influence on children's digital media (see section 2.2.4).

3.2.1.3 Gaps

Existing research describes reasons why teenagers use social VR [21] but does not quantify the number of minors and the frequency of use. While participatory observations [23] improved the understanding of social VR interactions between adults and minors, my work extends this by directly asking parents and non-parent adults about their feelings about interactions with children in Social VR, actions taken afterwards and the consequences on their children or their own use, as well as asking them to share notable first-hand experiences of encounters with children in social VR, interventions they currently use, and interventions they aspire to have.

3.2.2 Methodology

The questionnaire was distributed from May to July 2022. It gathered **149 participants**.

3.2.2.1 Study Design

The data was collected via an online mixed-methods questionnaire, as this allowed me to reach out to a wide number of social VR users and get both quantitative and qualitative data. Employing the critical incident technique [278], I collected anonymous stories of social VR users aged 18 or older. Similar to recent studies on social media [279] or educational VR [180], the approach entailed collecting data, currently lacking in the literature, from parents about their children's use of social VR, safeguarding practices and concerns.

3.2.2.2 Questionnaire Structure

The questionnaire was designed via Qualtrics [280]. It was divided into five blocks with a mix of open-ended questions and non-open-ended questions (sliders and Likert scales) figure 3.1. Some questions were shown only to participants who indicated that they are parents of children. The full questionnaire can be found in the Supplementary Material (STUDY 1/QUESTIONNAIRE).

Participants first completed standard demographic questions and were asked to indicate their familiarity with VR and social VR using a 6-point Likert scale (Block 0). In Block 1, participants were asked whether they were guardians or parents of children who use VR or social VR, and questions about the nature of children's use of social VR. In Block 2, I asked participants to write about a notable story that happened, without specifying in the question the focus on the negative aspects. Only after they answered the question were they able to express how they felt and categorise the events or perceptions into negative, neutral or positive. Block 3 focused on the participants' perceptions of appropriate minimum ages for social VR (3.1), and of safeguarding

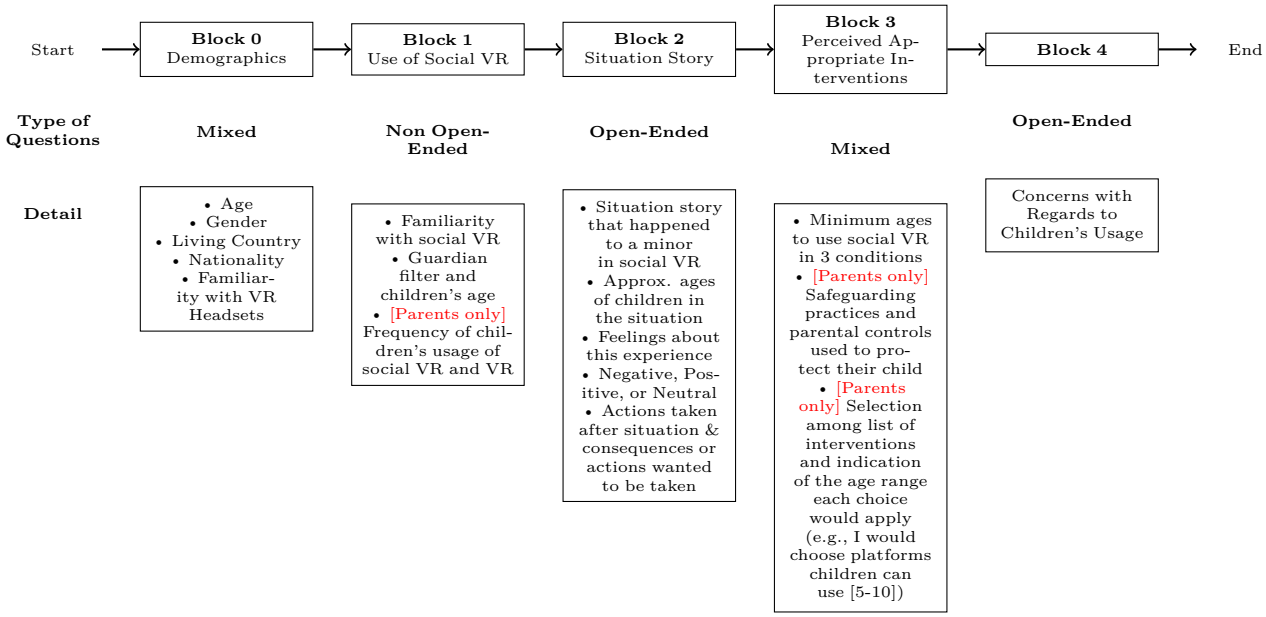


Figure 3.1: Questionnaire Blocks and Question Types

practices they currently use (3.2), and would seek to use (3.3). The interventions list proposed in Block 3.3 was under four categories: pre-planned controls (e.g., education, restrictions and negotiation of rules); knowledge of what happens as it happens (e.g., monitoring, receiving notifications); real-time interventions (e.g., remote actions, automated guardian); and after-the-fact (e.g., checking history). This non-exhaustive list of interventions was compiled based on existing approaches described in section 2.2.3.2, existing tools in section 2.3.4 and supplemented with original ideas from brainstorming for each category (see the full list in the Appendices section A.1). The aim was to integrate a range of potential interventions that could be implemented before, during, and after an incident, while grounding these in Livingstone's (2008) parental mediation model [112], which includes restrictive mediation, active mediation, parental monitoring, and technical mediation, and extending it to encompass potential AI-driven and automated safety mechanisms that could be relevant to social VR. Participants were shown the interventions and were asked to indicate the age range for which they would consider each appropriate. Text boxes for age ranges were left blank if they would not use an intervention.

3.2.2.3 Analysis

Quantitative I analysed the quantitative data using R statistical tools (Blocks 0, 1, 3.1, 3.3). The markdown with the code can be found in the Supplementary Material (STUDY 1/Analysis). The quantitative analysis involved data wrangling, summarising and visualising, and computing statistical tests. After verifying the assumptions of normality, homogeneity and the presence of extreme outliers to conduct a three-way ANOVA, the data failed the normality test (Shapiro-Wilk test). I therefore conducted a three-way Aligned Rank Transform (ART) ANOVA [281], a non-parametric statistical method that allows for factorial analysis, comparable to traditional ANOVA, even when key parametric assumptions, such as normality, are not met.

Qualitative I conducted the qualitative analysis with NVivo (Blocks 2, 3.2, 4). Participant responses regarding the situation story, concerns and safeguarding practices were coded using inductive thematic analysis [282]. This process consisted of the following four phases: data familiarisation, systematic data coding, generating initial themes and developing and reviewing themes. The first phase involved a thorough reading of the collected data. In the second phase, I employed inductive coding formation, with a single word as the coding unit and considered any new relevant detail as a new code. In phase three, the generation of initial recurring themes involved creating a spreadsheet and identifying initial topics (see Supplementary Material STUDY 1/Analysis). Phase four included three additional researchers who helped refine the themes through multiple discussions over eight weekly meetings.

While I indicate the count of participants who raised a particular theme, this is provided solely to offer contextual information. The focus is on the "keyness" of themes, i.e., defined by their relevance and richness in relation to the research questions, rather than by how frequently they were mentioned [283]. I value insights that contribute meaningfully to understanding user experience, regardless of how commonly they appear. Therefore, each idea even mentioned by one participant is considered important. **This applies to all studies in this thesis.**

3.2.2.4 Participants

Recruitment The mixed-methods questionnaire was released in May 2022 on multiple forums (Reddit social VR and VR communities: VRGaming, VirtualReality, OculusQuest, VRchat, Oculus, RecRoom, AltspaceVR (currently discontinued), metaverse; Oculus Community; AVForums; Twitter). This allowed me to target communities from a population of interest that has some knowledge of VR and social VR.

Demographics In total, 166 participants took part in the questionnaire, of whom 112 completed all of it. The questions addressed singular research outcomes (i.e., analysis of later questions was not predicated on analysis of prior questions). Therefore, I did not exclude partially completed responses, as many of them were relevant and valid. The following responses, however, were considered invalid: repeated responses in multiple unfinished questionnaires and irrelevant responses (i.e., words or sentences not answering questions). I omitted participants when all response fields were empty, either starting from the demographics section or from block 2. This left me with 149 responses (Mean age (in years) = 32.6, SD age (in years) = 6.9). Among participants, 107 are male, 38 are female, two are non-binary/third gender, and two preferred not to say. 51 fathers, 27 mothers, and one parent who did not disclose their gender, responded to the questionnaire (i.e., 79 parents). Among the 70 non-parent participants, there were 17 non-parent siblings or teachers. 110 are from the Americas (94 from the United States of America (USA), 10 from Canada and seven from South America), 25 from Europe, eight from Asia, three from Oceania, and two from Africa. The participants' age, knowledge and familiarity with general VR, social VR (6-point Likert Scale; 0='not at all'; 5='extremely' familiar) are summarised in table 3.1. Users reported the social VR platforms where the situation story occurred: 42 used Rec Room, 40 used VR Chat,

34 used AltspaceVR, one used Gorilla Tag, and one used Zepeto. Among parents, 15 parents were familiar with both VR and social VR, and 23 were less familiar with either. Among non-parent adults, 15 were less familiar, and 20 were familiar. Four parents have never used social VR.

Table 3.1: Questionnaire Demographics Table Summary. (N=149, 79 Parents, 70 Non-Parents). The questionnaire was mainly released on forums about VR/social VR; there may be self-selection bias, in which more engaged parents and non-parent adults in VR/social VR may have been more likely to participate in this study. However, this enabled me to get responses from users familiar with VR/social VR.

Groups		**Parents**	**Non-parent Adults**
N		79	70
Age (in years)	MEAN [SD]	32.59 [6.85]	26.64 [8.20]
VR headsets Familiarity 0-5 Likert scale ('not at all' to 'extremely')	MEAN [SD]	4.05 [0.80]	4.06 [0.88]
Social VR familiarity 0-5 Likert scale ('not at all' to 'extremely')	MEAN [SD]	3.88 [0.83]	4.19 [0.90]
Children Age (in years)	MEAN [SD]	9.69 [4.39]	NA

Ethics and Compensation Participants could optionally sign up for a lottery for one of two online shop vouchers. The study was approved by the university CoSE ethics committee (Approval number: [300210173]). Documents with the information sheet, consent and demographics questions (Block 0 of the questionnaire) can be found in the Supplementary Material.

3.2.2.5 Limitations

Sample Diversity The sample was not diverse in a number of aspects. Despite broad online recruitment on international websites, participants were mainly from the USA, and two-thirds were Male, only 38 were from Africa, Europe, Asia and Oceania.

Language Accessibility The questionnaire was released in one language only (English) on forums that may not have allowed a more global reach. However, the sample reflects the demographics of VR.

Novice Experience Lacking The participants are mostly engaged and familiar with social VR as table 3.1 shows. While this allowed me to gather insights stemming from actual lived experiences, it also means that I do not capture novice experiences, e.g., an experience of a user who tried social VR once and decided not to use it again or concerned parents with limited knowledge of social VR. This is important to address in future work, particularly to capture whether certain issues and concerns are amplified in these situations.

Sample Characteristics and Questionnaire Scope Approximately 50% of the sample is a parent, allowing me to gather valuable insight for the purpose of this study. I also elected to seek stories from participants aged 18 or older for ethical reasons. While I focus on negative experiences and perceptions, I did not directly ask participants to tell me about negative experiences but rather a "notable" situation, agnostic of valence. Accordingly, I did not explicitly ask about the benefits of social VR but rather left it open to respondents as to whether they would choose to report positive or negative experiences. Participants were asked to describe situations where they encountered or interacted with minors in social VR and to estimate their perceived ages. Consequently, the findings I report may on occasion be reflective of witnessed adult behaviours presented with a degree of immaturity. Subjective perceptions of age are nonetheless important and the best estimate available, as respondents would rely on a variety of contextual cues (context, actions) for the age approximation.

Qualitative Data Analysis Process Moreover, the initial coding of the qualitative data was conducted by one researcher, and then the resultant codes and themes were discussed and refined by a team of three researchers.

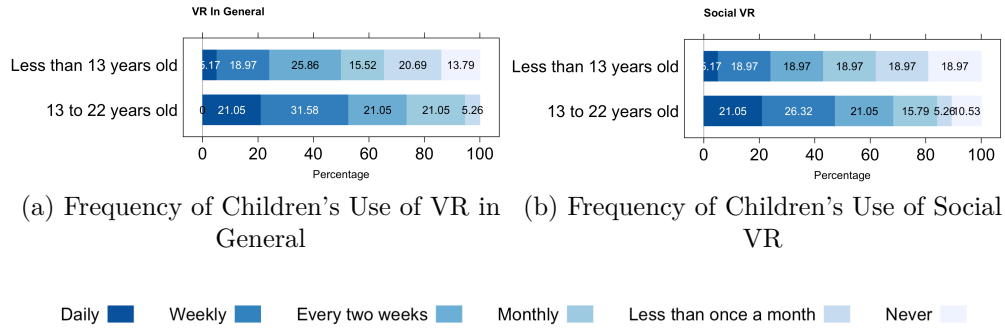
Future work would require reaching out to a more global and diverse audience as VR continues to see increasing adoption outside of Western demographics, and collecting children's perspectives on their social VR use and the interventions.

3.2.3 Results

3.2.3.1 Parent-Reported Children's Usage of Social VR and Appropriate Minimum Ages for Social VR from Parent and Non-Parent Adult Perspectives

Frequency of Children's Social VR Usage and VR in General Parents reported their children's frequency of social VR use. I divided the sample into two age groups: parents of children under 13 years old (N=58) and between 13 (inclusive) and 22 years old (N=19), and observed the frequency of use among participants' children per age group. I chose these two groups based on current recommendations of VR and social VR platforms use, many of which require users to be 13 or older [17, 284, 189, 186]. Among children of participants who are less than 13 years old (N=58), 43% use social VR at least every two weeks, whereas 50% use VR in general at least every two weeks. figure 3.2 shows the frequency of VR and social VR use, respectively, in each group.

Most VR headset manufacturers have restricted the age limit to 13 years old+, and most manufacturers warn against its use for children [17, 284, 206]. However, we can clearly observe that many children under 13 use VR and social VR. Social VR platforms (e.g., VRChat, Rec Room) mostly recommend 13-year-olds or above, and some provide junior accounts [186, 187, 185]. Industry mediation with age limits is not sufficient to restrain children's use, as most of the platforms rely on self-professed age [285, 114]. Other types of mediation are therefore needed to safeguard and supervise children.



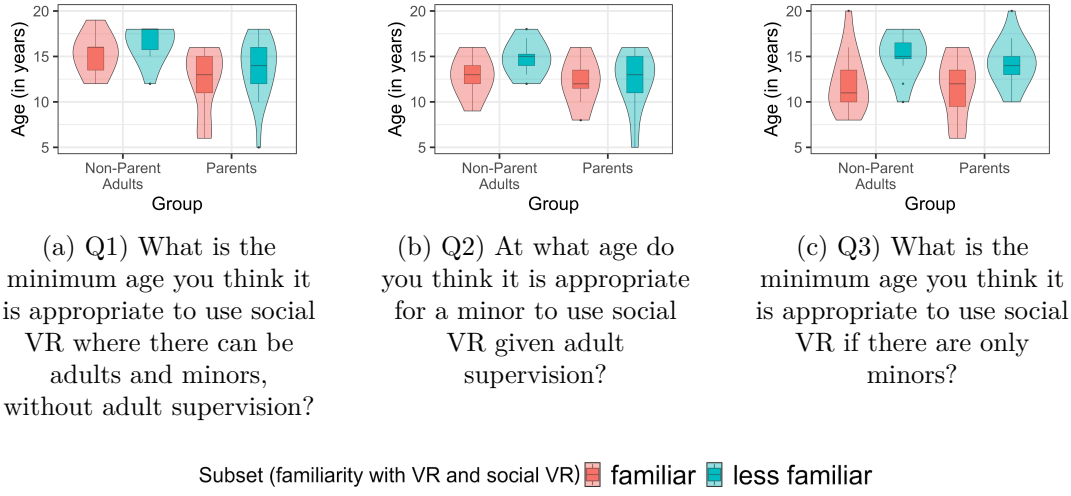
(a) Frequency of Children's Use of VR in General

(b) Frequency of Children's Use of Social VR

Figure 3.2: Likert Scales showing percentages of children using (a) VR in general or (b) Social VR in the sample. Among children under 13 years, 24% use social VR daily or weekly and 24% use VR in general daily or weekly. 43% use social VR at least every two weeks.

Finding 1: Despite age restrictions, some under-13-year-old children (43% in the data) use social VR regularly (at least every two weeks) against the recommendation of manufacturers and developers.

Appropriate Minimum Age in Social VR from Non-Parent Adults' and Parents' Perspectives While there are age recommendations from VR manufacturers and social VR platforms, children increasingly use social VR regardless of the guidelines [21]. I asked participants what the appropriate age is for using social VR when 1) adults can interact with minors with supervision, 2) adults can interact with minors without supervision, and 3) minors can only interact with minors. I grouped results based on parenthood and familiarity with VR/social VR (figure 3.3).



(a) Q1) What is the minimum age you think it is appropriate to use social VR where there can be adults and minors, without adult supervision?

(b) Q2) At what age do you think it is appropriate for a minor to use social VR given adult supervision?

(c) Q3) What is the minimum age you think it is appropriate to use social VR if there are only minors?

Figure 3.3: Violin and boxplots of inputted ages corresponding to the minimum appropriate ages in social VR for three different supervision states (a, b, c), by parent and non-parent adult groups, divided into familiar and non-familiar subsets.

I conducted a three-way ART ANOVA [281] (significance level 0.05) with age as the dependent variable and with parenthood groups (non-parent adult and parent), subsets (familiarity with VR and social VR) and questions (supervision states) as the factors involved. Significant results (significant p-values set at 0.05) with medium ($\approx$ 0.06) and large effect sizes ($\approx$ 0.14 or higher) partial eta squared η_p^2 [286, 287], were obtained for parenthood groups (non-parent and parent) ($F(1,54)=5.60$, $p=0.021<0.05$, $\eta_p^2=.1$, medium effect size), subsets (familiarity with VR and social VR) ($F(1,54)=5.62$, $p=0.022<0.05$, $\eta_p^2=.1$, medium effect size) and questions (supervision states) ($F(2,108)=13.49$, $p=5.9e\text{-}06<0.01$, $\eta_p^2=.2$, large effect size) with interaction effects also noted. In particular, from post-hoc comparisons, (1) parents familiar with VR and social VR are significantly more likely to let younger children use social VR without supervision compared to less familiar non-parent adults ($p=0.015<0.05$). (2) Compared to parents and non-parent adults that are less familiar with VR and social VR, those familiar with them are significantly more likely to let younger children use social VR ($p=0.022<0.05$). (3) Parents and non-parent adults are significantly more likely to let younger children use social VR with supervision ($p<0.0001$) and if there are only minors ($p=0.001<0.01$) compared to those without supervision.

Finding 2: Perceptions of age appropriateness are influenced by parenthood, familiarity with social VR and supervision in social VR. Parents in the sample, in particular those familiar with VR/social VR, have the tendency to choose lower minimum ages for social VR usage compared to non-parent adults less familiar with either. Participants are more likely to let younger children use environments with supervision or where there are only minors, compared to without supervision.

3.2.3.2 Experiences with, and Perceptions Towards, Social VR Children and Other Adults

Participants (N=149) were asked to recall and report instances from the following four types of encounters in social VR: 1) encountering users they suspect are children (60% of N), 2) encountering users that were definitely children (62%), 3) interacting (e.g., speaking) with children on social VR platforms (54%), and 4) encountering children in social VR in areas that were not age-appropriate (40%). Participants could select multiple options. I asked participants to tell us about a notable situation that happened to a minor or minors they knew or encountered in social VR. If the situation was selected as a "Negative situation" (among Negative, Neutral and Positive), they were asked what actions they took after the situation and what actions they would have taken or would have wanted to be taken. I gathered 63 responses: **40** negative, **15** neutral, and **8** positive. Stories are reported in Section section 3.2.3.3.

Participants were also asked to approximate the ages of the children they met or encountered in social VR during the situation they described. Most of the participants estimated the age through the child's voice or height. The mean estimated age was younger than 13 (Mean=11.42, SD=3.33). After coding the described situations, the following two categories emerged: **Experiences with Children**, where participants described a one-off situation, and **Perceptions/Views and Concerns**, where participants expressed a sentiment of use or concerns about the presence of children in social

VR. For both, participants described how adults or children were affected. Themes and counts of experiences and perceptions of said encounters are summarised in table 3.2. This model may allow us to better design interventions for different subjects and consider different perspectives, yet to be tested.

Table 3.2: Four-Quadrant Matrix presenting themes from social VR adults' and children's experiences and perceptions from non-parent adults' and parents' perspectives.

	A. Experiences with Children		B. Perceptions and Views	
		Counts		Counts
Children affected by Adults	A1 • Age-inappropriate behaviour that children encounter	18	B1 • Lack of Regulations, Safety and Supervision	10
	• Children sharing with adults	3	• Need of new safeguarding rules	2
			• Impact of VR and social VR	3
Children affected by Children	A2 • Children experiencing and dealing with abuse	4	*Not addressed in this survey*	
		Counts		Counts
Adults affected by Children	A3 • Adults dealing with children's immaturity	17	B2 • Disruption to adults' environment	3
	• Adults harassed by children	3	• Emotions: Positive **(n=3),** Negative **(n=17),** Neutral **(n=2)**	22

3.2.3.3 Notable Experiences Encountering Children in Social VR

Themes around experiences with children raised by both parents and non-parent adults are depicted in this section. They are generated from answers involving children as the main subjects, affecting children or adults. The count of responses on experiences categorised into themes is given in table 3.2.

Experiences that Affect Children in Social VR (A1, A2) Responses in this category were grouped into three themes: children encountering age-inappropriate behaviours, children facing abuse, and children self-disclosing to adults.

Age Inappropriate Behaviours That Children Encounter

Participants described situations in social VR where children were exposed to inappropriate content and conversations (e.g., sexual graphic content, erotic play, conversations that include drugs, sex and alcohol), forms of oppression (e.g., violence, racism, homophobia and verbal abuse) or had easy access to age-inappropriate environments/areas (e.g., VR nightclubs). The following is an example of inappropriate situations children can be exposed to, and also points to the high likelihood of encountering children in social VR:

> *"It's an extremely common occurrence to see children in these environments exposed to the sexually risqué avatars and actions of others, either directly or indirectly. In many cases, children will either start or join these experiences of their own volition. [...] With the added obfuscation a headset provides in seeing these age-inappropriate situations, it is very easy for a child to get*

> *into them without parental knowledge. As such, you would be hard pressed to *not* find a child in a lobby. Based on personal experience, I have never been in a Social VR lobby without a child present somewhere.“* [non-parent adult] (P13).

Participants also mentioned that children were being treated as adults or that children acted as adults in social VR, for example:

> *“People talking trash at a pool table [...] and one of the players revealing they were only 14. The mood changed, and some were a bit nicer, but the random minor playing had their feelings hurt and was treated like an adult (harshly).“*[parent] (P31)

> *“Someone I knew several months ago (I have them blocked for now until they turn 18) was a minor who was lying about their age and engaging in very much adult things (drinking and sexually charged activity) for several months before they confessed to it. They said they felt trapped and peer pressured into acting older after people initially assumed that they were of age.“* [non-parent adult] (P35)

Children Experiencing and Dealing with Abuse

Children would experience forms of abuse in social VR, similar to those in the real world, such as bullying by other children or verbal abuse from teenagers and young adults. Some children may ignore them, but others may be affected emotionally and physically or may want to act the same way, e.g.:

> *“My son frequently interacts with other people in social games who sound like late teens or young adults who use graphic sexual and homophobic language in “trolling”. They’re annoying, and he’s good at ignoring them. It reduces his enjoyment of the games, but it’s hard to see how to stop this. [...] He usually plays where I can hear him, and I talk about any abusive players and how to handle it.“* [parent] (P6)

Children Sharing with Adults In social VR, children may encounter adults who would be friendly and supportive of what they are going through or wish to share. This may be beneficial for children who may not want to share with someone they personally know. While some adults might support children by only listening to them, children may not be cautious and disclose personal information to malicious adults. This result is in line with current findings showing the prevalence of self-disclosure in social VR [192].

> *“I have also had conversations with a few kids about how they are happy that I talk to them [...]. They feel heard when people just listen to what they have to say. There was one kid who was going through a time when his parents were divorcing, and this was a way for him to get away from that. He talked about crying some nights, wishing it had never happened. [...]“* [non-parent adult] (P12)

Experiences that Affect Adults in Social VR (A3) Experiences affecting adults were grouped into two themes: adults dealing with children's immaturity and adults being harassed by children.

Adults Dealing with Children's Immaturity Adults face children's immaturity in social VR environments initially designed for adults and young adults. Adults typically complained about children's behaviour in social VR. Children create disruptive environments in *"mature crowd in public lobbies"* [parent] (P1). They like to be inappropriate, and children use these environments as a way to push boundaries:

> *"It feels like an outlet for children to misbehave or test the boundaries without facing real consequences."* [non-parent adult] (P61)

Children shout, yell and run around (e.g., a child started crying when a [non-parent adult] participant (P18) won a game). Participants mentioned other adults encouraging children to misbehave and adults shaming children. However, adults typically attempted to mitigate these disruptions. They would deal with these disruptions via: social distancing (adults leaving the VR rooms when in the presence of children or actively avoiding children and not engaging with them), blocking children and adults from having to turn down the volume and muting others. For example:

> *"One yelled/screeched at me for spawning next to them. It was a little unpleasant. [...] I turned my volume all the way down because I'm not using headphones, and my housemates can hear them. Once the sound was off, I navigated through the menu to leave the room."* [non-parent adult] (P56)

Among answers dealing with disruptions, adults also mentioned paying attention to their behaviour by watching their language in front of children and trying to be nice.

Adults Harassed by Children Adults may also be verbally harassed by children, for example:

> *"One example of many is seeing a sexually risqué avatar controlled by an adult being verbally harassed by children."* [non-parent adult] (P13)

Finding 3: 1) Both parent and non-parent adult social VR users in the sample have to deal with abuse and immaturity by children towards them. 2) Children are often exposed to age-inappropriate behaviours and activities, they face abuse, and sometimes use Social VR to seek moral support and self-disclose to adults.

3.2.3.4 Parent and Non-Parent Adult Perceptions about Youth in Social VR

Non-Parent Adult and Parent Perceptions Towards Children in Social VR (B1) Participants shared their views about children being present in social VR and raised their concerns and worries towards children (N=18). In particular, participants reported a lack of regulation and supervision in social VR, and how children may not know how to protect themselves and navigate safely in social VR. Responses also

described the effects and impacts of VR and social VR on children; these were subsequently grouped into negative and positive. Negative impacts included physical (e.g., eyestrain) and mental health impacts (e.g., mental development, trigger of negative emotions, negative influence). Examples of negative impacts that are unique to social VR include adopting or learning poor behaviour and becoming out of touch with social reality. Positive effects included the possibility of children having fun and being fully immersed, and teenagers willing to learn, socialise and introduce social VR to their friends.

> *"Concerned that a younger user might stumble into something inappropriate for them without realising it or meaning to. [...] Reminded me that many children play the game and makes me hope that more of them have good supervision/are willing to leave when people encourage them in case things get weird."* [non-parent adult] (P34)

> *"Minors [under] 13 years of age should [be] under adult supervision and must ensure intermittent use and rest. Long-term use should be avoided, as possible negative effects include: imbalance of hand-eye coordination, affecting the brain's ability to multitask, and adults should closely monitor the use of this product by minors."* [parent] (P64)

However, parents felt social VR could be comparable to real-world situations where parents have to trust their children in situations that may not always have supervision, for example:

> *"I think this constitutes standard exploratory behaviour by teenagers of adult environments, and with the added barrier of pseudonymity and non-physical interaction from VR, it is a less dangerous or simply real situation than e.g. someone in their mid-teens going to a house party."* [parent] (P65)

Non-Parent Adult and Parent Perceptions Towards Adults in Social VR (B2) While some responses were about how social VR can affect children as described above, other responses noted how the presence of children in these environments affected adults (N=6) and how children would need supervision (N=12). Adults face discomfort, annoyance, verbal abuse and harassment from children:

> *"I find these experiences deeply disturbing as I don't believe these children should have access to a VR headset without strict supervision. The sheer amount of swearing, insults and racial slurs is astonishing, both directed at and from these children. As my own children grow up, I will not be allowing them access to a VR headset without strict supervision. These games allow unfiltered contact between children and adults, which leads to inappropriate communication."* [parent] (P3)

> *"You see and hear children on platforms, such as VRChat, where the minimum age is 13. Why do parents give a young child access to an expensive piece of technology and put them on a platform with an age requirement that they fail to meet?"* [non-parent adult] (P51)

Based on Russell's Circumplex model polarities [64], I categorised emotions felt by users. For instance, negative emotions would include anger, frustration and stress, and positive emotions would include happiness, serenity and relaxation. Perceptions were mostly negative (17 negative emotions, 2 neutral emotions and 3 positive emotions). The analysis indicates that the feelings of users who noticed being observed were negative in the majority of cases (e.g., 'worry, concerned' (P36, P43, P75, P101, P139), 'bothered' (P3), 'disturbing' (P5), 'inappropriate' (P5, P36), 'annoyed' (P75), 'angry' (P87), 'embarrassed' (P58), 'isolated' (P99), 'dependent' (P99), 'felt bad' (P33, P98), 'sad' (P34, P103), 'surprised, shocked' (P118, P138). Some participants felt both 'happy and sad' (P14, P18), 'normal' (P26) or 'neutral' (P67). P104 felt 'happy', and P107 felt 'good'.

> *"I felt bad for the kid, but rejection is part of hanging out with new people. I think lobbies should be sorted out by relative player age so kids don't have to experience things they aren't ready for."* [non-parent adult] (P31)

Finding 4: Non-parent adults and parents in the sample typically perceive the presence of children in social VR as negative, affecting both children and adults negatively. They cite the lack of safeguarding and regulation in social VR spaces.

3.2.3.5 Parent and Non-Parent Adult Concerns about Children's Social VR Use

Both parents and non-parent adults are concerned about the freedom children have in these virtual environments where there can be malicious people (e.g., paedophiles) (N=8) and the lack of supervision, regulations and monitoring that can lead to several consequences: psychological (N=28) (e.g., addiction, mental health issues, skewed self-image, decreased real-world social awareness and learning skills) and physical (N=12) (e.g., eye-strain, impact on the body, anxiety). Parents expressed their concerns (N=33) with a particular focus on nurturing their parent-child relationship and wanting to build trust, for example:

> *"I'm no more concerned about VR than the internet in general. Maybe even less so, as almost all use in my house happens in front of me (due to space constraints, not a rule). Cultivating trust with your child and speaking openly about the rewards and dangers of online interaction is more important than surveillance (which can backfire and erode trust)."* [parent] (P6)

Among non-parent adults' concerns (N=34), a common concern referred to parents' lack of awareness regarding all the possibilities of interactions (both positive and negative) and the extent to which children are taking advantage of this freedom in social VR:

> *"[...] I think social VR can be a wonderful tool and a lot of fun, even for younger users, but I think that there are not really a lot of safeguards available for them at the moment on most platforms."* [non-parent adult] (P34)

> *"I am worried about the difference in knowledge between children and their parents about what occurs on VR platforms. It seems like there is a lot more freedom, and a lot more possibilities of interaction (both positive and negative) than parents believe."* [non-parent adult] (P166)

Finding 5: 1) Parents in the sample prioritise trust and conversations, whereas non-parent adults in the sample prefer enforcing rules. 2) Major concerns regarding children's use of social VR revolve around the issue of excessive freedom and psychological negative impacts (e.g., addiction, worries about children losing real-world social skills and learning abilities).

3.2.3.6 Safeguarding Online: Practices and Tools

Actions Taken, Consequences or Actions Wanted to Be Taken I investigated the actions participants took after the event they encountered and what actions they would have liked to be taken in similar events. Actions taken mostly involved active communication, explanation and guidance. Both non-parent adult and parent social VR users explained to children involved in the situations what went wrong soon after the incidents. Parents mentioned limiting or reducing their child's social VR usage and communicated about the impact of these inappropriate behaviours and disruptive situations before incidents. I also asked their views on what should be done and what they would have wanted to be done. One of the most wanted actions from non-parent adults is age verification, age restriction and limitations such as allowing minors to log in only the weekends. They also want improved moderation, education and suggested that developers should create safe places and protection and supervision systems. They expect parents to better moderate their children's social VR use and do not place the responsibility on the children. Parents would like strong authentications and time limits. A parent (P39) mentioned that children should not be allowed somewhere unattended, at least until they know how they process the information and know how to stop. Parents also expect more guidance and supervision of children in social VR.

> *"Minors should be accompanied by guardians."* [non-parent adult] (P132)

> *"Developers, publishers, and platforms should find ways to enforce their terms of service and make sure that safe places are safe, and open places are open. I do not blame the kids."* [non-parent adult] (P61)

> *"Parents are always monitoring their children. Before the children have no ability to distinguish themselves, parents still play the role of enlightenment teachers. In view of current VR, it is suggested that parents choose VR equipment and content reasonably. [...]"* [parent] (P163)

Regarding consequences, parents have restricted and limited children's social VR use, stopped children from using VR completely, watched and monitored:

> *"It used to be five times a week, but now it's once a week."* [parent] (P45)

Finding 6: 1) Parents in the sample want robust authentication and time limits. They stress that children should not be unsupervised. More guidance and supervision for children on these platforms is expected. 2) Non-parent adults in the sample strongly want age verification, age restrictions and improved moderation for social VR platforms. They urge developers to create safe spaces with better protection and supervision. They also expect parents to be more proactive in moderating their children's usage.

Parents' Safeguarding Practices I asked parents how they monitor or safeguard their children who use social VR, and what parental or standard privacy controls they currently use for social media, for example. Parents' answers on safeguarding included two main types: real-world safeguarding practices (external to the device, N=28) and virtual-world safeguarding practices (internal, via the device, N=12). The first type (real world) involved communication and education (i.e., education about anti-violence, communication on safety and privacy, rules, parents and guardians participating with children) and physical space and time limits (i.e., children permitted to use VR in the living room only, physical space safety to avoid children to hurt themselves when they use head-worn devices, presence of parents when they use VR, limiting the time of use).

> *"He usually uses it in the living room where I can hear what's going on. We have built trust around his online activities, he knows he can come to me to talk about anything that might bother him."* [parent] (P5)

The second type of safeguarding (VR world) involved: checking the history, watching over (e.g., via streaming), authentication, and downloading suitable applications. The tools and existing controls they use include blocking games/sites/rooms, privacy controls, surveillance, time limit checks, monitoring accounts history, muting and blocking, and mobile phone monitoring, social media privacy controls, and mobile phone to stream VR content.

Finding 7: Parents in the sample use a variety of safeguarding practices, both in the real-world (N=28) to put physical and time boundaries in place, but also through the virtual world (N=12) with existing tools that mostly include the mobile phone to stream VR content or controls for checks and authentication.

3.2.3.7 Parents' Interventions Selection and Corresponding Age Range

In this section, I summarise responses to questions in which parents were asked to 1) select from among suggested interventions they would put in place, and 2) give an age range (from minimum to maximum age) for which they think the selected intervention is adapted. I argue this list is an interesting slice of the intervention space, but may need to be refined in future work. The interventions were grouped into four categories: pre-planned (e.g., choosing the content and platforms), knowledge (e.g., monitoring via notifications received), real-time actions (e.g., blocking remotely) and after-the-fact (e.g., reviewing the history (figure 3.4).

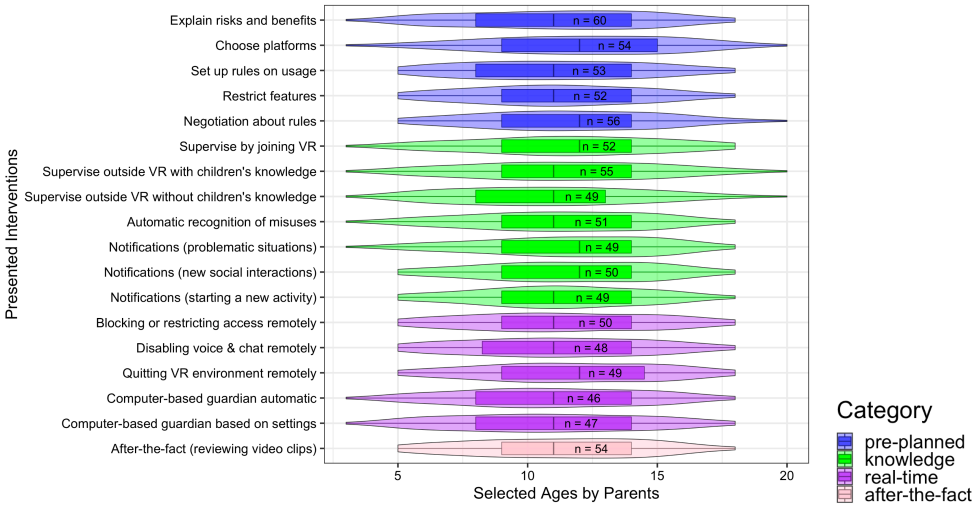


Figure 3.4: Interventions presented to the participants and the selected age range. The violin boxplots summarise age ranges selected for four categories of interventions (pre-planned, knowledge, real-time actions and after-the-fact). Each intervention was selected by n participants. The mean of the minimum ages given for all interventions is 10 years old, and the mean of the maximum ages is 15 years old.

1) Proportions of selected interventions I did not find significant differences between the number of selected interventions ($p > 0.05$ Chi-Square Test).

2) Ages given for interventions Age ranges given by parents among all intervention categories went from 0 years old up to 20 years in the categories of knowledge and pre-planned interventions. While there were no significant results that showed a difference between the number of times an intervention was selected, interventions in all four categories were generally selected with a mean age range of 10-15 years old. I detect outliers (maximum ages given of 20 years old) for interventions in the categories of 'knowledge as it happens' and 'pre-planned'. While these extremities are outliers, I argue that there could be a potential misuse of interventions.

Looking at minimum ages only, I found a significant difference between the minimum ages given for interventions ($F(17,752)=1.68$, $p= 0.041<0.05$, $\eta_p^2=.04$, small effect size, ART ANOVA). In particular, the minimum ages given for the intervention "Explain risks and benefits" were significantly lower than for the intervention "Negotiation about rules" ($p= 0.02<0.05$, Contrasts with Tukey adjustment). Moreover, the ages of the children of parents in the sample did not affect results, as I may have expected different opinions from parents who currently have younger children compared to those with older children. With interventions grouped into four categories, I observed significant differences between the minimum ages of pre-planned interventions and knowledge ($p= 0.01<0.05$, Contrasts with Tukey adjustment), with no effect from the parents' children's ages.

Finding 8: Parents in the sample seek to apply a breadth of safeguarding practices and interventions to protect their children, with limited evidence of adaptation based on minimum ages given, but no evidence of adaptation based on age ranges.

3.2.4 Discussion

A reflection across studies in Part II can be found in chapter 5.

3.2.4.1 Summary of Findings

I conducted an online mixed-methods questionnaire about children's usage of social VR, collecting data from the ecosystem of adult stakeholders of child social VR users.

S1-RQ1: Social VR Frequency and Appropriate Age of Children In the data, children less than 13 years old (N=58) use social VR and VR regularly, with approximately 43% using social VR daily, weekly or every two weeks. This result emphasises the importance of developing interventions to protect and safeguard children in social VR environments initially designed for adults. Findings also showed the impact of familiarity with VR and social VR, parenthood and supervision conditions in which children would use social VR (without supervision, with supervision and only with minors) on what participants would think is an appropriate age to use social VR. Specifically, the questionnaire showed that parents familiar with VR and social VR were significantly more likely to allow younger children to use social VR without supervision compared to less familiar non-parent adults.

S1-RQ2: Parent and Non-parent Adult Experiences and Perceptions Involving Minors' in Social VR Themes obtained from participant answers include the following: "dealing with harassment and bullying" (in minor-minor interactions), "Age-inappropriate behaviour that children encounter" and "Children over-sharing with adults" (in minor-adult interactions). Findings depict concrete experiences affecting adults or affecting children, to perceptions and concerns towards adults or children. In the data, most non-parent adults were concerned about the lack of supervision and the freedom children have in these environments, with mentions of how parents fail to meet age limit recommendations. There was an overwhelming number of negative answers and negative emotions compared to those that were positive or neutral, based on Russell's Circumplex model.

S2-RQ3: How Parents Choose to Moderate and Limit the Social VR Experience Parents' answers were polarised; some would practice strict supervision or completely stop their child from using social VR due to their negative experiences and perceptions, and the lack of regulation and intervention tools. Others would consider social VR similar to real situations where teenagers experiment and may not be under supervision when going to a party, requiring them to build a parent-child trusting relationship. These results exemplify a combination of the two approaches to parenting regarding digital media: "expressive empowerment" and "respectful connectedness" as described in section 3.2.1.2. Moreover, parents in the data seek to apply a range of interventions with little adaptation based on age. The minimum ages given for the intervention "Explain risks and benefits" were significantly lower than for the intervention "Negotiation about rules". The vast majority of interventions were applied indiscriminately with respect to age ranges. Proactive parents engaged to protect their children from risks are seen as "good" parents [130]. However, as Glassner (1999) re-

ports [288], people misassess risk, especially in new environments where they have little experience, thus leaning towards increased restrictions. It is therefore important to understand parental concerns when developing social VR interventions and regulations, but also consider children's perspectives.

The study's findings align with previous research [23, 22, 21, 192] but also contribute new perspectives, highlighting both advancements and knowledge gaps.

3.2.4.2 Addressing Negative Interactions in Social VR: Guiding the Adult-Minor Dynamic

Based on the findings and reflections, informed by social VR-aware parents and non-parent adults, I propose future directions for tackling negative interactions between adults and minors in social VR to create safer virtual environments.

Age-Matching and Private Social Virtual Spaces Non-parent adults recommended higher minimum ages compared to parents [Finding 2]. They raised concerns about the lack of parental supervision, the need for enforcement of rules and age restrictions in virtual environments where children enjoy considerable freedom [Finding 5], [Finding 6.2]. To tackle these challenges, social VR platforms could consider introducing private spaces that allow users to interact only with their friends, with their own rules. However, it is important to be aware that such spaces may also introduce the risk of new forms of abuse, such as adults pretending to be, or forming connections with, child users to join these private spaces. Additionally, implementing age verification measures to create age-matching virtual spaces that users can join, in addition to public spaces, is an option also proposed in recent work [193]. This would necessitate users sharing their identification or using ID authentication technology [289], however, potentially giving rise to privacy, potential leakage and ethical issues [290, 291]. Future research is needed to better understand how age verification and inclusion of users from various age groups in group instances can be effectively implemented in the context of social VR while tackling the limitations.

The Need for Behavioural Control and Access Control (Pre-Planned) Industry mediation with age limits is currently not sufficient to restrain children's use, as most of the platforms rely on self-professed age. While we may want to introduce biometrics to exclude children completely based on their voices or heights, for example, I argue that children may still be able to use social VR, and they may find ways around the mechanisms introduced, oftentimes with parental consent and support [292, 293]. Junior accounts and age-verification have been introduced [188]. However, children may choose (or attempt) to use a regular account so they can access the full version of the platform without limitations. More recently, Rec Room implemented a process that automatically converts a non-junior account to a junior account if the system detects that the user is under 12 years old. Users can appeal this conversion before it becomes permanent [188]. Other types of mediation are needed to safeguard and supervise children in social VR to make experiences safer for children. Social VR environments currently lack mature technical mediation, and there is an almost complete absence of supervision and safeguarding tools for parents on most platforms. However,

parents in the sample have been using "Behaviour Control" (i.e., based on defined rules and procedures) [114], in this case, by limiting children's use and "Clan Control" (i.e., acceptable behaviour is reinforced) [114], by communicating and explaining the impact of these disruptive situations, or by restricting their child from using social VR [See Section 4.6.2]. These controls are either done in the real world or via the virtual world. Regarding what participants would have wanted to moderate these disruptive situations, responses mainly included technical mediation for Access Control, in particular, age verification, login at weekends only, secure authentication, protection and supervision systems [see Section 4.6.1].

The Need for Oversight (Knowledge and Real-Time Interventions) Parents also want interventions to oversee (i.e., real-time monitoring) their children in social VR to understand what is happening in the virtual spaces. Parents may not be able to monitor by co-using social VR with their child via another VR headset. As an alternative, some parents use casting or stream the virtual environment via their mobile phones [See Section 4.6.2], but this requires them to focus on the video. If they do not pay attention, they may miss notable events. Parents need ways to become aware of problematic situations when their child is meeting someone or when they are starting a new activity. These alerts would facilitate gaining knowledge about what is happening in the environment. Parents also expressed their desire to be able to take action remotely without having to use a headset. However, I found that in the sample, parents less familiar with VR and social VR are more cautious than those familiar with them, in particular, if there is no supervision [Finding 2]. Therefore, oversight might be preferred by parents who are not aware of what usually happens in social VR and would want to get real-time notifications and interventions remotely, such as blocking or muting.

Recent work has proposed the implementation of automated moderation taking into account the unique affordances of VR and facilitating parental supervision (informing parents of negative encounters instantaneously), as an alternative to enforce rules, combat harassment and negative encounters in social VR [40].

The Need for Insight After-the-Fact As previously mentioned, parents may not always be able to follow the video cast when their children use social VR. Therefore, the possibility of reviewing the child's experience asynchronously (e.g., through video recordings) after the event would facilitate parental insight into social VR.

Parents have selected a wide range of interventions in the questionnaire with little adaptation based on age [Finding 8]. They mentioned wanting to build parental trust when moderating and safeguarding [See Section 4.6.2]. Trust echoes prior research in other technologies [114], highlighting the importance of parent-child trust.

Ethical Considerations Related to Parental Monitoring According to the Information Commissioner's Office (ICO) [294], children must have the right to know if their parents are monitoring them online, with an "obvious sign to the child" [294]. Transparency is paramount. However, parents still selected the intervention "Supervise from outside VR without children's knowledge" with the maximum age mean being 15 years old. Outliers were detected as age ranges were given with upper limits of 20 years old for interventions in the categories of 'knowledge' and 'pre-planned'. This result demonstrates potential misuse of interventions; for example, adults may use them without considering the best interest of their child, or without the child's consent. Therefore, we must consider exploitation risks and privacy concerns when developing interventions, ensuring these are in line with guidance from the ICO Children's Code and the five rights (5Rights framework [148]), regarding a child's evolving right to privacy and freedom of expression. There is also the potential that such interventions could be misused by adults applying them to other adults (e.g., for domestic abuse) [295, 296], thus, care needs to be taken as to what interventions are provided.

Potential Obstacles: Tension at Home and Lack of Trust The rise of technology such as smartphones and social media has shown to be a source of connection but also led to tension in families [22]. While children may tend to overestimate the opportunities and benefits, they underestimate the risks [297]. However, some children feel like technology can create tension, with the example of a boy not liking it when his father texts on his phone while the boy plays soccer [297]. While parents are typically concerned about their children's social VR use and the potential risks, some may want to overly restrict social VR use, which would impede parent-child trust. Future work is necessary to better understand parent-child relationships and how social VR may affect tension and bonding at home.

Perspectives and Parent-Child Relationship As shown in previous work and data, children's safety in social VR is an issue that requires attention and parents' involvement. I must also consider children's perspectives on the interventions to allow a parent-child trusting relationship and allow parents to intervene without taking away the child's sense of agency in the situation. I acknowledge that the choices of interventions may depend on parental strategies, cultures, and children's ages or maturity levels. Future work can more specifically investigate the reasons and factors that influence parents' choices of particular interventions.

3.3 Chapter Conclusion

This chapter focused on setting context, particularly adults' and parents' perspectives on the issues and concerns surrounding children's use of social VR, answering the following research question **(RQ1: What are the issues and concerns associated with children's use of social VR?)**. I aimed to answer this research question through a mixed-methods questionnaire with 149 anonymous responses, including responses from 79 parents. Participants shared their perspectives, experiences and social interactions involving children in social VR.

3.3.1 Contribution to Research Question 1

The data revealed that 43% of children under the age of 13 use social VR at least every two weeks despite age limits, highlighting the prevalence of young users in these environments. The findings show that parenthood, familiarity with social VR, and supervisory behaviour significantly shape adults' perceptions of whether social VR is age-appropriate. These perceptions are often linked to broader concerns about child safety, the need for parental responsibility, and the dynamics of trust between parents and children. Participants reported that children frequently encounter abuse, age-inappropriate behaviour and content, and engage in self-disclosure to adults within social VR. A recurring theme was the lack of sufficient supervision, regulation, and safety features, issues that not only endanger children but also disrupt adults' experiences in virtual spaces. In response, parents adopt a wide range of safeguarding strategies, both in virtual and real-world contexts. This points to a critical need for clearer guidance on mediation techniques for social VR.

Overall, this initial study builds on prior research by offering a deeper understanding of the risks children face in social VR, while also exposing significant gaps in safeguarding measures for children and parental mediation, as perceived by both parents and non-parents familiar with social VR. Given the importance of this topic and the exposed issues and concerns, it is essential to engage a wide range of stakeholders, not only parents, but also children and experts such as psychologists and child-safety specialists, to inform the design of effective, age-appropriate solutions. The next step is to gather and integrate these additional perspectives to ensure children's safety in social VR.

Chapter 4

Informing The Design of Child Safety Features in Social Virtual Reality

"Just having the potential to block one avatar at a time is something that is going to have to change in social VR." - [Study #2 Expert P6]

"[PointBlock] makes it fun when you're having to block people, it kind of removes a bit of stress because it can be really stressful to block people in games." - [Study #3 Child12]

4.1 Chapter Introduction

Adult social VR users and parents, as key members of the child social VR user's internal and external ecosystem (see figure 1.2), offered valuable perspectives on child safety in social VR, and chapter 3 centred on understanding these concerns and challenges. However, to move beyond identifying the problem and towards developing effective solutions, it is essential to also include the perspectives of experts trained in child development and safety and the children themselves. Experts can provide knowledge-based insights and guidance on potential interventions, while children's feedback is critical for capturing the nuances of their experiences and ensuring that safeguarding tools truly address their needs. Although these perspectives are crucial, they have often been overlooked in research assessing child safety tools in social VR. This chapter addresses the latter by focusing on the design space for social VR safety with experts and children.

4.1.1 Motivation

As more children engage with social VR, and given that VR can blur the boundaries between the real and virtual worlds, especially for children [38], the absence of proper safety mechanisms could expose them to inappropriate content, harmful social interactions, or harassment situations, as study #1 shows. This chapter investigates the design space for addressing child harassment in social VR by integrating expert insights with direct feedback from children. I aim to answer the following research question:

RQ2

What design considerations can help shape safety features to address child harassment in social VR?

4.1.2 Overview of Studies

4.1.2.1 Study #2

This study aimed to explore and better understand, from the perspectives of experts with relevant backgrounds, what safety tools are suitable and how they can be designed for children to protect them from both individual and group harassment in social VR.

Methodology I conducted one-to-one interviews with 16 experts specialising in child development and online safety in both academia and industry, as engaging experts in this manner allows for gathering in-depth insights, compared to participatory observation or surveys [298]. Experts can act as "crystallisation points" of insider knowledge and serve as proxies for broader stakeholder groups, especially in contexts where direct access is limited or sensitive [298].

Key Findings Among a list of safety tools across popular social VR platforms (Rec Room and VR Chat), Personal Space bubble, Block, Mute one or all & Human Moderators were ranked as top safety features by expert participants.

Expert design recommendations include:

- To tackle intentional harassment: enforce rules and community guidelines, consequences and stakeholders' responsibility.
- To tackle unintentional harassment: reflection, discussions with safety advisors, opportunities to explain and apologise and positive reinforcement.
- To tackle unwanted consequences: discreet ways of using safety features and enacting outcomes.
- Quick, easy, familiar and visible-grabbing-attention tools to be used when under pressure.
- Customisation of safety features based on the child's individual needs.
- Towards meaningful and balanced parental involvement in social VR: from the setup to real-time notifications to open conversations.

4.1.2.2 Study #3

In study #2, blocking was found as one of the most useful tools according to experts. However, previous research has not explored how it is used in practice with child users. This study, therefore, aims to examine children's perceptions of various blocking strategies in simulated social VR settings.

Methodology I conducted a user study with 40 children aged 8 to 16 years old. Drawing inspiration from the virtual watch and two of Rec Room's blocking features [187], as well as creating a third innovative blocking feature, I developed a VR prototype. This prototype integrated a similar watch with the three blocking tools within a simulated social virtual environment, including small and large group harassment scenarios featuring verbal, physical, and environmental harassment [31]. During the study, children acted as third-party observers but were instructed to imagine themselves as the victim in the centre of the group harassment scenarios. They were provided with three tools. This study focuses on the child perceptions and preferences among a) three blocking strategies on their virtual watches: 1) *[SelectBlock]*: Profile pictures of the avatars appear on the watch as buttons to click and block one at a time. 2) *[MassBlock]*: Button to block all avatars, a list to unblock avatars then appears. 3) *[PointBlock]*: Magic Wand Button appears, the participant clicks, the wand appears, and the participant "shoots" to block, b) two intervention perspectives (IP, on ground and above), and c) two group sizes of harassers (large and small).

Key Findings

- Regarding **Perceptions of Tools** *[MassBlock]* tool was noted for providing relief when managing harassment in large groups, while *[PointBlock]* was appreciated for its empowering, stress-reducing qualities, ease-of-use and particularly due to its fun, interactive nature. However, there were some negative aspects reported, particularly with *[SelectBlock]* and *[MassBlock]*, where children expressed concerns about accidentally blocking the wrong individuals. Additionally, some children noted the challenge of balancing the enjoyment and creativity of the wand with its physical demands.
- **Impact of Group Size** showed that harassment from larger groups intensified psychological distress and made blocking more challenging, reflected in both the qualitative and quantitative data. This suggests that safety tools should be easy to use and scalable to different group sizes.
- **Impact of Intervention Perspective** indicated that the majority of children reported feeling safer when placed in an elevated perspective compared to being on the ground. This effect was even more prominent in the larger group, as it made it easier to understand and control the situation. Statistical test shows a significant increase in calmness and sense of safety in the above perspective compared to the perspective on the ground. However, fewer children expressed discomfort with the height, finding it unnatural, and preferred the ground view. Overall, the above perspective was seen as empowering, as it reduced the intensity of the harassment.

4.2 Study #2 - Expert-Driven Design Considerations for Real-Time Safety Features in Social VR to Safeguard Children Against Individual and Group Harassment

4.2.1 Motivation

Safety features in social VR vary across platforms in terms of what protections are offered and how they are enacted, with little publicised evidence validating such differences. With the increase in children using social VR, the need to develop safety features designed around their needs, rather than targeting adult users, becomes crucial. Accordingly, I conducted in-depth interviews with experts (N=16) in child online safety and psychology to explore child-targeted safety features. I examined expert perceptions of ten existing *real-time* social VR safety features (i.e., as an incident happens, tools that allow immediate or short-delay intervention). These were drawn from two popular social VR platforms (VRChat [25] and Rec Room [27]). I then asked experts to propose recommendations for and to reflect on safeguarding children from harassment in social VR, giving specific consideration to a) individual versus group harassment, as the impact can be amplified and the social interaction dynamics differ; b) the perspectives of harasser versus victim.

In this study, I aim to address the following research questions:

S2-RQ1 What are experts' perspectives and considerations regarding existing social VR safety features to ensure child safety in one-to-one or group harassment?

S2-RQ2 What do experts recommend and propose, (beyond the existing ten safety features from S2-RQ1), for addressing one-to-one or group harassment targeting a child in social VR?

4.2.2 Related Work

This section synthesises related work on the specific topics addressed in this study. (A more comprehensive literature review is presented in chapter 2, with specific relevant sections: section 2.3.2, section 2.3.3 and section 2.3.4).

4.2.2.1 Existing Safety Features in Social VR and Challenges

Social VR platforms have incorporated existing safety social media tools, such as blocking and reporting a user, and developed new proxemics safety features to protect their users from physical and environmental harassment, such as personal space bubbles [204, 187, 205]. While little is known about their effectiveness and usage, a recent study involved researchers observing YouTube videos of social VR users' experiences and identified a list of major in-game real-time safety features across four social VR platforms, including VRChat and Rec Room [50]. They grouped social VR safety features into three categories: "Boundary settings" (e.g., intimacy proxemics, trust reputation, social spaces, shield levels, and controls for voice and avatars), "Quick-reaction tools"

(e.g., safety gestures, votekick, safety zones, and safety reports) and "Agreements" (e.g., codes of conducts, terms of service) [50]. The researchers highlighted that relying only on proxemics is inadequate since different safety threats require varied social cues due to the VR affordances. Additionally, current features lack quick reactive gestures (such as the "talk-to-the-hand gesture" used to block in Rec Room [187]), and adding these could enhance users' ability to respond to threats [50]. Most existing safety tools are reactive, designed to address issues after they occur. This puts pressure on victims to effectively use these tools while under distress, which may be especially challenging for children. The study advocates for a move towards preventive safety mechanisms, incorporating social cues to identify risks early, such as using signal processing data to detect gestures that signal safety concerns [50]. While the study examines the effectiveness of safety features through empirical observations, more research is needed to understand what should be considered for effective safety features, specifically for children, from expert perspectives.

Along with existing mitigation tools, there are human moderators to moderate and address incidents in real-time in the social virtual environments [31, 20]. Yet, only 24% of incidents observed in a study in social VR were addressed by human moderators, highlighting the lack of scalability for human moderators in a growing VR ecosystem and the need for new moderation tools [19]. Automated moderation can address this scalability issue and has, therefore, been proposed as an alternative safety tool to combat harassment in social VR [40].

4.2.2.2 Gaps

Prior work has shown that there are new forms of virtual harassment (e.g., physical and environmental) that require new safety features compared to 2D social media. This study aims to provide insights from the eyes of experts with relevant backgrounds in child online safety and psychology to better understand what safety tools are suitable and how they can be designed for children.

I explore safety-enhancing technologies for child users of social VR from expert viewpoints, diverging from prior studies focusing heavily on children or adult social VR users [23, 272, 193]. While child and guardian perspectives are crucial, expert insights can offer objective, evidence-based approaches, leveraging their knowledge for effective safeguards for children.

4.2.3 Methodology

This study was conducted in May-July 2023 with **16 experts**.

4.2.3.1 Study Design

One-to-one interviews were selected as the method for this study, as they are particularly suitable for in-depth exploration [299]. In exploratory research, interviewing experts offers a more efficient way to gather concentrated insight than methods like observation or surveys [298]: *"Talking to experts in the exploratory phase of a project is a more efficient and concentrated method of gathering data than participatory observation or systematic quantitative surveys. Conducting expert interviews can serve to shorten time-consuming data gathering, particularly if experts are seen as "crystallisation points" for practical insider knowledge and interviewed as surrogates for a wider circle of players. Expert interviews also lend themselves to situations where access to a particular social field may prove difficult or impossible (e.g. taboo subjects)."* [298]

In May 2023, I identified five common (i.e., they have the same end-goal) and five distinct (i.e., they have a different end-goal) real-time safety features enacted by the users or human and automated tools, also described as "Boundaries Settings" and "Quick Reactions" in prior work [50], after extracting all the safety tools obtained from the Trust & Safety webpages of VRChat [204] and Rec Room [187] (See table 4.1). VRChat and Rec Room are noted as top social VR platforms in January 2024 [300, 301], extensively studied in social VR research [50, 20, 302, 272] because of the capabilities of these platforms for enabling customisable, engaging experiences for children and adults, evident from the 2,530 papers found with the keyword "VRChat" on Google Scholar (in May 2023). Rec Room's design, in particular, has child-oriented design graphics that may be appealing to children and would require ensuring child safety. The five common tools available across both platforms also exist in other social VR platforms (e.g., Horizon Worlds [205] includes all five, Gorilla Tag [165] includes some of them). I chose ten for time constraints of the interview.

Table 4.1: Ten Safety Features in Social VR Platforms, five common and five distinct ones across VRChat [204] and Rec Room [187] (May 2023 - list is non-exhaustive for time and evaluation constraints).

Social VR Safety Feature	**Platform(s)**	**End-Goal [204], [187]**	**High-level Usage [303, 304, 305]**
Common Features			
Blocking	VRChat, Rec Room	Removes the user from the blocker's virtual environment.	In VRChat and Rec Room, you point at the user and click a "Block" button.
Reporting	VRChat, Rec Room	Allows users to report inappropriate behaviour that violates the platform's code of conduct.	Similar to blocking, there is a button "Report". The user then fills in a form.
Muting	VRChat, Rec Room	Disables a user's voice.	Similar to blocking, there is a button "Mute".
Personal Space Bubble	VRChat, Rec Room	Creates an invisible virtual barrier that prevents other players from getting close.	The user can create a personal space bubble by going to Settings (through their virtual watch in Rec Room and through the "Quick Menu" in VR Chat). The size of the bubble can be adjusted.
Human Moderators	VRChat, Rec Room	Volunteers or staff monitor interactions, review reports and enforce rules [186, 306, 19].	In both platforms, there is an internal team of human moderators but also community-based volunteer moderators in-game.
Platform-Specific Features			
Votekick	VRChat	Enables users to initiate a room-wide vote to decide on kicking disruptive members.	A user decides that another user is being disruptive and initiates a votekick. Other eligible users in the room are notified and vote "yes" or "no" in a limited time. If the majority votes yes, that player is removed.
Trust Rank	VRChat	Establishes user credibility based on behaviour and interactions within the platform.	The rank is based on activity time, content contributions and other indicators of participation. Higher ranks unlock more abilities. The rank appears on the user's profile, visible to everyone.
Gestures	Rec Room	Provides users with non-verbal communication options.	For example, a quick "talk to the hand" gesture the user can perform on another player to block them directly.
Automated Voice Moderation	Rec Room	Limit abusive or discriminatory language in public areas.	Uses algorithms to detect and manage abusive language in public rooms.
Enable / Disable Animations	VRChat	Allows users to disable custom animations of other users, which can prevent seeing disruptive animations.	The user can toggle a button in the "Quick Menu" to disable or enable custom animations for specific trust ranks.

4.2.3.2 Procedure

I conducted 16 semi-structured in-depth one-to-one online interviews with experts who have at least 5 years of experience (research and/or industry) in child development/educational psychology, psychiatry and social media/social VR online safety. More details about the experts are in section 4.2.3.5. Interviews lasted 20-30 minutes, from May to July 2023. I presented slides with the written questions and timers (varying from 2min to 4min) for each question. The interviews all started with a 5-minute presentation about social VR regardless of the experts' social VR knowledge and experience. I introduced social VR opportunities and risks and showed a 10-second video from YouTube of a child sharing their experience with another adult in social VR [307].

4.2.3.3 Interview Questions

After presenting the list of the ten safety features using illustrations on slides and a video demonstrating the safety tools from Rec Room (see Supplementary Material STUDY 2/INTERVIEW_SLIDES) and verifying whether they have understood all the tools presented, I questioned experts about a) their perception of existing tools, b) their applicability to individual/group harassment.

- Rank your top 3 most useful tools in your opinion, that address one-to-one harassment towards a child and the least useful. Please explain your top 3 and your least effective: What do you think their strengths and limitations are?
- Repeat for group harassment towards a child.

I then introduced one of two possible scenarios in social VR to provide the experts with context, keeping it broad enough to ensure the discussion remained open and unconstrained. I developed these scenarios based on reported events by participants in prior work [23, 22], capturing both individual and group harassment dynamics as well as including physical and environmental harms unique to social VR settings [31] (see scenarios below).

> **Scenario 1 - One-to-one Harassment** - (Adult player inappropriate towards a child: Example of environmental harassment). *A 12-year-old child named Tom joins social VR because, as someone with autism, he struggles with the real world. One day, in social VR, Tom meets an adult player who starts asking him personal questions, making suggestive comments and displaying graphic content on a virtual screen when he is present.*[1]

> **Scenario 2 - Group Harassment** - (Group of teenagers bullying a child: Example of physical harassment) *Jane, 10 years old, joins social VR. Jane loves social VR because she is quite shy in the real world; social VR really helps her be more confident. Soon after joining, Jane becomes friends with a group of teenagers who seem friendly. One day, a group starts to turn on Jane, teasing her, surrounding her, getting very close to her avatar, following her in the virtual room.*

[1] While prior work has identified situations in social VR settings where adults may be unaware that another user is a minor [23], I do not specify this here in order to keep the discussion open-ended.

I asked experts the following questions:

- From your experience, can you think of a new solution that would be most effective to tackle the scenario?
- How do you think the solution impacts the victim and the harasser?
- Would you want to adapt your solution for different groups? How and why?
- Can you reflect on your solution (gaps, weaknesses, etc.)?

4.2.3.4 Analysis

The analysis was completed in six phases, using thematic analysis [283]: data familiarisation, generation of initial codes, searching for themes, refining themes, defining and naming themes and writing the report. After transcribing the interviews, I reviewed the recordings and corrected any mistakes in the auto-generated transcripts. A second coder and I read and familiarised ourselves with the data. We created individual coding schemes independently using the online qualitative analysis platform QCAmap [308]. The codes generated are words or short phrases that describe an idea. We then collaborated to consolidate the two coding schemes into one combined scheme by collating or distinguishing between codes. A consensus was reached to adopt the final coding scheme based on the more detailed set of codes, adding missing codes or merging codes, before revisiting the transcripts with the final coding scheme.

We grouped codes into main categories by using a Miro Board during a meeting, where we organised individual codes on post-it notes and then clustered them into relevant categories as patterns emerged. (*The Miro board can be found via* `https://miro.com/app/board/uXjVN9xFlsM=/`.) The themes were developed based on the categories either as one full category developed or by combining one or more categories to reflect a higher-level aspect of the data (See Supplementary Material STUDY 2/CODING_SCHEME_ANALYSIS). The second coder, two researchers and I further refined both the names and descriptions of the themes, while also considering their grouping, such as combining or subdividing them into new thematic categories.

4.2.3.5 Participants

Recruitment Experts were recruited via LinkedIn from May to July 2023, in English. Participants were found via search with keywords related to disciplines involved in child safety, such as "Child Psychologist", "Online Safety", "Child Online Safety" and "Trust & Safety". After voluntarily signing up via Calendly [309], they received a Qualtrics [280] form with an information sheet, consent form to sign and demographics questions. I individually interviewed 16 experts.

Demographics The detailed demographics of the 16 experts are described in table 4.2 and Table table 4.3.

Table 4.2: Demographics of experts who participated in one-to-one interviews (Part 1). Experts had a Mean age of 43.3 years (SD=10.5).

P No.	Age	Gender	Parent?	Education	Profession	Child-related?	Country
P2	33	Female	No	Doctorate (PhD, Information Science)	Researcher in social VR harassment	No	USA
P3	54	Female	No	Doctorate (Psychology)	Academic Professor of Psychology, Security & Trust, Criminology	No	UK
P4	50	Male	Yes	University degree (Sport, business management)	Head of Child Safety Online - Online safety profession involved in CSE related content inc VR	No	UK
P5	30	Female	No	University degree (Marketing, management)	Child Safety Intelligence Analyst/Consultant	Yes	UK
P6	62	Male	Yes	Masters in Cybercrime Investigations and Forensic Computing	Trainer with Cybersafekids	Yes	Republic of Ireland
P7	48	Male	No	University degree (Social Psychology)	Child Safety Intelligence Analyst/Consultant (Online exploitation, AI, Metaverse, etc.)	Yes	Spain
P8	35	Female	No	University degree (Clinical psychology)	PhD student - Clinical Psychologist	No	Italy
P9	48	Female	Yes	University degree Linguistics and AI (and an MBA)	Online Safety Consultant, Clinical Psychologist	Yes	UK
P10	42	Female	No	Doctorate (PhD)	Head of online safety research	Yes	Malta / Worldwide
P11	42	Non-binary	No	Doctorate (Psy.D.)	Community Developer	No	USA
P12	59	Female	Yes	Doctorate (PhD)	Professor of Child Adolescent Psychiatry	Yes	UK
P13	27	Female	No	MSc Comparative Social Policy	Policy and Public Affairs Officer	Yes	UK
P14	42	Male	No	BA Italian	Director, Online Safety Child	Yes	UK
P15	46	Male	Yes	MA, PGCE:PSE, MBChB, MRCPsych	Clinical Senior Lecturer and Honorary Consultant Child and Adolescent Psychiatrist	Yes	UK
P16	49	Female	Yes	PhD (current) and MSc in Psychotherapy	Psychotherapist and cybertrauma Consultant	Yes	UK
P17	25	Female	No	Professional degree (Data & Society)	Digital Investigator, Social VR threat researcher, Threat Intel analyst	No	UK

Table 4.3: Demographics of experts who participated in one-to-one interviews: child bullying knowledge, child cyberbullying knowledge, Social VR awareness, Social VR experience, owner of VR headset, VR experience. Experts had a mean of Social VR Awareness of 4 (SD=1.5).

[1]While P8 mentions 'no knowledge' in child harassment and cyberbullying, they had 10+ years of experience as a Psychologist, with research focusing on video games, and had participated in a relevant workshop "Combating Toxicity, Harassment, and Abuse in Online Social Spaces".

P number	**Child harassment Knowledge**	**Child Cyber-bullying Knowledge**	**Social VR Awareness (1-5 scale)**	**Social VR Experience**	**Owner of VR headset**	**VR Experience**
P2	Extremely	Extremely	5	A lot	Yes. (Meta)	A lot
P3	Moderately	Moderately	5	A moderate amount	Yes. (Meta)	A moderate amount
P4	Very	Very	5	A little	No	A little
P5	Very	Extremely	5	A little	No	A little
P6	Very	Very	4	None at all	No	A little
P7	Very	Very	4	A moderate amount	Yes. (Meta)	A moderate amount
P8[1]	Not at all	Not at all	1	None at all	No	A little
P9	Moderately	Very	5	A little	Yes. (Meta)	A little
P10	Moderately	Very	2	None at all	No	A little
P11	Moderately	Moderately	5	A great deal	Yes. (Meta)	A great deal
P12	Moderately	Moderately	1	None at all	No	A little
P13	Very	Moderately	5	A moderate amount	No	A moderate amount
P14	Extremely	Extremely	5	A little	No	A little
P15	Very	Very	2	A little	No	A little
P16	Extremely	Extremely	5	A lot	Yes. (Meta)	A lot
P17	Extremely	Moderately	5	A great deal	Yes. (Meta)	A lot

The 16 experts recruited (10 female, five male, one non-binary/third gender) had a mean age of 43.3 years (SD=10.5).

Five experts held a Doctorate, eight had a University Degree, and three had a Professional Degree. Ten had professions involving child online safety (with five including social VR), seven had a background or profession in psychology / adolescent psychiatry, one in Policy and Public Affairs and one in Security, Trust and Criminology. Ten professionals practised within the UK, two in the USA, one in Spain, one in Italy, one based in Malta working Worldwide and one in the Republic of Ireland. Experts had a mean Social VR Awareness (1-5 scale, with 5: extremely aware; 1: not aware at all) of 4 (SD=1.5). All experts had at least *a little* experience in VR, seven owned a

VR headset, and 12 had at least *a little* experience in social VR. While some experts were limited in experience with social VR (although they had experience with VR), their expertise and background still offer informed insights, providing their vision of how to protect children in social VR, which future work can build upon. Four experts self-reported they were *extremely knowledgeable* in child cyberbullying, six self-reported being *very knowledgeable*, and six as *moderately knowledgeable*. (Expert participants are numbered P2-P17; P1 was excluded for lack of relevant input.)

Ethics and Compensation The protocol was approved by the university CoSE ethics committee (Approval number: [300220184]). Each expert was compensated with an Amazon voucher of £8. Documents with the information sheet, pre-survey with consent form and demographics questions can be found in the Supplementary Material.

4.2.3.6 Limitations

The rapid evolution of the social VR gaming landscape, where new safety measures are constantly being proposed, means that safety measures released after the study are not covered. While the tools proposed still represent some of the common safety features of popular social VR platforms, new tools merit further evaluation and engagement with experts as well as child users.

Users may not use these tools or may have additional needs that are not considered by experts, and longer reflections to tackle the broad range of topics covered would be needed. While some experts were limited in experience with social VR (although they had awareness of social VR and experience with VR), their expertise and background still offer informed insights, providing their vision of how to protect children in social VR, which future work can build upon.

Finally, the sample size and necessity of conducting the studies in English may limit the external validity with respect to the global population and applicability to other cultures.

Despite these limitations, the study proposes valuable groundwork for reflecting on social VR safety features for children through the lens of experts.

4.2.4 Results

4.2.4.1 A Comprehensive Review of Ten Existing Safety Features in Social VR for Ensuring Child Safety

Among the ten safety features presented, experts emphasised the usefulness for children's protection in social VR of **a personal space bubble, blocking and muting and human moderation**. Notably, the findings reveal the same top 3 among the list for one-to-one and group harassment (see figure 4.1). However, some experts noted that group harassment leads to amplified harm and should be considered as a more serious event. According to experts, the least useful safety features identified are the

"votekick" tool in a one-to-one situation and "trust rank" in group harassment (see figure 4.1) because of potential abuse against users. Experts' reflections and reviews regarding the reasons for choosing tools for one-to-one and group harassment are described in the following sections.

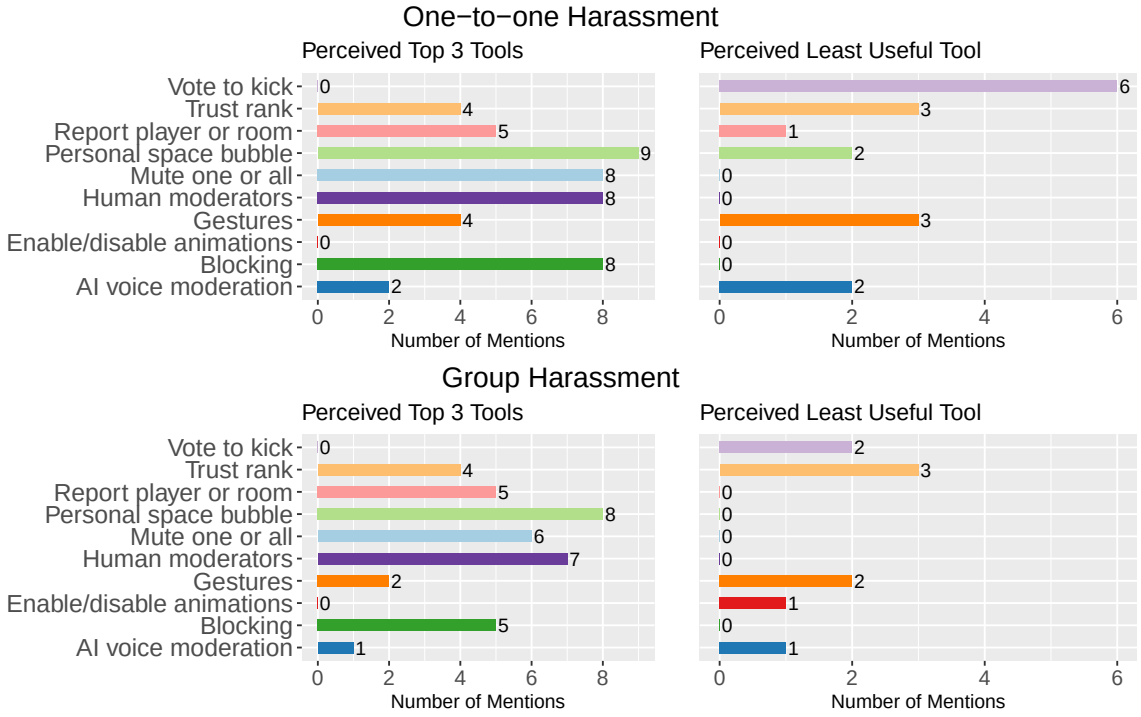


Figure 4.1: Counts of top 3 expert-selected tools for two harassment scenarios and the least useful tool perceived by experts: addressing one-on-one harassment and group harassment targeting a child.

How Existing Tools Contribute to Safety in Social VR Experts (N=6) noted that safety features in social VR can be effective when the child users have existing awareness from widespread usage in social media and gaming they may be familiar with, or by incorporating real-world metaphors, which can help increase quick-to-activate usage (e.g., using the hand as a stop-sign gesture).

> *"Muting and blocking are tools that have been used in content moderation for ages. I think they are the foundations of everything."* [Expert P7]

> *"It's also a little bit similar to what you do in real life, you might just be like "stop" (shows hand gesture) so I thought that was quite good."* [Expert P12]

Experts (N=4) also compared social VR environments to offline worlds and emphasised the need to remove the distinction between the online and offline world, where human moderators should be there to moderate children.

> *"I personally professionally think that there should be more human moderators online, particularly on these games that have been sold for children and towards children. We have to start thinking that the online world is not separate or different to the real world, like if you have children in a swimming pool in a gymnasium, in a football field. [...] We have to start thinking that the two are the same because they have merged."* [Expert P6]

Finding 1: Among the ten safety features presented, experts emphasised the usefulness for children's protection in social VR of **personal space bubbles, blocking and muting and human moderation**. According to experts, safety features in social VR are most effective for child users when they leverage existing experience and awareness from other forms of social media they may be familiar with and incorporate real-world metaphors (i.e., drawing parallels with physical world concepts to make usability safer, more tangible and relatable).

Seven experts added *human moderators* in their top 3 for both one-to-one and group harassment towards a child. They observed three noteworthy strengths of the presence of human moderators: first, the human presence amplifies the sense of protection and authority (N=7), giving children a dependable source to seek assistance from; second, a flexibility and nuance that allows for negotiation compared to more rigid software (N=4); and third, a deep understanding of the norms within the space (N=1). According to two experts, human moderators could also easily spot group harassment by scanning for group creation and ensuring safety. However, maintaining a human moderation team can be challenging and costly for social VR platforms.

Some experts suggested that automated moderation could serve as an alternative for scalability (N=2) or for handling minor tasks such as identifying instances of yelling or profanity (with the bot signalling to lower the volume) (N=1).

Finding 2: Human moderators within the game are an essential component, particularly for younger users, as they provide 1. presence and authority, 2. flexibility and nuance, 3. knowledge of social VR norms. However, maintaining a human moderation team can be challenging and costly. A more scalable alternative is automated moderation.

Experts (N=3) highlighted the importance of proactive and reactive tools like muting and blocking for the child users to **control others** in the space and the environment. However, in group harassment, the tools should allow users to moderate the group as a whole. Expert P2 highlighted the need to mute multiple users, and expert P6 suggested the development of blocking multiple users in a radius at once:

> *"Just having the potential to block one avatar at a time is something that is going to have to change in social VR."* [Expert P6]

Experts (N=12) specifically positively noted the possibility for a child to **control their own visibility, boundaries and space** while staying in the VR space as a preventive method, for example via a protective bubble that can make avatars disappear if they get close. They also underscored the significance of a child feeling comfortable and acquiring an understanding of personal space. Expert P11 suggested that the protective bubble could have privacy settings for different groups or friends. The child would start with maximum boundaries and adjust those boundaries for certain friends.

> *"People just disappear when they come close to you. So, having, maybe for children, a maximum, unless there's like a vetted friend or whatever that can be closer to your body is probably a good idea."* [Expert P11]

While the tools, "trust rank" and "votekick", were overall the least useful tools according to experts, two participants noted the strength of **sharing the problem with other users** who may have the same problem and resolving it as a group:

> *"For vote kick or trust rank, I would put them on the same level. A problem that, maybe, is just mine, can be a problem for the whole group."* [Expert P9]

Overall, experts (N=9) highlighted the need for proactive mitigation and instant, easy-to-learn-and-use tools (N=4) to give the child a sense of agency and control. Some experts noted gestures would allow faster actions, especially for beginners, compared to tools like blocking and muting complicated to find in the menus. In contrast to the finding above, three experts also noted the strength of tools when they inform the provider, such as reporting.

Finding 3: Experts considered safety features most effective for child users when they facilitate preventive and proactive user actions by **instantly and easily** giving back to the child the sense of agency as they control their experience: 1) restricting others, 2) controlling their own boundaries and space. For example, in group harassment situations, users should be able to mute and block multiple users at once.

Identified Child-Safety Risks and Weaknesses of Safety Features in Social VR As children transition from game to game, experts (N=2) emphasised the need for consistent safety across platforms to ensure children are acquainted with the tools and can effortlessly use them.

> *"All of the apps have different ways of pulling up these menus and actually accessing the controls. They're not governed by any kind of shared user interface or shared Paradigm across all of the social VR apps."* [Expert P2]

Children might not be equipped to react and block users in a disruptive situation in social VR. Expert P5 highlighted that trauma in VR can be very real and close to physical; the brain perceives the harassment in VR as real, and the child can freeze:

> *"Oftentimes in VR, your brain perceives this as reality. When you're verbally harassed even though the person is not there, people can have a reaction of just freezing, and the blocking would not come to their mind to protect themselves. [...] When it comes to children, they don't have the level of maturity to always block people."* [Expert P5]

Social VR platforms should also ensure children can address their concerns and take actions with discretion to avoid escalation or retaliation, according to three experts:

> *"My only concern with gesture-based tools would be the lack of discretion. [...] I would be worried about retaliation and escalation, especially for children."* [Expert P2]

> *"There's this idea of discreetly wanting to get help or raise a problem."* [Expert P8]

Another concern (N=2) revolves around children inadvertently entering inappropriate virtual environments, such as strip clubs, where inappropriate content is streamed, and the risk of adults, potentially untrustworthy, interacting with children in virtual rooms.

Finding 4: Experts identified key child-safety risks in social VR, including inconsistent safety features across platforms, potential immaturity of children in handling overwhelming VR environments, the importance of discreet tools to prevent escalation, the concern of adults and children mixing in virtual rooms, and the need for the establishment of social norms.

Experts reflected on the existing safety features and noted several factors that may lead to the ineffectiveness or weaknesses of those tools. The first factor is **Accuracy**. Experts (N=2) noted that human moderators might not be able to notice a disruptive situation is taking place, due to noise, subtle situations difficult to notice or a lack of resources and the number of moderators. Moreover, human moderators should be trained by the companies, according to two experts. While automated moderation may be an alternative solution in terms of scalability, four experts noted that the current stage of automated moderation tools lacks the necessary nuance and detection accuracy to be deemed effective.

The second factor is **Response-time issues**. Experts are worried about the response time of some tools (N=2), where the gestures might not be quick enough or where technical issues arise with the protective bubble:

> *"Gestures make it a bit more age-friendly. I just think it's too visible and it's not prompt enough. If you're in a really unsafe situation, someone's really close to you, and you put that gesture up to stop them from happening, that person's already too close. It needs to be much faster."* [Expert P13]

The third factor is **Abuse and Unwanted Consequences**. Safety features may not remove the problem (N=1). In contrast, they could potentially have the opposite effect, as users may abuse these tools to inflict harm on other players (N=5), or it could backfire, putting the onus on a target (N=6). Examples include gestures, "vote-to-kick" and "trust rank" tools, or situations where a child seeks assistance from an adult:

> *"It could backfire [...] like when a child talks to a teacher or an adult about something, and they say they were a crybaby, they went to speak to someone."* [Expert P10]

> *"Votekick worries me a little bit because I wonder if that could be used as a method of harassment. What happens if a group has voted to kick out kids who are vulnerable?"* [Expert P15]

The fourth factor is **Circumvention and Unawareness of Consequences**. Experts (N=3) highlighted the concern of children not fully grasping the outcomes following the use of certain tools, such as reporting, where the response to the player reacting and reporting is not always clear. The possible ways of circumventing should also be considered when developing a tool; for example, two experts (P13, P16) raised the concern of a player recreating an account after being blocked.

The fifth factor is **Privacy**. Experts (N=4) were concerned about privacy regarding human moderators actively listening or with automated voice moderation. VR can offer new forms of evidence that may not be only verbal compared to 2D social media or gaming, which also raises privacy concerns when there is the possibility of recording behaviours and gestures:

> *"There are privacy concerns with just recording everything that's happened in the whole space. Also, they don't have the operational capacity to just sit and watch hours of footage."* [Expert P2]

Finding 5: Experts highlighted critical weaknesses in current safety features for children due to: 1) accuracy detection issues with challenges of scale and resources, 2) response-time, 3) the risks of tools being misused and having unwanted consequences, 4) circumvention and 5) privacy.

4.2.4.2 Expert Propositions for Enhancing Child Safety in Social VR

Overall, experts proposed 1) reactive technical solutions, 2) preventive and educative solutions and 3) human-involved solutions.

One-to-One Harassment Towards a Child Experts proposed solutions (see table 4.4) for scenario 1 (see section 4.2.3.3).

For this scenario, experts (N=4) proposed to add **Age Verification and Age Assurance** and to allow users to interact or to be grouped with users in the same age range (N=4).

> *"One thing we might think about is like whether there are rooms where you ban people over a certain age, so that would reduce the likelihood of an adult being in that room and showing inappropriate material."* [Expert P3]

> *"Should a 12-year-old be on VR anyway. And if they are, they should only be allowed in certain areas. It can't be an open free-fall. Secondly, they shouldn't be able to interact with people, maybe of a certain age, older, so if you're a 12-year-old, you can only interact with up to 15-year-olds, but then we have the issue with age assurance and age verification."* [Expert P4]

> *"It should be suspicious if you have an older adult interacting so much with a younger account, and I'm sure that can be monitored."* [Expert P17]

Experts (N=2) proposed buttons to help the child with a **Panic/Pause button** or with a **"I don't feel comfortable" button** to take action instantly and pause the experience.

> *"An autistic child might know that they just don't feel quite right, but they are just not sure what it is. It might be nice for the child to have an "I don't feel comfortable" button which would pause the game."* [Expert P12]

An expert suggested educating children about social VR harms through real-world comparisons and metaphors:

> *"I would ask them "Would you go down the main street of your town and hand out photographs of yourself, would you stand outside the local supermarket and just hand out photographs of yourself to perfect strangers?". All the children say no, they wouldn't do that. So I say "Why do it in the online world" and then I'll talk to them about the people trying to get to know them. It could be somebody that is 62 years of age pretending to be a child that's 12 years of age, so you have to be careful who you're talking to in the online world unless you actually know that person."* [Expert P6]

If safety features are shown at the setup, it might also already be a preventive method to dissuade potential harassers (N=2):

> *"If you set that up in a way where even in small print it says to any game participant, "we have the parental invisible joining button". That in itself might police behaviour because there might even be an option where the parent can record and then that could be sent off."* [Expert P12]

Experts (N=2) propose human-involved solutions, setting in-game buddies that would help the child or the possibility of bringing in the parent as an invisible avatar when the child needs.

> *"If Tom is undergoing this kind of interaction and can't think about the gestures or the blocking, maybe someone else coming in and helping them would be a good thing, and a buddy could also be a parent."* [Expert P3]

> *"Children could invite their parent in to be there as a kind of invisible presence, so the other user doesn't know. It might be a nice thing with an autistic child to have that as an option where you would say to your child as a parent "look If you don't feel comfortable for any reason just press and it'll alert me on my phone and I can just join you in the game and the other person won't even see that I'm there." "* [Expert P12]

An expert also noted the need to focus on empowering the child, giving them the choice on what to do next.

Table 4.4: Solutions Proposed by Experts for Scenario 1 (Section 4.2.3.3): One-to-one harassment towards a child. (Half the participants had to propose a solution for scenario 1, while the other half had to propose a solution for scenario 2.)

Solutions Proposed for Scenario 1: One-to-one harassment	**Proposed by Experts**							
	P3	**P4**	**P6**	**P12**	**P13**	**P15**	**P16**	**P17**
Age Verification and Assurance	✓	✓			✓			✓
Age-Matching and Grouping	✓	✓			✓			✓
Buttons to Pause the Experience				✓			✓	
Panic Button				✓			✓	
'I don't feel comfortable button'				✓				
Prevention and Education			✓			✓		
Automated Moderation Flagging and History Checking		✓						✓
Buddies Assigned or Called (Bystanders, Friends or Parents) to Help Child	✓			✓				

Group Harassment Towards a Child Experts proposed solutions (see table 4.5) for scenario 2 (see section 4.2.3.3).

Table 4.5: Solutions Proposed by Experts for Scenario 2 (Section 4.2.3.3): Group harassment towards a child. (Half the participants had to propose a solution for scenario 1, while the other half had to propose a solution for scenario 2.)

Solutions Proposed for Scenario 2: Group harassment	**Proposed by Experts**							
	P2	**P5**	**P7**	**P8**	**P9**	**P10**	**P11**	**P14**
Quick, Discreet and Instant: Button and Gesture	✓	✓			✓			
Panic/Pause Button (to pause the experience)	✓	✓						
SOS Button (calls a moderator)		✓						
Discreet gesture to take action	✓	✓						
Pre-Set Suggested Prompts / Nudges					✓			
Automated Moderation Flagging / Biometrics and History Checking			✓	✓		✓		
Age Verification and Group Age-Matching			✓					
Preventive Solution: Rules, Expectations, Tips Reminding Community Guidelines				✓	✓			
Human-Involved Solutions			✓			✓	✓	
Safe Space / Timeout to Consult Safety Advisor						✓	✓	
Human Moderator Combined with AI			✓			✓		
Positive Reinforcement and Safe Space							✓	
Safety Bubble based on Privacy Settings								✓

Experts (N=3) proposed quick, instant and discreet actions as the child might freeze in a group harassment. Solutions include a panic button to pause the experience, an SOS button that calls a human moderator discreetly or pre-set suggested prompts to facilitate and ease the burden of conversation.

> *"In this scenario if there was an option for Jane to instead of exiting the game, flag this situation to one of the moderators to intervene without pointing at users and blocking them individually, [...] an easy gesture that's very discreet, to call for help, bring in the moderator from a different space to then resolve the situation, like an SOS button."* [Expert P5]

Other experts (N=3) also recommended solutions that involve pausing the experience and reflecting with a human, such as a safety advisor, to help the child. Two experts suggested combining the human moderator with AI or having a virtual online safety advisor.

> *"Human moderation combined with AI voice moderation that can detect if the terms they are using are harassment, but also building knowledge databases."* [Expert P7]

Experts (N=3) proposed the use of biometrics to monitor the age, if the child is in distress and develop a system that learns from previous iterations and interactions of the users.

> *"Biometrics are all the data that the headset is giving you in order to interact with the environment. Monitor biometrics because if someone is claiming that they are 13 years old when in reality they are 47, that is something that can be obtained thanks to biometrics."* [Expert P7]

Moreover, preventive and educative solutions were recommended (N=2) to ensure users understand the impact of harassment and know about the rules via immersive education or guidelines and tips sent via emails and ensuring the child understands when and how to take action.

> *"Young children are much more at risk online. So I'd be looking at Jane and what her skills are, and her resilience, like how confident she is in herself, whether she understands that she can get help, for example, and she knows how to do that."* [Expert P8]

> *"People could get emails with tips reminding them of the terms of service or perhaps the community guidelines about harassment, etc."* [Expert P10]

Age-matching was also suggested (N=2). One expert proposed, in addition to the existing personal space bubble, the possibility of adding privacy settings to your safety bubbles, where only your friends will appear visible.

> *"There could also be another version of that, which is sort of instant bubbles created through like privacy settings. So I'm playing with the friends of friends, I suddenly hit a button, and I move into another room, which just has my friends or people that I follow, where it's a little bit safer. It's a more closed group, almost like concentric circles where you're able to quickly move to a safer space and then back out again when you're comfortable."* [Expert P14]

Finding 6: Overall, experts recommended and proposed safety features designed to be: 1. easy-to-use under pressure (e.g., Panic/SOS Button, Pre-set prompts), 2. discreet in use and in the outcome (e.g., Discreet gestures), 3. offering guidance (e.g., prevention and education, buddies, human-involved), 4. providing immediate support (e.g., automated moderation), 5. tailored for VR physical, verbal and environmental harassment types (e.g., safety bubble based on privacy settings, biometrics), 6. enabling reflection and contemplation (e.g., pause button), and 7. delivering a sense of space, comfort, and empowerment (e.g., positive reinforcement).

4.2.4.3 Tailoring Child-Safety: Experts Advocate Personalised Social VR Tools for Children

While children are already a vulnerable group, experts (N=8) mentioned cases of neurodivergent children or children with autism who may need more help in tackling a harmful event and may not have the capabilities to take actions or to pick up appropriate social cues. Tools should be quick to activate and familiar, and should be explained to the users according to two experts. They should also consider variations in social skills and resilience (N=3), accommodating shy children who may find it challenging to assert themselves, as well as addressing the needs of children with neurodivergence.

> *"It's really difficult because no two people with autism are the same, and I think often parents are the best judge of where the child's vulnerabilities lie. [...] The parent would need to negotiate with the child and help the child understand."* [Expert P12]

> *"Some neurodivergent children might be more vulnerable to not picking up inappropriate social cues and people trying to harm them because they want to be friends, they want to fit in, and that's what they often struggle to do in the real world, so it might be a reason why they're here, and therefore they're vulnerable to using those tools."* [Expert P15]

Experts (N=3) suggested adapting safety features to the number of people co-present, with, for example, more immediate moderation tools and systems in the case of group harassment. It can be adapted depending on the configuration, i.e., a child harassed by a group of adults or a child harassed by children with easier language or more severe consequences for adults. Experts (N=2) suggested restricting the ability to communicate between unknown users via privacy settings, to only hear your friends, for instance. They also recommended the involvement of parents and guardians, especially in cases of group harassment.

> *"I think when it comes to group harassment, I'm not entirely sure we've seen all of the variants of this just yet, where it might be adults against children or mobs of children against children [...] It will be different for each child, and that's what I'm noticing."* [Expert P16]

The proposed tools should be tailored to children's age (N=9), with a focus on increased restrictiveness and enhanced comfort for younger age groups.

> *"The wording should be one that introduces features when the user signs up. If they're 13, the safety feature should be explained using one language, but if they're 10, the language should be different."* [Expert P5]

Additionally, the tools should be sensitive to the vulnerabilities present in various communities, including those related to gender-based concerns (N=1). Experts also noted that tools may need to be adapted according to diverse situations and severity of harm (N=3), personality traits (N=2), languages and countries (N=3), to ensure inclusivity and effectiveness.

> *"You could have some form of surveying on setup with what kind of things you enjoy, and there could be a logic that recommends your settings based on both age and personality. It's quite an interesting thing to dig into, but not something I've seen anyone do."* [Expert P14]

> *"I think it's mostly based on age or other characteristics such as disability or traits, a specific diagnosis, that could affect the way they interact with the world."* [Expert P5]

> *"Tools work perfectly in English, Spanish, Russian, Chinese, but when it comes to minority languages, they don't have the ability to detect the content."* [Expert P7]

One expert noted that child-friendly safeguards would be more easily adaptable for adult users:

> *"In therapy practice, explaining something to a 50-year-old as I would to a five-year-old really helps them grasp it. I think if we take a child-centred approach, it's applicable to everybody in that space."* [Expert P16]

Finding 7: Safety features should take a child-centred approach and should be adapted based on various factors: the child's age, social skills, neurodivergence or other disabilities, inclusivity, the number of individuals harassing and the severity and frequency of harm.

4.2.4.4 Reflection of Solutions for Children in Social VR

Weaknesses and Concerns of Proposed Solutions Experts noted limitations to the solutions proposed. For age verification and age grouping, as well as collecting biometrics, experts were concerned with the accuracy (N=2) and privacy issues (N=3). They also noted potential users bypassing the systems (N=2), as children want to be inclusive of older content, or malicious adults want to appear as child users. Moreover, age-grouping might lead to children not learning how to interact with adults and handle potential issues (N=1).

> *"Now, to have stronger Age-Assurance measures, you can face questions about GDPR and how you provide proof of identity. Is a platform allowed to hold this information? There are some challenges with that, but there's also no unique way to be able to identify whether someone is a child user or not."* [Expert P13]

> *"Younger people don't want to be restricted to a younger content, they want to be inclusive, so that could make users not want to do it, they would change their details to be over 18."* [Expert P4]

Solutions would need to be individualised based on the child's needs, but also generalised across countries, languages and platforms, which is a challenge (N=4). When solutions involve human moderators, buddies or parents, experts noted the need to coordinate time play (N=2).

Overall, there is a tension between costs (N=1), safety and privacy (N=2), and avoiding putting off users from using the platforms (N=1) and children with disabilities wanting to make friends (N=1). As bad actors evolve with the policy and safety features or hurdles to overcome, solutions need to be constantly updated (N=3).

> *"For the automated elements, we know how tech savvy harassers are, but we also know that oftentimes it's not as overt as the harmful behaviours. There could be codes, on the metaverse, there's lots of anime language that's used to indicate something very different, like sexual themes, and if you don't know the context, the moderation measures need to constantly be updated."* [Expert P13]

Finding 8: Safety features should aim to be accurate and constantly updated while considering privacy issues, children's engagement and potential ways of bypassing the tools. While human-involved-based solutions are crucial, coordinating time play is a challenge.

Through the lens of the victim Most solutions proposed are victim-based solutions where the victim has to take an action. Experts emphasised the need to teach the victim to self-moderate (N=3) by educating them on the safety features, teaching them to ask for help and remove themselves from a dangerous situation (N=2).

> *"(Solution safety advisor) For the victim, it would be an opportunity to learn some specific skills related to online safety and, through discussing with a virtual online safety advisor, what they could do differently."* [Expert P10]

> *"(Solution of prompts) It could be difficult for a child to initiate a confrontation with a person they know. If there is a list of pre-set phrases, that can be a way to help the child, especially if they're autistic."* [Expert P9]

Experts argued the tools should aim to support the victim quickly to feel more confident and relaxed (N=3). Children can lack maturity and self-awareness (N=4). They may be frightened of consequences if they tell someone (N=1), and they may freeze in a situation, so they need easy flag buttons or prompts to help them initiate a conversation to make the harasser stop (N=2). Bystanders or buddies assigned can help the children, as they may not be mature enough to take action (N=1).

> *"(Age-matching solution) I think it helps the victim in terms of empowerment, they can relax more in a situation where they know that adults are not in there."* [Expert P3]

> *"(Solution with buddies) With the buddy, it's kind of reassurance. [...] If two or three people go "yeah you're right, this isn't right" it really gives confidence that what you think is happening is actually happening."* [Expert P3]

Finding 9: In a situation of harassment, the victim's control is usually taken away. The solutions (safety features and education) should consider empowering the child again, giving back control while allowing the victim to be discreet. Considerations through the lens of the victim include: quick action and tools to 1. Help children to self-moderate 2. Help children who lack awareness or do not know what to do under pressure.

Through the lens of the harasser While safety features enable direct actions, they do not allow harassers to reflect on their actions, and they do not provide an opportunity to explain or apologise (N=4).

> *"(Human moderator pulled in bubble when nuclear button hit) The moderator can maybe help mediate and explain to these groups. [...] I want to allow ways for people to easily make amends, take accountability for their actions, apologise and be given the opportunity to step up and be a good community member."* [Expert P2]

> *"(New version of protective bubble pushing back harassers) You are pushed back a meter that may give the harasser a moment to consider that their actions are causing the person distress."* [Expert P16]

Experts made a distinction between intentional (N=2) and unintentional harassment (N=8). Not all harassment is intentional and may be due to a lack of self-awareness (N=8), leading to unwanted consequences (N=2).

> *"My concern is that a lot of these kids don't have a clue what's going on until it's way too late and well past the point of damage being affected, and they sometimes don't know how to get out of it. That would be my experience from a professional point of view. Once kids are locked into that abusive cycle, it becomes quite hard for them to get back out of it. Harassers rely on the victim not being able to advocate for themselves, or they make it such that the victim cannot advocate for themselves."* [Expert P15]

> *"In this case, we should really be careful to define the harasser, because even let's say an autistic person or a neurodivergent person is a person who sometimes can be doing things that are not considered normal, being very direct, for instance. So I think sometimes also the harassers should be protected in a way."* [Expert P9]

Harassment may also be intentional, and harassers may circumvent safety actions and consequences (N=8). Solutions should consider making harassers aware of the appropriate social norms, the seriousness of their actions and the consequences (N=2).

> *"If it's a deliberate thing, they will just move from that person to another target, another victim, or they might change the details to become another avatar."* [Expert P4]

> *"Unfortunately, parts of human nature are that once you realise that you can do something without repercussions, certain people will do more of it and they'll find ways to do it without repercussions."* [Expert P14]

While the core reasons of harassment might also need investigation and questioning of the design of the social VR environment (N=4) that currently allows mixing adults and children (N=3), anonymity (N=1) and group dynamics that may bring the worst or inner child out of people (N=1), safety features should detect proactively the harassment regardless of the reasons (N=1).

> *"If you were able to go steps backwards and do kind of reverse engineering and understand why they are doing it and how they are doing it, that will be the way to stop the content from the root, but we don't take the time to do that, we don't have the time to do that."* [Expert P7]

> *"I look at social VR as it brings out the child in all of us. I have seen 50-year-old executives of multi-billion dollar companies get in a headset and within minutes they are smashing pizza in your face and they are jumping and screaming."* [Expert P11]

Finding 10: Proposed solutions should prompt harassers to reflect on their actions. Considerations should include distinguishing between intentional and unintentional harassment. The solutions must detect and raise awareness of actions and their consequences, irrespective of the reasons behind the harassment.

4.2.4.5 Ways parents can be involved in social VR

Striking the Right Balance: Avoiding Over-Monitoring in Parental Involvement Experts (N=5) emphasised the importance of maintaining a balance in how parents engage with their children's use of social VR. While parental oversight plays an important role in ensuring safety, excessive control or intrusive monitoring was viewed as potentially harmful. One expert (P2) highlighted that constant surveillance could undermine a child's sense of agency, stressing the importance of allowing room for exploration and personal growth:

> *"No violation or extreme monitoring. I have seen so many parents who are doing this extreme monitoring, and they don't even tell their kids they're doing it, which to me is like an insane violation."* [Expert P2]

Instead of imposing tight restrictions, experts encouraged parents to show interest in their child's social VR experiences without infringing on privacy (N=3) *"I would say get very involved but on an interest level and a familiarity at the technology level."* [Expert P2] This requires parents to remain engaged while also respecting autonomy. Recognising that some parents may struggle to let go due to concerns about online safety, one expert acknowledged the emotional challenge of stepping back while still wanting to protect:

> *"This could be quite challenging for parents who are naturally concerned and want to maintain control over every risk."* (P8)

Education and Parental Awareness Experts highlighted the importance of parental knowledge and awareness and the ability to effectively educate children about social VR safety. A well-informed parent can better guide their child through the immersive VR environment without resorting to extreme restrictions, emphasising education rather than radical limitations (N=5).

> *"What we're trying to do is focus on education so maybe something that you don't have here is how some of these tools are educated with the parent and the child [...] Parents should be very involved in anything that a child does socially online and maybe even non-socially because there are things that the child will come into contact with that the parent should at least be aware of and have the ability to safeguard."* [Expert P14]

To achieve this, several solutions focus on educating children about online harms through relatable real-world comparisons and metaphors, such as explaining the risk of adults pretending to be children (N=1), and generally raising awareness about the dangers in social VR (N=2).

> *"We have to stop thinking that the online world is different to the offline world; we have to start thinking that they are the same because the two of them have merged in the last few years. They have merged and become one. We would have adult supervisors in a gymnasium, on a football field and in a swimming pool, and we need adult supervisors in the online world as well."* [Expert P6]

Importantly, parents need to understand social VR deeply themselves to provide meaningful support and to help children recognise potential trauma from immersive experiences (N=1).

> *"You feel like you're in the world, you're moving, so it gives you more sensation, so you are more likely to be left with the actual trauma from the interaction with users in VR. That's why parents should educate their children on how to behave and treat it as a discipline. You don't just let it happen."* [Expert P5]

Parental awareness involves self-learning and gaining knowledge about social VR safety tools and environments, allowing children to seek parental help when needed (N=4). However, some parents remain unaware of activities children do in social VR (N=2), underscoring the need for increased parental awareness.

> *"So many parents are not in the same vicinity as their child when they've got a VR headset on. They're not listening to the conversations that are taking place, and they're not necessarily watching to see what's happening."* [Expert P16]

> *"I think parents really need some more education on what goes on in social VR again and talk to people, they just have no idea of the dangers that children can be placed in."* [Expert P3]

Training resources like introductory videos explaining social VR and safety features for parents (N=1) and training videos to increase parental awareness of VR dangers (N=1) are essential. Furthermore, safety education embedded during onboarding helps set expectations and raise awareness from the start (N=1).

> *"I wondered if one way around it is actually to train parents in social VR so that they really do understand what's involved. Maybe you could provide a little training video for every game about "this is what the experience is going to like" and give them the option of either being in there with the child or just monitoring or not"* [Expert P12]

The education can be delivered not only through direct conversations with the parents but also by showing good practice with technology themselves as an example to the child and other players (N=2).

> *"[Having] conversations, demonstrating good practice yourself in terms of using technology, having fun, that kind of thing."* [Expert P8]

> *"Parental influence can not only educate the parent but also lead to more of a community ripple effect of everyone sort of being on their best behaviour because they know that there's someone there to safeguard."* [Expert P13]

Setup Experts (N=2) highlighted the importance of parents taking an active and hands-on role in the initial setup of social VR. An expert advocated for proactive behaviour, suggesting that safety settings must be *"set by a parent before the child enters the space"* [Expert P2]

Notably, P5 underscored that safety settings are only effective when controlled by the parent, rather than by the child. One expert suggested the need for "some sort of mappings" during setup that would make it easier for parents to understand and use parental controls, such as simplified configurations or suggestions tailored to the child's age or typical usage scenarios:

> *"Recommended settings, if your child is a teen or a pre-teen, then this is what we recommend."* [Expert P14]

Responsibility Experts (N=13) strongly emphasised that parental involvement is essential to ensuring children's safety in virtual spaces.

While some experts (N=3) pointed to the legal dimensions of this role, others stressed the moral and developmental obligations parents hold (N=4). Importantly, the role of parents was framed not as being a "friend" but as a guide who creates guardrails (N=1) for their child's behaviour:

> *"I don't believe that a parent should be a friend. I think that a parent is a parent and is the one who has to create the guardrails so their children don't get into trouble. The first thing that you should do as a parent, which I know that it's easier said than done, is to know what your child is doing on the apps that they are using."* [Expert P7]

An expert mentioned that parents are the ultimate guardians of a child's welfare (N=1) but others considered that safety in social VR also demands a broader collaboration (N=4): *"It is not just parents' responsibility but also community"* [Expert P13] and platform-based protections (N=3) *"Parents should be involved but it is mainly the platform's responsibility to ensure safety."* (P17). One expert (P7) suggested a 50% involvement with parental controls, indicating that balanced, cooperative strategies between parents and platforms may be the most effective approach.

From Time-Consuming and Restrictive Methods to Small Consistent Efforts Building a Trusting Parent-Child Relationship Experts noted that meaningful parental involvement is less about strict control and more about consistent engagement, communication, and relationship-building (N=12). However, some highlighted how parental involvement and engagement can be difficult for busy parents (N=3) *"I have a huge amount of sympathy for the difficulty that that could present for some parents. How can you be present doing all that stuff, working two jobs, looking after the younger siblings?"* [Expert P15] and the fast-moving development of tech makes it difficult for caregivers to stay actively involved (N=1). This was echoed by one expert describing the tension between parents being busy and staying aware of their child's VR usage.

Still, effort matters especially for the child-parent relationship (N=3). Parental involvement is most impactful when grounded in positive engagement (N=1) and relationship-building rather than restrictions (N=2). An expert noted that an ongoing, trust-based relationship between children and parents keeps them safe irrespective of environment: *"ultimately it is the relationship that children have with their parents that will keep them safe irrespective of which environment."* [Expert P10]. Parents should aim to find a balance between ensuring their child's safety and allowing them to enjoy the fun and opportunities that social VR offers (N=1).

Open communication is central (N=8). Experts strongly advocated for proactive conversations with their children (N=4) and open-ended discussions (N=2) where children feel comfortable sharing. Parents should create a safe space for the child to be able to talk (N=1):

> *"Creating a safe space where you can discuss, so they can report what happened and talk with the parent, and maybe the parent can be involved in giving meaning to what happens, without actively trying to solve the problem."* [Expert P9]

Expert P3 also suggested that parents and children should discuss restrictions transparently with negotiation between the child and the parent (N=1).

Real-time (During Incident) Experts highlighted the importance of enabling real-time parental involvement during incidents in social VR to help address safety concerns as they occur (N=10). One expert proposed that parents should receive notifications about *"potentially every flag the child makes, every user they block or every time they turn on the safety features or if moderators spot a child being harassed by someone."*[Expert P5]. Building on this, expert P5 emphasised the need for tools that allow parents to take action after receiving such alerts, offering a practical and immediate response mechanism: *"then the parent should decide to not allow the child to play that particular game anymore."* [Expert P5].

Some experts (N=6) focused on interface solutions. For example, monitoring through a mobile phone or being able to see the child's VR view (N=5). Some experts supported co-usage (N=3), where parents and children experience VR together, facilitating shared understanding and safer exploration. However, they were also cautious about overreach and excessive monitoring. Expert P12 proposed the option for parents to listen in when concerns arise, and expert (P5) noted that *"we shouldn't allow the parent to potentially enter the game and look at exactly what the child is doing, but only be involved when the child flags or blocks users because that's like a sign of concern."*.

These insights underscore the need for balanced, non-intrusive real-time involvement, giving parents the tools to respond to incidents effectively while maintaining the child's sense of autonomy and privacy.

Individualised Parental Involvement Parental involvement in social VR varies greatly on an individual basis (N=9), reflecting a spectrum from highly engaged parenting to raising children as independent citizens who are granted freedom and responsibility (N=1, P8). This involvement is age-dependent (N=9); younger children, especially under 13 years old (N=2), are generally seen as more vulnerable and open to parental support (N=3)

> *"The younger a child is, the more immature or vulnerable the child, the more they're going to need trusted adults around them to support them, whether that's online or offline."* [Expert P8]

As children grow older, they seek greater autonomy and independence, which parents should recognise and respect (N=2). Teenagers, in particular, are expected to have more freedom in their social VR activities, reflecting their distancing from parents and desire for self-governance (N=2). However, beyond chronological age, some experts (N=2) emphasise that parental involvement should also consider the child's stage of maturity and interest in technology, not just their age alone (N=2). This approach allows parents to tailor their supervision and guidance based on the unique needs and readiness of each child.

> *"Based on not just age but also their level of maturity and interest in technology, because sometimes we just use age and we know that age is not always the solution, like thinking of the cut-off age of 13."* [Expert P10]

> *"As children get older or grow in Independence and maturity, then naturally they're looking to make their own decisions and have their own autonomy over what they do."* [Expert P8]

Finding 11: All experts agree that parental involvement is essential for child safeguarding in social VR, but safety is a shared **responsibility**. Experts call for a **balanced approach**, with **parental knowledge and awareness** central. The **Setup** is critical. Trustful **parent-child relationships**, open dialogue, and real-time alerts **during incidents** are suggested. Rather than depending only on a child's age, involvement should be **adapted to each individual**, considering factors such as experience and maturity.

4.2.5 Discussion

A reflection across studies from Part II can be found in chapter 5 to situate this study's findings within the wider contributions of the thesis. Here we present a summary of the findings and references to corresponding guidelines in chapter 9.

4.2.5.1 Summary of Findings

S2-RQ1: Experts' perspectives and considerations regarding existing social VR safety features Of the ten safety features presented, experts most frequently ranked personal space bubbles, blocking and muting, and human moderation among their top three for protecting children in social VR in both one-to-one and group harassment. Experts agree that safety in social VR for children is most effective when it builds on their existing social media experience and uses real-world metaphors to make safety concepts familiar and relatable, or for parents to explain risks using real-world metaphors. Major design concerns of existing safety features include inconsistency across platforms. According to experts, human moderators can play an essential role by providing authority, flexibility, and an understanding of social VR norms. However, maintaining a 24/7 human moderation team is often impractical due to staffing and cost constraints. A more scalable alternative is automated moderation.

S2-RQ2: Experts recommendations and solutions As the responsibility is usually placed on children to enact reactive protection measures, who may have limited ability to manage overwhelming situations, experts recommend safety features that are quick and easy to use, discreet, supportive, and empowering, with options like panic buttons, preset prompts, blocking multiple users during group harassment and reflection tools such as pause features. Moreover, safety systems should be continually updated to ensure accuracy, privacy, and inclusivity while remaining resistant to misuse. These should adapt to children's individual needs, not only based on age but also on other factors such as social and cognitive abilities, as well as the severity of incidents. While solutions should support victims, they should also encourage harassers to reflect on their actions and understand consequences, whether their behaviour was intentional or not. Finally, parental involvement is considered critical, but should form part of a shared responsibility for safety. A balanced approach is encouraged, focusing on parent–child communication, informed setup, and real-time awareness during incidents, tailored to each family's needs.

(While some findings may affect all users, this does not diminish their relevance or value for safeguarding children.)

Based on the findings, I present practical design guidelines in section 9.2, specifically guidelines 1-5, 9-10, 12.

4.2.5.2 Comparing Expert Recommendations and Proposals with Prior Work and Emerging Social VR Safety Solutions

The findings align with points from prior work, such as the challenges presented by safety feature misuse and confusing tools in social VR [50]. The authors highlighted that confusing design elements can lead to safety risks and unintended consequences. The findings similarly advocate for familiar and real-world child-aware concepts for the design of safety features. Moreover, experts noted the potential of "votekick" and "trust rank" to be abused or misused, especially in group harassment scenarios. As mentioned previously, a study by Chen et al. (2024) also suggested that such safety features can create hierarchies or stereotypes that can foster discrimination [310].

Prior work has called for quick-to-activate safety measures, such as gesture-based features (e.g., covering ears or pushing away to trigger protective actions) [50]. While experts' recommendations echo this need by advocating for tools that are quick and easy to use under pressure, experts also recommend making safety features discreet for children to prevent escalation. The latter is a novel consideration not prominently addressed in previous literature.

Recent tools introduced by platforms, such as Horizon Worlds' wristband "Pause" feature [205] and VRChat's "Hiding All Avatars" option [204], resonate with the findings about empowering users to regain control, especially in group harassment scenarios. Furthermore, the recent advancements in automated voice moderation from Rec Room, improving in accuracy [311], align with experts' emphasis on continuous updates and accuracy in safety features.

Both the findings of this study and related work [40] emphasise the need for human involvement in moderation to create safer online spaces. This aligns with findings from experts who advocate for the involvement of human moderators, buddies, and parents via monitoring or SOS button to call a human. Yet maintaining continuous human moderation presents high cost and scalability constraints, as noted in prior research [19] and mentioned by experts. Consequently, human expertise should reinforce and complement automated approaches, which experts highlight as valuable for their scalability and cost-effectiveness in managing the tension between profit and safety.

4.2.5.3 The Many Routes to Harm: Intentional and Unintentional Harassment, and Unwanted Consequences

In this section, I discuss and reflect on the various types of harassment and victimisation identified by experts, which may influence the design of interventions in social VR.

A. Intentional harassment is facilitated in social VR due to anonymity, the lack of social norms, and the mix of adults and children, as well as associated peer pressure. Users may differentiate the online world from the real world as a place of freedom without rules, potentially bringing out the inner child or the worst in people. Experts emphasised the need to diminish that difference between the offline world social interactions and social VR interactions by, for example, establishing similar social norms and expectations around safety.

Moreover, there is a need to address the root causes of such harassment. The causes of bullying in a school were investigated via a survey with 176 Swedish students [312]. Teenagers attributed 69% of the reasons to the bully having inner flaws, social positioning or hostile feelings; 42% were attributed to the victim being deviant, weak or irritable, and 21% were attributed to peer grouping. Adult users may be harassers due to past trauma and deliberately aim to harass users. Experts noted the need to understand the reasons for harassment occurring so that they can be addressed. In contrast, experts mentioned the need to develop safety features that proactively detect and support reactions to harassment.

And intentional harassment is often a group affair. Safety features such as "vote-to-kick" and "trust rank", initially designed to maintain safety within social VR platforms, run the risk of being used against other users. A recent paper showed that the trust ranks can also lead to stereotyping, discrimination and conflicts among users [310]. Users, especially as a group experiencing peer pressure, may exploit these features to target a child without any valid reason or evidence of inappropriate behaviour, leading to an unfair and potentially damaging situation. The subjective nature of these tools introduces an element of bias into the safety measures. Users might base their judgements on personal opinions, social dynamics, or even misunderstandings, making the system prone to false accusations and creating an environment where individuals are unfairly targeted based on subjective criteria, or indeed fear targeting by larger groups.

B. Unintentional harassment was described by experts in the context of a neurodivergent child not understanding social cues, referring to situations where the child inadvertently engages in behaviours that may make others feel uncomfortable, offended, or harassed. Neurodivergent individuals, such as those with autism spectrum disorder, may face challenges in interpreting and responding to social signals in a typical manner. For example, a neurodivergent child might not recognise personal boundaries or understand non-verbal cues, leading to unintentional actions that others perceive as intrusive or inappropriate. Consequently, the child on the receiving end of the interaction might resort to blocking or reporting. In such cases, the neurodivergent child may miss the opportunity to learn, apologise, or adjust their behaviours.

C. Unwanted consequences for using a safety feature by a child would need to be considered when developing safety features. This could arise when a child uses a safety feature visible to the other users, such as the gestures introduced in Rec Room, when asking a human moderator for help in front of others or when appearing as blocked or reported. Design and implementation considerations should ensure that safety features do not unintentionally lead to visibly exposing the victims' actions in a way that might lead to further victimisation. In some instances, the child might simply be shy and prefer not to draw attention to themselves, resulting in them choosing not to use safety features.

Based on these findings, I present practical design guidelines in section 9.2, specifically guidelines 1, 2, 12.

4.2.5.4 The People Around the Child User: Harasser(s), Human Moderators, Friends, Bystanders and Parents

As described in the scope section 1.3, the child social VR user ecosystem encompasses a variety of individuals and entities that need to be considered, and experts suggest that safety features should account for these different groups. In a harassment scenario, the first direct entity is the harasser. Factors such as the number of harassers (group VS one-to-one harassment) and demographics (e.g., adult or child) are important when designing safety tools to shield the child effectively.

Human moderators are also considered crucial. However, due to costs, scalability and the typical tension mentioned between safety and profit, Human-AI solutions may be considered as an alternative solution [40].

Additionally, internal (in-VR) users such as bystanders and friends, and external (outside VR) stakeholders such as parents can play an important role in ensuring the child's safety. Further research should consider the role of bystanders and friends, for example, finding ways the platform can encourage them to become active upstanders or setting friends as "buddies" in case of a disruptive event, as experts suggested.

Based on these findings, I gather practical design guidelines in section 9.2, specifically guidelines 3, 5, 9, 10.

4.2.5.5 The (Victim) Child's Individual Needs: Age, Social Skills, Language, Vulnerabilities and Personality Traits

Safety features should enable child users to control their own boundaries based on their needs and preferences. They should also be inclusive to vulnerable populations, different countries and minorities.

Based on these findings, I gather practical design guidelines in section 9.2, specifically Guideline 4.

4.2.6 Conclusion

Drawing from interviews with 16 experts in child online safety and psychology, the study informs the factors contributing to the effectiveness and weaknesses of ten existing safety features from two of the most popular social VR platforms. Experts propose solutions for two given scenarios to aid experts in considering the potential context of harassment situations towards a child and to gain practical and concrete insights.

Transition to Next Study

I now have a better understanding of the benefits and limitations of existing safety features for children in both one-to-one and group harassment contexts, as well as expert-driven recommendations for enhanced safety and child-friendly tools (structured guidelines in section 9.2). However, there has been no validation or evaluation involving children, and no user-centred research has been conducted to assess their perceived effectiveness in practice. This motivates follow-up work to capture children's perceptions of tools in simulated social VR to provide a more complete understanding to answer the thesis **(RQ2: What design considerations can help shape safety features to address child harassment in social VR?)**.

4.3 Study #3 - Children's Perceptions of Blocking Strategies for Group Harassment in Social VR

4.3.1 Motivation

Harassment in social VR can occur as collective behaviour, where groups coordinate their actions. Recent events highlight the emergence of group harassment in social VR [22, 272, 20, 194, 313, 199, 198, 200, 201], "And more times than not, there would be one guy or a group of guys engaging in harassing behaviour because that's how they enjoy their weekend." [200]. A young girl under 16 was sexually attacked by a "gang" of adult men in 2024 [198]. In December 2021, a metaverse researcher's avatar was virtually raped by a group of male avatars [199]. An instance of a "flash mob" occurred in VRChat in January 2018, where echidna avatars known as "Ugandan Knuckles" would repeat a single phrase phonetically spelt "do you know de way?" [201]. This can have an amplified psychological impact than one-to-one harassment, leaving victims feeling outnumbered and powerless [31, 20, 314]. Although social VR platforms have been implementing safety features like blocking, reporting, and personal space bubbles [204, 187, 205], it is unclear how suitable existing tools are for children facing group harassment; and, as an expert mentioned in study #2:

> *"Just having the potential to block one avatar at a time is something that is going to have to change in social VR."* [Expert P6]

With blocking one of the most common safety tools in social VR [207, 315] and rated in the top 3 of the experts in Study #2, I address the following research questions:

S3-RQ1 How do children perceive the usability, effectiveness, and enjoyment of the three developed blocking techniques–*[SelectBlock]*, *[MassBlock]*, and *[PointBlock]*–during simulated social VR group harassment?

S3-RQ2 How does the harasser group size affect children's perceptions of usability, effectiveness, and enjoyment of the blocking techniques?

S3-RQ3 How do the child's intervention perspectives (IP) relative to the group of harassers influence those measures?

4.3.2 Related Work

This section synthesises additional related work on the specific topics addressed in this study. (A more comprehensive literature review is presented in chapter 2, with specific relevant sections: section 2.3.2 and section 2.3.3. In terms of methodology, the study relates to section 2.4.2 and section 2.4.3).

Group harassment is a prevalent issue in social VR, combining collective aggression with the immersive nature of virtual environments (detailed in section 2.3.3.2).

4.3.2.1 Group Size as a Factor in Evaluating Online Safety Tools with Children

While each person's harassment actions may seem minor, the cumulative effect can be overwhelming, as the victim is subjected to a greater volume of negative comments [314]. A study by Kim et al. [314] has shown that the extent to which a scale exceeds what an individual can effectively manage can be overwhelming and create a sense of helplessness. Networked harassment becomes much harder to intervene with the most common safety tools available [314]. One of their study's findings showed that providing scalable response measures such as mass-response features was effective in reducing anxiety and the feeling of helplessness in networked harassment [314].

4.3.2.2 Intervention Perspective as a Means of Diminishing Impact of Harassment for Children

VR's affordances, such as immersion and embodiment, amplify the emotional and physical toll of harassment. For example, "phantom touch", the phenomenon of feeling touch despite not receiving cutaneous stimulation (touched on the skin) [316], increases the impact of harassment, making the experience feel more real [316].

Safety tools that focus on increasing distance, such as "StayAway" [207], are particularly effective as they reduce the auditory and physical impact of abuse, provide victims with a clearer overview of the harassers' actions, and remove them from the immediate source of harassment. Among other tools evaluated, "StayAway" was the preferred response strategy in social VR due to its high efficiency, lack of restrictions, and freedom of operation, which allows users to leave the room at any time [207]. While the study has explored tools through surveys and interviews and highlighted the importance of distance as a factor, there has been no investigation into the role of elevation perspective, particularly with children and under controlled experimental

conditions. Prior work [317] demonstrates that spatial distance cues can significantly affect emotions and evaluations, suggesting that exploring elevation in VR could reveal how changes in spatial perspective, unique to 3D environments, influence the experience of harassment.

The psychological benefits of such tools align with the *"Construal Level Theory"*, which demonstrates that increased physical and psychological distance diminishes the emotional intensity of distressing experiences [318, 319].

Furthermore, non-confrontational distancing strategies are preferred by users because they allow victims to extricate themselves quickly and effectively without compromising their identity or interests [207]. A study in social media also showed that participants valued safety features that allowed them to distance themselves from the networked harassment or cyberbullying, both practically and emotionally [314].

4.3.2.3 Gaps

First, effective solutions to mitigate group harassment in social VR are lacking, and there is a noticeable absence of empirical investigations involving children to safely examine the use of existing tools in group harassment (see section 2.3.3.2). Second, research has overlooked group sizes in the context of harassment in social VR, potentially limiting the effectiveness of interventions (see section 4.3.2.1). Third, there has been no investigation into the role of elevation perspective, particularly with children in social VR (see section 4.3.2.2).

By introducing new blocking tools, exploring their effectiveness across small and large groups of harassers, and examining different intervention perspectives, this study aims to gain new insights into the under-researched area of designing child-focused safety interventions in social VR group harassment.

4.3.3 Methodology

The study was conducted in April-May 2024 with **40 children**.

4.3.3.1 Study Design

This study focuses on the child perceptions and preferences among 1) three blocking strategies, 2) two intervention perspectives (IP), and 3) two group sizes of harassers.

Variables: Three Blocking Tools The first tool *[SelectBlock]* allows to block one individual at a time from a list of surrounding users. Blocking one individual at a time is standard in most social VR platforms [187, 204, 205, 320] and was thus regarded as a baseline. I developed two additional blocking tools in group harassment situations: *[MassBlock]*, allows users to block everyone around them and then unblock specific individuals from a list (inspired by the "Mute All" from Rec Room [187] and "Pause" from Meta Horizon Worlds [321], aimed at taking a moment away from other people); *[PointBlock]*, enables selective blocking by pointing and "shooting" at avatars using a Magic Wand (inspired by the "Pointing feature" from Rec Room [187]), leveraging the predominant point-and-select raycast interaction technique commonly used in VR applications.

Two Group Sizes

The small group had three harassers, each simulating one of the three social VR harassment types (verbal, environmental and physical [31]). The large group had nine harassers, was a scaled-up version of the small group, tripling its size. As supported by research on networked harassment (section 4.3.2.1), safety tools designed for single harassers (e.g., one-user-at-a-time blocking) may fail to scale effectively for larger groups. Choosing groups of three and nine harassers creates a progression that allows investigating the impact of increasing group size. It allows me to examine how different intervention designs are perceived by children under varying pressures, ensuring that tools are effective in different situations.

Two Intervention Perspectives Given factors described in section 4.3.2.2, the IP was chosen as an independent variable to evaluate how different intervention strategies contribute to mitigating the harmful effects of group harassment, particularly in settings where children are vulnerable to amplified discomfort due to group harassment. I compared an elevated IP (above the group) with one at the ground level (figure 4.4) to understand how visual perspectives may influence the child's perceptions of the harassment scenario. An elevated perspective may decrease negative emotional valence and arousal, as well as increase the sense of safety. The above IP was developed to keep the user at the same distance from the child as when on the ground level.

Experimental Design *Tools* and *Group Size* were treated as within-participant variables with a 3 (*Tools*) x 2 (*Group Size*) design. In contrast, *IP* was evaluated for only one tool, *[SelectBlock]*, based on insights from the pilot tests, using a 1 (*Tool*) x 2 (*Group Size*) x 2 (*IP*) design.

The conditions were as follow:
(**C1**) Tool: *[SelectBlock]*, IP: *Ground*
(**C2**) Tool: *[MassBlock]*, IP: *Ground*
(**C3**) Tool: *[PointBlock]*, IP: *Ground*
(**C4**) Tool: *[SelectBlock]*, IP: *Above*

The two group sizes were counter-balanced, either starting with the small group (N=20) or the large group (N=20), and the three tools were counter-balanced using Latin-Square for the first three conditions: Order 1 [C1,C2,C3,C4] N=8, Order 2 [C1,C3,C2,C4] N=7, Order 3 [C3,C1,C2,C4] N=7, Order 4 [C3,C2,C1,C4] N=6, Order 5 [C2,C3,C1,C4] N=6, Order 6 [C2,C1,C3,C4] N=6.

4.3.3.2 Social VR Scenarios and Tools Implementation

The simulated social VR scenarios used to conduct the experiment were developed using *Unity* [322] (figure 4.2) (Videos in Supplementary Material STUDY 3/VIDEOS). The virtual scene was a house from an existing asset [323] and had eight avatars in the background from *Ubiq* Library [324], animated and talking to each other using a background audio of a small crowd talking [325], mimicking a social virtual environment. I aimed to reproduce each of the three types of harassment for each of the

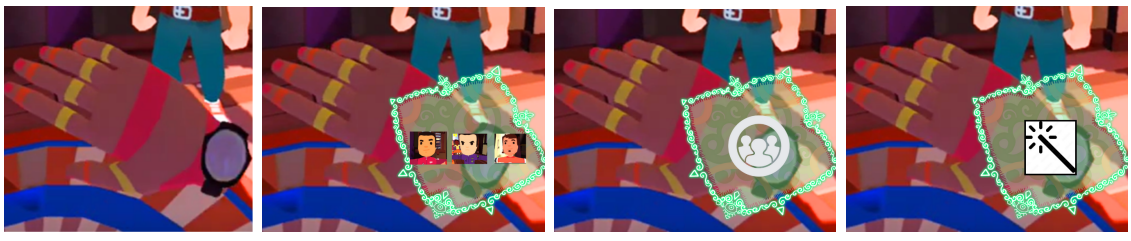
three harassers in the small group (replicated times three times in the large group) in social VR [31]: *verbal* (harasser saying 'loser' and mocking), *physical* (throwing virtual objects and surrounding the victim) and *environmental* (displaying a picture of the victim with added drawings as a form of mockery).



(a) (b)

Figure 4.2: (a) Three harassers (small group), one throwing objects at the victim, another showing a drawing of the child, and the last saying "Loser", with mocking laughs, (b) Nine harassers (large group) replicating the three types of harassment.

Finally, I developed the blocking tools, accessible through a virtual watch, similar to the one on Rec Room [187] (see figure 4.3). As participants look at the virtual watch on their hand, a green diamond appears with the button corresponding to the condition's blocking tool: *SelectBlock*: the profile pictures of the avatars appear on the watch as buttons to click and block one individual at a time, *MassBlock*: one button allows users to block all avatars, a list to unblock avatars then appears, *PointBlock*: a button of Magic Wand appears, the participant clicks, the wooden wand appears, and the participant shoots to block. In the practice session, all three blocking tools appear on the watch to practice using the tools.



(a) Virtual Watch (b) SelectBlock (c) MassBlock (d) PointBlock

Figure 4.3: As participants look at the virtual watch on their hand, (a) a green diamond appears with the button corresponding to the condition's blocking tool: (b) SelectBlock (profile pictures of users to block each individually), (c) MassBlock (block all users, unblock list appears) or (d) PointBlock (Magic Wand to point towards avatar and block). (Researcher-Developed using Unity.)



Figure 4.4: User's perspective in the large group harassment scenario (Ground Level IP VS Above IP). The left column shows the scene in a profile view, where the white camera represents the user's point of view. The right column shows what the user sees through their eyes in the large group harassment scenario.

4.3.3.3 Pilot tests

I conducted pilot tests, also under ethical approval, with four adults, followed by tests at a local library with four children aged 8-16. The experiment was considered too demanding, with 12 scenarios (3 tools x 2 group sizes x 2 IP), so I streamlined the design to focus on one tool when changing the IP. Despite being third-party observers, most children empathised with the victim, with some showing emotional reactions to the simulated scenarios.

4.3.3.4 Procedure

After refining the experiment based on the pilot tests, I proceeded with the finalised experiment. Children, either individually or in pairs and accompanied by an adult (like a parent or teacher), participated in the activity in an empty room, with the adult waiting behind a screen. The children signed a consent form upon arrival.

Before entering the VR scenarios, children were shown slides introducing social VR, bullying, and the concept of blocking (Supplementary Material STUDY 3/SLIDES). I explained the three blocking tools. Using a Meta Quest 2, they practised using the three tools in a simulated social VR environment where no harassment took place,

and static avatars kept reappearing to be blocked. Once comfortable with all three tools, children were informed that they would observe a child being bullied by both small and large groups of harassers. To avoid direct involvement, they were asked to imagine themselves in the victim's position and reflect on their feelings, reactions, and tools' effectiveness. This allowed them to act as third-party observers of the bullying, rather than experiencing the situation first-hand, but also be first-person users of the tools. This approach draws on research showing that empathy, defined as "emotional responses aligned with the emotions of others", emerges in infancy [270, 271], suggesting children should be able to emotionally engage with the victim's perspective even as third parties.

They experienced four rounds in each group. Following each round, children completed a questionnaire featuring the custom-developed Likert scales on a laptop. Sessions concluded with a 5-minute interview to gather participants' overall preferences and emotional experiences during the rounds.

4.3.3.5 Measures

I used both quantitative and qualitative methods to understand children's perceptions to the safety tools. Quantitative data involved ranking their responses to specific questions within the research scope, providing measurable insights. Qualitative data, however, gave children more freedom to reflect, offering deeper, more nuanced feedback. Combining methods added context and depth to the quantitative findings, as qualitative data provides valuable nuance [326].

Quantitative I assessed children's perceptions using a modified version of the Self-Assessment Manikin (SAM) scales, which I refined for children due to previous concerns about its clarity [327]. I improved features using colours [249], emojis [248] and avoided abstract concepts like the neutral point [250]. I also employed custom Likert scales to evaluate tool enjoyment, helpfulness, and ease of use, incorporating emojis and colours as well (Supplementary Material STUDY 3/QUESTIONNAIRES). After each round, children were asked to fill in the Likert scales. At the end of the first three conditions, they ranked the tools from the one they liked the most to the one they liked the least. And after the fourth condition, they were instructed to rank the IP.

Qualitative At the end of the experiment, I conducted short (5 minutes) semi-structured interviews:

- Can you describe how you felt putting yourself in the bullied child's (the victim's) shoes?
- If you were the bullied child in the middle, which tool would you use? Why?; What did you think about all the tools?
- Would you want to be moved up (seeing the group from the top) or further away when using tools? Why?
- How did you feel when there was the small group VS large group?
- What else could we do to help if you were being bullied in social VR?

4.3.3.6 Analysis

Quantitative Data were analysed using R. I used ARTool for Aligned ranks transformation (ART) ANOVA [281], a non-parametric approach for multiple independent variables, interactions, and repeated measures [281, 328] (Supplementary Material STUDY 3/R MARKDOWN). I set the significance level at 0.05, and interpreted effect sizes using partial eta squared η_p^2 with small effects (< 0.01), medium effects (≈ 0.06) and large effects (≈ 0.14 or higher) [286, 287].

Qualitative I used inductive thematic analysis [39, 283] to analyse audio-recorded interviews of children. After generating transcripts, I listened to recordings (∼5min each) and corrected mistakes in the automatically generated transcripts. I familiarised myself with the data before creating a coding scheme using *QCAmap* [308], line by line. Codes generated are words or short phrases that describe any new idea. Two other researchers reviewed 50% of the coding schemes (agreeing/refining codes) in an Excel sheet. The coding scheme was then updated by the initial researcher. All three researchers then developed a set of higher-level categories by bringing together related codes, discussing, and refining themes to generate a rich description (Supplementary Material STUDY 3/CODING_SCHEME).

A Miro board gaterhing themes can be viewed via `https://miro.com/app/board/uXjVKq8aQIg=/`.

4.3.3.7 Participants

Recruitment This research was conducted in English and in the UK. The children (8-16 years old) were recruited via their parents/legal guardians on a voluntary basis and required the parent/legal guardian's permission to participate. The study was advertised via university forums and mailing lists, social media and the local library. Consent and demographics questions were completed by the children's parents/legal guardians before the session. On the experiment day, children signed a consent form and were told they could take breaks or stop at any time. I assured participants that no wrong answer can be given and assured them of the confidentiality of their responses.

Demographics The demographics were provided by the parent or guardian of each participant. The study included 40 children (from 33 families) aged 8-16, with a mean age of 11.2 years (SD = 2.3), comprising 14 females and 26 males. There were three Arab, six Asian, one Asian British, one South Asian and Scottish, two South East Asian, 24 White/Caucasian, two mixed White & Latino, and one participant who preferred not to disclose ethnicity. table 4.6 summarises children's online gaming, VR and social VR habits.

Table 4.6: Participants' Online Gaming, VR, and social VR Habits (33 families, N=40 children)

Measure	Mean (SD)	Scale
Online Gaming Frequency	4.32 (1.28)	0 (Never) - 5 (Daily)
VR Usage	2.16 (0.97)	1 (None at all) - 5 (A lot)
social VR Awareness	1.93 (1.67)	0 (Not familiar at all) - 5 (Very familiar)
social VR Usage	1.44 (0.73)	1 (None at all) - 5 (A lot)
Households with VR Headset	6 households	Yes/No

Ethics and Compensation The protocol was approved by the university CoSE ethics committee (Approval number: [300220209]). I was registered with the national Protecting Vulnerable Groups (PVG) scheme to conduct this study with children. Children were exposed to 2-3 minutes of continuous VR at a time, following prior work [182] and child VR usage guidelines [255, 256]. A £10 Amazon voucher was given to the accompanying guardian, while children received a certificate as a token of appreciation. Documents with the information sheet, pre-survey with consent form and demographics questions can be found in the Supplementary Material.

Selection and Participation of Children Before participation, the details of the studies were explained to the children's legal guardians/parents via an information sheet sent after they expressed interest. A digital consent form was also provided to be signed before coming to the session, noting that the child's participation (aged 8 to 16 years old) was voluntary. At the start of each session, I explained the procedure to the child using slides (see Supplementary Material STUDY 3/SLIDES). The child was then given a paper consent form to sign, where they could agree or disagree (by ticking boxes), to be audio recorded for the interviews, and were informed that they could take breaks or stop at any point. I assured them that they could ask questions and that there were no wrong answers. Participants were found via university forums and mailing lists (researchers or lecturers with children), social media and the local library. All participants' personal data were stored securely, and all personally identifiable data were removed.

4.3.3.8 Limitations

For ethical reasons, children did not experience the harassment scenarios from a first-person perspective, but were instructed to imagine themselves in the victim's role instead (see section 4.3.3.4). This approach draws on research showing that empathy, defined as "emotional responses aligned with the emotions of others", emerges in infancy [270, 271], suggesting children can emotionally engage with the victim's perspective even as third parties.

Another limitation is the potential for some participants to feel overexposed to the group harassment simulations. Safeguarding measures were implemented if any child showed significant distress, including regular assessments of their emotional state and ensuring they could stop at any time. However, this did not happen during the studies. I also added a pause button in the VR scenarios, allowing users to pause quickly. Some children used it to recall how to use the blocking tools.

While repeated exposure to the same group harassment scenarios could lead to habituation, I prioritised consistency across conditions to avoid introducing new variables. To mitigate this, I used a balanced Latin Square design to counterbalance conditions, ensuring that any observed differences were due to the experimental conditions, not repeated exposure.

While this study does not assess the tools in existing social VR environments, where harassment behaviours may be less predictable, it still offers valuable insights into children's preferences and perspectives on these tools.

Although the study was conducted in English, within the UK, and most children were unfamiliar with social VR, the insights can inform the development of interventions tailored for new child users, who may lack experience and have greater needs compared to more experienced users.

4.3.4 Results

Since the analysis encompasses numerous themes, table 4.7 provides an overview of the categories and themes to help navigate the findings. table 4.10 presents a summary of the statistical results of the quantitative data. Where available, qualitative data is paired with corresponding quantitative data.

Table 4.7: Qualitative Findings: Categories, Themes, Sub-Themes with Counts, gathered from child interviews at the end of the user study.

Categories	Themes	Sub-Themes	Count
Category 1: Perceptions of the Simulated social VR Group Harassment Scenarios	**Theme 1:** Assessing the Validity of the Simulated Social VR Group Harassment Scenarios	Negative Emotional Valence Responses to the Group Harassment Incidents	32
		The Impact of Group Harassment in social VR	26
Category 2: Attitudes Towards the Proposed Blocking Tools	**Theme 2.1:** Positive Aspects	Impact of Usability on Child User Performance and Interaction	32
		Positive Child User Perceptions of PointBlock	14
		Psychological and Physiological Benefits from Effective Tools	8
	Theme 2.2: Negative Aspects	Usability Challenges and Design Limitations	23
		Unintended Negative Consequences of Tool Use	14
		Balancing Practicality with Creative Expression in Tool Design	5
Category 3: Impact of Visual Perspectives of Harassment Scenarios: The Above View and Further Away	**Theme 3.1:** Effect of Being Elevated on Child User Experience	Enhancement of Sense of Safety	15
	Theme 3.2: Usability and Dynamics of Being Elevated or Further Away from the Group Harassment Situation	Advantages	25
		Disadvantages	6
		Empowerment and Control when Elevated or Further Away	11
Category 4: Proposals for Improving Safeguarding Tools in social VR	**Theme 4.1:** Leveraging VR-Specific Features	Use of Physical Objects and Non-Verbal Cues	7
	Theme 4.2: From the Victim's and Harasser's Perspectives	Proposals that Impact the Victim's Perception of the Harasser	9
		Strategies to Influence and Alter the Harasser's Experience	3
		Suggestions for Enhancing the Victim's Perception of the Environment	1
	Theme 4.3: Seeking Assistance: Proposals for Human or AI Support Systems		8
	Theme 4.4: Usability and Practicality Considerations for Tool Improvement		3

4.3.4.1 Category 1: Perceptions of the Simulated Social VR Group Harassment Scenarios

Theme 1: Assessing the Validity of the Simulated Social VR Group Harassment Scenarios Negative Emotional Valence Responses to the Group Harassment Incidents As participants had a third-party perspective, positioned behind a child being bullied by the group, I asked them to share how they were feeling in the scenario, to assess the validity of the proposed methodology (See Limitations section 4.3.3.8). Children (N=32) described experiencing negative emotions when facing the simulated group harassment, prompting low arousal emotions (N=19) like being sad (N=15), hurt (N=2), uncomfortable (N=1), worried (N=1) and disrespected (N=1). *"I felt a bit sad. I imagined myself in the child's shoes." [Child21]* It also led to more intense high arousal reactions[2], including being annoyed (N=12), angry (N=4), scared (N=3), anxious (N=2), overwhelmed feeling surrounded (N=2) and intimidating (N=1). *"It's quite intimidating having all the people shouting." [Child26-B]. "I felt quite scared because everyone was laughing." [Child37]* The scenario also created a physiological impact (N=1): *"I felt calmer in the small group. In the large group, my heart would be beating really fast." [Child19-A].* Children were asked to put themselves in the victim's shoes when blocking the bullies. Children (N=19) managed to imagine being the victim when intervening, while others felt more like an observer (N=9), while still empathising (N=2) *"I did feel what the child was feeling, but it felt a bit limited because you're not actually seeing through their exact eyes." [Child40]*;

The Impact of Group Harassment in Social VR Children noted the amplified psychological impact of being harassed by a large group (N=26). They felt that the large group was more intimidating and less safe than the small group (N=26), amplifying the feeling of being harassed (N=1) and more specifically amplifying the verbal and physical harassment (N=1), *"The large group compared to the small group was more annoying because there were more people and more things thrown at the child like oranges and cards." [Child17].* In the smaller group of harassers, children (N=3) felt a sense of closeness where the harassers were compared to unpleasant friends: *"The large feels more ganged up, a bunch of people targeting you, rather than just a group of horrible friends." [Child12]*

Children (N=15) frequently mentioned the increased difficulty of blocking in the large group compared to the small group, noting that a higher number of harassers made blocking more challenging. *"The small group was easier because there were fewer people to take care of, but in the large group, there were more people, and it was harder." [Child24].* Therefore, children expressed wanting different tools to address different group sizes in harassment in social VR (N=25). For example, children would use the *[MassBlock]* if the group of harassers is large (N=8). Others (N=3) found that tools performed equally well across small and large groups in social VR harassment situations.

[2]Note that Safeguarding measures were put in place in case the child's negative emotional valence persisted, ensuring they could stop at any time.

This is also supported by the quantitative data. The two-way ART ANOVA with the independent variables as Group size and the first three Conditions (on ground level), with the emotional and perception ratings, revealed several significant effects of the size of the group across all Likert scales. Negative emotions and negative usability scores were significantly higher in the large group (see table 4.10). I suspect this is due to the amplified sense of pressure experienced in larger groups [314], which increases negative emotions and increases the difficulty of tool usage. The results are in line with children's feedback on the impact of the group harassment scenarios, suggesting that group dynamics play a critical role in shaping emotional and usability outcomes of safety features.

Finding 1: Children reported negative emotional responses to the simulated group harassment scenarios, with some experiencing physical reactions like a racing heartbeat. Many were able to empathise with the victim, imagining themselves in the child's position, although some felt more like observers. The large group of harassers was seen as more intimidating compared to the small group, making it harder to block individuals effectively. Quantitative data supported these findings, showing stronger negative emotions and usability challenges in large group settings.

4.3.4.2 Category 2: Attitudes Towards the Proposed Blocking Tools

Theme 2.1: Positive Aspects of the Studied Tools 1) Impact of Usability on Performance and Interaction: Children reported that some tools' aspects positively affected their performance by making it easier to block (N=32). For the *[PointBlock]*, specific advantages included ease of use (N=18), speed (N=9), accessibility (N=2) and the ability to block by simply looking around rather than navigating a menu (N=1). *"I would use the wand because it's so easy to completely eradicate the bullies" [Child8-A]"*; *"I like how the wand, wherever you put it, it just floats. So if anyone else starts being rude, you can immediately get it and block them." [Child4]*

With respect to *[MassBlock]*, children appreciated its efficiency in terms of speed (N=6) and found it straightforward to use, likening it to "switching off" (N=4). *"I would use just the [MassBlock], it's the same as just switching it off because no one can actually come in unless you invite them back." [Child33-A].* The *[SelectBlock]* was also noted for its speed in smaller groups (N=2).

2) Positive Child User Perceptions of PointBlock: The Magic Wand *[PointBlock]* was praised for its enjoyability (N=14), with many finding it fun to use (N=13). It was also recognised for its innovative and engaging design (N=1), visual appeal (N=2), and its interactive and imaginative approach to blocking (N=1). *"It's kind of fun to use, and it's kind of you saying like 'Abracadabra'. What you shoot out looks really cool when it flies." [Child3].*

This is also supported by the quantitative data. From a two-way ART ANOVA with the independent variables as Group Size (Small VS Large) and Tools (on ground level C1,C2,C3), there was a significant difference in terms of enjoyment between the three different tools (see table 4.10). While *[PointBlock]* enjoyment mean is higher (4.8, SD=1.2) compared to *[MassBlock]* (4.5, SD=1.2) and *[SelectBlock]* (4.5, SD=1.2),

the post-hoc test did not reveal significant differences between individual pairs. The enjoyment of tools was a key factor for children. I suspect this finding may be driven by personal preferences or the design of the tools. The shooting mechanic with a wand and engaging visual effects, which particularly resonated with some children, is reminiscent of popular games like *Fortnite* [329], *Hogwarts Legacy* [330], *Splatoon* [331], and *Wands* [332], where shooting is involved. Although no significant differences emerged between tool pairs, the overall enjoyment of tools remains a key factor for children, influencing their perception and interaction with the safety features.

3) Psychological and Physiological Benefits from Effective Tools: *[MassBlock]* was particularly effective in providing relief, as it alleviated the emotional burden of managing multiple harassers simultaneously (N=4). *"If it was a bigger group, I would do a group block, it was a sign of relief." [Child14]* Additionally, *[PointBlock]* contributed to a positive psychological impact, with children noting that it reduces stress levels. *"It makes it fun when you're having to block people, it kind of removes a bit of stress because it can be really stressful to block people in games." [Child12].* Furthermore, using the wand empowered children (N=2).

Theme 2.2: Negative Aspects of the Studied Tools 1) Usability Challenges and Design Limitations: Children highlighted various design challenges that hindered the effectiveness of *[SelecBlock]* and *[MassBlock]* (N=23). A common issue was a disconnect between recognising harassers and selecting them for blocking (N=6). *[SelectBlock]* was described as slow (N=16) and cumbersome (N=2) since it required individually selecting each harasser to block. *"I think they were all good, but the profile one was harder to use because you don't really recognise them from the profiles, you might not be looking at their avatars as much." [Child14].*

2) Unintended Negative Consequences of Tool Use: Negative aspects of tools included potential incidents of accidentally blocking or unblocking the wrong person (N=13) with *[SelectBlock]* and *[MassBlock]* (they could choose to unblock after blocking all harassers with *[MassBlock]* tool). *"The individual pictures are a little bit hard to use because it takes some time to look around and you don't want to block the friendly ones." [Child16].*

3) Balancing Practicality with Creative Expression in Tool Design

While most children valued fun and enjoyment of using a tool (see section 4.3.4.2), some (N=3) tended to value practicality over creativity, and would want prioritising ease of use and efficiency *"The wand [PointBlock] is quite fun but I don't think that's the priority if you're being bullied and treated badly. It's not about having fun, it's about stopping them from doing that." [Child26-A].* A notable negative aspect was the physical demands of using *[PointBlock]*, which some found tiring and impractical for extended use (N=2).

Finding 2: Children responded positively to the blocking tools, highlighting ease of use, speed, and efficiency, especially with the *[PointBlock]* (wand) and *[MassBlock]* tools. The *[PointBlock]* was found to be enjoyable and engaging, while *[MassBlock]* provided psychological relief during large-group harassment. However, children also

identified usability issues, particularly with *[SelectBlock]*, which was seen as slow and cumbersome. Mistakenly blocking or unblocking the wrong users was a concern with *[SelectBlock]* and *[MassBlock]*. Some children preferred practical over creative features.

4.3.4.3 Category 3: Impact of Visual Perspectives of Harassment Scenarios: The Above View and Further Away

Theme 3.1: Effect of Being Elevated on Child User Experience Enhancement of Sense of Safety Being elevated increases child participants' sense of safety (N=15), allowing them to separate from the situation physically (N=3) and emotionally (N=3). It is harder for harassers to hurt, and it can impact the harassers' perspectives. It led to psychological distancing and diminished experience of harassment (N=2), making the harassers appear silly and less intimidating (N=2). *"On the ground, I felt anxious it was not very nice but on the top [...] I felt like I was separating from the situation." [Child27]*; *"The ability to see them from above is way better because they're not eye level staring at you, shouting. [...] You can see them clearly, they're just small and irrelevant. It felt freeing not having to look them in the eyes." [Child40].*

This is also supported by the quantitative data. The two-way ART ANOVA with the independent variables as the Group Size and the IP perspective (two conditions C1 and C4), shows a significant increase in calmness and sense of safety in the above perspective (C4) compared to the perspective on the ground (C1) (see table 4.10). As the qualitative data shows, this is likely due to the above perspective providing a greater sense of oversight and control, which contributes to feelings of calmness and safety compared to the ground-level perspective. This suggests that IPs can positively influence emotional and psychological responses.

Theme 3.2: Usability and Dynamics of Being Elevated or Further Away from Group Harassment Situation Advantages Children preferred the above view (N=23) compared to the ground IP. The above view has usability benefits as it provides a more comprehensive and helpful perspective to see (N=21), making it easier to block (N=4) and control the situation (N=2). This perspective enables less noticeable blocking actions, reducing the visibility of intervention (N=2). *"I would prefer to be up top because then it would be less noticeable if I were trying to block people and have a better and efficient aim. I wouldn't want them to see. I felt I could see it more so then I knew exactly who was doing it." [Child36].* It alleviates the pressure of time by freezing the moment, allowing for more thoughtful responses (N=2). *"I want to be moved up because it's easier to have some time to think about what you're going to do." [Child39]*

Disadvantages However, some children disliked being above (N=4) due to discomfort with heights (N=2) or feeling weird or unnatural (N=1). *"It felt quite strange, a body experience looking down, so I didn't enjoy that. I just felt weird." [Child26-B]*; *"I felt a bit scared because I'm scared of heights." [Child25-B].* It also created surprise (N=2).

Some children preferred the ground-level view as it was easier to see and navigate compared to the above perspective (N=3), and some would just want to stay on the ground while being further away (N=5). *"I want to be further away but on the ground because there might be other people who can help you." [Child13].*

Empowerment and Control when Elevated or Further Away The above view enhanced the child's sense of empowerment and control by providing a commanding perspective over their environment (N=7). *"I think I felt I had a little more control over the situation from above." [Child26-A]. "I thought it was quite cool, but also I felt I had power." [Child8-A].*

Finding 3: Children felt safer and calmer when viewing the harassment scenario from above, as the elevated perspective created emotional and psychological distance, making the bullies seem less threatening. Many preferred the above view for its usability benefits, such as better visibility, easier targeting, and more control over the situation. It also allowed discreet intervention and more time to think. Some children felt empowered and in control from this perspective. However, a few disliked being elevated due to fear of heights or feeling disoriented, and some preferred staying on the ground but positioned further away for comfort and support. Quantitative data showed a significant increase in calmness and sense of safety in the above perspective.

4.3.4.4 Category 4: Proposals for Improving Safeguarding Tools in social VR

Theme 4.1: Leveraging VR-Specific Features: Use of Physical Objects and Non-Verbal Cues Participants proposed creative tools that offer fun and playful alternatives to existing methods (N=2). For instance, they suggested throwing objects, such as snowballs [Child3], to block or unblock avatars (N=2), as well as using items from a virtual pocket (N=1) "Something you can just pop out of your pocket" [Child19-B]. Another idea involved hand gestures with labels on top for quick blocking (N=1). Additionally, a sword, similar to the wand, was proposed (N=1). A touch-based blocking mechanism, where children could block avatars simply by touching them, was also suggested (N=1). These proposals aim to introduce engaging and interactive elements to improve user experience in VR environments.

To further improve the proposed blocking tools (N=2), children proposed VR-unique additions. A child recommended implementing a message or ghost avatar to clearly indicate which avatars have been blocked or unblocked (N=1). For the unblocking process, a child suggested displaying a list of avatars who were within the radius at the time of unblocking, with their pictures highlighted on the list to prevent accidental unblocking of harassers (N=1).

Theme 4.2: From the Victim's and Harasser's Perspectives Several strategies were proposed to influence and alter the harasser's experience effectively (N=9). These included implementing censorship and filtering mechanisms, such as voice detection for profanity (N=2) and creating servers restricted to friends only (N=1), which would limit interactions with strangers and reduce opportunities for harassment.

Other suggestions focused on enhancing the victim's ability to manage interactions, such as providing options to mute (N=1) or block bullies from all games (N=1). Additionally, participants suggested reversing the unblock process by allowing victims to select individuals to keep unblocked while automatically blocking the rest (N=1).

Children also proposed several strategies to alter the harasser's experience (N=3). An idea was to isolate bullies by placing them in separate rooms (N=2), which would limit their ability to interact with others and potentially reduce their disruptive behaviour. Another suggestion was to discourage the harassers' actions through retaliation (N=2), for example, by implementing deterrents such as loud music played in the bullies' ears (N=1): *"It could be a tool that plays loud music in their ears and makes their ears pop out." [Child17].* For a more restorative approach, a child recommended initiatives to foster empathy in bullies, making them less inclined to engage in harmful behaviour (N=1).

To improve how victims perceive their virtual environment, a proposal was to introduce a pause button (N=1), allowing victims to temporarily halt interactions and regain composure during stressful situations.

Theme 4.3: Seeking Assistance: Proposals for Human or AI Support Systems Children proposed several ideas for leveraging human and AI assistance to enhance safety and support in virtual environments and offer help and guidance in a group harassment situation (N=9). *"I would also want someone else to protect, just anybody that would stick up for me." [Child19-B].* Children suggested incorporating mechanisms for receiving help from others, including both humans and AI, to protect users from harassment (N=8). *"I feel like even if it's not a real person, anything that helps is just better." [Child32-B].* A child proposed developing tools that use AI to detect mean or problematic behaviour and provide recommendations on who to block (N=1): *"I would say because it takes the stress off that person for having to do everything themselves." [Child40].*

Theme 4.4: Usability and Practicality Considerations for Tool Improvement Children also proposed several usability improvements to enhance the blocking tools (N=3). One suggestion was to include a search function to look up usernames from the unblocking list, making it easier to manage specific avatars. Another idea involved replacing the watch with clickable buttons or a wand that appears for blocking actions, streamlining the process (N=1). Although the watch was found to be easy to use after some practice, a child recommended incorporating a menu for easier access when looking to the side (N=1). Additionally, there was a call for tools that are quick and efficient to use, emphasising the need for quick and responsive features that facilitate seamless interaction (N=1).

Finding 4: Children proposed several ways to improve safeguarding tools in social VR, suggesting playful, VR-specific features like throwing objects, using hand gestures, or pulling items from virtual pockets for blocking. They recommended adding visual cues, such as ghost avatars, to indicate which were blocked. From the victim's perspective, suggestions included muting bullies, pausing the experience, and blocking all except those deemed to be trusted. To deter harassers, children proposed isolating them, using sensory deterrents like loud sounds, or promoting empathy. Many emphasised the need for human or AI support to guide and protect users during harassment. Usability improvements included search functions, quick-to-activate menus, and faster, easier blocking methods.

4.3.4.5 Overall Ranking of Tools, Emotional Responses and Usability Preferences

Ranks In the interviews, the children mentioned preferring the *[PointBlock]* wand (N=25) and the *[MassBlock]* (N=9), particularly in the larger group harassment situation. Quantitative data supports this preference: in both the large and small groups, *[PointBlock]* has the highest weighted score (see table 4.8)[3]. Additionally, the Above IP was preferred in both group sizes, with the highest total weighted scores (see table 4.8).

While children had similar overall ranks in both the large and small groups, the total weighted scores show *[MassBlock]* was preferred in the large group compared to the small group, and *[SelectBlock]* was preferred in the small group compared to the large group (see table 4.8). Moreover, the ART ANOVA tests indicated that the perceived difficulty and unhelpfulness of the safety tools were significantly higher in the large group compared to the small group table 4.10. Additionally, in the interviews (section 4.3.4.1), children expressed that using the blocking tools in the large group was more demanding. This suggests that safety tools should be easy to use and scalable to group sizes.

[3]Children's overall ratings are based solely on preference, reflecting how much participants liked each option, rather than perceived effectiveness or emotional perceptions.

Table 4.8: Weighted Ranking of Tools Based on Favourite, Second Favourite, and Least Favourite Scores. The ranking is based on a scoring system where favourite votes are worth 3 points, second favourite votes are worth 2 points, and least favourite votes are worth 1 point, with the total weighted score being the sum of these points.

Rank	Group Size	Tools	Favourite Score	2nd Favourite Score	Least Favourite Score	Total Weighted Score
1	**Large**	**PointBlock**	20 x 3	16 x 2	4 x 1	**96**
1	**Small**	**PointBlock**	15 x 3	22 x 2	3 x 1	**92**
2	**Large**	**MassBlock**	15 x 3	15 x 2	10 x 1	**85**
2	**Small**	**MassBlock**	17 x 3	10 x 2	13 x 1	**84**
3	**Small**	**SelectBlock**	8 x 3	8 x 2	24 x 1	**64**
3	**Large**	**SelectBlock**	5 x 3	9 x 2	26 x 1	**59**

Table 4.9: Weighted Ranking of Views Based on Favourite and Least Favourite Scores. The ranking is based on a scoring system where favourite votes are worth 2 points and least favourite votes are worth 1 point, with the total weighted score being the sum of these points.

Rank	Group Size	View	Favourite Score	Least Favourite Score	Total Weighted Score
1	**Large**	**Above**	33 x 2	7 x 1	**73**
1	**Small**	**Above**	31 x 2	9 x 1	**71**
2	**Small**	**Ground Level**	9 x 2	31 x 1	**49**
2	**Large**	**Ground Level**	7 x 2	33 x 1	**47**

Likert Scales The Likert Scales' results for the emotional responses and usability for each condition are presented in figure 4.5. Overall, among the first three tools (C1,C2,C3), C3 *[PointBlock]* has the highest percentages in terms of all happiness, safety, calmness, helpfulness and enjoyment, in both the Large and Small group. [C1: top view] appears to be the one with the highest safety percentage.

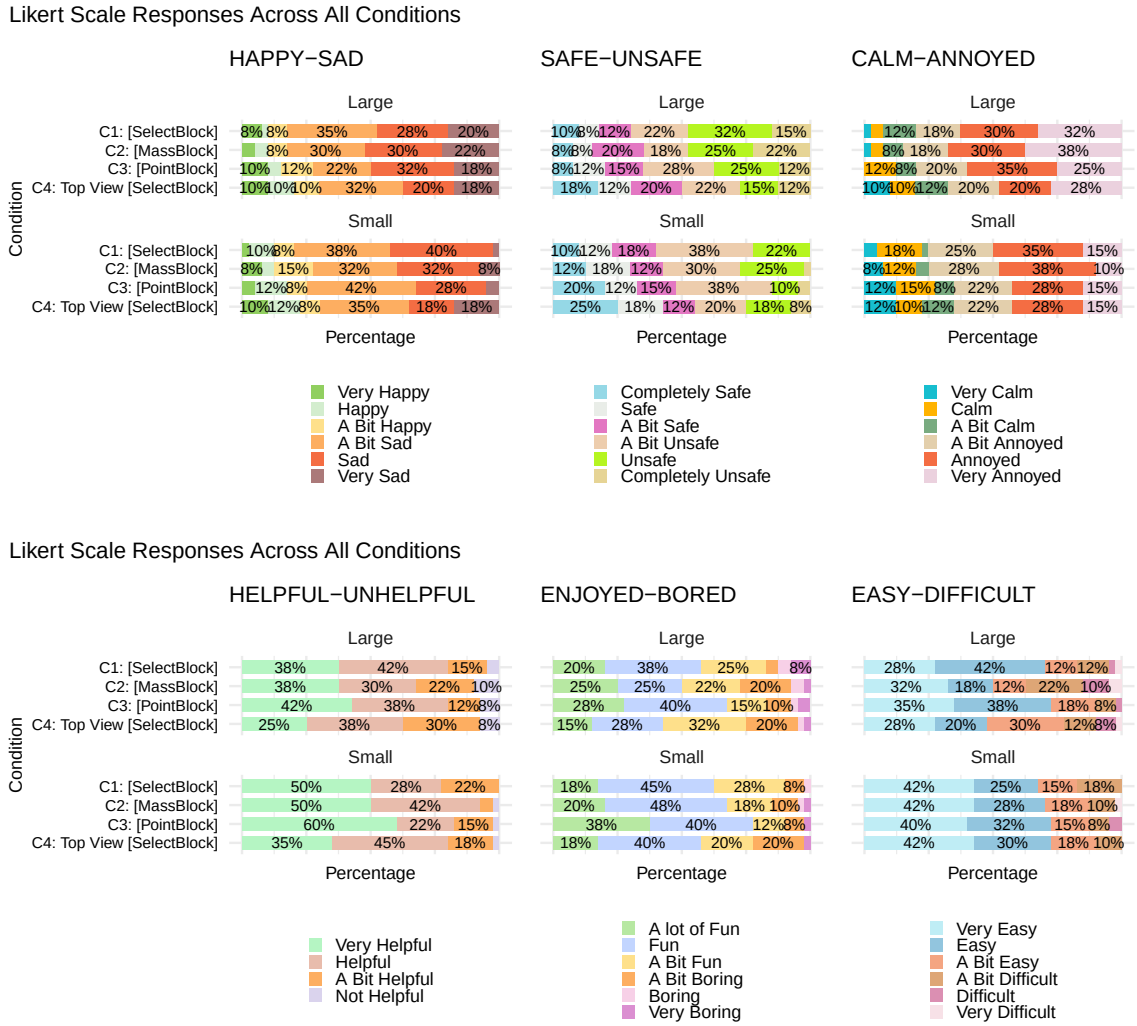


Figure 4.5: Likert Scale Responses

Statistical Results A summary of the statistical tests with significant results is presented in table 4.10. Non-statistically significant results are in the Appendices section B.2.

A table summarising means, medians and standard deviations can be found in the Appendices section B.1.

Finding 5: Children preferred the *[PointBlock]* tool across both large and small group harassment scenarios, with the highest weighted scores in the rankings. *[MassBlock]* ranked second, while *[SelectBlock]* ranked lowest overall. The above-view perspective was also ranked higher in both group sizes. Likert scale responses confirmed that *[PointBlock]* scored highest for happiness, safety, calmness, helpfulness, and enjoyment. Statistical tests further showed that emotional and usability ratings were significantly worse in large group conditions. The above view significantly increased feelings of safety and calmness compared to the ground view.

Table 4.10: Two-way ART ANOVAs results 1) with the independent variables as Group Size and Tools (C1,C2,C3) 2) with the independent variables as Group Size and IP (C1 VS C4). While [PointBlock] enjoyment mean is higher (4.8, SD=1.2) compared to [MassBlock] (4.5, SD=1.2) and [SelectBlock] (4.5, SD=1.2), the post-hoc test did not reveal significant differences between individual pairs. (Small effect: ≈0.01, Medium effect: ≈0.06, Large effect: >0.14).

Independent Variables: Group Size and Tools			
Dependent Variable	**Effect Size (η_p^2)**	**F-statistic**	**p-value**
Tools - Group Size (Small VS Large) Effect			
SAD-HAPPY Scale (1-6)	Small (0.04)	$F(1,196) = 7.92$	$p = 0.015$
UNSAFE-SAFE Scale (1-6)	Large (0.2)	$F(1,196) = 50.9$	$p = 1.9e^{-11}$
CALM-ANNOYED Scale (1-6)	Medium (0.1)	$F(1,196) = 22.0$	$p = 5e^{-6}$
UNHELPFUL-HELPFUL Scale (1-5)	Medium (0.09)	$F(1,196) = 20.3$	$p = 1e^{-5}$
BORED-ENJOYED Scale (1-6)	Small (0.04)	$F(1,196) = 7.9$	$p = 0.0053$
DIFFICULT-EASY Scale (1-6)	Small (0.05)	$F(1,196) = 11.1$	$p = 0.001$
Intervention Perspective (IP) - Group Size (Small VS Large) Effect			
UNSAFE-SAFE Scale (1-6)	Medium (0.06)	$F(1,118) = 8.0$	$p = 0.0055$
CALM-ANNOYED Scale (1-6)	Small (0.04)	$F(1,118) = 5.2$	$p = 0.025$
UNHELPFUL-HELPFUL Scale (1-5)	Small (0.04)	$F(1,118) = 5.4$	$p = 0.022$
DIFFICULT-EASY Scale (1-6)	Medium (0.06)	$F(1,118) = 7.0$	$p = 0.009$
Tools - Conditions Effect			
ENJOYMENT (PointBlock vs. MassBlock, SelectBlock)	Small (0.03)	$F(1,196) = 3.3$	$p = 0.04$
Intervention Perspective (IP) - Conditions Effect			
UNSAFE-SAFE Scale (C4: Above View vs. C1: Ground)	Medium (0.09)	$F(1,118) = 11.1$	$p = 0.0009$
CALM-ANNOYED Scale (1-6)	Small (0.03)	$F(1,118) = 4.1$	$p = 0.046$
UNHELPFUL-HELPFUL Scale (1-5)	Small (0.05)	$F(1,118) = 6.5$	$p = 0.001$

4.3.5 Discussion

A reflection across studies from Part II can be found in chapter 5 to situate this study's findings within the wider contributions of the thesis. Here we present a summary of the findings and references to corresponding guidelines in chapter 9.

4.3.5.1 Summary of Findings

The simulated group harassment scenarios elicited a clear emotional impact (4.3.4.1), as evidenced by both quantitative and qualitative findings of negative emotional valence, with the majority perceiving themselves as the victim and putting themselves into the victim's shoes. The emotional engagement underscores the scenarios' validity in evoking emotional responses from children without exposing them to the direct, first-person harassment.

Regarding **S3-RQ1: Perceptions of Tools** *[MassBlock]* tool was noted for providing relief when managing harassment in large groups, while *[PointBlock]* was appreciated for its empowering, stress-reducing qualities, ease-of-use and particularly due to its fun, interactive nature (section 4.3.4.2). However, some negative aspects were

reported (see section 4.3.4.2), particularly with *[SelectBlock]* and *[MassBlock]*, where children expressed concerns about accidentally blocking the wrong individuals. Additionally, some children noted the challenge of balancing the enjoyment and creativity of the wand with its physical demands.

S3-RQ2: Impact of Group Size showed that harassment from larger groups intensified psychological distress and made blocking more challenging, reflected in both the qualitative data (sections 4.3.4.1) and the quantitative data (see sections 4.3.4.1, 4.3.4.5 and 4.10). This suggests that safety tools should be easy to use and scalable to different group sizes.

S3-RQ3: Impact of Intervention Perspective indicated that the majority of children reported feeling safer when placed in an elevated perspective compared to being on the ground (see section 4.3.4.3, and table 4.10). This effect was even more prominent in the larger group, as it made it easier to understand and control the situation (section 4.3.4.3). However, fewer children expressed discomfort with the height, finding it unnatural, and preferred the ground view (see section 4.3.4.3). Overall, the above perspective was seen as empowering, as it reduced the intensity of the harassment.

4.3.5.2 Usability and Design Considerations of Safety Features for Children in Social VR

Elevated IP Benefits: Detachment from Harassment Situations A first design consideration is the ability to elevate children's scene view while keeping the harassers' perspective at ground level. The above IP, where children viewed the harassment group from a higher vantage point, played a critical role in enhancing their sense of safety significantly, as both the quantitative and qualitative results showed. Children reported feeling an increased sense of control and detachment from the harassment, which contributed to a sense of empowerment. The perspective manipulation provided a psychological buffer, reducing the direct confrontational nature of the harassment and allowing children to view the situation from a more detached perspective of the situation. Being above, therefore, diminished the harassment experience and reduced the power dynamic, making the harassers seem more trivial rather than threatening. Children felt empowered and less intimidated. This finding underscores the potential benefits of integrating perspective adjustments in social VR environments for children when using safety features to enhance emotional safety and mitigate the negative effects of harassment. This supports prior work described in section 4.3.2.2.

Moreover, this effect aligns with the principles of the Proteus Effect, wherein an individual's behaviour and self-perception can be influenced by the characteristics of their virtual avatar [169]. Research demonstrated that participants assigned taller avatars in virtual environments exhibited increased confidence and assertiveness compared to those with shorter avatars [169]. While the Proteus effect has been observed in adult populations [169, 333], its applicability to children remains an area for further exploration. However, the observed responses in this study suggested that similar mechanisms may be at play, indicating the potential for avatar-based interventions to influence children's perceptions of safety and empowerment in virtual environments.

Immediate, Scalable and Stress-Relieving Tools Children in both studies reported experiencing psychological relief when using the tools to address group harassment in real time, particularly in larger group settings where managing multiple harassers was more challenging and overwhelming. This supports prior work described in section 4.3.2.1. The effectiveness of *[MassBlock]* in alleviating stress highlighted the importance of scalable solutions that reduce emotional distress in children. Similarly, *[PointBlock]* was noted for its role in lowering stress levels, reinforcing the need for immediate interventions that are immediate, easy to use in group scenarios and stress-relieving.

Balancing Safety, Practicality with Enjoyment of Safety Features Children appreciated the enjoyable aspects of safety tools, which suggests that incorporating fun can enhance their use and support the development of autonomy and resilience. Moreover, the "Magic Wand" *[PointBlock]* complements the fantasy of the virtual environment, without interrupting immersion and presence, while allowing space for imagination and personal interpretation of what happens to harassers once they disappear. Nevertheless, safety and practicality must remain the primary focus. Achieving this balance requires ensuring that safety features both effectively safeguard children and engage them. However, overly engaging tools may lead to potential gamified misuse and enact abuse.

Based on these findings, I gather practical design guidelines in section 9.2, specifically guidelines 5 and 6.

4.3.6 Conclusion

Group harassment in social VR presents challenges for child users, particularly in terms of how they can quickly protect themselves or seek help. This user study reveals that children value interventions that are engaging, scalable, and accurate in addressing group harassment in social VR, with an elevated perspective notably enhancing their sense of safety.

4.4 Chapter Conclusion

This chapter investigated design considerations for safety features aimed at addressing both one-to-one and group harassment targeting children in social VR environments **(RQ2: What design considerations can help shape safety features to address child harassment in social VR?)**, by integrating insights from experts (section 4.2) and direct feedback from child users (section 4.3). Through this multi-stakeholder approach, critical gaps were uncovered in existing safety mechanisms, highlighting significant opportunities for innovation and improvement.

4.4.1 Contribution to Research Question 2

Specifically, the findings underscored the critical insight that existing safety solutions place responsibility on children to independently handle harassment scenarios in social VR. Experts stressed a balanced integration of human-involved solutions, providing clear examples such as buddy systems, transparent parental supervision mechanisms respecting child autonomy and privacy from the setup to open conversations, and educational initiatives to help increase parental awareness and guide responsible usage. Additionally, broader systemic responsibilities were emphasised, highlighting the roles of platform companies in deploying automated moderation solutions and preventive safeguards (e.g., age verification, age matching and enforcing rules, expectations, tips reminding community guidelines and training tutorials for parents).

Furthermore, the findings revealed the necessity of creating child-friendly, real-time safety tools such as quick-access, discreet and engaging features (e.g., Panic Buttons, Discreet gestures, *[PointBlock]* "Magic Wand" and [MassBlock]) and intervention perspectives that position users above the harassment situation, thereby significantly reducing distress. These solutions should be characterised by their ease of use, rapid deployment in high-stress scenarios, scalability to manage large groups of harassers effectively, and customisability tailored to diverse developmental needs and individual preferences of children.

Consequently, this chapter provides a detailed foundation and precise guidance for designing innovative, balanced, and child-centred safety solutions to effectively safeguard children against one-to-one and group harassment in social VR.

Chapter 5

Summary & Reflection of Part II

5.1 Introduction

Children under the age limit platforms have set are using social VR. Study #1 showed that 43% of children under 13 years (in the sample) old use social VR at least every two weeks, and this reaffirms prior work [21, 22].

News articles frequently spotlight some of the issues, including harassers using social VR to abuse children [334, 33, 335, 336], as well as recent papers [195, 210]. Research into child use of social VR is in its early stages, and does not take into account the pertinent stakeholder groups, such as parents and experts, and how children themselves perceive and respond to these harms [8]. What responsibilities do each of them bear in shaping safer immersive digital environments?

Addressing this gap in stakeholder insight in part II, I examined the child social VR ecosystem through three empirical studies, each focusing on a key stakeholder group: adults, including parents and social VR users, child psychology and online safety experts, and children themselves. These studies investigated both the nature of the risks **(RQ1)** and the potential for intervention **(RQ2)**. The goal was to surface not only what is happening with children within these spaces, but also how various actors make sense of, experience, and attempt to address child harassment in social VR.

This chapter critically reflects on the findings from these three studies, each of which contributes a distinct yet interlocking multi-layered understanding of the challenges and prospective mitigations for child safety in immersive social environments. These three studies contributed:

Study #1 A mixed-methods questionnaire with social-VR aware adults (70 parents and 79 non-parents) revealing child usage, concerns, and safeguarding strategies;

Study #2 Interviews with 16 experts in online safety and child psychology, exploring strategic gaps and intervention solutions;

Study #3 A user study with 40 children 8-16 years old, exploring first-hand experiences, design preferences and needs of a common safety tool, blocking.

While some perspectives diverge, they mostly converge or complement each other. Concerns were articulated by adults and parents, whereas experts focused on interventions, and children provided empirical grounding by testing the proposed tools. When taken together, these viewpoints expose not only what the risks are but also suggest how interventions might be designed to be both feasible and meaningful.

5.2 Converging and Diverging Insights from Four Stakeholder Groups

This section considers the perspectives of four stakeholder groups (adult users, parents, experts and children), in particular where they strongly converged or diverged. In some cases, divergence is arguably an artefact of differences in methodological approach; however, this variation is nonetheless valuable as each method was appropriately tailored to the stakeholder group involved, enabling contextually relevant and meaningful contributions.

I argue that effective safety mechanisms in social VR cannot be built from any one perspective in isolation. Rather, they must be co-constructed, drawing on experiential and professional insights to develop safeguards that are not only protective, but also empowering and responsible to immersive embodied experiences.

5.2.1 Converging Insights

There is a shared understanding that current safety features are insufficient in social VR, there is a lack of parental awareness of supervision, and there is broad agreement on the need for robust child-friendly safety measures.

Challenges for Children in Social VR Across all stakeholders, there was strong agreement that harassment in social VR can be challenging and can adversely affect children's well-being. In study #3, children described experiencing negative emotions and, in some cases, physical reactions such as a racing heartbeat, when exposed to simulated group harassment scenarios (Study #3 finding 2). Experts echoed these findings, noting that children often may not have the maturity, skills or confidence yet, to protect themselves or navigate such spaces safely, especially with current safety features developed without a child perspective (Study #2 finding 4). One expert highlighted that trauma in VR can be very real and close to physical; the brain perceives the harassment in VR as real, and the child can freeze (see section 3.2.3.5). Parents and adults noted the effects and impacts of VR and social VR on children. Negative impacts included mental health effects (e.g., mental development, trigger of negative emotions and negative influence) (see section 3.2.3.4).

Shared Concerns About Safety Risks Across parents, non-parent adults, and experts, there was consensus that children face significant safety risks in social VR environments. Both parents and non-parent adults expressed concern about the level of freedom children have in these spaces, where they may encounter malicious individuals, including potential predators, and the absence of consistent supervision, regulations, and monitoring (Study #2 finding 5). Experts reinforced these concerns as social VR lacks established social norms. They have identified risks such as inconsistent safety features across platforms, children's limited ability to cope with overwhelming VR experiences, and the dangers of adults and children mixing in shared virtual rooms (Study #2 finding 4). These stakeholders all advocated for age-verification and age-matching social VR spaces.

Importance of Supervision and Effective Moderation Parents and non-parent adults are significantly more likely to let younger children use social VR with supervision, and if there are only minors compared to those without supervision (Study #1 finding 2). Parents also consistently stressed that children should not be left unsupervised and need clear guidance while using these platforms (Study #1 finding 6), and non-parent adults urge developers to create safe spaces with better protection and supervision (Study #1 finding 6). Experts agreed, emphasising that human moderators are an essential safeguard, particularly for children, as they can provide context-sensitive support and intervention. At the same time, experts noted that automated moderation systems offer valuable scalability and rapid response capabilities (Study #2 finding 2). In fact, both experts (Study #2 finding 3) and children (Study #2 finding 1) agree that scalability is crucial in addressing harassment in social VR. Large groups of harassers pose greater challenges, making it difficult to block individuals effectively. Therefore, safety tools must be designed to handle harassment at scale, with features that enable users to quickly and effectively manage both small and large groups, or automated moderation, to reduce negative impact and improve usability.

Parental Awareness and Informed Engagement Among non-parent adults, a common concern was parents' limited awareness of the range of interactions in social VR and the ways children may exploit this freedom (see section 3.2.3.5). Experts similarly emphasised that parental knowledge is critical, urging parents to be curious and informed through self-learning or training (Study #2 finding 11). Both perspectives converge on the need for informed engagement, emphasising that parents maintain active understanding and awareness of social VR spaces.

Trust and Communication Between Parents and Children Both parents and experts emphasise the value of trust and communication. Parents prioritise ongoing conversations with their children (Study #1 finding 5). At the same time, experts advocate replacing restrictive methods with a trusting parent–child relationship built on positive engagement and open dialogue (Study #2 finding 11). Together, these perspectives point to trust-based, communicative approaches as central to guiding children's social VR use. This aligns with prior work underscoring the importance of fostering open communication to acknowledge both benefits and risks, while supporting parental supervision and control [210].

Child-Friendly Safety Tools for Harassment Response Experts emphasised the need for quick, easy-to-use safety tools that empower children to regain control during harassment (Study #2 finding 5, Study #2 finding 9). Features like "panic" buttons and simple blocking tools help children respond fast. Children reacted positively to the tools proposed, especially the *[PointBlock]* wand, which was engaging, and the [MassBlock], which offered relief during large-group harassment (Study #3 finding 2).

Discreet Safety Features Empowering Children Experts recommend safety tools that are discreet in use and outcome, such as subtle gestures and that provide a sense of space, comfort, and empowerment through positive reinforcement (Study #2 finding 6). Children echoed this, reporting that viewing harassment from an elevated perspective made them feel safer and calmer. This vantage point created emotional distance from bullies, improved visibility and targeting, enabled discreet intervention, and gave them more time to respond thoughtfully (Study #3 finding 3).

5.2.2 Diverging Insights

Balancing Trust, Monitoring, and Privacy in Parenting Approaches Parents in the sample, in particular those familiar with VR/social VR, tend to choose lower minimum ages for social VR usage compared to non-parent adults less familiar with the technology (Study #1 finding 2). Both non-parent adults and experts, however, stress the importance of parental awareness and involvement to ensure safe experiences. Non-parent adults often advocate for stricter rules, improved moderation, education, and safe virtual spaces, expecting parents to take a stronger supervisory role (Study #1 finding 6). Experts agree that parental involvement is essential but recommend a balanced approach that avoids excessive monitoring, respects children's autonomy, and focuses on education over rigid restrictions (Study #2 finding 11). As mentioned in section 5.2.1, some parents prioritise trust and open conversations with their children as they would in everyday social settings and see social VR as similar to real-life situations and are more indulgent (Study #1 finding 5). However, others may be more intrusive, actively monitor and even stream their children's social VR use, alongside tools like robust authentication and time limits to manage social VR (Study #1 finding 7). Experts suggest finding a balance of involvement, for example, via real-time alerts during incidents, but without being overly invasive. Safety features must respect privacy, and consider children's consent and participation, but also possible misuses and ways to bypass protections (Study #2 finding 11). Indeed, it is essential to consider risks of exploitation and privacy concerns, ensuring alignment with the ICO Children's Code [294] and the Five Rights framework [148], which address a child's evolving rights to privacy and freedom of expression. Additionally, there is potential for these interventions to be misused by adults in other contexts, such as domestic abuse [295, 296]; therefore, careful consideration must be given to the design and deployment of such measures.

Addressing Harassment Through Both Victim and Harasser Perspectives Unlike parents and non-parent adults who mainly describe harassment, risks and concerns from the victim's perspective, experts proposed solutions that should also encourage harassers to reflect on their behaviour, distinguishing between intentional and unintentional harassment (Study #2 finding 10). Although beyond the scope of this thesis, experts further noted that interventions should also support harassers, drawing on restorative justice approaches rather than relying solely on punitive measures. Prior work indicates that restorative practices offer a constructive alternative to anti-bullying laws and zero-tolerance policies, mitigating their drawbacks while fostering the emotional well-being and relationships of both victims and harassers [337].

5.2.3 Integrating Stakeholder Views

When perspectives diverge, whose guidance should we follow: the adult social VR user, the parent, or the expert? Each offers a valuable, complementary viewpoint that should be considered [242, 338, 339, 340]. There are multiple possible ways this could be addressed. For example, it could be that some views are less or more biased, and this would need to be taken into consideration. Although an expert's perspective may be informed by a larger evidence base and specialised knowledge compared to that of a parent, it remains susceptible to biases and subjective interpretation. It could also be that future work is needed to expose the different stakeholders to the different views and have them discuss and reach a common conclusion. This would require further work and research.

What emerges here is a diverse range of perspectives, each shaped by different forms of knowledge, responsibility, and lived experience. Adults tend to frame risk through the lenses of supervision, trust, and control, as they grapple with their own familiarity with social VR platforms. Experts, in contrast, adopt a broader objective view considering both the family and the system level. They advocate platform accountability through design mechanisms such as age-verification, scalable moderation, child-friendly safety tools, autonomy-respecting parental controls, as well as open conversations and preventive solutions while considering child learning, agency and privacy. Children, often excluded from design and development phases, contribute a perspective that highlights the realities of being targeted in immersive environments and the immediate need for responsive, easy-to-use, in distress, safety tools.

5.3 Limitations and Future Work

While the limitations of each study have been discussed in the chapters, I now consider the broader limitations of the overall work in Part II.

One challenge lies in the diversity of methods across studies and uneven sample sizes across stakeholder groups, which makes direct comparisons difficult. These differences sometimes meant that insights emerged in parallel rather than in direct conversation with one another. However, this diversity remains valuable, as it enabled the collection of insights tailored to each stakeholder group: surveys captured perspectives of adult social VR users, expert interviews provided depth, and user studies with children offered practical understanding. In the future, co-design sessions that bring together multiple stakeholder groups could help bridge these gaps, allowing participants to collectively discuss priorities and design considerations (see section 6.3).

As discussed in Section section 1.3 and illustrated in figure 1.2, multiple stakeholders are involved, with social VR adult users, children, and guardians being those most directly connected to child users of social VR. Future work should extend to additional stakeholders, such as developers and designers, who may provide valuable insights into the feasibility and technical requirements of proposed solutions. While experts often propose solutions considering the harasser's perspective, most studies to date have focused on the victim's perspective, in line with the scope of this thesis (section 1.3. Recent research has begun exploring the bully's perspective, highlighting the import-

ance of including them when evaluating the design of intervention tools [341]. This may require testing tools within actual social VR platforms, rather than simulations, to assess their long-term impact on reducing harassment and recidivism [342, 343, 344, 345].

The landscape of harassment and the technologies through which it occurs is constantly evolving [20, 31]. This creates an inherent tension: recommendations and guidelines risk becoming outdated unless they are revisited and adapted over time. Longitudinal research is therefore essential, not only to track the persistence and evolution of harms but also to evaluate whether safety interventions retain their relevance and effectiveness. Yet platforms may be reluctant to support such studies, as they raise complex regulatory and ethical challenges. Sustained access for research requires handling sensitive personal data, securing informed consent from users, and navigating strict data protection obligations (e.g., UK GDPR [346]). At the same time, platforms must also comply with broader regulatory frameworks, including Ofcom's oversight responsibilities and the duties imposed by the Online Safety Act 2023 [347], which may make them cautious about granting or maintaining long-term access to data.

This research also highlights the need for customisation in safety features and adapting safeguards to the individual needs, not only based on age but also social skills and disabilities (see (Study #2 finding 7)), but these ideas remain under-explored. Future work should investigate how children themselves perceive and prioritise different safety options, similar to recent investigations with adult social VR users [9], and how these can be implemented in ways that are both empowering and accessible. Likewise, the scope of safety tools examined here was narrow, focusing mainly on reactive measures such as blocking. Broader experimentation with proactive, preventative, and community-driven interventions (e.g., with bystanders) could reveal alternative or complementary strategies.

Ultimately, by engaging a wider set of stakeholders, experimenting with diverse safety tools, and committing to longitudinal studies in social VR platforms, we can better understand and more effectively design for child safety in social VR.

5.4 Implications Informing the Next Part of the Thesis

Drawing from the findings in Part II, it is evident that current safety measures in social VR are insufficient for protecting children. Age limits are often disregarded. Existing tools, such as blocking, muting, and reporting [187, 204] place the burden of protection on children themselves, and parental awareness of risks and supervision remains inconsistent. The immersive nature of social VR amplifies harassment impact compared to 2D social media and exposure to inappropriate content experienced by children, as it is highly realistic and distressing. Stakeholders across the three studies consistently highlighted the need to develop robust, child-centred safety solutions that combine effective intervention, scalability, and mechanisms for involving guardians. There is a clear need to move beyond child-dependent safety tools toward proactive, child-friendly and context-sensitive interventions.

5.4.1 Leveraging AI-Moderation

With the rise of cyberbullying and online harassment on social media, automated moderation has become increasingly common [101]. This approach deploys various artificial intelligence (AI) techniques to filter and process user-generated content, applying predefined rules to approve or reject images or text posted online [102]. Recent advances in conversational agents and large language models (LLMs) have demonstrated remarkable capabilities for automating tasks and driving innovation across diverse applications, such as AI-educational tutors [348, 349, 350].

While AI moderation shows promise, it still faces limitations. These include challenges in handling ambiguous content, high rates of false positives [102], and the potential to reinforce existing biases. Ethical concerns also arise, such as lack of transparency [106], visibility of moderation processes [107], threats to privacy, and reduced user agency [108, 109]. A study involving interviews with 59 teenagers found that moderation interventions can impact children's rights and underscored the importance of involving young users in AI design decisions for online platforms [110].

In social VR, some platforms rely on human moderators to address misconduct [31]. However, they are unable to manage most incidents, and their availability is limited. A recent study reported that only 24% of disruptions in social VR environments were addressed by human moderators [19]. Automated moderation has therefore been proposed as an alternative safety tool for combating harassment in these environments [40, 4]. Unlike traditional social media, social VR is accessed via head-mounted displays capable of capturing additional biometric data. Such data, including non-verbal cues and physiological signals such as heart rate and eye-tracking, could potentially be leveraged to detect stress or other indicators of harassment [7].

5.4.2 Leveraging Embodiment

One of the unique affordances of VR, embodiment, i.e., *the representation of bodily states such as posture, movement, and facial expressions* [160] (see section 2.3), offers new opportunities for safeguarding. Prior work has shown that embodiment can shape social interactions, foster connectedness, and influence behaviour through non-verbal cues [10, 159]. This suggests the potential in safety contexts for interventions with embodied presence to communicate authority or provide support and comfort.

Specifically for children, embodiment and socially responsive virtual AI characters in VR could have a notable impact on the way they experience, and react to, social interventions [37]. Embodied avatars could leverage an influential appearance and relationship: prior work shows that children's perception of such figures (e.g., Disney characters [67], Teachers VS Peers [42]) influences their behaviour, highlighting the need to carefully consider the avatar's representation as it may lead to negative (decreased agency) or positive (inspire good behaviours) impact [67, 42]. Few studies have examined social embodiment in VR among children [37, 36, 161, 4], leaving questions regarding what features are suitable, demanded and preferred for an embodied avatar safeguarding children from harassment in social VR spaces.

5.4.3 Proposed Solution

In response to the challenges and opportunities with AI and embodiment, I propose *Automated Embodied Moderators* (AEMs): *AI-driven, embodied safety agents that combine automated moderation with the affordances of VR embodiment.* AEMs are designed to detect and respond to harassment in real-time, visibly safeguarding via an embodied avatar within social VR. AEMs are envisioned as supportive co-actors that enable graduated intervention, from passive presence to active involvement, while maintaining opportunities for children to respond, participate and develop resilience as prior work notes its importance [351, 352].

This proposed solution aligns with criminological perspectives that emphasise the importance of "capable guardianship" [134, 135] (see section 2.2.5.1), where harm is less likely to occur in the presence of an active and engaged guardian. In social VR, such guardianship has yet to be meaningfully integrated, leveraging AI and embodiment, to safeguard children exposed to harassers in unsupervised environments [136].

The following part will focus on exploring AEMs, understanding the impact and implications of automated moderation that leverages the power of embodied experiences in a virtual space on children, how it is perceived by the stakeholders (parents, children and experts), what the design factors and actions are for an effective system and how parents and children should be involved.

PART III
Safeguarding Children from Harassment in Social VR using Automated Embodied Moderators

Chapter 6

Exploring Child, Guardian, and Expert Perspectives on Automated Embodied Moderators

"I felt much safer when Big Buddy was there and a lot more protected." - [Study #4 Child32a]

"The AEM can be kind of funny, kind of whimsical but actually it communicates the rules in a way that you don't dismiss it but it could make you laugh at the same time so that you actually see it as a piece of the environment and a part of the game and it's not annoying." - [Study #5 Expert P17]

"My point is it's not infallible. There's going to be workarounds." - [Study #6 Parent P3P-2]

6.1 Chapter Introduction

This chapter presents foundational research into how parents, children and experts perceive my proposed solution: AEMs, AI agents that embody a virtual form and take action to mitigate harmful interactions, as a potential safety mechanism in social VR. To investigate this, in this chapter, I report on the development and testing of the first implementation of *"Big Buddy"*, a Wizard-of-Oz AEM system, to evaluate children's real-time reactions to a disruptive scenario. At the same time, their parents viewed the disruption on an iPad. The emphasis is on the intervention, with the expectation that automated detection will eventually be developed (e.g., [208]). This was followed by studies involving key stakeholders (children, parents and experts) to further unpack the requirements, expectations, and concerns related to the feasibility and acceptability of AEMs in terms of safeguarding children in social VR. Together, these studies explore both experiential and conceptual aspects of AEMs to assess their potential role in child safeguarding. The goal is to gain a better understanding of not only how such systems are received but also what design considerations may support their suitability, effectiveness, trustworthiness, and child-appropriateness.

6.1.1 Motivation

As described in part II, existing real-time safety features in social VR have significant limitations. First, they place the responsibility for safety on potentially ill-equipped children. Second, safety tools are inconsistent across platforms [9], both in how the same action is implemented (e.g., blocking may be performed through a virtual watch or hand gestures [187], or via a menu [204]) and in the additional functionalities they provide (e.g., trust ranks [204]). Third, most tools overlook unique aspects of social VR, such as scalability, embodiment, and immersion. And while human-based moderation can help establish norms for appropriate behaviour [31], it faces limitations. Users may lack trust in moderators due to possible personal or subjective biases, and such approaches may only be effective in small-scale environments [20]. In fact, human moderators addressed only 24% of incidents observed in social VR [19]. Given the scale and complexity of online interactions alongside the rise of AI, automated moderation approaches are increasingly used in forums, social media, and games [353, 354, 355]. Recent work based on interviews with adult social VR users has highlighted both the opportunities and limitations of AI moderation, showing its potential to create safer and more inclusive social VR environments [40].

On the one hand, AI-based moderation has demonstrated promise (see section 5.4.1). On the other hand, embodiment in VR has proven to be highly influential, particularly for children (see section 5.4.2), and could be leveraged in safety-focused applications. However, we do not know how AI moderation can be integrated into the unique embodied and immersive context of social VR, how children might perceive AEMs and what parents and experts would consider important in such designs.

This chapter, therefore, addresses the following research question, from the perspectives of children, parents and experts:

RQ3

How are Automated Embodied Moderators (AEMs) perceived as a solution to safeguard children from harassment in social VR, and what factors are considered important to the prospective usage of AEMs?

6.1.2 Overview of Studies

6.1.2.1 Study #4

This study introduces Big Buddy, a Wizard-of-Oz AEM, designed to safeguard children from harassment in a simulated social VR environment. The goal was to explore whether automated moderation embodied approaches are effective and how they are perceived by both children and parents, providing insights to guide the future design of such systems.

Methodology A total of 43 children (ages 8–16) participated in a simulated social VR tower-construction game that I developed, and they completed Likert-scale questionnaires after each condition. In addition, 17 parents were interviewed after observing their child's VR experience on an iPad while the child engaged in the custom-built social VR game. The latter involved fictitious competitors disrupting the participant's game in a virtual classroom mimicking a social VR environment. I evaluated children's emotions and perceptions towards Big Buddy and the interventions.

Key Findings

- Child Perspectives: The presence of Big Buddy significantly affected children's reactions to disruptions. Its absence led to negative feelings, while applying punishments with exclusion of the competitor significantly reduced sadness. Children perceived Big Buddy as a referee ensuring safety and fairness, but some found the punishments unfair or too harsh.
- Parent Perspective: Parents empathised with their children when watching the recording of the disruptions experienced in the social VR game. The presence of Big Buddy reassured parents. They suggested graduated interventions and customisation based on repeated disruption or severity of incidents. However, they expressed wanting to understand the process and stay involved to some extent.

6.1.2.2 Studies #5 and #6

Building on the prior study #4, I aimed to explore the benefits and concerns, proposed interventions and embodiment features of AEMs in social VR through multi-stakeholder perspectives.

Methodology I conducted interviews with the same 16 experts from Study #2 and 5 design workshops with 8 guardians (5 parents and 3 grandparents) and 13 children aged 8 to 16 years old in total. By merging perspectives, I identified areas of consensus and differing views around the suitability, role, presentation, interactions and actions of the AEM. The findings inform the development of effective and child-centred AEMs, providing the foundations for the development of safer social VR through automated embodied moderation.

Key Findings This study provides an important first step in reflection around, and recommendations for, the design and use of AEMs to safeguard children in social VR. While there is consensus across stakeholder groups regarding most recommendations and preferences around AEMs, there is a contrasting perspective regarding the extent of AEMs' involvement: children favour a more passive role for the AEM, intervening only when necessary, with discretion and immediate actions, whereas adults anticipate a broader range of engagement with discussions and activities beyond intervening when an incident happens. There is also a tension between making the AEM friendly without being intimidating, while still maintaining authority and being credible.

6.2 Study #4 - Big Buddy: Exploring Child Reactions and Parental Perceptions towards a Simulated Automated Embodied Moderator for Social VR

6.2.1 Motivation

This study addresses the gap in understanding how AEMs should be integrated into a social VR environment from children's and parents' perspectives. To explore the effectiveness and perception of AEMs by children and parents, I introduced "*Big Buddy*" as an example of AEM, employing the Wizard-of-Oz prototyping approach in a simulated social VR environment.

I explore the impact of the presence of the AEM and the visibility of its interventions on child and parental perceptions and sense of safety of the moderation experience. 43 children (8-16 years old) played a researcher-developed VR tower-block-construction game. It involved fictitious competitors disrupting the participant's game in a virtual classroom, mimicking a social VR environment. Parents watched a video of what the child was seeing under the VR headset before semi-structured interviews. I evaluated children's emotions and perceptions towards Big Buddy and his interventions, as well as parents' perceptions. In this study, I aim to answer the following questions:

S4-RQ1 How is children's emotional valence impacted in provocative situations by the presence of the AEM (Big Buddy)?

S4-RQ2 How do children and parents perceive the embodiment of the AEM (Big Buddy) and moderation actions (punishments)? Will children feel safer and/or inhibited given its presence?

S4-RQ3 How do children and parents envision an AEM in social VR?

S4-RQ4 How involved would parents want to be with the AEMs?

6.2.2 Related Work

This section synthesises related work on the specific topics addressed in this study. (A more comprehensive literature review is presented in chapter 2, with specific relevant sections for context: section 2.3.2, section 2.3.3, section 2.3.4, section 2.2.2.2; for the proposed solution of the AEM: section 2.2.3.2, section 2.2.2.2, section 2.2.5.1; and for methodology: section 2.4.2, section 2.4.3).

The proposed solution of AEMs is based on prior work showing that AI holds promise in addressing moderation challenges on social media (section 2.2.3.2). The research also underscores the importance of embodiment in social interactions and VR, particularly for children (section 2.3) and the influence an authority figure can have on children (section 2.2.2.2), highlighting its potential to enhance automated moderation actions with a visible embodied character. Moreover, as the criminology capable guardianship theory showed, the sole presence of a guardian can be an effective way of deterring offences (section 2.2.5.1).

This study design is based on multiple prior studies to ensure the realism of the VR game and create a socially disruptive situation that is ethical but still evokes an emotional response. The VR game and provocative scenario are based on a study which evaluated children's aggressive behaviour, in particular aggressive social information processing (SIP) [44]. Researchers evaluated how emotionally engaged children were in disruptive situations via interactive and realistic VR scenarios. Boys aged 8 to 13 years old (N=32) (from regular and special education) were individually tested in a silent room. The VR games included building a tower of blocks with six scenarios: starting with a practice, one neutral scenario where no engaging event occurred, two instrumental gain scenarios (participants could choose to steal a block or ball from the virtual peer to obtain additional points and participant could win the game by sabotaging the virtual peer's game) and two provocative situations (participants were refused to join a game by two virtual peers and participants' game was ruined by the virtual peer). Results showed that peer provocation led to more anger, hostile intent attributions and revenge goals than the instrumental gain and neutral contexts, and more aggressive responses than the neutral context. The VR game and provocative situation are inspired by this study, as it was shown that the game was challenging enough but not too difficult for children aged 8-13 and the disruptive situation was designed to provoke an emotional response. The measures of emotional reactions in studies with children were taken using SAM scales [63]. Furthermore, there is a class of interventions proven to work in schools and to reduce cyberbullying in social media [61, 88, 90, 93], from which I base the interventions (see section 2.2.2.2). Additionally, the presence of visible authority figures and parental involvement can affect the effectiveness of interventions [42, 67, 70, 71, 98]. Given the feasibility of automatically detecting harassment in social media [103, 104, 105], and as new forms of harassment in social VR have become a growing concern [31, 21, 22, 20], it is important to investigate suitable interventions.

6.2.3 Methodology

This study was conducted in November-December 2022 with **43 children and 17 parents**.

This study focuses on the design and perceptions of an AEM for social VR, evaluated in a social VR gaming experience with children and parents. I simulated a social VR environment game with disruptive situations based on prior research [44] and implemented a Wizard-of-Oz AEM, Big Buddy, who put in place interventions when a disruption occurred based on the class of actions shown to be effective in prior work described above [61, 88, 90, 93]. I measured child reactions using SAM scales [63], and designed Likert scales and interview questions to measure child and parental perceptions towards Big Buddy. The protocol was approved by the university ethics committee.

6.2.3.1 Study Design

Experimental Design The study follows a Within-Participant design with Big Buddy's interventions as the independent variable of 4 levels:

(C1) [No BB] - Big Buddy is absent, no intervention is taken.

(C2) [BB, Reset Points] - Big Buddy is present and intervenes: resets harasser's points back to 0.

(C3) [BB, Reset+Notify Parents] - Big Buddy is present and intervenes: resets harasser's points back to 0 and notifies harasser's parents.

(C4) [BB, Reset+Notify Parents+Exclusion] - Big Buddy is present and intervenes: resets harasser's points back to 0, notifies harasser's parents and harasser is excluded from the game.

Each level represents a mix of punishments that increases in intensity, from no punishment to highly restricting actions. Big buddy announced these punishments verbally (via generated audios), and punishments were written in a bubble (see figure 6.2 c), ensuring that the child participant is aware of the actions taken. The punishments were based on the top three of the perceived effective sanctions in a school, according to children and parents [61]. The four conditions were counter-balanced with Latin Square: [C2,C3,C4,C1; N=10], [C3,C2,C1,C4; N=10], [C4,C1,C2,C3; N=13], [C1,C4,C3,C2; N=10] (as 5 participants were removed (see section 6.2.3.6). The steps of the experiment and conditions can be seen in figure 6.1. The provocative situation was the same throughout the game (i.e., destroying the participant's tower). Therefore, I randomised the number of games (2 to 4) for each round and the moment when the competitor would sabotage the game, randomising the fairness within each of the four conditions (fair win/loss, unfair loss) to reduce the effect of repetitiveness of sabotaging.

6.2.3.2 Social VR Game and Avatars Implementation

The simulated social VR game used to conduct the experiment was developed using Unity [322]. Its design was based on the game used to evaluate the aggressive behaviour in boys in prior work [44]. Based on the latter study, I simulated provocative situations where a virtual character sabotages the participant's game by destroying their built tower. I constructed avatars for the Big Buddy and the 4 competitors using the Ubiq library [324] with animations and a voice generator. To distinguish Big Buddy from the other avatars, its size was increased, and it was given a more noticeable appearance (see figure 6.2 e). The avatar featured a pre-made, male-oriented robotic outfit (see Limitations section 6.2.3.7). To match this visual design, Big Buddy was characterised as a robotic artificial moderator and used a monotone synthetic voice. This choice helped avoid adding variables, such as variation in tone or intonation, while leveraging the available visual assets from the Ubiq library.

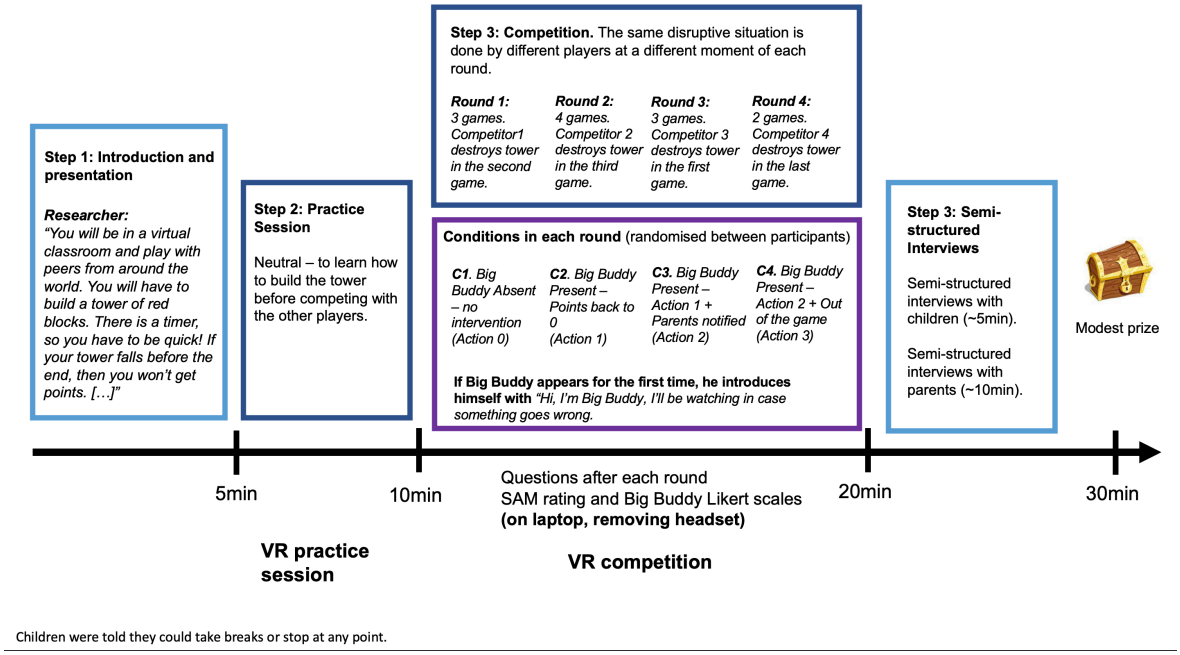
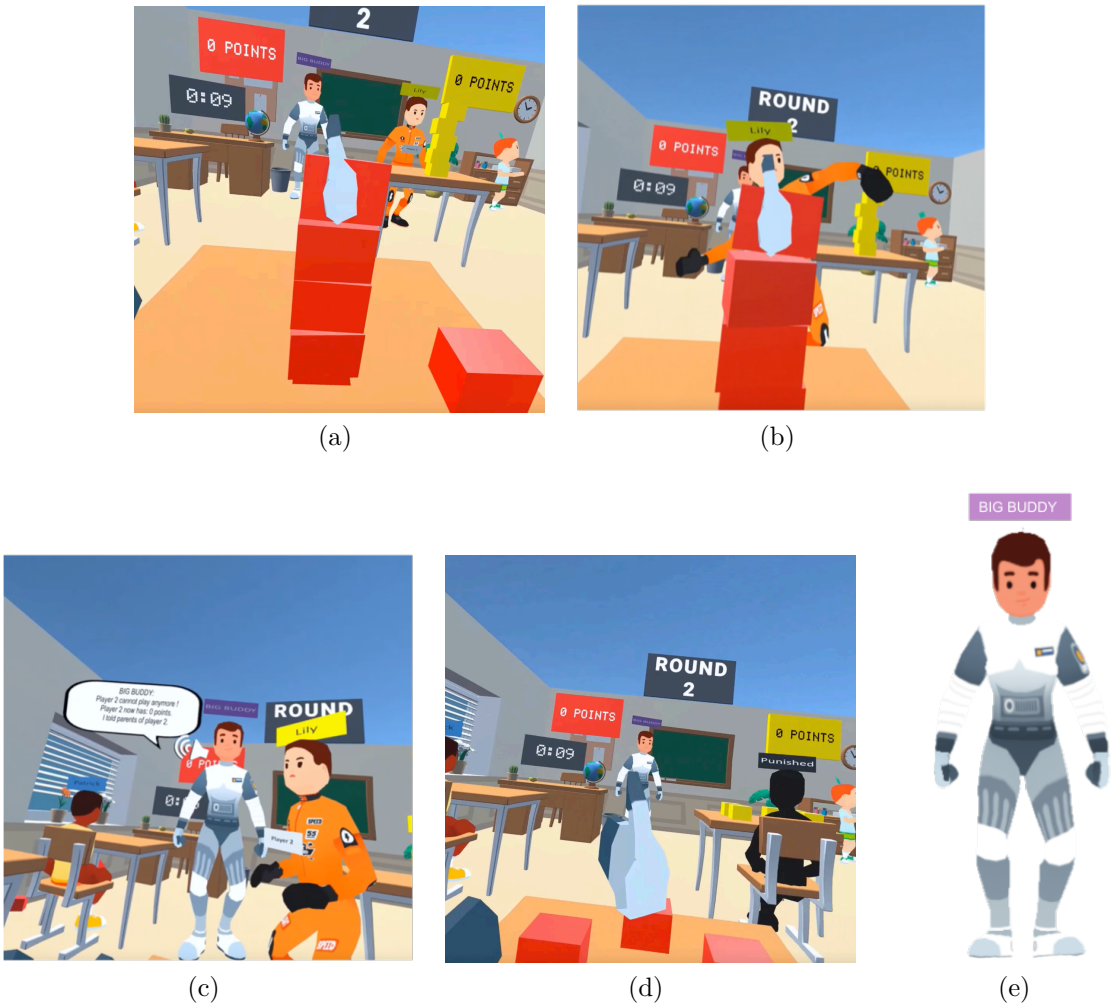


Figure 6.1: Experiment chronology and VR game conditions.

After the provocative situations were approved by a Professor of Child and Adolescent Psychiatry at the university, the game was pilot tested with eight adults and one child, before starting the experiment with children. Pilot tests primarily allowed me to note and fix technical issues (see Supplementary Material STUDY 4). A sample scenario in VR (from the participants' perspective) that occurs in one of the rounds for C3 is shown in figure 6.2. A video figure with the prototype can be found in the Supplementary Material STUDY 4/VIDEO.

6.2.3.3 Procedure

Children (one or two participants per session) were accompanied by their parent or by their teacher, and I welcomed them in an empty and quiet room. While the children played the VR game, the parent waited at the other side of the room, separated by a screen, where they watched a video recording of what the children would be experiencing in the VR game before the interviews. Before playing the game, the children were presented with slides showing what they would be doing (see slides in Supplementary Material STUDY 4/SLIDES). They were told they would be in a virtual classroom competing with real players from all around the world. To verify their understanding of the three SAM Likert scales, they were asked how they were feeling in that moment and to give a score: from sad (1) to happy (5) (Mean = 4.4, SD = 0.7), calm (1) to angry (5) (Mean = 1.4, SD = 0.5) and scared/intimidated (1) to safe (5) (Mean = 4.5, SD = 0.7). Children used a *Meta Quest 2* VR headset with an adjustable strap for more comfort. They were standing but did not need to move around the room. They were immersed in a virtual classroom with non-player virtual characters (e.g., peers sitting or talking in the background) to mimic a social VR environment. The game consisted of a practice session and four rounds during which they competed with each



(a) (b) (c) (d) (e)

Figure 6.2: Example of the VR game scenario occurring in round 2 with (C4) [BB, Reset+Notify Parents+Exclusion] from the user's eyes. (a) Tower Building Game. (b) Provocative situation. (c) Big Buddy intervenes: points reset to 0, parents notified and exclusion. (d) Competitor appears punished. (e) Big Buddy Close-Up.

of the four fictitious players. The goal was to build a tower of five blocks as quickly as possible and ensure it remains stable until the time limit ends. Points were awarded to players who successfully met this goal. If they had N blocks stacked together, they would earn N points. They started with a practice session where they could build the tower, and practise until they felt confident. I checked that they had managed to get at least 4 points (4 blocks stacked) before starting the competition, as the sabotage would mainly affect the tower-building. Overall, based on 4-point scales (0 'a little' to 3 'extremely'), children found the VR game a little easy (Mean = 1.6, SD = 0.8), requiring a little effort (Mean = 1.1, SD = 0.7) and the headset was very comfortable (Mean = 2.3, SD = 0.9). After each round, they had to fill in a form on a laptop (see Supplementary Questionnaire STUDY4/QUESTIONNAIRES). At the end of the game, the child and parent were interviewed individually.

6.2.3.4 Measures

Quantitative *After the practice session*, children were asked to give a score on 4-point scales (0 'a little' to 3 'extremely') for ease of use (standard SUS questionnaire item [356]), effort (standard NASA-TLX item [357]), comfort and their performance by giving the maximum points they got. *At the end of each round of the game*, children were asked to self-rate the emotions felt when the other player destroyed their tower, using the three 5-point SAM Likert scales [65], re-adapting the third one (from scared/intimidated to safe) (see figure 6.3). If Big Buddy was present in the round, they had to give a score on three 4-point scales (from 'a little' to 'extremely') focusing on the sense of agency: 1) Seeing Big Buddy when playing, 2) Feeling seen by Big Buddy, 3) Big Buddy helped with a fair punishment) and select among choices, the corresponding punishments that were put in place. *At the end of the game, during interviews,* children and parents were separately asked to give scores for preferred physical and social attributes, using 5-point scales from -2 to 2 (e.g., authoritarian to liberal, non-humanised to humanised, visible to invisible, amateur to expert). The latter scales were designed based on items used to create a tutor social robot [358], to gain a better understanding of what children and parents would want as a preferred AEM in social VR (See figure 6.4). The questionnaires can be found in the Supplementary Material STUDY 4/QUESTIONNAIRES.

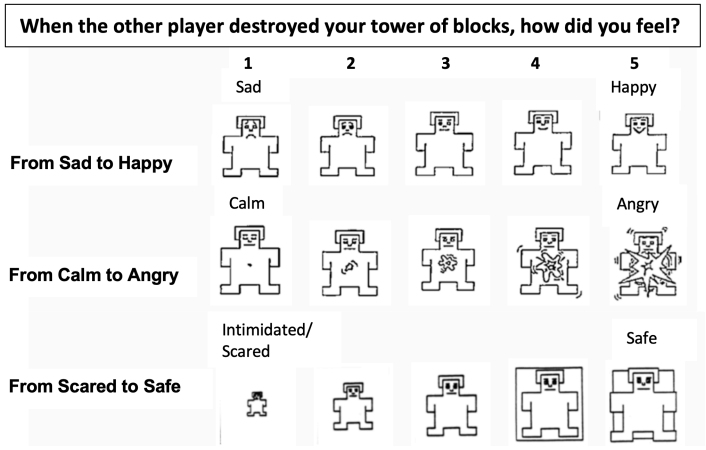


Figure 6.3: Self-Assessment Manakin (SAM) Likert scales. Re-adapted the third scale from 'not in control-in control' [65] to 'scared-safe'.

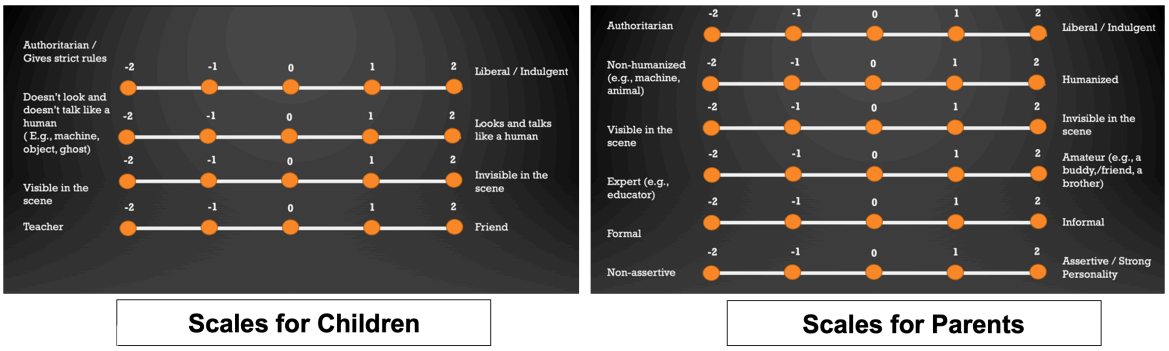


Figure 6.4: Attributes Scales For Preferred Embodied Moderator in social VR for children and parents.

Qualitative *At the end of the game, during semi-structured interviews*, children and parents were separately asked questions around their perceptions of the harasser and Big Buddy, their sense of safety, and potential customisation of Big Buddy. For example, children were asked to describe Big Buddy's role, how they felt when Big Buddy was present compared to when absent, and their perceptions of the different punishments.

Children's Interview Questions:

- Did you feel surprised in all the rounds when the other destroyed the tower, or did you start seeing it coming?
- What was Big Buddy's role?
- How did you feel when Big Buddy was absent compared to when he was present?
- What do you think of the different punishments made by the Buddy?
- If you were bullied in other ways (e.g., someone making fun of you or insulting you), would you feel safe if Big Buddy were there? (Why?)
- Let's say you can customise Big Buddy, you can change something to make you feel safer. Is there something you would change?

Parents were also asked about how they felt when Big Buddy was present, their perceptions of the punishments and how involved they would want to be given such a system.

Parents' Interview Questions:

- How did you feel watching the disruptive situations when the other player destroyed the tower? (Why?)
- What do you think of the Big Buddy's presence VS absence? (Why?)
- What do you think of the different punishments made by Big Buddy? (Why?)
- How involved would you want to be, if there is a virtual guardian moderating social VR?
- How would you want to customise the virtual guardian?

6.2.3.5 Analysis

Quantitative I analysed the quantitative data using R statistical tools. In particular, using ARTool for Aligned ranks transformation (ART) ANOVA, a non-parametric approach for multiple independent variables, interactions, and repeated measures [281, 328]. I conducted one-way ART ANOVAs (significance level 0.05) and post-hoc comparisons with Tukey adjustment with 1) SAM score as the dependent variable and conditions as the independent variable, and 2) scores for Big Buddy scales (seeing Big Buddy/feeling seen/fairness of punishment) as the dependent variable and conditions as the independent variable. I also used K-means clustering analysis to analyse preferred physical and social attributes of Big Buddy [359]. The optimal number of clusters was chosen based on the elbow method (the optimal k is the point on the graph where the rate of decrease starts to slow down significantly, which appeared to be at 3 clusters). Clusters were visualised using fviz_cluster() in R, along the first two principal components (Dim 1 and Dim 2), which together captured the largest proportion of variance in the data. The first principal component represents the direction of greatest variation in the data, and the second principal component represents the direction of the second greatest variation that is orthogonal to the first component. Participant IDs were used to trace back each cluster's corresponding scale scores. This enabled the characterisation of clusters in terms of participants' overall AEM preferences, highlighting distinct patterns. The analysis and codes can be found in the Supplementary Material STUDY 4/DATA_ANALYSIS.

Qualitative I employed inductive thematic analysis techniques [39, 283] to analyse parents' and children's perceptions from audio-recorded interviews. I did not seek inter-rater reliability because researchers may interpret the meaning of codes differently [360]. After generating transcripts, the experimenter listened to all recordings (∼5min each) and corrected any mistakes in the automatically generated transcripts. A second coder and I read and familiarised ourselves with the data. We both then created individual coding schemes independently using NVivo, line by line. The codes generated are words or short phrases that describe an idea. Then, the second coder and I collaborated to consolidate the two coding schemes into one combined scheme by collating or distinguishing between codes. We created a set of higher-level codes by bringing together related codes. (See Supplementary Material STUDY 4/CODING_SCHEMES). The second coder, two other researchers and I collaborated in an iterative process to discuss, combine, and refine themes and features, generating a rich description.

6.2.3.6 Participants

Recruitment The children (8-16 years old) were recruited via their parents on a voluntary basis and required the parent or legal guardian's permission to participate. The consent form and demographics questions were completed by the children's parent/legal guardian before booking a slot. The first batch (N_children = 20, N_parents = 14) was found via the university forums and mailing lists. The second batch (N_children = 28, N_parents = 3) was from a school where I had spent two days. This required permission from the headteacher and the parents, filling in the same form with the information sheet, consent form and demographics question, but they were not required to be present; instead, teachers and school staff accompanied the children to the experiment. I asked for a verbal agreement from the children to participate and informed them that they could take breaks or stop at any time. Due to technical issues and as two children did not want to play anymore, I had to omit in total of five child participants from the data analysis. The total number of participants is therefore: **43 children and 17 parents.**

Demographics The participants' age, awareness and frequency of use of general VR, social VR and online games (6-point Likert Scale; 0=never; 5=daily) are summarised in table 6.1. **Children:** Among the 43 child participants, 20 are female, 21 are male, 1 is non-binary/third gender, and 2 preferred not to say. Regarding ethnicity, there are 29 White Caucasian, four Arab, two Asian, one Hispanic/Latino, one Mixed (Scottish-North African), Four chose "Other" and two prefer not to say. 40 children were interviewed at the end of the game. Children were from 31 families. **Parents:** 18 mothers, 10 fathers, and two parents who did not disclose their gender responded to the form. There were 22 White Caucasian, three Arabs, two Asian, one Black/African, one Hispanic/Latino, and two preferred not to say. Among the 31 parents, there were 17 who accepted to participate in the interview (12 female, four male and one undisclosed gender; 10 White/Caucasian, three Arab, two Asian, one Black/African and one Hispanic/Latino). One parent mentioned their child had Asperger's Syndrome (Child31a), but could communicate comfortably.

Table 6.1: Table summarising demographics of parents, parents who were interviewed and children. (N_parents = 31, N_parents(interviewed) = 17, N_children=43). MEAN [SD].

Groups	**All Parents**	**Parents Interviewed**	**Children**
N	31	17	43
Age (in years)	42.37 [5.20]	42.35 [5.13]	11.49 [2.09]
Social VR Awareness (0-5 Likert scale 'not at all'-'extremely')	3.63 [1.10]	3.41 [1.12]	NA
VR Use Frequency (0-5 Likert scale 'never'-'daily')	0.33 [0.55]	0.35 [0.49]	0.68 [1.04]
Social VR Use Frequency (0-5 Likert scale 'never'-'daily')	0.27 [0.78]	0.41 [1.00]	0.40 [1.08]
Online Games Use Frequency (0-5 Likert scale 'never'-'daily')	NA	NA	3.54 [1.87]

Ethics and Compensation The protocol was approved by the university CoSE ethics committee (Approval number: [300210315]). I was registered with the national Protecting Vulnerable Groups (PVG) scheme to conduct this study with children. An £8 Amazon voucher and commute costs were compensated to adults for their time, and a token of appreciation (sticker/keychain) was given to the children. Documents with the information sheet, pre-survey with consent form and demographics questions can be found in the Supplementary Material.

Selection and Participation of Children I followed a similar selection process to the study by Evropi Stefanidi et al., published in the Interaction Design and Children (IDC) Conference 2021 [361]. The first batch (N=20) was found via university forums and mailing lists (researchers or lecturers with children). The second batch (N=28) was invited through the school's teacher, who also filled in a form giving consent and contacted colleagues who had children in the school. All participants' personal data were stored securely, and all personally identifiable data were removed. Before participation, the details of the study were explained to the children's legal guardians/parents in the information sheet sent after they expressed interest and gave their consent, in which it was noted that the child's participation is voluntary. At the start of each session, I explained to the child what they would be doing, with slides and a video clip of the VR game. I asked for a verbal agreement to participate and informed them that they could take breaks or stop at any point.

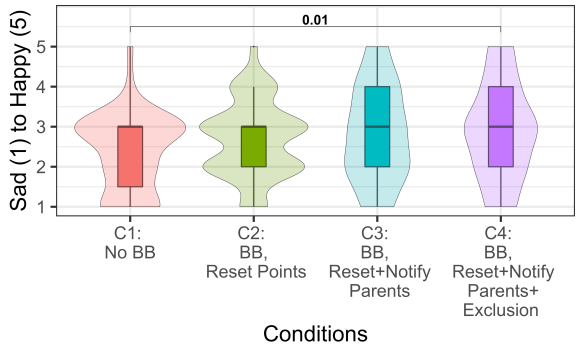
6.2.3.7 Limitations

Demographics: Participants were mostly White/Caucasian, and among the parents, those who were interviewed were mostly female. The study was done in English in the UK, which may introduce cultural bias. **Locations:** The experiment was carried out in two different locations; the first batch of children was accompanied by parents, whereas the second one was not and was in a school environment. However, in both cases, children were separated from the teacher or parent to reduce possible influence. In some sessions, the participants were in pairs playing the VR game at the same time (families with two children or in the school to avoid taking some of their valuable time), which may have influenced their responses. Nevertheless, I ensured they could not talk to each other, and they answered their questions quietly on a separate device. **Counter-balancing:** As I omitted 5 participants (due to technical issues and two children not wanting to play anymore), the Latin Square randomisation of the four conditions is imbalanced. As the study is within-subjects, there are some limitations as they might carry on valence effects (e.g., participants still angry from C1 when indicating their anger levels after C2), but this is more ecologically valid as such a situation could occur in social VR. I counter-balanced the win-loss of the games (fair win or loss and unfair loss) as well as the number of games per round to reduce this effect. **Surprise Effect:** While participants may have expressed that they were very surprised in the first round but felt less surprised from the third round when the other player destroyed their tower, it is not the surprise or expectation of the attack that changed their perception of lack of fairness and need for interventions. **Big Buddy Design:** The avatar used was male-oriented and selected from Ubiq's pre-designed library, which may have influenced children's perceptions. Future research should explore alternative designs and forms of embodiment to better understand their effects.

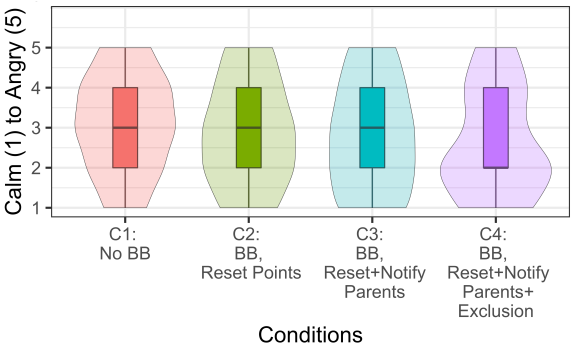
6.2.4 Results

6.2.4.1 Children's Emotional Reaction after the Disruptive Situation

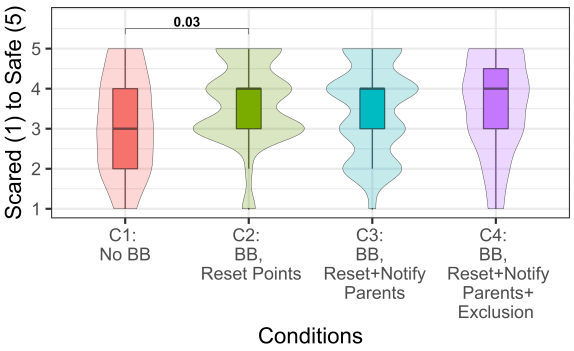
Overall, children felt more negative valence emotions (sadness) and felt less safe when the other player destroyed their tower and when Big Buddy was absent compared to when Big Buddy was present (see figure 6.5). However, in terms of arousal, children's anger was relatively the same throughout all the rounds, which shows that even if the disruption was repeated, their level of anger did not particularly change. Significant results (significance level set at 0.05), interpreted using partial eta squared η_p^2, with medium (≈ 0.06) and large (≈ 0.14 or higher) effect sizes [286, 287], were obtained for the different conditions for sadness ($F(3,117)=3.55$, $p=0.02$, $\eta_p^2=.08$, medium effect size, ART ANOVA) and for safety ($F(3,117)=3.22$, $p=0.03$, $\eta_p^2=.08$, medium effect size, ART ANOVA). In particular, children felt significantly sadder ($p=0.01$) in the round with (C1) [No BB] compared to the round with (C4) [BB, Reset+Notify Parents+Exclusion]. Moreover, children felt significantly safer ($p=0.03$) in (C2) [BB, Reset Points] than in (C1) [No BB].



(a) SAM scores (1 'sad' to 5 'happy')



(b) SAM scores (1 'calm' to 5 'angry')



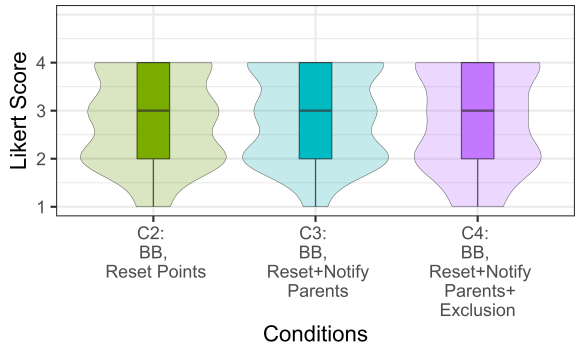
(c) SAM scores (1 'scared' to 5 'safe')

Figure 6.5: Violin-boxplots of three SAM Likert scales' scores for each condition. (a) Children felt significantly sadder in the round with (C1) [No BB] compared to the round with (C4) [BB, Reset+Notify Parents+Exclusion]. (b) No significant effect. (c) Children felt significantly safer in (C2) [BB, Reset Points] than in (C1) [No BB]. Significant pairwise comparisons are labelled.

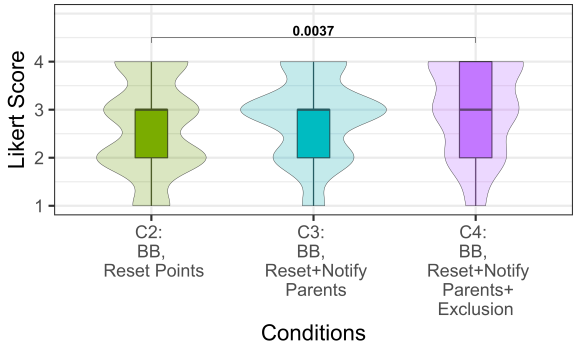
Finding 1: Children felt significantly sadder in (C1) [No BB] compared to (C4) [BB, Reset+Notify Parents+Exclusion]. Moreover, children felt significantly safer in (C2) [BB, Reset Points] than in (C1) [No BB].

6.2.4.2 Perceptions towards the AEM (Big Buddy)

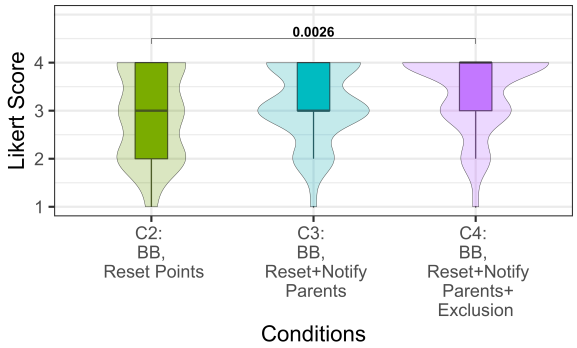
Children's Perceptions towards Big Buddy From self-ratings (sense of agency), children felt they were 'seeing Big Buddy' while playing similarly in all conditions where Big Buddy was present. Nevertheless, there was a significant difference in scores across conditions when answering if they 'felt seen by Big Buddy' when playing ($F(2,84)=5.58$, $p=0.005$, $\eta_p^2=.12$, medium effect size, ART ANOVA). More specifically, they felt that Big Buddy could see them significantly more in (C4) [BB, Reset+Notify Parents+Exclusion] than (C2) [BB, Reset Points] ($p=0.0037$) (see figure 6.6).



(a) "I could see Big Buddy while playing" (1 'not at all' to 4 'a lot')



(b) "I felt Big Buddy could see me while playing" (1 'not at all' to 4 'a lot')



(c) "Big Buddy helped with a fair punishment" (1 'not at all' to 4 'a lot')

Figure 6.6: Violin and boxplots of the three 4-point scales' scores for Big Buddy conditions. (a) No significant effect. (b) Children felt that Big Buddy could see them significantly more in (C4) [BB, Reset+Notify Parents+Exclusion] than (C2) [BB, Reset Points]. (c) Children felt that the punishment in (C4) [BB, Reset+Notify Parents+Exclusion] was significantly fairer than in (C2) [BB, Reset Points]. Significant pairwise comparisons are labelled.

Big Buddy's Role Seven children described him as being like a teacher, seven described his role as keeping the game fair, seven mentioned he is there to keep them safe, six said he is there to punish and 14 described him as someone to help regulate other users' actions: *"His presence helped to regulate what the people did"* [SchoolChild17]. *"I think it was to help if any of the other children were mean, then it meant that they would get the punishment that they deserved"* [Child22a].

Impact of Big Buddy's Presence VS Absence Children felt that Big Buddy's presence was reassuring and they felt safer (N=10): *"I felt much safer when he was there and a lot more protected"* [Child32a]. However, two children had mixed feelings; one of them mentioned it was comforting but strange at the same time [SchoolChild14]. Children also linked Big Buddy's presence with the interventions and moderation; they knew interventions would only occur if Big Buddy were shown in the scene (N=5): *"Well, when he was present, he made sure that if they did something wrong, they would get in trouble. When he wasn't there, nothing happened"* [Child28a]; *"I didn't like it when he wasn't there because they'd knocked down my tower and nothing happened"* [Child25b]. Eight children were annoyed by the absence of Big Buddy due to the lack of punishment. Twenty-five children responded that Big Buddy's presence would make them feel safer in other bullying situations (e.g., name-calling or someone making fun of them). Children noted he would be able to help them (N=9) *"He'd be there to like protect me and he'd give them a fair punishment".* [Child32a]. Big Buddy's absence mostly led to making the child feel negative feelings (N=13), including: being annoyed (N=3), nervous (N=2), afraid (N=1), not safe (N=1), sad (N=1), tormented (N=1), having less fun when he was absent (N=1). *"I felt like I had more fun when he was there because he was making it fair"* [SchoolChild01b]. However, one child mentioned feeling better without Big Buddy *"when he was present, it was just weird because all I felt was him staring at me, not looking at the other student, but when he was not there, I felt better."* [Child01a] and found Big Buddy's presence inhibiting fun.

Some children did not really notice him or pay attention to him as they were focused on their game until the disruptive situation and punishments were put in place (N=9), other children did not realise when he was not there (N=3), and some mentioned they were not looking at him, but they knew he was there (N=3).

Finding 2: Children felt significantly more "seen by Big Buddy" in (C4) [BB, Reset+Notify Parents+Exclusion] than (C2) [BB, Reset Points]. Most described Big Buddy as keeping the game fair, ensuring safety, and punishing unfair behaviour. His presence was generally reassuring, while his absence often led to negative emotions; a few found him strange or did not notice him.

Parents' Perceptions towards Big Buddy As parents watched the recorded video of the VR game, they were asked about their first impressions of the disruptive situations and how they felt knowing their child went through these events. Parents felt negatively (N=17), including the feeling of anger (N=2) *"Maybe a bit of anger [...] there was nothing I could do [...] I wasn't sure he's been in such situations before [...]"* [Parent33], annoyance (N=2), confusion (N=2), frustration (N=3), worry (N=2), feeling sorry for the children (N=1), feeling surprised or shocked (N=5), discomfort (N=1), unfairness (N=1), feeling strange because it is not real (N=1), feeling passive as it is just a game (N=1), it reminded them of their own childhood experience (N=1) or hoping that their parenting allowed their child to be able to deal with these situations (N=1). For example: *"I hope that we've helped our child to be able to deal with that sort of thing. If they want to play, then they need to be prepared for these things happening [...]"* [Parent30]. Parents also shared how they would imagine their child feeling: children would feel angry or upset (N=2), not a pleasant experience (N=1).

Impact of Big Buddy's Presence Parents noted the usefulness of an embodied safeguarding system such as Big Buddy to increase their children's safety and well-being in virtual social environments. Parents mentioned Big Buddy's presence was reassuring (N=9), as someone was watching them (N=4) with quite a reassuring appearance (N=1). They would have more fear if he were not there (N=1) and felt it would make the place safer for their child (N=3), where they can feel more confident (N=1). Parents also thought Big Buddy's presence was good and important (N=8) and would want him to be there rather than having no supervision (N=8). One parent noted: *"I think it depends on how much of a presence there is and how much it feels like you're being supervised. I suppose in some instances it would feel a bit weird to have someone looking over your shoulder all the time, but in the same sense because that other player had already destroyed the tower [...]"* [SchoolParent01]. Some parents found Big Buddy helpful and useful (N=7) *"I think it's a helpful system not just not just suitable purely in terms of how they will be punished but to have that as a safety measure I think might be nice for everyone so that if something goes wrong there will be consequences [...]"* [Parent30]. Parents, however, indicated a few dislikes about Big Buddy: too intimidating (N=1), too passive (N=1), potentially not effective, its effectiveness may depend on children's age and if they take him seriously (N=2).

Finding 3: Parents reacted negatively when seeing the disruptive situations. They generally valued Big Buddy as a reassuring and useful presence for safeguarding children in social VR. Most preferred his presence over no supervision, though a few found him intimidating, passive, or questioned his effectiveness, depending on the child's age and perception.

6.2.4.3 Perceptions towards Punishments (Fairness of AEM)

Children's Perceptions towards Punishments As part of self-rating, children evaluated if Big Buddy put in place a fair punishment using a 4-point Likert scale. Across conditions, results' scores were statistically significantly different ($F(2,84)=5.94$, $p=0.004$, $\eta_p^2=.12$, medium effect size, ART ANOVA). Children felt that the punishment in (C4) [BB, Reset+Notify Parents+Exclusion] was significantly fairer than in (C2) [BB, Reset Points] ($p=0.0026$) (see figure 6.6 c).

Preferences have been raised for punishments during interviews. Children considered that the punishments were a way to increase safety (N=3). Among the three actions taken, children had different opinions regarding which punishment or combination of punishments was the fairest. Six children noted that putting in place the three punishments ((C4) [BB, Reset+Notify Parents+Exclusion]) was better than putting just one ((C2) [BB, Reset Points]). Seven highlighted that the fairest punishment is having the player out of the game. Some children thought that the combination of punishments was a bit extreme and severe (N=7), and that points reset to 0 was fair (N=13). In particular, Child01a and SchoolChild15 both suggested having at least a first warning before the harasser is banned. However, resetting the competitor's points felt useless to some children (N=2). Two children preferred when the harasser's parents were contacted. The latter was particularly disliked by others: feeling weird and uncomfortable (N=1), seemed as an unfeasible punishment (N=1), or not necessary (N=1). Children pointed out that punishments should have been the same as the disruption was the same (N=5).

Finding 4: Children rated punishments as significantly fairer in (C4) [BB, Reset+Notify Parents+Exclusion] than (C2) [BB, Reset Points]. While many preferred multiple punishments for fairness and safety, some found them too severe or favoured warnings before bans.

Parents' Perceptions towards Punishments Similar to children, preferences and pros and cons have been raised for each punishment. Overall, parents found the punishments adequate and reasonable (N=8), and one noted that all is covered *"I thought that's about as much as you could do in that situation, that's probably the most you could do"* [Parent03]. Parents particularly liked the immediate consequence of having the harasser's points reset to 0 (N=2). Some parents were positive about the time-out punishment (N=6) and liked the label showing that the player is punished (N=1). Others, similar to some children, found the latter punishment severe (N=4), and a parent suggested having a warning before exclusion. Regarding the punishment of parents being notified, parents were sceptical as to whether it is effective (N=4) in terms of other parents' parenting, and if they would notice, it might worry parents, and for practicalities, you would need parents' contact details. Parents noted that punishments would depend on children's age, as there might be an age limit for punishments' efficiency (N=2) *"I think that's probably going to reach a point where your kids are old enough they don't really mind about that [...]"* [Parent17]. Finally, parents suggested a progression of punishments based on repeated bad behaviour or level of bad behaviour (N=3) and making sure children understand Big Buddy's intentions (N=1).

Finding 5: Parents generally viewed punishments as adequate, valuing immediate consequences like point resets, though some questioned the effectiveness of notifying parents and suggested age-appropriate and gradual increase of punishments.

6.2.4.4 Customisation of Big Buddy

How children envision the AEMs. While eleven children mentioned they liked Big Buddy the way it was presented in the experiment's game, eleven others indicated wanting Big Buddy more like a realistic human and less like a robot and AI-looking, wearing normal clothes (N=10) and having a normal name (N=1). They found Big Buddy a bit intimidating and would prefer it to be more friendly (N=3). They would also prefer visual and audio characteristics that tend towards more real-life authority figures (teacher/parents) (N=2), familiar/positive-related figures (Game Characters) (N=2), or their friends (N=1). In particular, two children did not like the robotic voice and found it uncomfortable. Children suggested having the possibility of personalising it (N=3), *"I think different people would want to see themselves or like some of their interests or something"* [SchoolChild14]. SchoolChild08 suggested having Big Buddy with positive-related clothes or gestures, for example, wearing a t-shirt with a motivating and helpful note "You will be fine" and having him give thumbs up. Children's visions of a virtual supervisor based on the items of the scales were nuanced. Using K-means clustering analysis, I was able to observe three main clusters figure 6.7. Among the three clusters, I observed preferred attributes for cluster 2 leaning towards a more realistic authority figure: authoritarian, visible, humanised, teacher, whereas cluster 3 preferred a more friendly, indulgent supervisor, but still visible and humanised and cluster 1 would want it non-visible.

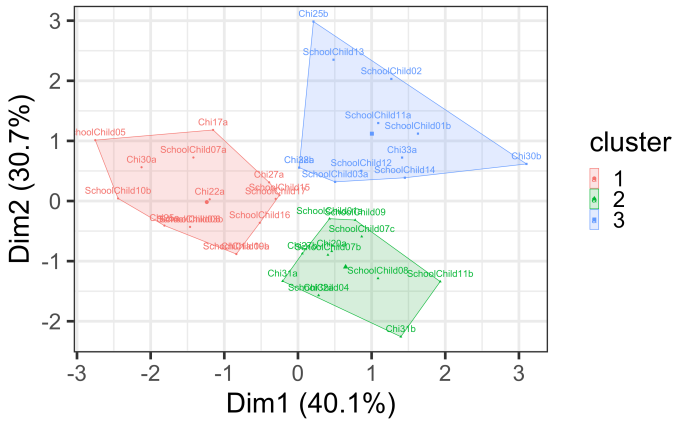
Finding 6: Children's views on Big Buddy's appearance were split: some liked him as-is, others preferred a more human-like, friendly figure with normal clothes, a natural voice, and customisable features. Suggestions included making him resemble familiar authority or positive figures. Cluster analysis revealed three archetypes: 1) non-visible, 2) authoritarian, visible, humanised, teacher, 3) friendly, indulgent but still visible and humanised.

How parents envision the AEMs. In terms of parents' preferred social and physical characteristics for an AEM, preferences among parents were shared again. Some said it should be context-dependent, based on the game children play (N=1), children's age and their preferences (N=1), and the situation (N=2). Four parents found Big Buddy too bulky, like a *"big law enforcing robot"* [Parent22] and would prefer someone more friendly (N=1), like an older kid as opposed to an adult (N=1), smaller and more discrete (N=2), with the interventions occurring in the background rather than in front of everyone (N=1). In contrast, it was suggested that Big Buddy names the player who did something wrong *"because they deserve to be like at least having something against the reputation"* [Parent32]. Parents also wanted something more familiar like a parent (N=1), a teacher (N=2), with a more natural voice (N=1), and more inclusive (e.g., gender, skin colour, personalised voice) (N=2): *"I would make them more um inclusive like showing white person and black person being supervisor*

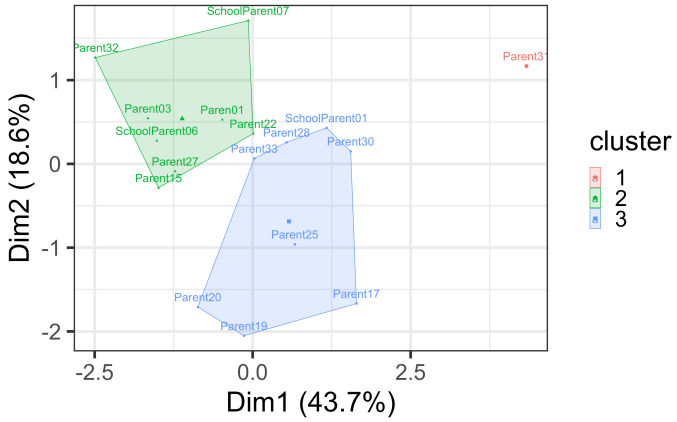
and maybe personalise as well their tone of voice" [Parent01]. Parent30 suggested it could be dressed differently for different/special roles, but highlighted that it would not be the parents' role to choose the appearance etc., but it should be personalised by the player. However, SchoolParent01 questioned whether making the supervisor similar or completely distinct from other players would be helpful (N=1). Another would prefer if it is non-human like (N=1) *"I would prefer to be honest to not be like a human being, I would like it to be someone else, like an alien or something different because [...] now it's like having someone who is outside of the humanity".* [Parent20]. Big Buddy could also be more informative so that the children understand its purpose (N=2).

Using K-means clustering analysis, I was also able to observe three main clusters among the sample of parents (see figure 6.7). Preferred attributes for cluster 3 leaned towards a non-humanised and indulgent figure, but formal and assertive to some extent, whereas cluster 2 preferred a more realistic humanised expert and authoritarian supervisor and cluster 1, an outlier (Parent31) would want it as a complete opposite of the latter (indulgent, non-humanised, more friendly being an amateur, informal and not too assertive). Parent31 mentioned still wanting some regulation to some extent, but would want to prioritise being a relaxing and fun experience for the children. Overall, preferred attributes from children and parents' perspectives can be visualised in figure 6.8. The reasons for their given scores are summarised in table 6.2 and table 6.3.

Finding 7: Parents' preferences for Big Buddy's appearance varied. Some wanted a friendly, smaller, more discreet figure; others preferred a familiar authority such as a teacher or parent, with a natural voice and inclusive features. A few preferred non-human forms (e.g., alien-like). Many supported personalisation by the player. Cluster analysis revealed three archetypes: 1) indulgent, non-humanised, amateur, informal, not too assertive, 2) humanised, expert, formal, authoritarian, visible, assertive, 3) non-humanised, visible, indulgent but formal and assertive.

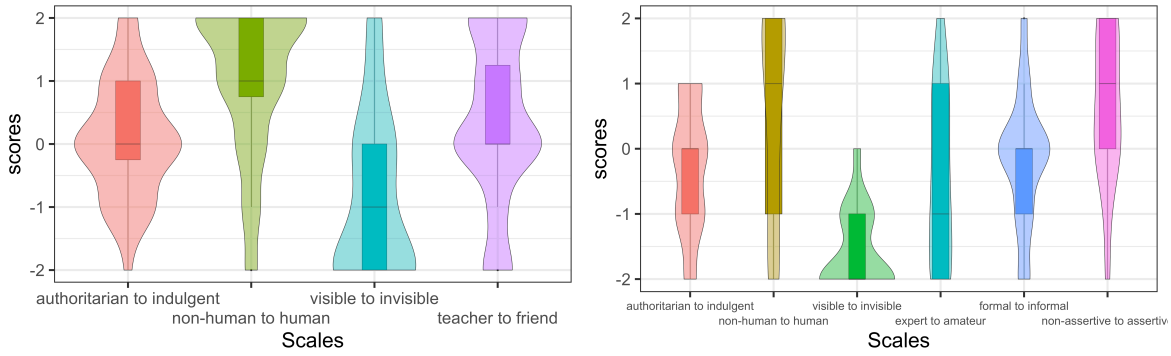


(a) Children K-Means clusters with 4 variables. Cluster 1 red: non-visible; Cluster 2 green: authoritarian, visible, humanised, teacher; Cluster 3 blue: more friendly, indulgent supervisor but still visible and humanised.



(b) Parents K-Means clusters with 6 variables. Cluster 1 red (outlier): indulgent, non-humanised, amateur, informal, not too assertive; Cluster 2 green: humanised, expert, formal, authoritarian, visible, assertive; Cluster 3 blue: non-humanised, visible, indulgent but formal and assertive.

Figure 6.7: Clusters for an AEM attributes scores given by children and parents. The first principal component (dim1) represents the direction of greatest variation in the data, and the second principal component (dim2) represents the direction of the second greatest variation that is orthogonal to the first component. Labels inside clusters correspond to participants' IDs.



(a) Children Scores (b) Parents Scores

Figure 6.8: Violin-Boxplots of children's and parents' scores given for the attributes (social/physical) of the scales for a preferred AEM.

Table 6.2: Summary of children's qualitative explanations for their AEM scores. Percentages indicate the proportion of responses in negative, neutral, or positive categories for each scale.

Scale	**Negative**	**In the middle**	**Positive**
From authoritarian (-) to indulgent (+)	*25%*: "A bit strict but not unpleasant to be there." [SchoolChild17] "If alone, I'd rather feel they're a bit strict." [Child17a] "To get everyone to do the right thing." [Child32a]	*45%*: "If too strict, not fun anymore." [Child33a] "You'd want like a strict rule that puts them in place, but not too strict so that it's still kind of a competition" [SchoolChild16] "Should have strict rules, but give chances." [SchoolChild15]	*30%*: "Not strict unless something happens." [SchoolChild14] "If he gave really strict rules, it wouldn't be very fun." [SchoolChild11b]
From non-human (-) to human (+)	*12.5%*: "Fluffy monster" [SchoolChild02] "he would look like Blue" [Child33a] "Maybe a frog." [SchoolChild01b]"because when it was a human, it just felt like it was real life." [Child25b]	*12.5%*: "I wouldn't really mind because they'd do the same thing." [Child30b] "I play VR to kind of escape the real world so if everything's like not real you know it's better but I don't I think a ghost would not help" [SchoolChild01a] "I feel like I'd want I'd want them to talk like a human but then looks like an animal that sometimes help in those situations" [SchoolChild11a]	*75%*: "Because I can relate to humans." [Child32a] "If human, feels like they could actually do something." [SchoolChild12] "More natural" [Child17a] "I think a human's probably the best one because it's more like us [...] because if they were a ghost or a machine it would be a bit weird" [SchoolChild09b]
From visible (-) to invisible (+)	*62.5%*: "So that I know that it's there to keep me safe." [Child32a] " I'd rather him be there and listen to what happened" [Child01a]	*22.5%*: "Something that appears at the start so you knew it was there but then disappeared during the actual gameplay until needed" [SchoolChild17] "Sometimes appears." [SchoolChild14] "Because it's a game, but it's kind of powerful". [Child32b]	*15%*: "I'd want to know they were there, but I wouldn't want to be able to see them." [Child17a]
From teacher (-) to friend (+)	*20%*: "I wouldn't want them too casual." [Child17a] "They'd listen to a teacher for authority." [SchoolChild10b]	*37.5%*: "A friendly teacher." [SchoolChild17] " I would feel safer around both." [SchoolChild03b] "because I wanted it to be friendly but also an expert" [SchoolChild01a]	*42.5%*: "I would want a friend but older." [SchoolChild14] "A friend you know is more reassuring" [SchoolChild11b]

Table 6.3: Summary of parents' qualitative explanations for their AEM scores. Percentages indicate the proportion of responses in negative, neutral, or positive categories for each scale.

Scale	**Negative**	**In the middle**	**Positive**
From authoritarian (-) to indulgent (+)	*41.2%*: "To avoid damaging the kids." [Parent32]	*35.3%*: "Children might rebel if too authoritarian; if too liberal, they might ignore it." [SchoolParent01] "Playfulness affected." [Parent33] "Someone to intervene when something not nice happens." [Parent30]	*23.5%*: "Should be relaxing; not authoritarian by default, sterner if misbehaviour occurs." [Parent31]
From non-humanised (-) to humanised (+)	*29.4%*: "Easier to trust someone not simulating to be human." [Parent20] "Non-human for younger kids." [Parent31]	*11.8%*: "Maybe children want different things." [Parent30]	*58.8%*: "Has more authority and respect." [Parent32] "Relevant for the society that they live in. They should interact with human beings rather than machines." [SchoolParent07]
From visible (-) to invisible (+)	*94.1%*: "As reminder of behaviours." [Parent25] "Depends on age but good for everyone." [Parent30] "Important to see someone's there." [Parent03] "Nice for them to see that somebody is there." [Schoolparent07]	*5.9%*: "Impersonal just like a pop-up or something." [Parent31]	*0%*: NA.
From expert (-) to amateur (+)	*52.9%*: "Authority figure so they know there's a consequence." [SchoolParent06]	*17.6%*: "Should focus on training rather than labels." [Parent30] "Someone child feels comfortable with." [Parent01]	*29.4%*: "Gaming environment to escape educators." [SchoolParent01] "Less senior/authoritarian, just facilitator." [Parent31]
From formal (-) to informal (+)	*29.4%*: "Goes along with authority." [SchoolParent06] "If informal, children may ignore." [SchoolParent01] "Introduce formality for the child to listen." [Parent33]	*52.9%*: "Kids react against too formal." [Parent17] "Presence informal but formal when drawing line." [Parent30]	*17.6%*: "Generally informal by default but formal when reminding rules." [Parent31] "Wouldn't want to put children off the game. A figure that is friendly but in charge." [Parent25]
From non-assertive (-) to assertive (+)	*11.8%*: "Gentle approach more effective." [Parent17]	*35.3%*: "Fun is important but some supervision needed." [Parent25]"Not like controlling police but minimum interference." [Parent30]	*52.9%*: "Assertive but still fun." [SchoolParent06] "Children will pay more attention." [SchoolParent01]

6.2.4.5 Parents' involvement in Safeguarding

While parents noted the usefulness of Big Buddy and how reassuring it might be, they would nonetheless remain in the supervision loop with different levels of involvement:

Trusting The System and Staying Involved to Safeguard Their Children Parents mentioned they would want to be more involved, specifically in the beginning (N=8), to see how it works and if it is efficient before trusting the system:

> *"If I could see it operating okay and knew that it was effective for a certain period of time, then I could probably get to the point where I say okay, I know that this does work [...] I wouldn't trust it like I wouldn't trust anybody that I don't know."* [Parent22]

> *"I'd like to know what the system is and how it works, but once I've done that, I just think it's good for kids to be able to work things out between them so long as it's a fairly genuine system where all the kids will actually be genuine and they actually exist and so on."* [Parent30]

Parents would not trust the system or lack trust to some extent (N=7) until it has proven its effectiveness (N=2).

Real-Time Notification of Disruptive Events in Social VR Eight parents mentioned they would want to be notified in real-time about the disruptive situations and the outcomes of the behaviours if there were a notification system. Parent15 and Parent19 pointed out that they would want to be involved without real-time notifications, as it may be difficult to manage and take care of real-time notifications, and the latter system might pressure and inhibit children's fun and expression, according to Parent15. Parent19 suggested receiving one notification or a report at the end of the day on what happened. Furthermore, parents suggested it would depend on different factors: SchoolParent01 highlighted that the use of the notification system (e.g., content and recurrence of notifications) may also depend on the age of the child, and Parent25 pointed out it would depend on the level of the situation, if it is repeated, for example.

Finding 8: Parents valued Big Buddy's potential for safeguarding children, but would want to remain involved at the setup, real-time with notifications of the disruptive events or with a report at the end of the day.

6.2.5 Discussion

Reflection around the AEM and future work are provided in chapter 8 to situate this study's findings within the wider contributions of the thesis. Here we present a summary of the findings and references to corresponding guidelines in chapter 9.

6.2.5.1 Summary of Findings: The Children's Perspective

S4-RQ1: How is children's emotional valence impacted in provocative situations by the presence of Big Buddy? - Big Buddy's presence and interventions have significantly impacted children's reactions to the disruption. Its absence induced negative feelings (e.g., sadness, frustration), whilst applying the full set of punishments (C4) significantly reduced sadness compared to no intervention (C1). **S4-RQ2: How do children perceive the embodiment of the AEM (Big Buddy) and moderation actions (punishments)? Will children feel safer and/or inhibited given its presence?** - Interestingly, C2 has also been perceived as substantially fairer than C1, which led us to believe that perceived fairness is a potential determinant of children's post-harassment reactions. Big Buddy's perceived role is congruent with these results, as children compared him with a referee to keep people safe, punish bad behaviours and ensure game fairness. Its presence is associated with active punishments and increased perceived safety. Regarding punishments, some children underlined the unfairness of applying different punishments for the same action. They also criticised the ban and notifying parents as too harsh. However, others thought it was appropriate and preferred the most severe interventions. **S4-RQ3: customisation of moderator** - Children nonetheless raised concerns about Big Buddy's appearance and shared their design preferences. They can be clustered in three trends: 1) a more realistic authority figure, 2) a more friendly, indulgent moderator, and 3) a non-visible moderator.

6.2.5.2 Summary of Findings: The Parents' Perspective

S4-RQ2: How do parents perceive the embodiment of the AEM (Big Buddy) and moderation actions (punishments)? - Parents showed empathy to their children experiencing the attacks and said they felt reassured by the presence of an AEM such as Big Buddy. Parents provided similar feedback to children's and added that notifying parents may not be ideal. Its effectiveness relies on others' parents, and bullies may not be sensitive to it. Some proposed graduated interventions and an initial warning before intervening. **S4-RQ3: customisation of moderator** - They expressed similar design preferences as their children but mentioned that its appearance and type of intervention would depend on various factors (e.g., gaming context, children's preferences, bad behaviour characteristics). **S4-RQ4: parents' involvement** - They were not completely convinced by letting an AEM ensure the safety of their children in Social VR and expressed their desire to better understand the process and remain in the loop.

6.2.5.3 Implications for Designing Embodied Moderation to Safeguard Children in Social VR

Child and Parental Preferences are both Important Grounded in the findings, children's and parents' answers were quite similar. Parents valued what their children would want, their needs and preferences and noted that enjoyment is important. Most children and parents felt reassured having a system such as Big Buddy to safeguard social VR. This means the design of AEMs should involve the perspectives of both parents and children. Overall, the majority of children and parents would prefer an AEM that is visible, looks and talks more like a human, with a middle-ground between authoritarian and indulgent, friendly but that is still respected with an assertive personality. However, a few children and parents noted possible feelings of discomfort or strangeness being overly watched and would prefer something more discreet that appears only to intervene. Some parents and children found Big Buddy too intimidating, like a "law-enforcing robot". Both parties mentioned wanting something more familiar and inclusive. In terms of punishments, some children and parents valued fairness and found that resetting the harasser's points to 0 was fair, noting that exclusion might be too severe, and a warning before exclusion and progression of punishments based on the level of bad behaviour or repetitiveness would be fairer and more effective.

Leveraging the Opportunities brought forth by AI-driven and Embodied Moderation Systems This study reveals a number of particularly novel and useful features of AEMs that should be capitalised on when designing such systems.

Immediate Consequences: One of the many opportunities enabled by systems like Big Buddy is allowing "*Immediate Consequences*". Indeed, some parents highlighted they particularly appreciated how the sabotage was detected and dealt with in real-time. This is in contrast to typical moderation systems, where decisions are delayed until they are reviewed by human moderators. Future systems can use social signal processing to detect when children are in distress and intervene accordingly [362, 158].

Customisation: Another advantage is that embodied moderation systems are "*customisable*". This allows for personalisation in multiple ways as it can be programmed to be perceived differently by different users [11]. For example, the embodied avatar can be customised to represent an authority figure from the perspective of the harasser, while at the same time representing a friendly figure from the victim's perspective. This way, each user would see two versions of Big Buddy: a version that is on their side and a version that is effective in stopping said user from performing further offensive actions. The question remains as to which attributes should be personalised by the child to feel more comfortable and safe, which settings and rules can be personalised by the parent, whether parents would make good arbiters of how Big Buddy should act and be presented, and which features would be fixed for all the players in social VR. Customisation raises concerns about the ability of parents and children to make informed decisions and the extent to which parents should be able to influence

punishments for their child as a harasser and/or a victim, particularly if they may be promoting harmful behaviour. Future research may be required to determine which aspects of moderation should be applied universally, tailored to specific demographics of children, or customised for individual families.

Parents can stay involved: Alongside the results, the experimental design showed that having parents view the recording of the VR game in real-time could allow parents to offer sympathy and empathy to their children. Warm and positive parent-child relationships are important for the growth and development of children and need to be maintained through moderation and monitoring. In particular, some parents valued giving space to their children, letting them have a relaxing experience and being able to express themselves and still have fun. They may trust the AEMs only to an extent until it has proven its effectiveness. However, they would also want to be aware and know what is happening and were positive about receiving real-time notifications if something bad occurs. There are also tensions around feasible parental involvement. Despite the benefits of real-time observation (which is often impractical), we need more research into asynchronous forms of parental involvement that can still give them the same opportunity for insight, oversight, and support.

Perceived Effectiveness of AEM Based On Multiple Factors: This study shows the *potential* usefulness of an AEM. However, its real-world usefulness and effectiveness may vary based on the children's age, their maturity, background, education and their preferences. It may also vary based on the social VR platform used and the situation in the game, as parents pointed out. Parents suggested that for younger children, someone visible and familiar would most likely be more effective than for older teenagers. Therefore, there is a need for personalised AEMs. For social VR, a promising approach is the creation of pre-designed, validated AEMs using proven intervention strategies as a necessary first step to adapt for different contexts (e.g., age groups, games and situations).

Importance of Safe Virtual Spaces for Children: Considering Context and Real-World Applications This study on AEMs in social VR settings is crucial, given the increasing number of children in these environments and the potential for new forms of harassment. However, the effectiveness of AI moderation systems depends on various contextual factors such as the type of social VR experience, age range of users and type of threat (e.g., verbal, physical or environmental). To design safe virtual spaces for children, I need to consider these factors and explore potential redesigns that prevent harassment and foster positive interactions.

This study may also have implications for real-world bullying interventions. The principles used to create safe and positive virtual spaces may be applied to real-world situations to prevent bullying and harassment. By understanding the contextual factors that contribute to harassment and bullying in virtual environments, we can identify similar patterns in real-world settings and apply similar intervention strategies. This highlights the potential of virtual environments as a testing ground for interventions that can be used in both virtual and real-world settings.

6.2.6 Conclusion

Given the increased use of social VR platforms by children and the emergence of harassment enabled by embodied social experiences, the need for moderation, parental awareness, and effective and accepted safety-enhancing technologies for children and parents is growing. Through this study, I explored the extent to which an AEM-prototype, Big Buddy, is perceived by children and parents as helping increase feelings of safety, comfort and reassurance and ensuring the game's fairness. Regarding the preferred embodiment attributes of Big Buddy, I observed three main clusters among children and parents. Both groups shared a cluster characterised by a visible, human-like, authoritarian figure. Children additionally preferred a friendly, indulgent, humanised, visible figure, as well as a non-visible figure. Parents showed two distinct preferences: a non-humanised, visible, indulgent yet assertive figure, and a non-humanised, indulgent but not overly assertive figure. Parents would still like to remain in the supervision loop at different levels of engagement, through notifications, settings or direct parental watching, mainly due to a lack of trust in the system. Parents proposed progressive punishments based on repeated bad behaviours and including a warning phase. Preferences for the moderator's attributes were ambivalent. The findings provide formative guidance around child and parental perceptions towards the design of AEMs for safeguarding child use of social VR. This work is crucial if we are to appropriately and effectively leverage and act upon advances in AI detection of harassment to create safer social VR environments with AEMs.

Transition to Next Study

Based on the prior study findings, AEM has emerged as a potential solution from child and parental perspectives that promises the advantages of automated approaches, alongside the benefits of human moderators like relatability to the moderator, sense of safety, and improved explainability and engagement.
However, as opposed to prototype-anchoring (as conducted in study #4), I needed to capture broader perspectives of the extent to which AEMs are suitable and applicable, what features should be customised to enhance child users' engagement and enhance credibility, as well as intervention approaches that can be put in place by AEMs.
Studies #5 and #6 invited children, parents, and, for the first time, experts in child psychology and online safety to imagine and critique AEM designs.

6.3 Studies #5 and #6 - Collecting Multi-stakeholder Perspectives on Automated Embodied Moderators for Child Safety in Social VR

6.3.1 Motivation

The following studies are motivated by the need to explore further the design space of embodied automated moderation in social VR. Driven from one-to-one interviews with experts and workshops with families and by merging perspectives, I identify where there is consensus and different views around the suitability, role, presentation, interactions and actions of AEMs.

I investigate the following research questions from the perspectives of experts, guardians[1] and children:

S5&6-RQ1.1 What are the benefits and concerns associated with automated moderation for children in social VR from expert, guardian and child perspectives?

S5&6-RQ1.2 What are the benefits and concerns associated with embodiment features of AEMs for children in social VR from expert and guardian perspectives?

S5&6-RQ2 What are the proposed intervention approaches of AEMs from expert, guardian and child perspectives?

S5&6-RQ3 What are the proposed embodiment features of AEMs from expert, guardian and child perspectives?

S5&6-RQ4 To what extent and how should parents be involved according to experts, and does this align with the extent of involvement desired by parents?

The findings inform the development of effective and child-centred AEMs, providing the foundations for the development of safer social VR through automated embodied moderation.

6.3.2 Related Work

This section synthesises related work on the specific topics addressed in these studies. (A more comprehensive literature review is presented in chapter 2, with specific relevant sections for context: section 2.3.2, section 2.3.3 and section 2.3.4; and for methodology: section 2.4.1, section 2.4.2).

[1] In this section of the thesis, I include both parents and grandparents under the term "guardians", as both groups participated in the workshops (Study #5).

6.3.2.1 Automated Moderation in Social VR

There is an important need to safeguard social VR platforms where children meet in 3D embodied virtual environments, leading to amplified harm [20]. Existing safety tools [187, 204] aim to protect users from disruptive individuals. However, these mechanisms place the responsibility on child users, who might lack the necessary understanding or familiarity with the technology [122, 114, 363, 115], including children and their parents. Furthermore, they do not facilitate remote parental supervision or inform parents about their children's negative encounters, whether as targets of bullying or as bullies themselves. While some platforms recruit human moderators to intervene in cases of misconduct [31], they cannot address the majority of incidents, and their availability is limited. A recent study found that only 24% of disruptions within social VR environments were managed by human moderators [19]. Automated moderation that takes into account the unique affordances of VR has, therefore, been proposed as an alternative safety tool to combat harassment in social VR [40]. While promising, research is needed to better understand the impact and implications of automated moderation on children, how it is perceived by child users, their parents and experts, and what the design factors and actions are for an effective system.

6.3.2.2 Embodiment of Moderator Agents to Safeguard Children in Social VR

As VR enables a myriad of avatar representations, from human-like to non-human-like, it raises questions about how this impacts user perception. Prior work found that children perceived and described the embodiment and personification similarly across three forms of embodied characters (human, animal and anthropomorphised creature) [161]. They treated these virtual bodies as if they were physical bodies, regardless of their realism. It was concluded that the types of characters do not affect how children perceive the realism of VR [161]. There are further potential benefits to such AEMs. Children have displayed increased social responsiveness to VR characters, as opposed to their counterparts on 2D television [36], and research shows that individuals tend to alter their behaviour when they are aware of being observed due to the so-called "Audience Effect", resulting in actions for the betterment of others [68, 69].

6.3.2.3 Gaps

There is an important need to develop suitable safety-enhancing tools for children in social VR. While there is potential in AEMs (as study #4 showed), it is unclear how suitable these are and what the key factors and actions are for an effective system that not only removes the source of harassment but also provides support. Moreover, further research is needed on how AEMs can be presented for effective safeguarding, what appearance and customisation features (e.g., personality, communication) children may change and personalise to feel safer without removing enjoyment and the credibility of the moderator.

6.3.3 Methodology - Collecting Multi-stakeholder Perspectives on Automated Embodied Moderators for Child Safety in Social VR

The aforementioned gaps are filled by interviewing experts (Study #5) and running participatory workshops with families (Study #6), to get insight into recommendations and preferences of how AEMs should be presented and what interventions they should take.

6.3.3.1 Studies Design

After gathering practical and experiential insights from children and parents in (Study #4), I conducted exploratory qualitative research through interviews with experts (Study #5) and workshops with families (Study #6) to get more in-depth ideation about AEMs design considerations [283]. Although experts and families did not participate in the same sessions, each group received a similar introductory briefing (see slides in Supplementary Material STUDY 4/SLIDES). The subsequent questions and activities were then tailored to their respective roles (see section 1.3), considering specific questions for experts, parents and children as well as child-friendly activities for children. Workshop activities included brainstorming via post-its for both parents and children, and hand-drawn storyboarding for children [242], based on design workshops featured in prior work which were used to develop mediation tools for 2D social media [238, 239], conversational agents [226, 240] or robots to foster anti-bullying [241].

6.3.4 Methodology - Study #5 One-to-one Interviews with Experts

The interviews were conducted in May-July 2022 with the same **16 experts** as in study #2.

6.3.4.1 Procedure

I conducted 16 semi-structured in-depth one-to-one interviews with experts that met the criteria defined in section 1.2, via Zoom, that lasted 20-30 minutes. I presented slides with timers (varying from 2 minutes to 4 minutes) for each question. The interviews all started with a 5-minute presentation about social VR regardless of the experts' social VR knowledge and experience. I introduced social VR opportunities and risks and showed a 10-second video from YouTube of a child sharing their experience with another adult in social VR [307]. I then listed existing safety tools and a video of Big Buddy, a current example of what an AEM could look like and actions it could take (from the above study #4). Big Buddy is introduced as a Wizard-of-Oz automated agent in a simulated social VR game and intervenes when a fictional player disrupts the child's game. It has a robotic voice (text-to-speech), a robotic/astronaut suit and appears as a tall White/Caucasian man. When he intervened, he could take one of the three actions or a combination, spoken verbally and written in a speech bubble: reset points of the harasser back to 0, notify parents of the harasser, and/or

exclude the harasser for a day. The three actions were chosen based on punishments on children's ratings of school punishments: 'information being sent home', 'teacher explaining what is wrong with their behaviour in front of the class' and 'being stopped from going on a school trip' as the top three of the most effective punishments [61].

6.3.4.2 Interview Questions

During the interview, I asked questions related to Big Buddy, with follow-up questions when further precision or clarification was required, or where points of interest were noted. These questions explored topics including:

- First impressions regarding the suitability of AEM to safeguard children from harassment in social VR;
- AEM's strengths and weaknesses, alternative actions, alternative features (physical, social);
- If and how interventions/features can be adapted for different groups, individual VS group harassment (i.e., if a group harasses a child).
- The importance of guardians involvement with AEM and ways they can be involved.

The slides with the introduction and full list of questions can be found in the Supplementary Material STUDY 5/INTERVIEW_SLIDES.

6.3.4.3 Analysis

The same analysis as described in section 4.2.3.4 was used, with a different second coder. The coding scheme can be found in the Supplementary Material STUDY 5/CODING_SCHEME.

6.3.4.4 Participants

Recruitment and Demographics The same experts from Study #2 participated in this study, as part of the same interview session. The full summary of the demographics of experts can be found in section 4.2.3.5.

Ethics and Compensation The protocol was approved by the university CoSE ethics committee (Approval numbers: [300220161]). Participants received an £8 Amazon voucher. Documents with the information sheet, pre-survey with consent form and demographics questions can be found in the Supplementary Material.

6.3.5 Methodology - Study #6: Workshops with Guardians and Children

The workshops were conducted in July-August 2023 with **eight guardians and 13 children**.

6.3.5.1 Procedure

I conducted five workshops (approximately 75 minutes each), with a total of eight guardians (five parents, three grandparents) and 13 children. The table 6.4 and table 6.5 provide detailed breakdowns of the workshop sessions. The first workshop was conducted in a Social Community Centre with three child participants (13-16 years old). I led the workshop, which was assisted by a librarian and a social worker for logistics. The other four workshops were conducted in the local Library from which families were recruited on a voluntary basis. Guardians and children (8 to 16 years old) participated. Guardians and children were seated on opposite sides of the room to minimise influencing the children's responses. A second researcher and I led the sessions. As I had the PVG scheme, I facilitated the activities with the children, and the second research facilitated the brainstorming sessions with the guardians. Both tables were audio recorded with written consent. The librarian helped with logistics and primarily sat at the children's table to help them feel at ease, as many of the children were familiar with her.

6.3.5.2 Workshop Activities

Introduction and VR Demonstrations The workshops started with a similar introduction as described in section 6.3.4.1 before featuring live VR demonstrations to introduce the concept of social VR and the concept of AEMs. I used the same simulated virtual social game experience with Big Buddy described in Study #4, which was showcased in Study #5 through a video. The demonstrations aimed to provide a better understanding of the topic at hand.

Guardian Brainstorming Guardians engaged in a collaborative group brainstorming session to identify pertinent aspects to consider for an effective AEM. The questions posed to guardians were similar to those posed to the experts regarding *suitability, strengths and weaknesses, alternative actions, alternative features* (see section 6.3.4.2).

Children Brainstorming Questions for the children focused on attitudes towards, and the anticipated interactions with, AEMs:

- *the appearance they would prefer for their AEM*;
- *when* and *where* in the scene they would want their moderator to appear;
- *actions and phrases* their AEM would employ.
- In the fourth workshop, older children (13 to 16 years old) mentioned not particularly wanting an AEM that is fully automated or a character. I therefore re-tailored their questions: *Why they did not want an AEM, its strengths and weaknesses, actions/features they would recommend.*

Each stakeholder group received an A3 paper per question, along with stationery and post-its for their responses, which I collected at the end of the workshop. The overall discussion covered both the embodiment and the system of the moderator. This aimed to generate ideas and perspectives from multiple viewpoints.

Design Activity: Designing an AEM via Drawing and Storyboarding The next step involved children designing their AEM, including its name, its appearance, and outlining its actions in a particular scenario, building upon the ideas they had previously brainstormed. Children were given a sheet of A3 paper for storyboarding and one of two child harassment scenarios of their choice based on existing possible harassment scenarios [31, 23, 22] (1. Group harassment scenario: a group of teenagers harassing a child in social VR, 2. One-to-one harassment: an adult being inappropriate to a child). They could draw their AEM, give it a name and actions it would take to tackle the scenario chosen. The slides used to introduce the tasks are in the Supplementary Material STUDY 6/SLIDES.

Table 6.4: Workshops Schedule Breakdown for Guardian and Child Participants.

Workshop Breakdown	**Guardians**	**Children**
Welcome and Social VR Introduction	10-15 min	
VR Big Buddy Demonstrations	20 min	
Brainstorming Post-its	40 min	20 min
Design activity (Children only)		20 min

Table 6.5: Table Summarising participant groups' attendance at workshops. Whilst the presence of guardians varied, there was always a responsible adult accompanying children (the librarian was present in all workshops). P(Participant Number)(P parent or G grandparent)-(Workshop#) (e.g., P3P-2) with corresponding child as P(Participant Number letter)(C)-(Workshop#), a letter is added if more than one child per guardian came (e.g., P3aC-2, P3bC-2).

Workshops	**Children (participants)**	**Guardians (participants)**
Workshop 0	3 (P1, P2, P3) (group: 13-16 yo)	0
Workshop 1	3 (P1C-1, P2aC-1, P2bC-1) (group: 8-12 yo)	2 (P1P-1, P2P-1) (2 parents)
Workshop 2	4 (P1C-2, P2C-2, P3aC-2, P3bC-2) (group: 8-12 yo)	3 (P1P-2, P2G-2, P3P-2) (2 parents, 1 grandparent)
Workshop 3	1 (P1C-3) (group: 8-12 yo)	1 (P1G-3) (1 grandparent)
Workshop 4	2 (P1C-4, P2C-4) (group: 13-16 yo)	2 (P1G-4, P2P-4) (1 parent, 1 grandparent)

6.3.5.3 Analysis

The process of analysis and theme development mirrored the one described in section 4.2.3.4 with a different second coder. This led to four separate coding schemes (two separate ones for guardians and two separate ones for children) (See Supplementary Material STUDY 6/CODING_SCHEMES). The second coder and I then collaborated to consolidate the coding schemes into one combined scheme for guardians and one for children, by collating or distinguishing between codes using the compare option on QCAmap [308]. There were just a few codes from one researcher that were not reflected in the second researcher's code scheme. This informed how I combined the schemes by incorporating the missing codes and then revisiting the transcripts with the final code scheme. I created a set of higher-level codes and themes by bringing together related codes. I identified the common themes across experts' and families' results. Once the Findings Section was written, the second coder, two other researchers and I reviewed themes.

6.3.5.4 Participants

Recruitment Families were recruited via a local library in the UK, in English, on a voluntary basis between May and July 2023. Parents' or legal guardians' consent was required. Once they signed up on Eventbrite, they were sent a Qualtrics form with an information sheet, a consent form to sign and demographics questions (parents' and children's social VR/VR/online games experience, age, gender, ethnicity, and education). The latter was completed by guardians only. Children were all accompanied to the workshop by their legal guardian. Once at the workshop, both parents and children had to sign a consent form to make sure both agreed to participate. At the start, I assured participants that no wrong answer could be given and assured them of the confidentiality of their responses.

Demographics Guardians included five parents (four female, one male) and three grandparents (two female, one male). Parents had a mean age of 45.4 (SD=8.4); two identified as Black/African and three as White/Caucasian. Grandparents had a mean age of 63.5 (SD=2.8), all identifying as White/Caucasian. Guardians had a mean social VR awareness (level of understanding and knowledge of what social VR is on a 1-5 scale, with 5 being extremely aware) of 1.9 (SD=1.4), while regarding Attitudes towards social VR (1-5 scale, with 5 being a very positive attitude), they scored a mean of 2.6 (SD=1.1). Children (seven female, six male) had a mean age of 11.5 (SD=2.7). Eleven children had at least *a little* VR experience, two had *none at all*, and one was not known. The full summary of demographics can be found in tables in table 6.6, table 6.7 and table 6.8.

(*Note: I used a format where family's IDs were represented for guardians as P(Participant Number)(P or G)-(Workshop#) (e.g., P3P-2) with corresponding child as P(Participant Number letter)(C)-(Workshop#), a letter is added if more than one child per guardian came (e.g., P3aC-2, P3bC-2)*).

Table 6.6: Demographics of parents and grandparents who participated in the workshops. Parents had a mean age of 45.5 (SD=8.4). Grandparents had a mean age of 63.5 (SD=2.8).

P#	Guardian	Age	Gender	Ethnicity	Education	Profession
P1P-1	Parent	58	Male	White/Caucasian	Professional degree	Employed full time
P2P-1	Parent	39	Female	Black/African	Professional degree	Student
P1P-2	Parent	37	Female	Black/African	Professional degree	Unemployed
P2G-2	Grandparent	NA	Male	White/Caucasian	Prefer not to say	Retired
P3P-2	Parent	44	Female	White/Caucasian	2 year degree	Employed full time
P1G-3	Grandparent	65	Female	White/Caucasian	Some college	Retired
P1G-4	Grandparent	61	Female	White/Caucasian	Prefer not to say	Prefer not to say
P2P-4	Parent	49	Female	White/Caucasian	4 year degree	Other: Housewife

Table 6.7: VR ownership, experience, and attitudes of the guardians who participated in the workshops. Guardians had a mean Social VR awareness of 1.9 (SD=1.4), and a mean attitude towards social VR of 2.6 (SD=1.1).

P#	VR headset Owner	VR Experience	Social VR Awareness (1–5)	Social VR Experience	Attitude towards Social VR (1–5)
P1P-1	No	None at all	2	None at all	4
P2P-1	No	None at all	1	A little	2
P1P-2	No	A little	1	A little	3
P2G-2	No	None at all	1	None at all	3
P3P-2	No	A little	4	A little	1
P1G-3	Yes, Meta Quest 2	None at all	1	None at all	2
P1G-4	Yes, Meta Quest 2	A little	4	None at all	4
P2P-4	No	A little	1	None at all	2

Table 6.8: Demographics of Children who participated in the workshops. Children had a mean age of 11.5 (SD=2.7). These were reported by the guardians.

P#	Age	Gender	Ethnicity	VR Experience	Social VR Experience
P1C-1	11	Female	White/Caucasian	unknown	unknown
P2Cb-1, P2Ca-1	12, 10	Female, Female	Black/African, Black/African	A little, None at all	None at all None at all
P1C-2	8	Female	Black/African	None at all	None at all
P2C-2	8	Male	Other: African/Scottish	A moderate amount	A little
P3Cb-2, P3Ca-2	11, 8	Male, Female	White/Caucasian, White/Caucasian	A little, A little	None at all None at all
P1C-3	10	Male	White/Caucasian	A lot	A little
P2C-4	13	Male	White/Caucasian	A little	None at all
P2P-4	15	Male	White/Caucasian	A great deal	A moderate amount
P1	16	Male	Black/African	A little	None at all
P2	14	Female	White/Caucasian	A little	None at all
P3	13	Female	Prefer not to say	A little	None at all

Ethics and Compensation The protocol was approved by the university CoSE ethics committee (Approval number: [300220177]). I was registered with the national PVG scheme to conduct this study with children. Children were offered a certificate of attendance at the end of the workshop as a token of appreciation. Documents with the information sheet, pre-survey with consent form and demographics questions can be found in the Supplementary Material.

Selection and Participation of Children Please refer to table 4.3.3.7, as the same selection procedure was followed.

6.3.6 Limitations and Future Work

The sample size and necessity of conducting the studies in English in the UK may limit the external validity with respect to the global population and applicability to other cultures. The design activity of the workshops captured children's initial reactions in one session, rather than refining proposed designs over multiple sessions. Consequently, the proposed AEM designs by children serve as formative, imaginative insights that may not be practical or actionable, but reveal much about children's attitudes towards, and perceptions of, AEMs. Further iterative co-design activities, engaging both adult and child users as well as experts, should be considered to refine these designs into those that could be implemented and deployed in practice. Experts' in-depth remarks resulted in significantly longer transcripts and more detailed answers when compared to those of the children. Future research should include a more balanced representation of families and may also require a more specific criterion for experts' recruitment if I need to focus on the psychological aspects or technical aspects, for example. Still, the exploratory studies provide invaluable and novel insights into an emerging research field. The interviews were conducted online, which meant that experts experienced Big Buddy through videos, rather than in the immersive VR environment. This could lead to missing recommendations that could not have been noted solely from video-based observations. Nevertheless, experts reported finding the introductory presentation insightful and informative. All experts were already familiar with VR and had a high level of awareness of social VR, with a mean of 4 (1-5 scale, with 5 being extremely aware) (SD=1.5), and 75% had practical experience with social VR. Therefore, their perspectives still offer valuable contributions. Furthermore, I present findings around AEMs, but acknowledge that the only experience participants had of AEMs was through the Big Buddy example, potentially influencing the results. However, the use of such an example was also a strength as it offered a more practical and hands-on experience of what could be an AEM.

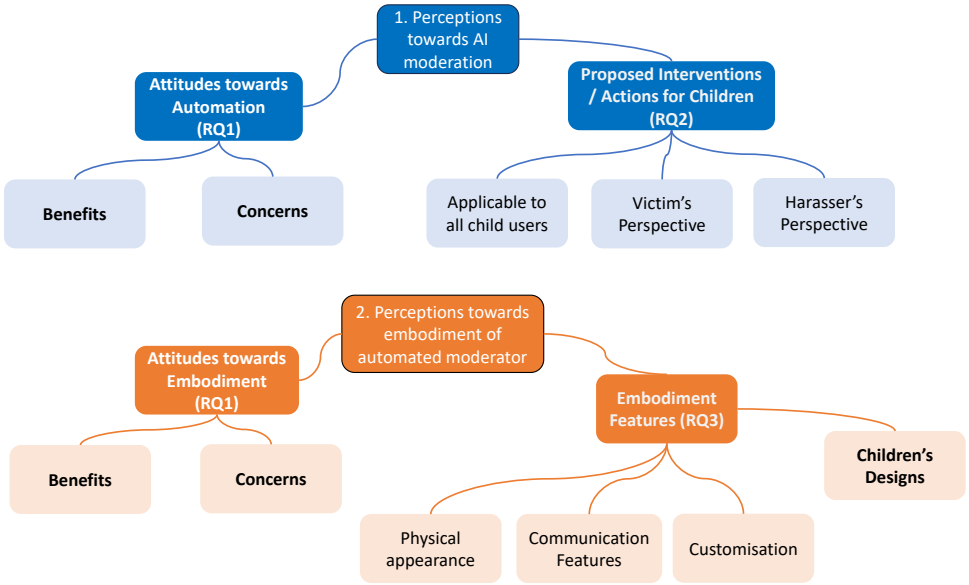


Figure 6.9: Findings structure across all three stakeholder groups (experts, guardians and children).

6.3.7 Results of Studies #5 and #6

In this section, I present the combined findings from Studies #5 and #6, providing a comprehensive overview that integrates all three perspectives simultaneously. *A Miro board gathering results can be found via:* `https://miro.com/app/board/uXjVMFSwws4=/`.

Participants shared their thoughts on Big Buddy, an example of an AEM designed to intervene in child harassment situations in social VR (see study #4). Findings were grouped into two categories across all coding schemes: **1) Perceptions towards the automated moderation**, i.e., the components involved in ensuring effective moderation through automation and **2) Perceptions towards the embodiment**, i.e., the physical and social components that include the moderator's appearance, ways of communicating and interacting with child users. The structure of the findings' section is shown in figure 6.9. Categories obtained from coding schemes are highlighted in grey. The following section describes themes across all three perspectives: experts, guardians and children.

6.3.7.1 Attitudes towards the Automated Moderation to Safeguard Children in Social VR

Benefits of Automated Moderation **AEMs as a Guide and Exemplar** Experts found the AEM useful for guiding, educating and reminding users of rules in an easy and understandable way (N=5), addressing the lack of socialising norms in social VR. It could educate harassers and instil good behaviour: *"It's a great easy way to understand what someone is doing wrong, what the codes are, social codes for operating in the space"* [Expert P13], *"The idea is there because there are no real true safety measures in the metaverse at the moment"* [Expert P4].

Immediacy of Response Experts also valued that an automated system can deliver immediate action (N=9) in contrast to current methods like report delays and human moderation, and particularly in the case of group harassment, as it can be easier to detect. It can also offer support (N=2): *"Having somebody to come and help"* [Expert P12]. Some noted that the system could be particularly useful in supporting new users and adults.

Guardians also found one of its strengths is the fact that it can take immediate actions (N=3) and it can enforce rules (N=5), there are consequences for actions, and it can restrict access to inappropriate content. It can also give helpful advice (e.g., take breaks) *"I'm just wondering if Big Buddy could also see if a child has been logged on to the game for a long time. They could say: How about taking a break?"* [Parent P2P-4].

An Intermediary Between Parents and Children Moreover, AEMs can potentially help with the parent-child relationship (N=3) as they can facilitate conversation between children using social VR and their parents [Grandparent P1G-4] and help children disengage from prolonged VR usage. Some parents mentioned that as their children might be indifferent or not listen when they ask them to stop, Big Buddy could be an effective solution to stop the child from playing: *"The kids aren't going to really care about their parents being told"* [Parent P3P-2]; *"A big buddy could just see your time and the game is finished [...] because I can't stop them. Especially when they're younger.* [Parent P2P-4]. Finally, guardian P1G-4 felt it would be useful if the AEM is linked to guardians (N=1) via parental influence over the AEM (e.g., settings overriding the AI system) and parental insight into AEM actions and the events occurring (e.g., via real-time smartphone notifications).

Fair Moderators Children from the 13-16 year-old group (workshop 4) (N=2) reflected on the AEM and noted that, unlike human moderators, the AEM would not be able to intentionally abuse their power and can act fairly if correctly trained. They also mentioned that AEMs would lower cost and would not need to take breaks, allowing them to be present in the virtual environment at any time.

Finding 1: Participants viewed AEMs as valuable for guiding, educating, and reminding users of rules, addressing the lack of social norms in social VR. Experts emphasised their ability to instil good behaviour, detect issues like group harassment, and take immediate action, which current reporting and human moderation systems often lack. AEMs could also support new users and adults. Guardians valued the system's immediacy, rule enforcement, and ability to restrict inappropriate content while offering well-being prompts. They also saw AEMs as a potential mediator between parents and children. Parental control and insight via notifications were seen as beneficial. Teenagers noted that AEMs, if well-trained, could be fairer than humans, always available, and cost-effective.

Concerns Around Automated Moderation **Technical Feasibility and Contextual Understanding** Experts expressed concern about an AI moderator's ability to address mistakes and technical challenges (N=7) with evolving online harms and social norms. To resolve this, they suggested updates to keep up with emerging and evolving harms and online social norms. It would also require the tool to gather contextual information to accurately detect disturbances. Experts mentioned the automated moderation needs to be context specific (N=4) for different games and environments.

Transparency and Trust An expert highlighted a pitfall of moderators administering judgment without a visible reflection process: *"The weakness is the moment that any kind of automatic or human moderator comes in, enacts a judgment, and does not create a questioning phenomenological process for everyone to understand how they were all contributing to that. You destroy the group safety, and that's going to be hard to train into a moderator"* [Expert P11].

Half of the guardians (N=6) would partially trust such a system. A guardian mentioned they would trust it as it would be impartial and not malicious [Parent P2P-4], and another would trust it *"as long as there is open communication and I know what is going on"* [Grandparent P1G-3]. However, the other half of the guardians lacked trust due to concerns that it would not be infallible.

Taking the Wrong Actions and Fallible AI Some experts questioned how the system would address false positives: *"With any kind of automated tool I'm always worried about false positives"* [Expert P2], and that inaccurate judgments may feel intrusive:

> *"I think that your inaccuracy levels or anything automated is gonna make that feel intrusive a little bit, so unless it's 100 accurate, which really you can only do with the human still, I think that it's quite intrusive."* [Expert P14]

Some experts also mentioned that an AEM may be more suitable for group harassment, as it might be easier to detect, but interventions may be more effective in a one-to-one situation, as the resolution can be individualised and personalised.

Guardians had very similar concerns regarding *technical challenges* (N=6). The child can potentially circumvent or outsmart the system (N=3) *"My point is it's not infallible. There's going to be workarounds. The kids can work out to get around it."* [Parent P3P-2]. This susceptibility is a concern specifically for secondary school children who might exploit the AEM's limitations and weaknesses, and find ways to be unsupervised and bend rules, according to Guardian P1G-4. Moreover, mistakes and a lack of context recognition (N=4) by the AEM introduces risks, as it may not always detect problematic content or understand the nuances of different situations, potentially leading to false positives or negatives. Thus, some guardians mentioned that human-in-the-loop (N=2) is necessary and *"we must be able to shut it down"* [Grandparent P2G-2]. The AI system requires continuous updates and management (N=2) to adapt to the changing dynamics of supervision as a child grows.

Teenagers mentioned concerns about the moderator being AI-based (N=2), as children may abuse the automated system and it might be too sensitive, leading to false positives or not accurate enough with languages other than English [Child P1C-4]. They wished for AEMs to be carefully controlled [Child P2C-4] and suggested human involvement and, thus, a semi-automated approach. Child P1C-4 suggested different lines of moderation: the AEMs detect events and intervene, the human moderators check, and the admins finalise the decision.

Impact on Children Experts questioned the long-term impact on children (emotional and ethical issues) (N=9). Specifically, they questioned the emotional ramifications of harsh tactics, saying:

> *"I think it can actually be quite demeaning to the harasser [...] because I'm 99[%] certain that Big Buddy might make a mistake over something that wasn't deliberate or perceived [...] that's because we're still doing the learning in terms of what's acceptable behaviour."* [Expert P16]

Unintended consequences of AI-moderation may lead to risks of self-harm or extreme reactions from children *"My sense of unintended consequences would be massive meltdowns if there was a sense of injustice".* [Expert P15]. Moreover, there are consequences of how the AEM is designed if it goes beyond just a moderator, for example, acting as a tutor or friend:

> *"I suppose there is a question about how we feel about that child engaging with a tutor. [...] I wonder what Big Buddy's role is because what we've seen is Big Buddy comes when something's gone wrong. Does Big Buddy come in when things are going right as well?"* [Expert P8]

This highlights that the role and scope of involvement of the AEM are not yet defined. Additionally, experts discussed the potential for future and real-world consequences. While the example AEM (Big Buddy) gives a punishment and removes the harasser, it does not support the victim on how to deal with future situations. Expert P11 also critiqued that setting a player's points back to zero could have a damaging and toxic effect.

Guardians expressed a concern about the negative impact on children (N=3), including their lack of skills to handle incidents, potential negative effects on mental health due to automated moderation, and children may feel devastated according to Guardian P1P-2, and possibly agitated, resulting from warnings and restrictions [Grandparent P2G-2]. Moreover, children (N=2) were worried about not being able to refute and resolve an argument.

Privacy Child data protection (N=2) was mentioned by experts as an important consideration, as investigations and automated interventions would require accessing private content. Experts also had surveillance concerns (N=2) and how it may take away the child's agency. In contrast, children could also circumvent and find a way around it (misusing, provoking or using it as a harassment tool). Expert P17 mentioned the need for long-term research to see if behaviour changes and insights into how children interpret the AEM over time.

Finding 2: Experts questioned whether AEMs can keep pace with evolving harms, handle context-specific harassment situations and avoid false negatives and false positives. They called for transparency, continuous updates and human-in-the-loop approaches. Privacy and surveillance concerns were also raised. Guardians echoed these concerns, noting that children could circumvent the system and that mistakes or lack of nuance could undermine its effectiveness. Teenagers feared over-sensitivity, language limitations, and misuse, preferring a semi-automated model where humans are involved.

6.3.7.2 Proposing Suitable Interventions of the AEMs for Child Users, Victims, and Harassers

Interventions Applicable to All Child Users Several approaches and actions for the AEM were suggested to safeguard children and promote safe social VR environments for all users:

Intervention Theme 1: **Rules and Fairness**

1.1) Set expectations and enforce rules Experts (N=9) recommended the reinforcement of consequences for bad actions with visual and verbal reminders clearly indicating and explaining expectations and rules. Expert P2 proposes disturbance tests with dummy players to show new users what kind of sanctions can be taken for certain situations. Users should be made aware that a moderator is present in the room.

> *"To me, the most important thing it does is just tell people more information about the expectations of the space and serve as that visible reminder [...]. So just serving as a reminder that this is still a social space and like any other social space, there are expectations, there are rules, and there are going to be consequences for violating those rules."* [Expert P2]

Experts also suggested that children set the rules at the start, either individually or collaboratively. This can include selective positive actions from a list to receive positive feedback from the AEM. Guardians emphasised the importance of enforcing rules (N=3), with sanctions and real-time notifications to guardians from both parties (victim and harasser). They highlighted the significance of reinforcing rules through verbal reminders (N=2). Child P2C-1 want the AEM to set the rules and expectations; it should make sure users follow the rules (N=1). **1.2) Applying Principles of Procedural Justice** Experts (N=2) noted that consequences should be consistent and fairly applied across all ages, applying principles of procedural justice:

> *"I could see a lot of issues surrounding feelings of injustice or unfairness.[...] Looking at the principles of procedural justice and how people perceive the fairness of offline governance systems and criminal justice systems, I would want to apply the same sort of principles to a system like this so that we can ensure that it's as fair as possible."* [Expert P2]

Intervention Theme 2: **Positive Support**

2.1) Positive reinforcement over negative punishments Experts (N=5) discuss the benefit of positive reinforcement in behavioural change, which can be accelerated when compared to negative punishments. This can be done by giving more attention to positive behaviours, shifting the focus from negative punishments, to promote behaviour changes, by giving positive feedback to young people engaging in appropriate behaviours, for example, *"The good thing about switching it so that it praises for positive behaviour is that it's likely to extinguish the behaviour of the negative player much faster and it's likely to model good behaviour"* [Expert P12]. Positive reinforcement is also universally accepted, according to Expert P12, which reduces the need for adaptation between groups. Experts noted that negative punishments remove enjoyment, while promoting positive behaviour through behaviour can increase enjoyment. Statements should be predominantly positive and encouraging to emphasise the opportunity to try again and improve. A guardian also suggested positive reinforcement *"with positive rewards or points for those who follow the rules"* [Grandparent P1G-3].
2.2) Reflection and learning experience According to experts (N=7), the AEM should give the opportunity for reflection, repair and rehabilitation. This can be done by pausing the ongoing interaction to facilitate reflection or by notifying both parties (victim and harasser) with post-resolution reflection messages and an explanation of wrongdoing to educate children about their actions. Child P2C-4 would want the AEM to give a tutorial on how to use the safety tool and AEM at the start of the game via a pop-up. **2.3) Offer Psychological Support to all Child Users** Experts (N=2) suggested strategies that involve not only enforcing rules but also providing psychological support to both the victim and the harasser, by asking follow-up questions (e.g., "Are you okay?" [Expert P3]) or, for example, linking them to child helplines [Expert P10]. Expert P8 suggested monitoring self-harm as well, considering mental and physical health and safety concerns. However, some experts also raised doubts about the role of the AEM beyond a moderator enforcing rules as this could lead to ethical concerns. **2.4) Encourage open discussion and conversations** Experts (N=2) mentioned that the AEM should facilitate open conversations, helping the child to discuss the incident.

Intervention Theme 3: **Adapting to Context and Dynamics**

3.1) Interventions proportional to severity and frequency Experts (N=8) recommended reasonable and proportional outcomes based on behaviour, depending on frequency, context and the level of harassment and proportional support by the AEM (e.g., more check-ins). Sanctions should depend on the severity but also be balanced with educational intervention. More generally, some experts recommended building on the history of interactions, for example:

> *"If it just came up with the same answer every time they'd start to sort of ignore it um and I think that's somewhat similar for the harasser I think you can you know when we talk about sanctions we talk about a matrix and an incremental path through a sanctions you know based on how many times you and what you've done the that the responses could be different and they would feel more in response to the actual situation."* [Expert P14]

To transition from warnings to more stringent enforcement, guardians (N=2) suggested employing strike warning systems. Children also suggested the use of a strike system: from warning to banning to increasing banning time (N=6) where the AEM gives a certain number of warnings before taking an action, such as banning the user. If the inappropriate behaviour is repeated, the ban gets longer. Additionally, some children would want the AEM to gather evidence when an incident occurs so that it can be re-watched to finalise the outcome. For instance, audio recorded within a radius of the child within the VR scene or a video clip summary sent to a human moderator. This would also allow the child to more effectively engage with their parent regarding the incident in specifics, rather than a parent just seeing the end result. **3.2) Interventions adapted to one-to-one harassment VS group harassment** In group harassment, i.e., a group harassing a child, experts (N=10) discuss approaches that include encouraging groups, and in particular those leading the groups, to reflect on their behaviour, for instance:

> *"Group dynamics are very different, there's always an alpha person that is leading the group and the others are followers so it should be customised depending on who you are dealing with."* [Expert P7]

Body language monitoring of the victim, harasser but also bystanders could be used to detect group harassment according to Expert P11:

> *"In group dynamics there's usually an aggressor and a victim, everyone else is often bystanders and their body language is going to shut down, freeze and get very awkward. I think there could be a great opportunity to be monitoring for that kind of body language."*[Expert P11]

Thus, taking more immediate actions in the case of group harassment was recommended. However, they also suggested focusing on individuals to break the group effect, balancing individual warnings and group approaches. **3.3) Private VS Public conflict resolution** Experts (N=2) question the effectiveness of public or private resolution. An expert suggested involving bystanders anonymously by posing questions that prompt them to reflect on the situation and contribute to a different outcome.

Child P2P-4 also mentioned not wanting punishments of actions that are acceptable between friends.

Intervention Theme 4: **Human and User Involvement in Decision-Making**

4) Human-in-the-loop Experts (N=2) suggested the involvement of a human, with either the user being able to call the AEM in case it does not pick up the issue, or the moderator being automatically triggered and the intervention or outcome being decided by a human. They noted a tension around how, and to what extent, humans work with AEMs and question if the AEM should be the first line of defence or if it should make decisions on interventions.

Intervention Theme 5: **Non-Verbal Interventions**

5) Non-verbal Interventions Experts (N=2) suggested the AEM to act as a *"buffer"* [Expert P9] and separate the victim from the harasser, *"putting their avatars in opposite sides of the room"* [Expert P17]. Some children (N=3) also proposed non-verbal actions taken by the moderator: *"a virtual hug"* [Child P2aC-1] to comfort the victim or *"a virtual punch"* [Child P1C-2], *"slap 3 times"* [Child P1] to the bully.

Intervention Theme 6: **Immediate Safety**

6) Immediate Actions and Notifications to Guardians Guardians proposed immediate actions (N=3) that include shutting down the source of harassment, blocking, bubble the harasser and the victim, immediate time out and debrief to parents. Children also suggested that the AEM takes direct action (N=8) in a child harassment situation. These direct actions include blocking, reporting and banning. Children also want notifications and messages (N=2) when an incident happens to be sent either to parents, to the victim, saying the moderator went to the harasser and then the outcome to the bully.

Additional approaches and actions proposed by Experts

An expert mentioned there should be the possibility to complain about the AEM and switch it off (N=1). There should be a balance between offering guidance and allowing user autonomy (N=1), light touch with occasional interventions. In contrast, one expert suggested maintaining a constant moderator presence (N=1), creating a situation where users cannot easily distinguish when they are being monitored, similar to how speed cameras are not always active. Experts also suggested age-verification (N=1).

Although this research focused on child users, it was mentioned by experts that for adult harassers targeting a child, consequences should be more severe (N=3). Experts suggested adapting interventions to age groups. Other experts recommended adapting interventions based on development stages and emotional skills rather than just age (N=4).

Finding 3: Experts, guardians, and children suggested strategies for AEMs to ensure safe VR environments:

1. **Rules & Fairness** – Clearly set and visibly enforce rules; apply consistent, age-appropriate consequences; involve children in rule-setting; notify guardians in real-time.
2. **Positive Support** – Prioritise positive reinforcement, opportunities for reflection, rehabilitation, and psychological support; encourage open conversations.
3. **Context-Responsive Actions** – Tailor interventions to severity, frequency, and context; use strike systems; adapt responses for group vs. one-to-one harassment; consider public vs. private resolution.
4. **Human Oversight** – Include human moderators in decision-making; allow user appeals.
5. **Non-Verbal Interventions** – To embody and illustrate actions.
6. **Immediate Safety** – Take instant protective actions (block, bubble, timeout) and notify guardians and victims promptly.

Additional ideas included: complaint and opt-out options, balancing autonomy with monitoring, constant or unpredictable moderator presence, age-verification, and harsher penalties for adults targeting children.

Interventions from Victim's Perspective Several approaches and actions for the AEM were suggested specifically to victims.

Victim's Intervention Theme 1: **Victim-Specific Psychological Support**

1.1) Emotional Support, Empathy and Comfort For the victim, experts proposed emotional support and empathy (N=9). For example, the AEM could offer emotional assistance and check-ins to the victim, asking how they are feeling, if they want to continue, communicating with personal messages and helping build back their confidence. The AEM could offer a reflection space and connect the victim to resources (e.g., child helplines) recognising potential trauma and ensuring the child's well-being. Overall, the moderator should be empathetic. Guardian P1G-4 suggested to comfort the victim, asking the victim how they felt and what they think should happen and administer a ban level based on responses. They suggested to bubble the victim.

1.2) Supporting Language Children (N=4) emphasised the importance of supportive phrases being used with victims. This included informing the victim that the moderator is going to intervene in a situation involving a harasser, according to child P3aC-2. There were also phrases designed to assist and provide advice to victims in handling difficult situations. Messages like *"The Player has been banned"* [Child P3aC-2] were introduced to reassure victims, and there were phrases that instructed victims on how to report an issue *"You can report, press this button to report"* [Child P1C-3]. Furthermore, phrases like *"Please wait while we talk to the player (bully)"* were suggested to maintain transparency during interventions [Child P1C-3].

Victim's Intervention Theme 2: **Empowering Victims through Voluntary and Discreet Active Engagement**

2) Empowering Victims

Experts also suggested empowering the victim with decision-making (N=8). On the one hand, the AEM could automatically be called to offer assistance with options of actions (N=4) to take such as blocking or protective bubble. On the other hand, it could be encouraged that the child reaches out to the AEM (N=4) rather than the moderator stepping in, invoking it when needed. Experts also noted that victims can become harassers, therefore, supporting the victim is important to avoid perpetuating cycles of aggression (N=2). Children also proposed to involve children in the decision-making with discretion (N=3). Child P1C-3 suggested:

> *"I think what you should do if somebody does that to you, a pop up that says "this is a request to kick this person or report this person" and if you request yes then you can report them that's only when you can report them, only if the computer senses that something went wrong."*. [Child P1C-3]

Another child suggested a button to report and call the AEM:

> *"Izzy [Made-up victim name] says: "I'm reporting and pressing the button". [...] Then there is this notification saying "Hey Izzy I went to tell Rob [Imaginary harasser] off, I will notify you when he is banned." "* [Child P3aC-2].

Child P2C-4 suggested, however, the possibility to change the outcome temporarily, for example unblocking someone when joining a private room with friends:

> *"If you have a friend who invites you in a room with the person you've blocked, you should be able to choose to unblock temporarily just for that room and then after immediately they are blocked again."* [Child P2C-4].

Finding 4: Victim's Perspective Interventions – Suggestions focused on psychological support, empowerment, and discreet engagement. *Psychological Support:* Offer emotional assistance, empathy, and confidence-building through check-ins, personal messages, and connections to resources (e.g., child helplines). Use supportive language and transparent updates during interventions. *Empowerment:* Enable victims to discreetly engage with the AEM. Provide decision-making opportunities to the child. Support victims to prevent cycles of harassment.

Interventions from Harasser's Perspective Several approaches and actions for the AEM were suggested specifically to harassers.

Harasser's Intervention Theme 1: **Stepwise Interventions from Warnings to Repair to Sanctions**

1.1) Stepwise approach The use of a stepwise approach (warnings before sanctions, severity increased if repeated) was recommended by experts (N=4), with the opportunity for behaviour to improve before sanctions. If the inappropriate behaviours are repeated, the severity of sanctions can also be increased. Experts suggested that warnings should be private first, as children can become emotionally invested in these games and might react strongly in public. In fact, expert P11 mentions that there could be potential dangers of completely excluding the harasser directly, leading to triggering real-world negative consequences and inappropriate behaviours. Children suggested in cases where a victim had reported the bully, private warnings like "You have been reported; your fate will be decided" [Child P1C-3] or giving more chances before consequences such as being banned.

1.2) Notified with Reasons for their Consequences Experts also suggested they should be notified of their actions with reasons for their consequences (N=3). Children proposed phrases that explicitly highlighted code of conduct violations *"You broke code of conduct/tos [terms of service]* [Child P3aC-2].

1.3) Opportunity for Reconciliation Experts recommended an opportunity for the harasser to explain and apologise (N=5) before the outcome. This can be done with personal messages of the decision to the harasser with the opportunity to explain themselves if they disagree with this decision. Similarly, children proposed messages

that pointed out the negative impact of their actions on victims. **1.4) Considering victims of the past** Expert P9 reflected on the harassers' background, mentioning they are usually victims of the past so directly negatively punishing is not the best solution:

> *"If you just use the moderator to convey punishment, I don't think it's very effective for the harasser because there are some problems with the fact that you might be harassed in your family so you just repeat this chain so you don't know how to you know interact with others and you're just repeating."* [Expert P9]

Harasser's Intervention Theme 2: **Focusing on Reflection and Educating to Improve Behaviour**

2) Reflection and Education The moderator could also provide a space for reflection, training and educating the harasser (N=8). The latter can be done through a test, game or any interactive learning process that add motivational opportunities (e.g., regaining points through participation of a test before re-entering the game). Gamifying the process of reflection for harassers could help them better understand and rectify their inappropriate behaviour. One suggestion was putting the harasser in a safe zone as a reflective zone education break with the AEM which they must complete or receive a time-ban. Another example of interactive learning process was suggested, where the AEM guides the harasser to step into the shoes of their victim promoting empathy:

> *"I think that interventions where the AEM asks the harasser to say "hey you did this thing before and I really want to know what that feels like" and they switch into that other person's avatar and they have that harasser do that instance with them and then they make an observation."* [Expert P11]

It can help harassers understand the impact of their actions, which could, for example, be of use for neurodivergent children: *"Particularly if they've got learning difficulties or autism, there is a way that you could give an explanation"* [Expert P16]. Additionally, the harassers can be encouraged to reflect on their actions and the moderator can show empathy and explain the reasons for the consequences. Experts anticipate that the latter intervention can have a positive impact on their real-world behaviour as well.

A child also proposed a follow-up before the harassers come back in the game *"They get banned for a day. Before joining the game, they have to say they won't do it again. The fate will probably be a ban for one day. Then, they go back in the game, but they have to say "I won't do that again"* [Child P1C-3].

Harasser's Intervention Theme 3: **Immediate Actions against Harasser Prior to Providing an Explanation**

3) Immediately Bubble Harasser In contrast to experts suggesting reflection and explanation prior to the consequence of action, guardians and children suggested a direct time-out for the harasser before repairing. For example, a grandparent would want the AEM to *"blackout harassers straight away, then contact the harasser to debrief*

and ask them for an explanation of their actions" [Grandparent P1G-4]. Children also suggested time-out and putting harassers in a bubble or restricted area, outlined in phrases, such as a 20-minute timeout *"You will be timed out for 20 min"* [Child P2] or a 30-minute cool-down period *"Calm down 30min then come back."* [Child P3].

Finding 5: Harasser's Perspective Interventions *Stepwise Approach:* Begin with private warnings, escalating to sanctions for repeated behaviours, while explaining reasons and allowing opportunities for reconciliation. Consider that some harassers may be past victims, making purely punitive measures less effective. *Reflection and Education:* Provide empathy-building activities, interactive learning, or gamified reflection (e.g., empathy exercises, safe-zone "education breaks") to promote behavioural change, including tailored approaches for neurodivergent players. *Immediate Containment:* Some preferred instant protective actions (e.g., bubble or timeout) before dialogue, followed by explanation and opportunities to repair harm.

6.3.7.3 Attitudes towards the Embodiment of the AEM for Children in Social VR

Benefits Experts identified several strengths of an AEM for children. Experts refer the role of AEM (Big Buddy) as modelling good behaviour (N=11) similar to a *"teacher", "referee", "football coach", "police officer", "brother" or "parent figure".* It can model civility, kindness and can exemplify good behaviour via its communication and interaction. The "*embodied version has value over a just a voice*" [Expert P9], being beneficial as its visibility and presence offers a sense of protection, comfort and authority (N=5). Its physical presence can also be used for physical interaction dynamics, posture and position in the scene, such as a "buffer" between the victim and harasser.

Guardians noted that the embodiment of the AEM has notable strengths (N=4), including being friendly and hero-like, and instils a sense of protection and authority in users. This presence of an authority figure also contributes to feelings of security and comfort, *"similar to the reassurance provided by CCTV surveillance"* [Parent P2P-1], leading to self-moderation among users.

Concerns Experts had concerns related to the embodiment, such as potential systemic bias in design, where it may reinforce stereotypes, as well as worries that children might feel watched, not behave authentically and lack agency and control. They worried that the robotic voice and appearance of the moderator would lead to Credibility issues (N=5) and provoke skepticism from users about its effectiveness and seriousness. Finally, an expert suggested that the embodiment should blend more seamlessly into the social space, and some experts mentioned it should engage more like a human to enhance credibility.

Guardians also raised weaknesses associated with the embodiment (N=3).

Guardian P1G-4 mentioned the "Panopticon Effect" the sensation of being constantly watched, can diminish the enjoyment of the child's experience. According to guardians, the effectiveness of embodiment would vary with age (N=2). Guardians find that it may not be suitable for teenagers, *"too childish"* [Guardians P1P-1, P1G-

3], and guardians feel that it needs to be more depersonalised for older age groups. It was viewed as suitable for certain age groups, particularly for children from early ages up to 11 years old, where participants felt it would help maintain order and considered it appropriate. However, its perceived effectiveness diminishes for older children, as they may not respect it. There was also a concern about the potential reinforcement of stereotypes of behaviours and authority figures.

Finding 6: Benefits of Embodiment - Experts and guardians valued the embodied AEM as a visible, authoritative, and protective presence, modelling good behaviour and kindness. It can act as a friendly, hero-like authority figure and buffer between the victim and the harasser. **Concerns of Embodiment** - Concerns included potential bias, stereotype reinforcement and reduced agency. Credibility issues may arise from robotic appearance or voice. Guardians noted a "Panopticon Effect" of constant surveillance and age-appropriateness concerns, effective for younger children but potentially "too childish" for older users.

6.3.7.4 Proposing Embodiment and Design Features of the AEMs

Embodiment and Design Features: Experts and Guardians Physical Appearance Considerations. Experts' opinions on general appearance considerations for AEMs highlighted several key points. Neutrality (N=4) was emphasised to build trust and avoid reinforcing stereotypes. To achieve this, using non-human representations like wise owls or fantasy creatures is suggested. Additionally, neutral features resembling emojis or familiar cartoon characters are recommended for broad cultural acceptance. The idea of an authority figure was mentioned, stressing, however, that it should not reinforce stereotypical traits *"It does not have to be a man or a human"*. Approachability and relatability (N=3) are also mentioned to be important for effective engagement with the moderator. Striking a balance between approachability and authority was advised. An expert raised a question between robot-like and human-like moderators, considering emotional transference and user perceptions, as well as the impact of the moderator's height on user perception. Visual feedback (N=1) also plays a role in appearance considerations. Suggestions include using different emotional faces as visual cues, such as empathetic faces for victims and sad faces for harassers.

Communication Considerations Experts highlighted communication considerations of the AEM. A friendly, warm and encouraging (N=6) moderator was recommended rather than an authoritarian enforcer to enhance enjoyment, foster inclusivity and unity. It should also be more empathetic and can use playful moderation: *"It could separate users in a fun way if it's detecting a lot of yelling between two different people it would be kind of funny if he could just immediately put people to the opposite sides of the room."* [Expert P17]; or whimsical communication of rules: *"[...] kind of funny, kind of whimsical but actually communicates the rules in a way that you don't dismiss it but it maybe it could make you laugh at the same time so that you actually see it as a piece of the environment and a part of the game um and it's not annoying"* [Expert P17].

Moreover, a natural and human-like, less robotic approach (N=6) was suggested, focusing on decision-making like a human and giving the opportunity to reply, as well as using understanding language and slang to connect with users. For a more natural approach, it should also be able to adapt responses to avoid redundancy. An expert suggested the use of a voice with personality and a natural conversational tone to improve user engagement. Based on the uncanny valley theory [364], "Uncanny aspects" that include features or behaviours that seem too artificial or not quite human-like can make users, especially children, feel uncomfortable or reluctant to trust and interact with the system, according to Expert P9. Considerations also include consistency of communication characteristics across victims and harassers, and tailored approaches for vulnerable children to improve trust towards the moderator. However, Expert P17 suggested a stern tone when engaging with the harasser.

A guardian suggested incorporating playfulness and positivity (N=1) into moderation to encourage rule adherence: *"Adding playfulness (Fun theory) to follow rules. Again just positive"* [Grandparent P1G-3]. Another guardian advocates for communication features (N=1) like using a soft tone and recognising the presence of adults, which would enable the moderator to adapt its communication style based on the audience, whether they are adults or children.

Customisation Customising moderators based on age groups (N=15) is a prominent theme. Experts stressed the need to treat older children and teens (13-16 years old) differently from younger ones (early ages to 12 years old), while maintaining the concept of a protector. Customisation of appearance and speech can enable age-appropriate communication: *"If it's a childlike environment, you would speak and act in one particular way. If it's a slightly older environmen,t it's speaking and acting in a different manner as well so it's understanding the audience"* [Expert P4]. In terms of age-dependent appearance (N=4), this includes age-appropriate customisable appearance from fantasy characters to established ones such as *"cute little dog, cute monster, paperclip" [Expert P14], "little imaginary friend" [Expert P2], "Clippy" [Expert P17].* In terms of communication (N=11), they suggested adapting voices, languages and tones for based age groups, developmental stage and preferences. For example, users can opt for gentle versus blunt approaches. The moderator should adapt its behaviour to the environment; if there are children present, different cultures, languages and accents. Options could be given at the start for children to choose the way the moderator guides and interacts with them. In terms of approachability and trustworthiness (N=7), experts suggested non-intimidating non-verbal approaches, such as having the moderator meet the child's eye level to reduce intimidation, and using of gentler and softer ways of communicating across ages, specifically, with a more authoritative stance for teenagers. They also proposed customisable approaches based on trust, with a questionnaire at the start to help tailor the moderator's interactions and appearance based on the answers.

Additionally, guardians recommended options for customisation (N=2), allowing users to personalise the moderator's features for a tailored experience (colours, animals, gender, non-human etc.).

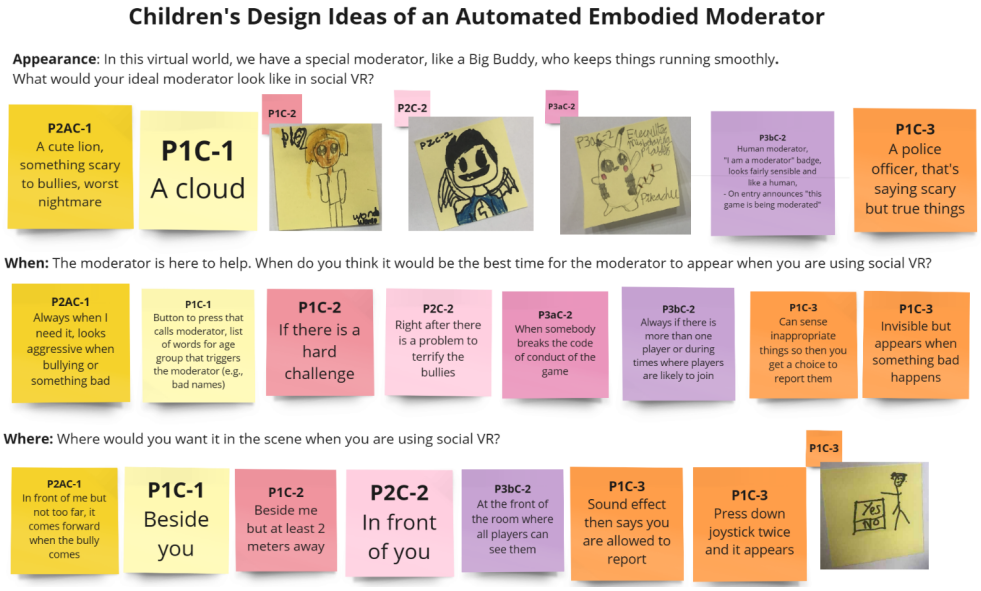


Figure 6.10: A selection of children's post-its transferred to a digital Miro board, regarding physical appearance, when and where their moderator would appear.

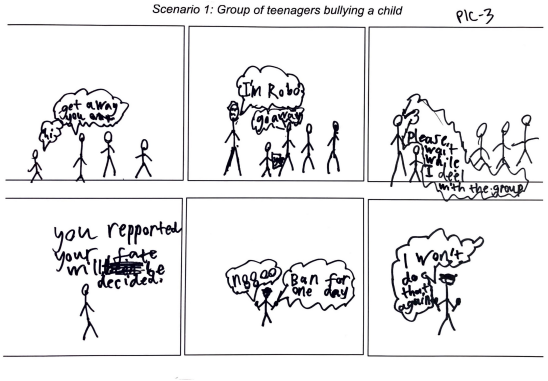
Embodiment and Design Features: Children Findings mainly encompass children's ideas regarding: appearance of the moderator, when they would like their moderator to appear, where they would like to appear. A portion of children's answers regarding physical appearance, when and where their moderator would appear, are displayed in figure 6.10, highlighting creative designs and drawings. Children's storyboarding designs can be found in figure 6.11, figure 6.12 and figure 6.13.

Appearance and identity Children suggested a Human-like appearance (N=1, Child P3bC-2) for the AEM. Some proposals leaned towards an Animal-like (N=5) form, including options such as *e.g., a snake [Child P1], their pet [Child P2, P3], a cute lion for the victim transitioning from cute to a scary lion for the bully [Parent P2aC-1].* There were also suggestions for designs that did not fit neatly into either category, encompassing original creations like *"a cloud"* [Child P1C-1]. These diverse form factor ideas highlight the creativity and variety in envisioning the AEM's appearance.

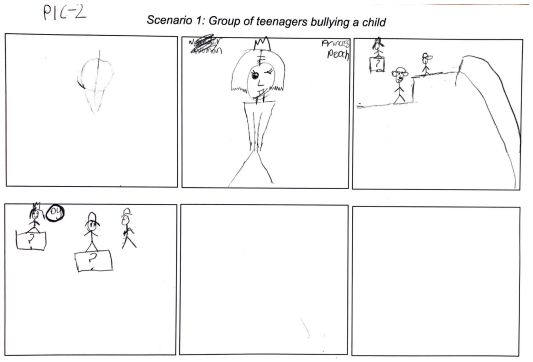
Children proposed incorporating Pre-existing Media Characters into the moderator's appearance, including figures from cartoons and superhero comics *e.g., Wonder Woman [Child P1C-2] or Superheroes Comics [Child P2C-2]*, and fictional game characters (N=3) *e.g., "Pikachu would electrocute people who are misbehaving" [Child P3aC-2], "Princess Peach" [Child P1C-2].* In contrast, there were suggestions for Original Characters that would be unique creations specifically designed for the role of a moderator. Some designs leaned towards a Naturalistic approach, with suggestions like a tall, realistic police officer figure [Child P1C-3]. In contrast, the Supernatural design category featured suggestions like AEMs for younger children and victims and creatures that are scary but not petrifying *"something scary but not petrifying and realistic: not a monster or a teacher"* [Child P1C-3].



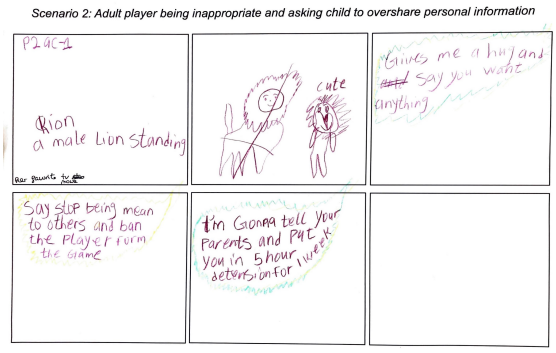
(a) Design Child P3aC-2 "This guy, Rob, is bullying Izzy: "give me your address so we can date" and because it's against the rules to try and date minors, Izzy says "I'm reporting and pressing the button". Then there is this notification saying "Hey Izzy I went to tell off Rob, I will notify you when he is banned". Pikachu the moderator says to Rob "I'm angry you broke TOS" then "Rob has been banned permanently". I chose Pikachu because he would electrocute people who are misbehaving"



(b) Design Child P1C-3 "As soon as they log in, children, to progress, they have to give it a name. The victim being harassed says "Go away" then the moderator appears, it's taller than everyone, says "I'm (NAME), here Robo". "You" is the victim. Robo says: "Please wait until I talk to the bully". You have to press the button for the group to be able to see the moderator saying "You've been reported your fate will be decided". The fate will be a ban for one day. They go back in the game but they have to say "I won't do that again". If it happens again, they get 3 warnings then a ban for much longer."

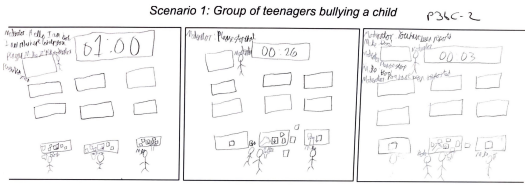


(c) Design Child P1C-2 "My moderator from the Mario (Princess Peach). When something happens she says "Oy, why are you (Luigi) hurting (Mario)" and then Luigi runs away".



(d) Design Child P2aC-1 "Name of the moderator: Kyle, it's from my favourite cartoon, the Lion King. It looks like a Lion. If the person becomes mean it becomes scary."

Figure 6.11: PART 1: Examples of children's storyboarding designs with their created automated AEM in a scenario of harassment.

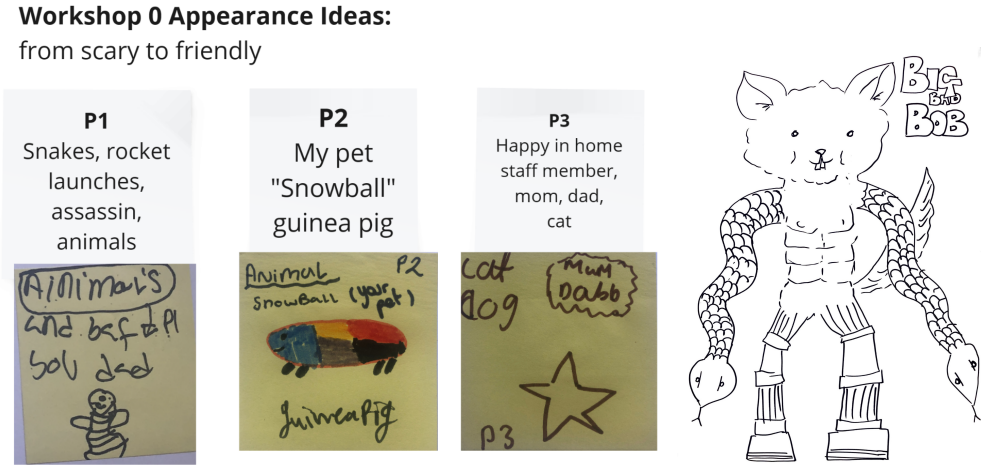


(a) Design Child P3bC-2 "The moderator announces they are here. Player Mike joins and then player Bob joins. Bob's knocking Mike's tower over. The moderator says "Can you please stop that" and then he does it again, the moderator "you've been reported"."



(b) Design Child P2C-2 "Name of moderator is Monitor (from Dark Monitor Superman comics) a guy with wings and asks the bully to stop mocking the other child."

Figure 6.12: PART 2: Examples of children's storyboarding designs with their created AEM in a scenario of harassment.



(a) Physical Appearance Post-its Workshop 0.

(b) Combined Design from Children in workshop 0 (cat, snake, assassin)

Figure 6.13: Designs from workshop 0 combining children's moderators' ideas into one final design (cat, snake, assassin).

Relationship to User Participants proposed a Friend Figure design (e.g., their best friend [Child PC2-1] for the AEM, suggesting it could take on the role of a supportive companion or their own pet as mentioned above with the same name. Another child suggested a Family Figure [Child P2] to provide a sense of familial connection and comfort. Additionally, the concept of an Authority Figure emerged, with suggestions ranging from powerful animals (Lion) [Child P2aC-1] to police officers [Child P1C-3], all designed to enforce rules and deter inappropriate behaviour, scaring the bully. Child P1C-3 also suggested that players can choose the name of their moderator to feel safer.

When the AEM should appear

Children expressed various preferences regarding when they wanted their AEM to be appear. Some children advocated for the moderator to be Always (N=4) there, particularly when there were multiple players involved, ensuring a constant presence to oversee interactions. Others favoured the moderator's intervention When something happens (N=7) with options for it to stay invisible during periods of inactivity and to be triggered by specific words or incidents. Additionally, there was a consensus that the moderator should only become active once a child approves or decides (N=3), with features like a button for calling the moderator or its appearance after the child clicks a report button. A child also suggested implementing the moderator during "Peak times based on the number of player" (N=1, Child P3bC-2) to manage higher activity periods.

Child P1C-3 preferred a more subtle or invisible presence for the moderator. This could involve it being *"Always there but invisible or behind you"* or having a small pop-up and a larger buddy on the side to avoid disrupting the gaming experience. There was also a suggestion for the moderator to become visible once it detects inappropriate behaviour.

Finding 7: Experts and guardians generally suggested neutral, non-human, approachable yet authoritative AEMs that communicate warmly and adapt to age, culture, and context. Customisation by age and user input is important, with playful and empathetic communication encouraged. Children proposed more imaginative diverse forms (animals, familiar characters, friends, or authority figures) with options to name and personalise. They want the AEM either always present or activated when needed, sometimes preferring subtle or invisible appearances.

6.3.7.5 Guardian's Role and Involvement with the Automated Embodied Moderator

Reflection on Guardian's Involvement with the System Parents In-the-Loop It is generally considered that guardian involvement is important for both victims and harassers. Experts compared the AEM to a teacher or football coach keeping guardians in the loop (N=3). Parents know their child well and might be able to adapt the AEM interventions and interactions to their child's needs or preferences, especially at younger ages. Yet, children also need their privacy and autonomy. Therefore, guardian involvement should be balanced with children's autonomy and privacy (N=6).

Experts expressed concerns about possible excessive and unhealthy involvement (N=4), with parents overly monitoring and restricting the child's experience through AEMs. Experts had negative feelings towards real-time parents' notifications (N=2), saying it's "creepy" [Expert P2] or feeling "nervous about notifying parents" [Experts P2, P3].

Accommodation of Parenting 'Styles': Moreover, some experts highlighted the challenge of accommodating different "parenting styles" (N=2). It cannot be assumed that every parent is appropriately involved or good at parenting.

> *"There are those parents who are on it, they understand what their children are doing, they take an interest; there are those that want to go down the punitive route, and there are those that really don't care. I think the phrase is 'plotting fog', trying to appeal to every parental style. [...]. I'm not sure you would be able to adapt to every kind of parenting style."* [Expert P16]

Fostering Independence and Resilience The system should also take the extent of guardian involvement into consideration, and the impact this has on children's resilience and ability to deal with incidents independently:

> *"Our children do need to be more resilient, they do need to know what the world is, it can be a tough place, and if they know that the parent is standing over their shoulder there to protect them all the time, it's not really good for the child. Resilience is something that we have to teach our children and something that they need to discover for themselves, so that's why I just take the notification to the parent, 'this happened', is adequate".* [Expert P6]

Finally, some experts mentioned potential guardian disengagement or even laziness (N=3) leading to parents potentially outsourcing their responsibility to AEMs. However, the moderator also allows guiding parents in supporting and educating their children. Either way, experts agreed that legal guardians must be aware and knowledgeable of what is happening to some extent and understand how the moderator works.

Ways Guardians Can Be Involved with the System Both experts and guardians highlighted multiple opportunities for parental involvement in supporting children's safety within social VR environments. Experts identified key moments for engagement, including during initial setup (N=10), in real-time while the child is using social VR (N=9), after incidents or events (N=5), through ongoing education and guidance to enhance guardian awareness (N=6), via personalisation based on the child's developmental needs (N=3), and during the testing phase of the AI moderator (N=1).

Similarly, guardians expressed a desire for flexible and practical ways to engage with the automated moderation system, reflecting diverse levels of availability and digital literacy. While some guardians acknowledged being too busy for continuous monitoring: *"I can see [the AEM supervising] really appropriate for my younger children. I'm thinking as I'd realistically be getting on with something else."* [Parent P2P-4]. Others pointed to a lack of technical knowledge as a barrier (N=2). Despite these challenges, all recognised the importance of access to clear evidence, open communication, and increased awareness to better protect their children online. Their preferences echoed expert recommendations, with particular interest in involvement at the setup (N=4), in real-time (N=3), after incidents (N=2), and through fostering open dialogue (N=2). Grandparents (N=3) also found the AEM valuable but wished to be reassured with some real-time feedback, pre-briefs and debriefs. Together, these insights underscore the need for adaptable systems that offer educational resources, transparency, and personalised support to empower informed and effective parenting in social VR.

At the setup Experts want guardians to understand the moderator's functionality and be aware of the technological options (N=6). The latter can be done through a tutorial involving parents and children explaining the code of conduct, the AEM's role and consequences for actions.

> *"If the kid was using the [virtual] room for the first time, there could be like a tutorial that the parent could watch with them to say "yeah this is how we're going to keep everyone safe and happy in this room and play nicely", and it might reassure a parent."* [Expert P3]

> *"You'd still want some training for parents. You'd want parents to understand what the role of Big Buddy was so you might want to have a little kind of introductory video for parents."* [Expert P12]

The goal is for parents to become familiar of the technological options and to negotiate appropriate settings for their child. Parents (N=4) would want to learn about the AEM and get information about social VR. They would choose settings (e.g., time limits, rules) with the AEM which would depend on the age. A parent suggests implementing a *"Parental Permission Button per Session"* [Parent P1P-1] that allows them to grant consent for each VR session their child participates in.

Experts also suggested a Notifications setting (N=6) to choose when to receive notifications and debriefs from the AEM and where (e.g., messages, emails, etc.). Additionally, they believe parents should set initial choices for effective intervention methods (N=2), given their understanding of their child's needs. This could involve selecting effective phrases and strategies that assist their child, creating a familiar environment, or selecting a range of good behaviours, customisable to align with cultural norms and individual preferences.

> *"Let's say I'm a parent and I have my own ways to you know to talk with my child whenever they're upset. And instead of saying "are you okay" maybe I have different phrases I use that is really effective. So I think the fact of maybe I can suggest phrases as a parent in this in the design of Big Buddy."* [Expert P9]

Real-time insight while the child is playing social VR Some experts recommended real-time notifications (N=7) when an incident occurred with an explanation about what happened to parents of both parties (victim and harasser).

> *"[Big Buddy] could send a message to the parents to let them know that there was action taken. I think if these are registered children then the app knows that it is children, and when something happens to them that necessitated Big Buddy intervening. I think a message to the parents just letting them know exactly what happened."* [Expert P6]

Guardians (N=3) proposed constant monitoring and parental notifications, enabling them to actively stay informed about their child's social VR activities. They suggested having the ability to check what is happening in VR at any time and receiving immediate notifications when issues arise (whether their child was the victim

or the bully): *"Get email/text when something is happening"* [Parent P3P-2], *"Instant notifications when something happens"* [Parent P1G-3], *"Notifications to both parties' parents. I think, certainly, there should be notification to the parent advocate, but also notification to the parent to whom the harassment or bullying happened."* [Parent P1P-1]. They also suggested parents to take part in the decision-making of the outcome (N=4) with options or potential responses the AEM can give (e.g., block rooms; report, pausing or preventing the child from entering this room again). They could also give another chance for the child to change their behaviour. However, the outcome should be directly taken care of by the AEM in severe instances and disruptions, according to experts (N=2).

After-the-fact oversight Experts (N=5) recommended that parents receive debriefs after the child gets a certain number of warnings in the game. They also suggested that parents have open conversations and discussions, providing explanations to the child without blame. Guardians (N=2) would want to receive feedback via email or text message, not only regarding an incident that happened but also the child's overall online experience:

> *"I think we mentioned earlier that if there are any incidents, we receive some kind of notification [...] but it's also "did you enjoy that?" because one of the problems is speaking about that to kids once they get to about 10, 11 and 12. It's like: Did you have fun? And this is as far as it goes. So it would be interesting if you could get something like that, some kind of not necessarily important, but what they really did enjoy it and to what extent they enjoyed it.* [Parent P1P-1]

Education and guidance to enhance guardian awareness The lack of guardian awareness to adjust settings in the context of using an AEM in child-oriented settings is a concern. To address this, experts (N=6) noted a need to educate less knowledgeable parents about best practices and offer suggestions for interacting with the AEM. This should include explaining the consequences of different settings and providing discussion tips to parents: *"The notification to parents could include some tips as to what the parents can discuss with their child as a result of this, in terms of why this action was taken."* [Expert P10]. Guardian awareness and consent for the use of AEMs are crucial, and there's a need for educational tools and support to help parents handle conflicts with their children effectively. Providing guidance and advice can enhance parents' knowledge in this regard.

Guardian involvement should be personalised Experts suggested having a continuum of versions of parental involvement (N=3), from non-guided to guided depending on different groups, not only age but also based on social abilities, developmental stages or considerations of neurodivergent children (e.g., autism spectrum, ADHD). The level of interaction and intervention can be adapted based on the child's needs and the parents' availability; they can change the settings over time.

During the testing phase of the AEM. Expert P7 suggests involving parents during the testing phase when developing the prototype to ensure effectiveness, usefulness and preferences of being involved with the system.

Open communication and facilitator of conversation Guardians (N=2) would want the AEM to help facilitate child-parent conversations by giving children the opportunity to communicate with the parents about their online interactions. They proposed that the AEM introduce the rules to children, which can help improve the parent-child relationship while ensuring a safer virtual environment for children. *"The AEM could highlight situations to parents, it could facilitate conversations around bullying or the game with parents and children".* [Grandparent P1G-4].

Grandparent involvement While grandparents (N=3) mentioned their lack of understanding on the technology (N=1), they would like to receive reassurance about the AEM's competence, feedback and weekly reports on their grandchild's behaviour, as well as instant notifications of serious events.

> *"It's kind of dependent on what it is, if it's something very serious, I would like to know because the child may be distressed, if it's a serious incident, they're in, then you need to know."* [Grandparent P1G-4]

A grandparent noted they would want to use calm language in rule reinforcement.

Finding 8: Guardian involvement with the AEM in social VR is seen as important. Experts warned against excessive monitoring, urged parents to build understanding and awareness, and noted challenges in accommodating diverse parenting styles.
Ways Guardians can be Involved: Guardians should engage during initial setup (understanding the AEM, setting preferences), real-time (notifications and input on outcomes), and after incidents (debriefs and open discussions). Experts suggested personalised involvement where the level of interaction and intervention from the parent with the AEM is adapted to the child's needs and parent's availability. Parents would want AEMs to facilitate parent-child dialogue and inform parents. Given limited tech familiarity, grandparents would want reassurance about the AEM's reliability, simple updates, instant serious-event alerts, and the use of "calm" language for enforcing rules.

6.3.8 Discussion of Studies #5 and #6

Reflection around the AEM and future work are provided in chapter 8 to situate these studies' findings within the wider contributions of the thesis. Here we present a summary of the findings and references to corresponding guidelines in chapter 9.

6.3.8.1 Summary of Findings

This work provides an important first step in reflection around, and recommendations for, the design and use of AEMs to safeguard children in social VR. I merged perspectives from experts, guardians and children, providing an enriched and holistic set of insights. Summary maps of findings from each perspective on *automated moderation* and *perceptions towards the embodiment of AEMs* are displayed respectively in figure 6.14 and figure 6.15.

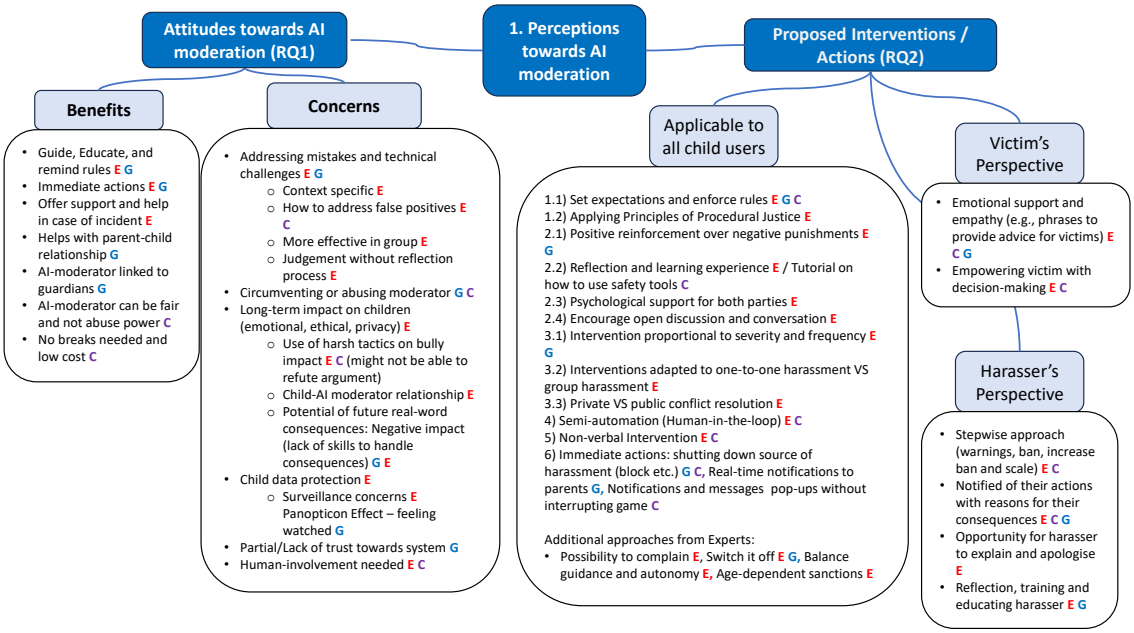


Figure 6.14: Summary Map of Findings regarding Automated Moderation with corresponding perspectives (E for Experts, G for Guardians and C for Children).

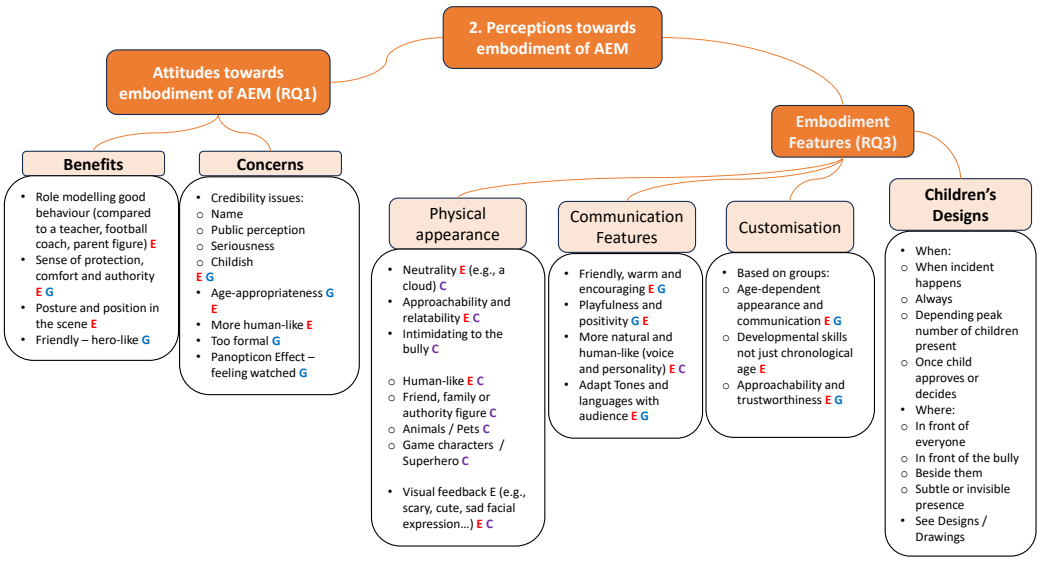


Figure 6.15: Summary Map of Findings regarding Embodiment of AEMs with corresponding perspectives (E for Experts, G for Guardians and C for Children).

While there is consensus across stakeholder groups regarding most recommendations and preferences around AEMs, there is a contrasting perspective regarding the extent of AEMs' involvement: children favour a more passive role for the AEM, intervening only when necessary, with discretion and immediate actions, whereas adults anticipate a broader range of engagement with discussions and activities beyond intervening when an incident happens. There is also a tension between making the AEM friendly without being intimidating, while still having authority and being credible.

Children and adults choose to join social VR to enjoy fun immersive and embodied social experiences, to escape from reality [22, 21] and benefit mental health [365]. However, it is a space lacking social norms and effective moderation, especially child-centred solutions. Assuming a problematic event occurs in VR and the automated moderation system can detect said problematic event (e.g., via the biometrics and social/contextual signals that VR headsets and platforms can provide [7, 50]), this work shows that AEMs appear to be suitable for children and desired by experts and guardians, potentially able to tackle most challenges current safety tools and human moderators have (see section 6.3.2.1). The motivation behind AEMs is to create *safety-enhancing technology* that can react instantly to harassment or problematic incidents, offering help and a sense of protection. And with the affordances VR offers, we can not only introduce AEMs, but also tailor their appearance and interactions for each user.

Moreover, guardian involvement, with the AEM in social VR is seen as important for protecting child victims and guiding harassers, but it must balance children's privacy, resilience, and independence.

Based on these findings, I gather design guidelines, see Guideline 14 section 9.2.

6.3.8.2 Specificity of Findings to VR versus Non-VR Social Media

Compared to non-VR social media (e.g., Instagram), VR offers the feeling of presence, with users inhabiting social virtual environments with embodied interactions, replicating face-to-face experiences, where almost all senses are engaged (vision, hearing, phantom touch [366]). Therefore, *embodied* automated moderation is particularly unique to social VR, and brings with it notable benefits and challenges compared to how moderation is enacted on non-VR social media.

Differences of the Findings in Social VR Compared to Social Media Firstly, the capacity of surveillance in VR, raised as a significant concern by the participants, is increased, given the additional movement, physiological and behavioural data available [367], and participants speculated this could impact the perceptions towards moderation over time through the "Panopticon Effect" [368] where users feel like they are being watched.

Challenges related to circumventing automated moderation also become more pronounced in social VR compared to social media, where interactions are predominantly text- or image-based. In social VR, users may resort to forms of embodied, real-time harassment (e.g., physically, environmentally), which makes automated detection (previously relying on, e.g., textual analysis [107]) more challenging. Unlike social media, which retains users' preferences over time, AEMs in social VR may also depend on new sensitive biometric data related to a child's behaviour, emotions, and physical reactions. This includes biometric data, such as facial expressions or physiological responses, raising new privacy concerns.

Additionally, VR offers the unique advantage that AEMs can be perceived as being *socially present* in similar ways to real authority figures, which the participants noted could give a sense of protection and potentially support AEMs being 'role models' to children. This characteristic draws a more direct parallel to traditional figures such as teachers, coaches, or parent figures and adds a layer of symbolism and authority that is challenging to replicate in 2D social media moderation.

With respect to interventions, while some proposed from the findings (Intervention Themes 1-4, Victim's Interventions and Harasser's Interventions) may also be relevant to social media, the way these are presented differs in VR due using embodiment, immersion and continuous real-time interactions, which lead to an increased importance of interventions by the AEMs being synchronous rather than asynchronous (Intervention Themes 5,6, Victim's Intervention Theme 2, Harasser's Intervention Theme 3). Non-verbal interventions stand out as specific interventions to VR, with the AEMs acting as a 'buffer' to resolve a conflict, or offering 'virtual hugs' to comfort the victim. Dealing with harassment and abuse in social VR also requires immediate actions, due to it being more confronting and continuous compared to 2D social media. Such actions may include promptly shutting down sources of harassment or sending real-time notifications to parents.

The distinction between resolving conflicts privately and publicly is harder to manage in social VR, due to the complications of real-time face-to-face interactions in virtual rooms. However, this may be closer to moderation of live streams in Discord, for example [369]. In contrast to social media, which primarily handles issues privately, the dynamics in social VR can mirror a classroom-like scenario where a teacher might address an issue in front of all classmates or opt for a private discussion with a student during recess. This underscores the need for nuanced conflict resolution strategies tailored to the different social dynamics in VR environments.

Finally, social VR must be accessed through a headset, which makes parental oversight more challenging and highlights the need for real-time notifications. Furthermore, the AEM can serve as a valuable guide and mediator, helping bridge the gap between parents and children.

Similarities of the Findings in Social VR Compared to 2D Social Media
Studies have revealed children's openness to automated and semi-automated monitoring approaches [370, 220] and emphasise the importance of incorporating risk-coping strategies, including victim-oriented and quick interventions to improve their well-being, such as watching cat videos or suggesting activities like playing with a sibling [220]. Interestingly, these prior investigations also raised concerns similar to those in the study regarding the feasibility of automation. Children expressed worries about the app potentially misclassifying conversations as risky and unnecessarily notifying parents, thus escalating situations [370].

The findings also highlight the importance of designing proactive solutions to safeguard against risk exposure, not just after-the-fact, but at every stage of the risk (before, during and after-the-fact) and designing guidance for the teenagers with live assistance to help them respond after the risk [221]. Researchers similarly identified the potential of restorative justice in understanding and addressing adolescents' needs in online harm [371]. However, in social VR, restorative justice can be enacted differently, via non-verbal communication, 3D interactive simulations and spatial engagement.

The need for parental involvement in social media has been widely researched, underscoring its importance and highlighting effective approaches such as parental interest [372], parental awareness [373], and trust-building [374], while balancing children's privacy [375], independence and self-efficacy [376, 377].

6.3.9 Conclusion

I have presented two empirical studies to address the gap in understanding AEMs to help children in harmful experiences in social VR. The studies serve as groundwork to gain initial insights into the design space of AEMs. Driven by individual interviews with 16 experts in child online safety and psychology and workshops with 13 children and 8 guardians, I collected recommendations and reflections regarding the automation and the embodiment of AEMs to safeguard children in social VR. Experts, guardians and children see benefits of AEMs for children, such as enforcing rules and 24/7 availability in a virtual space that lacks social norms. Intervention approaches should focus on clear communication of expectations, positive reinforcement, and adaptive responses to harassment. Victims need emotional support and decision-making capabilities, while harassers should be held accountable with educational and repair opportunities. For AEM embodiment, neutral or player-customised forms are suggested, with adaptable communication and natural voices. Future interdisciplinary research remains to tackle design challenges raised.

6.4 Chapter Conclusion

This chapter examined how AEMs are perceived as a safeguard for children in social VR and identified the important factors considered to prospective usage of AEMs **(RQ3: How are Automated Embodied Moderators (AEMs) perceived as a solution to safeguard children from harassment in social VR and what factors are considered important to the prospective usage of AEMs?)**. Through experimental research with children, workshops with families, and interviews with experts, I have built a holistic understanding of how AEMs could intervene in harmful interactions while maintaining children's agency, enjoyment, and sense of safety.

Study #4 makes a number of contributions to CCI. It presents the first, to my knowledge, experimental research of a VR-constructed Wizard-of-Oz AEM that aims to mitigate the harasser's actions, with the anticipated benefit of making children feel safer and more comfortable in the simulated social virtual environment through visible safeguarding interventions. I get first impressions and perceptions from both children

and parents. Secondly, I gathered the first data about design features an AEM would need to not only safeguard children but also be perceived positively and reassuringly to children and parents without completely removing children's sense of agency and enjoyment or overly restricting their freedom.

Building on this foundational study #4, I then gathered considerations into the design space of AEMs uncovered by Studies #5 and #6 to better understand their potential and applicability. These contribute insights into the design of AEMs to help children in harmful experiences in social VR that and balance the three stakeholders viewpoints (children, guardians and experts). Driven from one-to-one interviews with experts and workshops with families, I synthesise their concerns, preferences and proposals for the design and use of AEMs.

From these two phases (along with the studies from the next chapter), I produced design considerations towards eventual realisation of AEMs to safeguard children in social VR, around the lifecycle of a moderation incident (see chapter 8).

6.4.1 Contribution to Research Question 3

While participants highlighted AEMs' benefits, including immediate actions, fairness, presence and victim support, experts and parents noted that complete automation could lead to unintended consequences, inaccurate harassment detection and reduced child independence in tackling incidents. Moreover, children may need to be more proactive in a bullying situation to feel empowered. Prior work highlights the importance of building resilience, as it helps shield children from distress during school bullying situations [351]. Similarly, parental involvement, especially through real-time notifications, was mentioned in prior chapters. Study #8 examines how varying levels of parental engagement, indirect (facilitated by the AEM, Big Buddy) or direct (via in-VR embodied presence), affect the experiences of both parents and children after simulated social VR harassment incidents.

Next steps build on these insights by exploring practical human-in-the-loop AEM deployments that include 1) varying degrees of child participation in decision-making with the AEM and 2) different levels of parental involvement with the AEM.

Chapter 7

Deriving Requirements for Automated Embodied Moderators from Child and Parent Experiences

"That's good, it has options because I get to choose if I want the wand or the AEM. I like having a choice." - [Study #7 Child5 - (C3: AEM gives options)]

"I think this digital notification is a great way to notify whenever there's some kind of social VR bullying or any kind of disruptions." - [Study #8 Parent P7]

7.1 Chapter Introduction

Chapter 6 highlighted the potential benefits of AEMs as seen through the eyes of key stakeholders (children, guardians and experts). While the promise of immediate intervention, fairness, and victim support was clear, concerns about over-automation, inaccuracies in harassment detection, and impacts on children's autonomy remain. Furthermore, the critical role of parental involvement, particularly through timely notifications and active engagement, emerged.

7.1.1 Motivation

Building on these findings, this chapter aims to investigate how AEMs can be practically integrated with child and guardian participation. Specifically, this chapter explores the dynamics of human-in-the-loop AEM deployments, focusing on two key dimensions: varying levels of child participation in decision-making alongside the AEM, and varied degrees of parental involvement, ranging from indirect support via the AEM to direct, embodied presence within social VR environments. The two user studies presented build on theories of child participation [352] and parental mediation [112, 117, 378].

Therefore, in this chapter, I aim to answer the following research question:

RQ4

Informed by RQ3, what are the requirements of the Automated Embodied Moderators to safeguard children in social VR while still empowering them?

7.1.2 Overview of Studies

7.1.2.1 Study #7

The optimal automation level to maximise AEM's effectiveness, children's well-being and agency still requires further investigation. In studies #5 and #6, experts and guardians mentioned the importance of victim empowerment in harassment situations and the children's involvement in the decision-making (studies #5 and #6, [6]). The extent to which they should be part of the decision-making with AEMs remains unexplored.

Methodology In simulated group harassment social VR scenarios (from study #3) with 40 children aged 8-16 years old, I investigated children's sense of agency and safety when initiating (identifying the need for an intervention) and deciding on taking action (blocking a group of harassers). Using Roger Hart's Ladder of Children's Participation [352], which outlines levels of decision-making authority, control, and power that adults can grant to children and youth, I explored four levels of child participation/AEM's automation.

Findings Children valued tools that gave them control, but also valued being assisted by the AEM if they were part of the decision-making. Therefore, designing tools that let children control their level of involvement in moderation is essential for balancing personal autonomy with automated assistance tailored to their needs and preferences. This approach can enhance agency and empowerment in social VR when confronting group harassment.

7.1.2.2 Study #8

While parents acknowledged the usefulness of AEMs in Chapter 6, they emphasised their desire to be more involved, particularly by receiving real-time notifications about disruptive incidents. Experts also recommend a degree of parental involvement via mobile apps and notifications, acknowledging that parents are often busy and may not have direct access to the VR environment. Consequently, any safety-enhancing companion tool should provide meaningful insight and oversight without resorting to intrusive real-time monitoring, such as continuously streaming the child's social VR activity video feed.

Methodology In study #8, I designed a four-condition VR study (N=34 parent-child dyads) that varied parental involvement using a custom-built mobile companion app (with Big Buddy) to receive real-time notifications and child support in a social VR simulated game scenarios based Study #4: no support, AEM-only support, indirect parental support via parent-written messages spoken by the AEM, and direct parental support via parent voice messages delivered by the parent's avatar in VR.

Findings Findings showed that parents and children both prefer some degree of parental involvement in social VR safety, with stronger support conditions (Indirect and Direct Parental Support) rated significantly higher than no support. Parents generally favour direct involvement and value concise but informative notifications with video clips of incidents and control options, especially in serious incidents. However, they adjust their preferred level of involvement based on their availability, often opting for the AEM's support when hands-on participation is not possible, and some consider their child's needs and consent. Some parents acknowledged their own limited knowledge of the VR environment and expressed a reliance on Big Buddy, trusting the AEM's protection capabilities, even sometimes with a degree of blind trust. At the same time, others showed self-awareness about the need to be more actively involved to ensure their child's safety. Children, meanwhile, felt safer and more supported with higher levels of parental involvement and parental knowledge, but often prefer less intrusive forms, such as indirect support or messages through Big Buddy that respect their autonomy and privacy. Both groups wished for customisable options for parental presence and in-VR parent avatar design to avoid embarrassment or disruption while maintaining connection and safety. Parents desired communication features that would enable two-way interaction (including brief, low-disruption feedback from the child, like emojis), greater transparency and context information about the incident and AEM's actions, as well as more control over interventions, while children would want agency over when and how parents are involved.

7.2 Study #7 - Exploring Levels of Automation in Social VR: From Manual to Automated Interventions in Group Harassment Targeting a Child

7.2.1 Motivation

Automated moderation holds promise in addressing moderation challenges in social VR, particularly for children. However, children need to be empowered and proactive in a harassment situation in social VR. This study builds on these gaps by exploring the initiation and decision-making processes behind blocking actions in group harassment scenarios. By addressing the role of initiation and decision-making, the study aims to provide valuable insights for developing effective and child-centred intervention strategies.

Based on these gaps, through this study, I address the following research question:

S7-RQ1 How does the intervention automation level affect emotional perceptions, safety, agency and perceived effectiveness during simulated social VR group harassment?

7.2.2 Related Work

This section synthesises related work on the specific topics addressed in these studies. (A more comprehensive literature review is presented in chapter 2, with specific relevant sections for context: section 2.3.2, section 2.3.3, section 2.3.4; and for methodology: section 2.4.2, section 2.4.3).

7.2.2.1 Automated Moderation in social VR to Safeguard Children and Child Participation

In social VR, safety tools like blocking, personal space bubbles, muting, and reporting [187], aim to protect users from disruptive individuals [204], but are not always adequate for children. They place responsibility on children, who might lack the necessary understanding or familiarity with the technology [122, 114, 363, 115]. AEMs designed with VR's unique affordances show promise for child safety in one-to-one harassment (see study #4). Studies #5 and #6 investigated AEMs' potential in terms of the interventions and the embodiment features (studies #5 and #6, [6]). While participants highlighted AEMs' benefits, including immediate actions, fairness, presence and victim support, experts and parents noted that complete automation could lead to unintended consequences, inaccurate harassment detection and reduced child independence in tackling incidents (studies #5 and #6, [6]). Moreover, children may need to be more proactive in a bullying situation to feel empowered (studies #5 and #6, [6]). Prior work highlights the importance of building resilience, as it helps shield children from distress during school bullying situations [351].

With the evolving rights to privacy and freedom of expression for children, complete automation moderation poses challenges [294, 148]. However, children may still require guidance. Roger Hart's ladder model of child participation offers frameworks to assess and improve child engagement [352]. Proposed rungs include: 1) Child-initiated, shared decisions with adults, 2) Child-initiated and directed, 3) Adult-initiated, shared decisions with children, 4) Children are not involved in the decision or action [352]. These directly informed the conditions for this study.

7.2.3 Methodology

The study was conducted in July-August 2024 with **34 children**.

7.2.3.1 Study Design

From study #3 methods, I used the same simulated social VR, the large group harassment scenario, and Likert Scales refined (section 7.2.3.4). The procedure of the experiment and the recruitment were similar.

Variables I compared child perceptions towards three ways of manually blocking harassers and two IPs. Based on the results, I selected the most preferred tool (*[PointBlock]*) and IP (from above). In this study, I investigated how the initiation and decision of blocking, whether made by the child or by an AEM, impact children's sense of agency, safety, emotional perceptions, and perceived usability.

Experimental Design This study follows a within-participant design with one independent variable (four levels of automation), based on Roger Hart's ladder model of child participation described in section 7.2.2 [352]:

(**C1**) *[Manual]* **Initiation: Child, Decision and Action: Child.** The child independently initiates the intervention using *[PointBlock]* from study #3, as it was the most preferred tool (see section 4.3.4.2 and section 4.3.4.5).

(**C2**) *[Call AEM]* **Initiation: Child, Decision and Action: AEM**. The child calls the moderator by clicking on a button on their virtual watch. The AEM then takes action.

(**C3**) *[AEM gives options]* **Initiation: AEM, Decision: Child, Action: Child or AEM.** The AEM initiates the intervention and presents the child with three options ("It seems something wrong is going on. Do you want me to block the bullies?"): 1. "Yes", the AEM proceeds to block the harassers automatically, 2. "No", the AEM stands down, 3. "Wand to block" allows the child to manually block the harassers.

(**C4**) *[AEM fully automated]* **Initiation: AEM, Decision and Action: AEM**. The AEM automatically detects the need to intervene, makes the decision to block harassers automatically without any involvement from the child.

Conditions were counter-balanced using a Latin Square design. The figure below describes the condition dimensions of initiation and decision of action from children and AEMs figure 7.1.

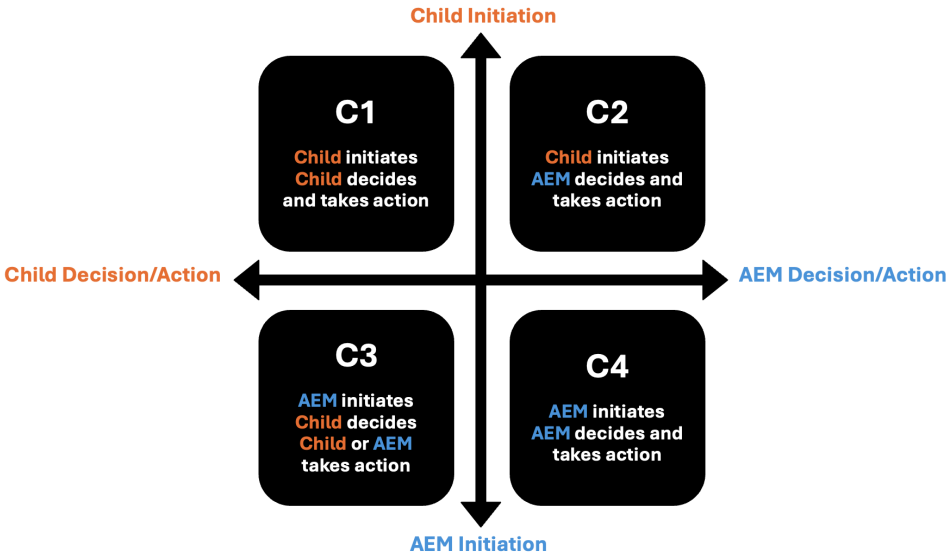


Figure 7.1: Diagram plotting the four conditions of Study on two axes: on the y-axis, the initiation (who starts the intervention) and on the x-axis, the decision/action (who decides and acts). The diagram highlights the varying levels of child and Automated Moderator (AEM) involvement across conditions.

7.2.3.2 Social VR Scenarios and Interventions Implementation

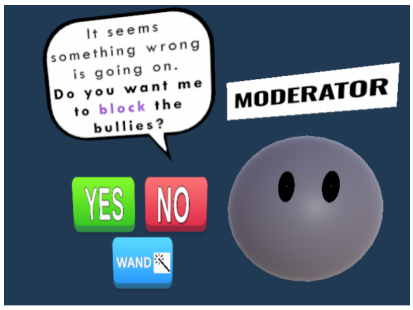
I used the prototype from study #3 section 4.3.3.2 as a foundation, with the large group harassment scenario. Based on study #3 results, I incorporated *[PointBlock]* for the manual tool C1 and the AEM for conditions C2, C3 and C4 (see figure 7.2). In all conditions, the user was elevated above the group during the interventions. The AEM was designed to be as neutral as possible in terms of embodiment features, to minimise the likelihood of triggering varied responses among children based on personal embodiment preferences (e.g., some may prefer animal-like AEM, while others favour human-like AEM) (studies #5 and #6, [6]) and maintain the focus on the main variable (levels of automation). Therefore, the AEM was a grey sphere with two eyes and had a label with 'Moderator' figure 7.3. The voice was chosen to have a neutral tone and free from gender associations as described by the platform *SpeechGen* Gender-Neutral Voice Generator [379]. The AEM's statements also appeared in written form inside a speech bubble. (Video in Supplementary Material STUDY 7/VIDEO).



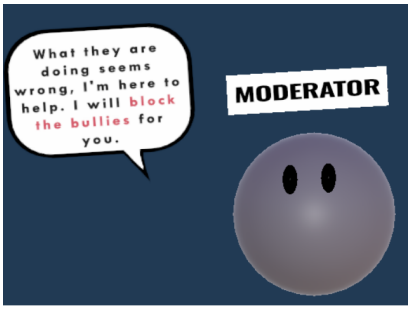
(a)
C1



(b)
C2



(c)
C3



(d)
C4

Figure 7.2: The four conditions presented. a) C1: Participant uses their watch to click on the button and uses the [PointBlock] tool from study #3 to block harassers' avatars. b) C2: Participant uses their virtual watch to click on the button that calls the Automated Moderator (AEM). The AEM then blocks the bullies. c) C3: The AEM initiates the intervention, asking the child if they want the AEM to block the bullies, and gives the child three options: yes, no or wand. Child clicks on the chosen button. d) C4: The AEM initiates and blocks the bullies automatically.



Figure 7.3: The Automated Moderator in the virtual scene.

7.2.3.3 Procedure

I conducted pilot tests, also under ethical approval, with four adults, then tested with an 8-year-old. Although the VR sessions went well, the questionnaires lacked clarity, as children would ask if they should answer how they felt before or after using the tools. To address this, I split the questions into two sections: before and after the bullies were blocked. After refining the prototype and questionnaires, I ran the experiment.

Initially, I conducted the study with eight children at a summer camp, but only four completed the full experiment. The remaining 30 children participated at the university (individually or in pairs), accompanied by a parent who waited in the room, separated by a screen. Children signed a consent form.

Children were then introduced to social VR, bullying, blocking and the AEM through slides (Supplementary Material STUDY 7/SLIDES). I explained two of the blocking tools: the "Magic Wand" *[PointBlock]* (used in Condition C1) and "Call the Moderator" (used in Condition C2). Using a Meta Quest 2, children practised those tools in a simulated social environment with fictional avatars that repeatedly appeared, allowing them to practice blocking them until they reported feeling comfortable using the tools.

They were then told that they would be observing a child being harassed by a group of nine harassers (Study #3's large group section 4.3.3.1). To avoid direct involvement, they were asked to imagine themselves in the victim's position and reflect on their feelings, reactions, and the interventions' effectiveness. They experienced four rounds, one round per condition. Following each round, children completed the questionnaire featuring the custom-developed Likert scales on a laptop. Sessions concluded with a 5-minute interview to gather participants' overall preferences and emotional experiences during the rounds.

7.2.3.4 Measures

In this study, I used the same approach as study #3, combining both quantitative and qualitative methods.

Quantitative Building on the refined scales from Study #3, I have improved the Likert scales by adding the option 'Other' to allow participants to describe a neutral or different feeling. This provides a compromise between an abstract neutral point and having the child select an emotion from the scale (Scales in Supplementary Material STUDY 7/QUESTIONNAIRE). Before the experiment, I also checked their understanding of the scales with simple questions (Supplementary Material STUDY 7/QUESTIONNAIRE). All 34 children said there was nothing they wanted to change about the scales. "No. I think there are enough options to choose from, including others." [Child4-A]. *After each round*, children were asked to rate how they felt in the bullying situation (before and after bullies were blocked). *At the end of the experiment*, children were asked to rank the four conditions from the one they liked the most to the one they liked the least.

Qualitative Children were asked to explain 'why' for each of the above questions using typing or audio-recording, based on the child's preferences. *At the end of the experiment*, I conducted semi-structured interviews to gain insights from the children:

- Can you describe how you felt putting yourself in the bullied child's (the victim) shoes?
- Imagine you are the victim being bullied, what do you think of the wand? What do you like? What do you not like?
- Imagine you are the victim being bullied, what do you think of calling the moderator? What do you like? What do you not like?
- Imagine you are the victim being bullied, what do you think of that assistance (when it arrives and gives 3 options)? What do you like? What do you not like? Which option would you choose and why?
- Imagine you are the victim being bullied, what do you think of the moderator coming and directly blocking the bullies? What do you like? What do you not like?
- Would you want the moderator to look differently? Please describe.
- Would you want the moderator to speak differently? Please describe.
- Would you want the moderator to behave/act differently? Please describe.
- If you were the bullied child in the middle, would you want assistance from the moderator? Why?
- If you were the bullied child in the middle, how much assistance from the moderator? Why?

7.2.3.5 Analysis

This study followed the same analysis methods as study #3 (see section 4.3.3.6) for both the quantitative and qualitative data. Coding schemes with categories can be found in the Supplementary Material STUDY 7/CODING_SCHEME as well as the R markdown in STUDY 7/DATA_ANALYSIS.

A Miro board gathering themes can be viewed via `https://miro.com/app/board/uXjVKq8aQIg=/`.

7.2.3.6 Participants

Recruitment The children (8-16 years old) were recruited in the same way as in study #3 (see section 4.3.3.7).

Demographics The demographics were provided by the parent or guardian of each participant.

This study involved 34 children (from 26 families) aged 8-16 years old, with a mean age of 11.47 years (SD = 2.35). The sample included 16 females and 18 males; with two Arab, seven Asian, two Hispanic/Latino, one Mixed, one Mixed Race, and 21 White/Caucasian. Four children had also done study #3. table 7.1 summarises children's online gaming, VR and social VR habits.

Table 7.1: Participants' Online Gaming, VR, and social VR Habits (26 families, N=34 children)

Measure	Mean (SD)	Scale
Online Gaming Frequency	3.26 (2.06)	0 (Never) - 5 (Daily)
VR Usage	2.18 (0.83)	1 (None at all) - 5 (A lot)
Social VR Awareness	1.94 (1.52)	0 (Not familiar at all) - 5 (Very familiar)
Social VR Usage	1.45 (0.94)	1 (None at all) - 5 (A lot)
Households with VR Headset	5 households	Yes/No

Ethics and Compensation The protocol was approved by the university CoSE ethics committee (Approval number: [300220209]). I was registered with the national PVG scheme to conduct this study with children. A £10 Amazon voucher was given to the accompanying guardian, while children received a certificate as a token of appreciation. Documents with the information sheet, pre-survey with consent form and demographics questions can be found in the Supplementary Material.

Selection and Participation of Children Before participation, the details of the studies were explained to the children's parents via an information sheet sent after they expressed interest. A digital consent form was also provided to be signed before coming to the session, noting that the child's participation (aged 8 to 16 years old) was voluntary. At the start of each session, I explained the procedure to the child using slides (see Supplementary Material STUDY 7/SLIDES). The child was then given a paper consent form to sign, where they could agree or disagree (by ticking boxes), to be audio recorded for the interviews, and were informed that they could take breaks or stop at any point. I assured them that they could ask questions and that there

were no wrong answers. Participants were found via university forums and mailing lists (researchers or lecturers with children), social media and the local library. All participants' personal data were stored securely, and all personally identifiable data were removed.

7.2.3.7 Limitations

The same limitations from section 4.3.3.8 apply for this study.

Another limitation of this study is that some participants took part in both Study #3 and the current study, which could have introduced familiarity effects. However, this overlap was minimal as only four children (Child1, Child9, Child17, Child26 from this Study #7 and Child22, Child7, Child36, Child21 from Study #3) participated in both studies. Moreover, the three-month gap between the studies and the qualitative data provides no evidence of any learning effects.

7.2.4 Results

To navigate categories and themes, please find table 7.2. Moreover, table 7.4 presents a summary of the statistical results of the quantitative data. *Note that children from this study are referred to with new participant IDs which do not correspond to the same children from study #3.* Where available, qualitative data is paired up with corresponding quantitative data.

Table 7.2: Qualitative Findings - Categories, Themes, Sub-Themes with Counts, gathered from child interviews at the end of the user study.

Categories	Themes	Sub-Themes	Count
Category 1: Perceptions in Simulated social VR Group Harassment of Interventions from Manual to AEM	**Theme 1.1:** Group Harassment Scenario Validity		28
	Theme 1.2: Emotional and Psychological Impact of Interventions from Manual to AEM		17
Category 2: Attitudes Towards the Proposed Interventions from Manual to AEM	**Theme 2.1:** Positive Aspects	Usability, Interactive and Creative Design	22
		Interventions Create Positive Emotional Valence and Relief	3
		AEM's Behaviour and Actions	32
	Theme 2.2: Negative Aspects	Balancing Fun with Practicality	13
		Usability Issues	12
		AEM Understanding Context and Accuracy	11
Category 3: Embodiment of AEM (Physical Form, Speech, and Behaviour)	**Theme 3.1:** AEM's Physical Form		26
	Theme 3.2: AEM's Voice and Communication		25
	Theme 3.3: AEM's Movement and Gestures		21
Category 4: Child Needs Regarding AEM's Involvement and Sense of Agency	**Theme 4.1:** AEM Involvement in Helping [**strong assistance**]		8
	Theme 4.2: Agency and Collaboration with AEMs [**medium assistance**]		30
	Theme 4.3: Child wants Full Agency and Control [**low assistance**]		6

7.2.4.1 Category 1: Perceptions of Interventions, From Manual to AEM

Theme 1.1: Group Harassment Scenario Validity Similarly to study #3, I assessed how the child user felt in the group harassment scenario from a third-party perspective. The harassment scenario was largely perceived as valid by participants, eliciting negative emotional and empathetic responses (N=16) (See Limitations section 7.2.3.7). Many experienced negative valence emotions due to the group harassment situation, highlighting the emotional impact of being targeted by multiple harassers (N=4). Some children were able to put themselves in the victim's shoes (N=6), where some answered the questions using "I" when talking about the victim being bullied (N=2), while others did not feel a similar level of empathy (N=6). A few children also described themselves as more of an observer than an active victim (N=6), using phrases like "If I were in such a situation, I would feel ..."; *"I would be really sad because I would be bullied by people." [Child3].; "If I were that child, I would be quite upset." [Child8-A].* A child took on a protective role, acting as a defender of the victim (N=1).

The scenario's realism was noted by several children (N=4), with one child finding the prototype harassment scene particularly realistic (N=1). The sense of being outnumbered in the bullying situation was described as intimidating (N=4), although one child clarified that they did not experience physical harassment, only verbal (N=1).

These varied responses reflect the scenario's ability to evoke both emotional and empathetic reactions, while also highlighting differences in how children perceived their involvement (see Limitations section 7.2.3.7 reflecting on methodology's validity).

Theme 1.2: Emotional and Psychological Impact of Interventions from Manual to AEM The AEM was perceived as a supportive presence by some children, with one viewing it as a friend (N=1): *"I felt left out and no one was helping, the moderator was my friend." [Child10].* This sense of companionship and safety extended to how children felt during the experience. Many felt reassured and safe knowing that the AEM were there to help them (N=5): *"I felt safe because the moderator was right in front of me, he would see if I needed help or not." [Child8]*, *"I felt left out and no one was helping, the moderator was my friend." [Child10].* The tools would help (N=2), and several children reported feeling safe simply by seeing the AEM (N=2). The belief that the bullies would not return or that they were effectively blocked contributed to their sense of safety (N=4). *"I felt safe because they were blocked and couldn't come back." [Child5-B].*

Other factors contributing to this feeling of safety specifically included the muting of their voices (N=2) and the physical absence of the bullies (N=1). Some children expressed happiness in response to the tools and the AEM's assistance (N=3): one child felt happy knowing the AEM would help protect them, while another found joy in using the wand to successfully defend themselves. Additionally, a child felt a sense of happiness from helping others within the virtual environment. Children (N=2) also reported feeling calm and less stressed due to the AEM's help in (C4).

Finding 1: Children generally found the harassment scenario realistic and emotionally impactful, with varied responses ranging from empathy and identification with the victim to more detached, observer roles. The AEM interventions fostered feelings of safety, reassurance and led to calmness and reduced stress.

7.2.4.2 Category 2: Attitudes Towards the Proposed Interventions from Manual to AEM

Theme 2.1: Positive Aspects Usability, Interactive and Creative Design

1) Fun and Interactive Features Create Positive Emotional Valence and Relief: Children found the wand *[PointBlock]* (N=10) and calling the AEM fun to use (N=2). The ability to block in a playful and innovative way made the tools both effective and enjoyable. *"It's quite fun because then you can kind of you just get to get rid of the bullies, just like at the touch of your fingers, it's easy to use." [Child23-A]*; *"I love the wand. I live and breathe for the wand." [Child16].*

Children (N=8) appreciated the metaphor of "shooting to block", which made the process of managing harassers feel more dynamic and engaging. This design added a playful element to an otherwise stressful situation, turning the act of blocking into a more enjoyable interaction. A child mentioned being reminded of a video game played *"Point and shoot like every video game I've played." [Child1]*, while another child felt like the wand was an effective retaliatory feature *"What I like about the wand is if you're getting bullied, you can shoot them." [Child10].*

This is supported by the quantitative data. There was a significant difference in terms of enjoyment (see table 7.4), with children enjoying more [C1: manual] compared to [C2: Call AEM] (post-hoc test adjusted with Tukey $p<.0001$) and [C4: fully automated] (post-hoc test adjusted with Tukey $p<.0001$). Children also enjoyed [C3: AEM gives options] compared to [C4: fully automated] (post-hoc test adjusted with Tukey $p=0.0114$). These results suggest that the more interactive and controllable the experience, the higher the enjoyment, with children preferring conditions that offer more direct involvement rather than passive or fully automated interactions.

Interventions Create Positive Emotional Valence and Relief *2) Relief* Children also expressed feelings of happiness and relief when the AEM successfully intervened in blocking bullies (N=2). For instance, one child [ChildSCA] noted feeling happy when the AEM took action, while another reported joy upon receiving help from the AEM.

3) Immediate, Responsive and Practical: A key usability feature noted by the children was the tools' immediate responsiveness (N=15). The tools worked promptly, allowing them to receive unprompted support in C4 (N=4) *"Regarding [C4: AEM fully automated] it was good because he just came in and helped out without having to be asked or anything." [Child1]*; *"You can just sit back and watch the show." [Child6 - (C4: AEM fully automated)]* and to block harassers quickly (C2,C3,C4) (N=8). This quick reaction led to greater satisfaction with the tools' performance. Simplicity was a significant factor in how children perceived the usability of the tools. They appreciated the straightforward design in (C1, C2, C3) (N=11). They also noted they appreciated (C2 and C3) as it did not require much effort to block the bullies (N=3) *"I think [(C2) calling the AEM] is quite good as well because you just call when you want it and the bullies are gone, it's nice and easy." [Child23-A]; "Just clicking a button and *poof* they're gone is quite helpful." [Child5-A - (C3: AEM gives options)].* Children liked the tools being practical and satisfying rather than just fun (N=2).

This is also supported by the quantitative data. From a one-way ART ANOVA with the independent variable as the interventions, there was a significant difference in terms of difficulty to use (see table 7.4), with children finding more difficult [C1: manual] compared to [C4: AEM fully automated] (post-hoc test adjusted with Tukey $p=0.0047$).

As described in the qualitative findings above, this is due to the increased cognitive load and physical effort required in the [C1: manual] condition, which may have made it feel more challenging compared to the [C4: AEM fully automated] condition, where the system handles the blocking of a large group.

AEM Behaviour and Actions *4) AEM Giving Children a Choice:* Children particularly expressed positive feedback with (C3: AEM gives options), where, rather than imposing actions, the AEM allowed children to participate in the decision-making process of blocking (N=19). *"I like it because you actually get to do something, but I do like that it does quite a lot of stuff for you." [Child3].* They valued having multiple options available, allowing them to tailor their response based on the context, enhancing their sense of control (N=17), feeling empowered and more confident (N=2).*"It feels like you have more power over the situation." [Child20-A].* Offering more than two choices was also seen as beneficial, providing children with a range of responses that could be adapted to different scenarios (N=1). The ability to choose from multiple options aligned with children's desire for independence. *"That's good, it has options because I get to choose if I want the wand or the AEM. I like having a choice." [Child5].* Even if they may not always use the options, having them available made them feel more in control (N=1).

5) Perception of AEM's actions and help: Children appreciated the dynamic presence of the AEM during bullying incidents (N=3). The balance between visibility before blocking and removal after blocking was appreciated. Specifically, a child felt comforted seeing the AEM during the harassment situation *"It was helpful because I saw it there before I clicked and I knew I could just call him." [Child5-B]* and others felt reassured when the AEM moved into the scene to detect and address the bullies, but also valued its subtle disappearance once the situation was resolved (N=2), *"It was just there, moving around. I think it was fine because it disappeared after." [Child4].*

6) Children Wanting to be Helped by AEM: Children appreciated when the AEM was proactive and took action on their behalf in situations where they felt scared or overwhelmed (N=3). One child noted that the automatic appearance of the AEM made them feel like it understood their emotions without them having to express them *"I felt really good because I felt like the moderator [AEM] knew how I felt." [ChildRAC].*

Finding 2: 1) Children enjoyed tools like the wand, especially the "shoot to block" metaphor, finding it playful and effective. 2) Successful AEM interventions in blocking bullies brought relief. 3) Interventions were valued for their quick response, ease of use, and low effort, increasing satisfaction and reducing difficulty compared to manual blocking. 4) C3 (AEM gives options) was appreciated for allowing children to decide how to respond, fostering a sense of empowerment. 5) Children liked the AEM's visible and dynamic presence during harassment, and its subtle withdrawal once the situation was resolved. 6) The AEM's ability to act without being asked (C4), especially when overwhelmed, made children feel understood and supported.

Theme 2.2: Negative Aspects Balancing Fun with Practicality

1) Importance of Fun: Children highlighted the importance of fun in the interventions designed for blocking the group of harassers (N=7). Some found the AEM's blocking actions monotonous and less engaging in C2 and C4 (N=6) *"I don't think it [C4: AEM fully automated] was that fun because it was just over and done with quite quickly." [ChildSCA]; "It [C2: call AEM] is not really fun because it kind of does everything for you, but it's nothing really wrong with it." [ChildSAM]*, with feedback indicating that the process was not enjoyable compared to the excitement of actively seeking out and blocking harassers with the wand *[PointBlock]* (N=3) *"I kind of felt like looking for them was kind of more fun." [Child5-B].*

2) Sense of Agency: They expressed dissatisfaction with tools where the AEM handled all actions, leaving them feeling passive and disengaged (N=2). Tools that allowed children to actively participate in the blocking process and make their own decisions were preferred. *"I like that it notices, but I would rather do it myself, or else I feel like I would just never stand up, it would just be someone else coming in and doing it for me." [Child18].*

3) Social Acceptability: The social acceptability of tool interactions was also important. A child was concerned about the appropriateness of certain actions, like shooting avatars to block them, which they felt might be perceived as inappropriate or uncomfortable (N=1), *"I think it's nice just having a wand but it's but it's also kind of like weird just tapping on people and they just go away like they're being blocked." [Child2].* The credibility of the AEM and its actions also played a role, with a child finding the AEM's appearance or behaviour to be "silly" (N=1).

4) Interactive Tool Overload: Excessive or complex options could lead to confusion and stress. A child reported feeling panicked by having to choose and preferred tools that balanced interactivity with ease of use: *'When the options came, yes, no and wand, so I was panicking, so I was just panicking so I quickly chose the wand." [Child8].*

Usability Issues *5) Missing Options or Features:* Children noted several issues related to missing options and features (N=3). A child expressed a desire for more options to manage the group harassment scenario. The AEM's inability to offer options before its intervention (C2 and C4) was problematic, as it might block all bullies when the child might have preferred to block only certain individuals (N=2). *"I don't think I want to instantly block all of them, I may want to block some of them." [Child26 - C4: AEM fully automated]*; *"I wish it offers like options. I want to call it so it comes up and says "do you need help yes or no". Then like after it if you choose no you have the wand." [Child - C2: Call AEM]. 12) Undo actions:* The lack of options to confirm or undo actions was also mentioned, especially in the case where the AEM does everything in C4 (N=1).

6) Confusing Design or Functionality: Several aspects of the interventions were confusing for children (N=4). There were difficulties in understanding how to use the tool effectively (N=1) and issues with the functionality of the 'no' button in C3, which did not seem to make sense to them (N=1 [Child10]). Additionally, the overall process of the AEM doing everything was strange to them, requiring time to grasp what was happening (N=1). A child found it confusing to figure out how to block the bullies (N=1, [Child12-A]): *"I found it strange, like it just came and I was confused what was happening." [Child12-A].*

7) Speed of Response: The speed of the intervention's responses was another concern (N=7). The intervention's speed was noted to be either too slow (using wand C1) (N=3), delaying the blocking process, or too fast (C4), not allowing enough time to think (N=2). *"Before you could actually block them, it took a while to actually block, so sometimes it could be a hit or a miss." [Child17 - C1: manual].*

8) Visibility Leading to Persecution: The discretion of the AEM was a concern (N=2) *"I feel like that one might be a bit more up to like the discretion of the moderator and I'm not sure how I feel about that." [Child23-A - (C4: AEM fully automated)].* A child felt that calling the AEM could potentially lead to further bullying rather than offering protection (N=1) *"If you're the child, not really because I think I might give other people a chance to bully me more." [ChildRAC - (C4: AEM fully automated)].*

AEM Understanding Context and Accuracy

9) False Positives: Children expressed concerns about the AEM's ability to grasp the context of interactions (N=11), particularly with nuanced or sarcastic remarks. More specifically, children were concerned that the AEM might mistakenly block (in C2, C4) innocent users, initiating blocks in non-harassment situations, and confuse friends' behaviours with bullying (N=5). *"The moderators might think oh they're bullying me but then they block a random person." [Child16].*

10) AEM Restorative Actions: A child mentioned that the AEM could take an action that would give the opportunity to the bullies to stop before taking direct action to block them: *"It was good, but maybe it should give the bullies some kind of other chance." [Child2 - (C4: AEM fully automated)].*

Finding 3: 1) AEM-led blocking (C2, C4) felt less engaging than using the wand (C1). 2) Automation reduced participation; children preferred making their own decisions. 3) Some children were concerned over "shooting" avatars, AEM credibility, and appearance. 4) Too many options (C3) can cause stress; balance is needed between choice and ease. 5) Children noted missing features, such as undo actions. 6) Some found certain tools and buttons unclear, or were confused by the AEM actions without explanation. 7) Responses could be too slow (C1) or too fast (C4), limiting effectiveness or decision-making time. 8) Calling or perceptibly using the AEM could risk further bullying and would need more discretion. 9) Children were concerned that the AEM might misinterpret situations, block innocent users or mistake friendly behaviour for bullying. 10) AEM should give warnings first before blocking.

7.2.4.3 Category 3: Embodiment of AEM (Physical Form, Speech, and Behaviour)

Theme 3.1: AEM's Physical Form The physical form of the AEM was an important factor in its acceptance and effectiveness (N=26). Many children found the proposed physical form of the AEM to be suitable (N=14) and non-intimidating (N=3), which helped them feel comfortable and it made sense to them (N=1). A child mentioned they were confident about the AEM's ability to block the harassers (N=1).

Visibility and Vibrancy: Some children mentioned wanting the AEM to be more visible (N=8), by adding colours (N=6), making it distinct from other players to make it easier to spot (N=2) and less bland (N=1). *Friendly Appearance:* The AEM's appearance was also assessed based on its friendliness. Some children would prefer a more friendly and engaging form compared to a simple orb (N=2). *Human vs. Non-Human Forms:* Preferences varied regarding whether the AEM should have a human-like appearance or a non-human form. Some children would prefer a realistic human-like form, to help with relatability (N=4). Conversely, others found that a non-human, animal-like form could also be engaging and suitable (N=3). The choice between human-like and non-human forms highlights the need for flexibility in design to cater to different user preferences.

Theme 3.2: AEM's Voice and Communication AEM's voice was found to be clear (N=4), although a child was confused by what the AEM said. Some children found the voice too monotone and wished for a more engaging tone or of same gender as the child (N=3). Preferences for the AEM's voice varied, with some wanting it to sound more human-like (N=1), while others were comfortable with its robotic nature (N=20). The recognition of the AEM's voice was compared to familiar voices like "SIRI" (N=1 [Child5-A]) and "news announcers" (N=1 [Child23-B]).

Theme 3.3: AEM's Movement and Gestures Children appreciated when the AEM's actions, like moving towards the user to help the child, were helpful and in line with its purpose (N=15). However, some found the AEM's approach too direct or abrupt, preferring a more gentle or non-intrusive manner (N=2). Some children would want the AEM's behaviour to fit its appearance (e.g., if looks like a frog, jumps like a frog, if looks human, walks like a human) (N=3) *"If it looked a bit like an animal, maybe let's say frog, it would be like a bit jumping." [Child6].* A child suggested incorporating AEM non-verbal expressions (e.g., thumbs up).

Finding 4: Many children found the AEM's design suitable and non-intimidating. Suggestions included making it more visible and colourful, and preferences varied between friendly, human-like forms for relatability and non-human/animal-like forms for engagement, highlighting the need for customisable design. The AEM's voice was clear, but for some, it was too monotone. Children would want a more engaging voice. Helpful and purposeful movement was valued, though some wanted a gentler or less intrusive approach. Children suggested behaviour should match appearance, and proposed adding non-verbal cues.

7.2.4.4 Category 4: Child Needs Regarding AEM's Involvement and Sense of Agency

Theme 4.1: AEM Involvement in Helping [Strong Assistance] Some children desired strong assistance from the AEM (N=8), valuing its proactive role in blocking bullies and managing disruptions. They would prefer scenarios where the AEM would automatically step in to provide help and intervene when needed (N=7) *"As much as I could get because you never know what [the bullies] would be doing." [Child1].*

Theme 4.2: Agency and Collaboration with AEMs [Medium Assistance] The majority of children would want to maintain their sense of agency while still benefiting from the AEM's capabilities at different levels of involvement (N=23). Children mentioned wanting to have the choice and be given options to choose from, like in condition 3 (N=9). A child suggested a setting to choose how much assistance they received from the AEM, allowing them to balance between taking action themselves and relying on the AEM's support depending on the child preferences and needs, which would make the children feel calmer (N=2) *"I like having it, but I wish you could do some of it and some not. I could do half, and it'll do the rest because it's easier, I'd feel calmer. [...] The moderator picks how much you need: a quarter, half, or however much." [Child10].* Another way of AEM's involvement and collaboration was suggested, where the AEM would help them by showing them how to block the users or by blocking the users alongside them using the [PointBlock]. *"It feels better to have someone who understands how you felt and helps you. I just want it for some things, like to tell me how to block them." [ChildRAC].*

Children also indicated a preference for first invoking the moderator and then receiving options (N=7). This approach allows them to initiate help and subsequently explore different choices for managing the situation, giving them more agency *"You get to decide and also get to like decide what to do. I think that would probably be a very good idea." [Child16].*

Theme 4.3: Child wants Full Agency and Control [Low to No Assistance at all] Some children want to stand up for themselves (N=3) and have no help from the AEM (N=4) *"I like the wand the most because it feels like I'm the one taking action instead of just sort of something else swooping in." [Child18]; "I could usually just stand up for myself. I want to have control. If they were really mean, then I would want help, but if I was just there and they weren't that mean, I would just do it by myself or just walk away." [Child2].* Some still mentioned wanting the possibility of getting help if the situation is severe (N=2). However, a teenager [Child12-A] considered the AEM more suitable for younger children.

Finding 5: The majority of children preferred collaborative support, maintaining agency while choosing the level of AEM assistance. Others wanted either strong proactive help or full control, with the option to seek help only in severe situations.

7.2.4.5 Likert Scales and Statistical Results

Ranks The final ranks table 7.3[1] based on the weighted scores for the conditions are as follows: (C1) [manual] scored the highest, followed by (C3) [AEM gives options], (C4) [AEM fully automated] and (C2) [Call AEM]. (C1) [manual] was the second least favourite, and (C3) [AEM gives options] was ranked first as the second favourite. In the qualitative findings, the majority of children (N-23) would still want medium assistance (see section 7.2.4.4), and the Likert scales showed C3 has the highest percentage of "Very Happy" (38%) and "Completely Safe" (38%) after bullies were blocked (see figure 7.5). This suggests that while children preferred the manual tool overall, they still valued some level of assistance, highlighting their desire for a sense of agency and acknowledging the need for support when necessary.

Table 7.3: Rankings and favourite children's interventions with score breakdowns and total weighted scores. The ranking is based on a scoring system where favourite votes are worth 4 points, second favourites 3 points, third favourites 2 points and least favourites 1 point, with the total weighted score being the sum of these points. Ranked as top 1 is (C1) [manual], followed by (C3) [AEM gives options], (C4) [AEM fully automated], and last, (C2) [Call AEM].

Rank	Condition	Favourite Score	2nd Favourite Score	3rd Favourite Score	Least Favourite Score	Total Weighted Score
1	**C1: manual**	18 x 4	5 x 3	2 x 2	5 x 1	**96**
2	**C3: AEM gives options**	7 x 4	10 x 3	8 x 2	5 x 1	**79**
3	**C4: AEM fully automated**	4 x 4	9 x 3	4 x 2	13 x 1	**64**
4	**C2: call AEM**	1 x 4	6 x 3	16 x 2	7 x 1	**61**

Finding 6: Children ranked manual tools (C1) highest overall, followed by AEM gives options (C3), AEM fully automated (C4), and Call AEM (C2). However, C3 had the most "Very Happy" (38%) and "Completely Safe" (38%) ratings, indicating a preference for agency with supportive options.

Likert Scales The Likert scale results for emotional responses and usability in each condition are presented figure 7.4 and figure 7.5. All four interventions had a statistically significant impact on emotional responses after the bullies were blocked (see table 7.4). C3 appears to have the highest percentage of "Very Happy" (38%) and "Completely Safe" (38%) after bullies were blocked (figure 7.4). Moreover, C4 appears to be the most helpful, C1 the most enjoyed and C4 the easiest (figure 7.5).

[1]Children's overall ratings are based solely on preference, reflecting how much participants liked each option, rather than perceived effectiveness or emotional perceptions.

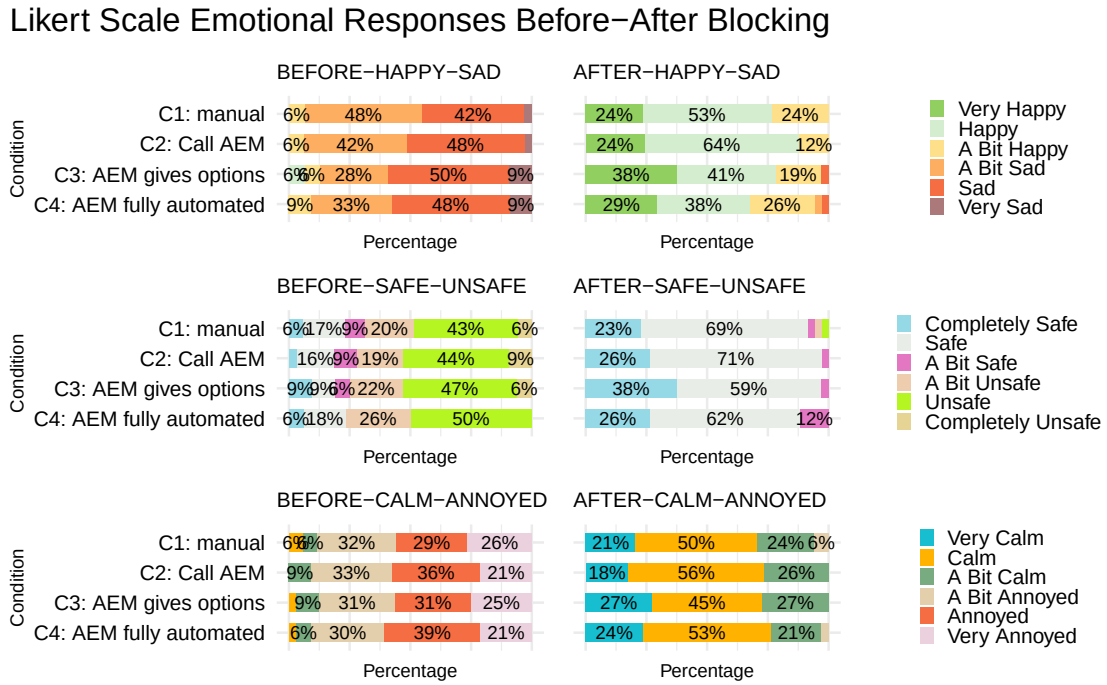


Figure 7.4: Emotional Responses Likert Scales showing results from child participants for each condition. Responses were provided for both before and after the bullies were blocked. (Colourblindness note: Legend labels are in the order of blocks from left to right on the plots.)

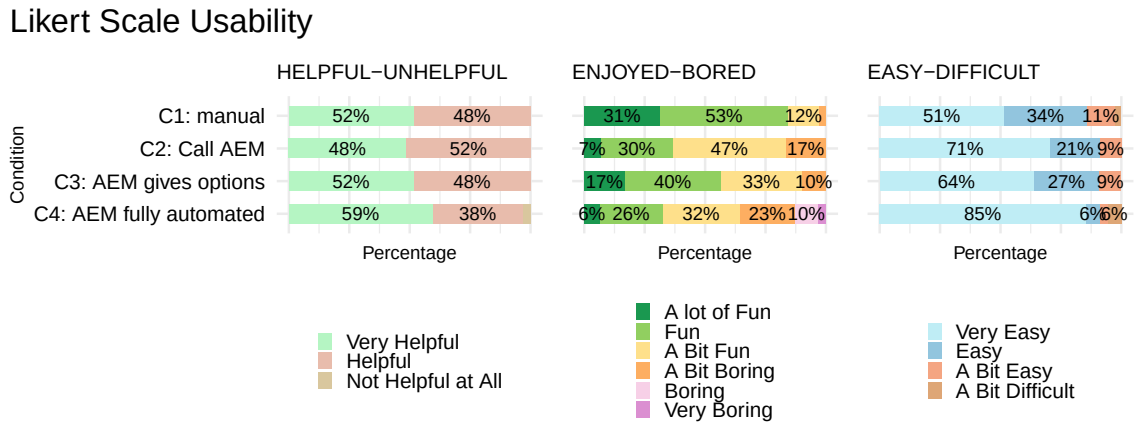


Figure 7.5: Usability Likert Scales showing results from child participants for each condition.(Colourblindness note: Legend labels are in the order of blocks from left to right on the plots.)

Statistical Results A summary of the statistical tests with the significant results is in table 7.4. Non-statistically significant results are in the Appendices section C.2.

Table 7.4: One-way ART ANOVAs results 1) with the independent variable as the Interventions (C1, C2, C3, C4) 2) with the independent variable as Time (Before VS After bullies blocked). (Small effect: <0.01, Medium effect: ≈0.06, Large effect: >0.14)

1) Independent Variable: Interventions

Dependent Variable	Effect Size (η_p^2)	F-statistic	p-value
DIFFICULT-EASY Scale (1-6)	Medium (0.11)	$F(3,99) = 4.1$	$p = 0.008$
BORED-ENJOYED (1-6)	Large (0.3)	$F(3,90) = 13.1$	$p = 3.5e^{-7}$

2) Independent Variable: Time Component (Before-After Blocking)

Dependent Variable	Effect Size (η_p^2)	F-statistic	p-value
SAD-HAPPY Scale (1-6)	Large (0.75)	$F(1,228) = 685.9$	$p < 2.22e^{-16}$
CALM-ANNOYED Scale (1-6)	Large (0.78)	$F(1,233) = 802.75$	$p < 2.22e^{-16}$
UNSAFE-SAFE Scale (1-6)	Large (0.63)	$F(1,233) = 390.81$	$p < 2.22e^{-16}$

Summary tables of quantitative results can be found in section C.1.

7.2.4.6 Usability

From a one-way ART ANOVA with the independent variable as the interventions, there was a significant difference in terms of difficulty to use (see table 7.4), with children finding more difficult [C1: manual] compared to [C4: AEM fully automated] (post-hoc test adjusted with Tukey p=0.0047). The qualitative data in Theme 2.1 - Immediate, Responsive and Practical is in line with these results.

7.2.4.7 Enjoyment

There was a significant difference in terms of enjoyment (see table 7.4), with children enjoying more [C1: manual] compared to [C2: Call AEM] (post-hoc test adjusted with Tukey p<.0001) and [C4: fully automated] (post-hoc test adjusted with Tukey p<.0001). Children also enjoyed [C3: AEM gives options] compared to [C4: fully automated] (post-hoc test adjusted with Tukey p=0.0114). This is in line with children's feedback on how they experienced the tools (see Theme 2.1 - Sub-theme: Fun and Interactive Features & Theme 2.2 - Sub-theme: Importance of Fun).

Finding 7: Intervention type significantly affected ease of use (medium effect, C1 more difficult than C4) and enjoyment (large effect, C1 and C3 more enjoyed than C2 and C4), while blocked bullies led to large increases in happiness, calmness, and safety across all conditions.

7.2.5 Discussion

Reflection around the AEM and future work are provided in chapter 8 to situate this study's findings within the wider contributions of the thesis. Here we present a summary of the findings and references to corresponding guidelines in chapter 9.

7.2.5.1 Summary of Findings

Similar to Study #3, the simulated group harassment scenarios elicited a clear emotional impact (section 7.2.4.1).

The study explored **S7-RQ1: Impact of Automation Level**, revealing that children significantly enjoyed C1 (manual) and C3 (semi-automated) more than C4 (fully automated) (see table 7.4). Moreover, C1 emerged as the most preferred option, followed by C3 (see table 7.3). Qualitative data showed that C1 was praised for its hands-on, interactive approach to block (see section 7.2.4.2). However, C3 was the second most preferred tool while C1 ranked lowest in second preferences and only four expressed a preference for minimal to no support in the interviews (see section 7.2.4.3). The majority of children, 23 out of 34, expressed a desire for medium assistance (see section 7.2.4.3). The combined quantitative and qualitative findings highlight a dual preference between manual, engaging tools and AEM assistance while being involved in the decision-making, which underscores the necessity of designing systems that cater to varying levels of control and assistance based on child preferences and needs.

Based on these findings, I gather practical design guidelines in section 9.2, specifically guidelines 7 and 14.

7.2.5.2 Children's Participation with Semi-Automated AEMs

As outlined in the related work (see chapter 2), the 1989 UNCRC affirms that children have the right to express their opinions [56]. The increasing use of digital media by children at very young ages and the rise of automated or AI-driven moderation [107, 101], including the design of AEMs, must therefore take these rights into account.

Moreover, stakeholders in this thesis highlighted the importance of empowering victims and ensuring meaningful human-in-the-loop involvement in automated interventions, in study #2 (Expert P8 noted: *"As children get older or grow in independence and maturity, then naturally they're looking to make their own decisions and have their own autonomy over what they do."*), study #3 (section 4.3.4.4), studies #5 and #6 (item 3, section 6.3.7.2).

Although an AEM, fully automated, may be especially helpful in situations where a child has limited maturity or agency to take action, findings from this study indicated that most children prefer to participate in decision-making. Mainly, condition C3 was preferred, where the AEM would propose possible courses of action. This preference aligns with the top five rungs of Roger Hart's Ladder of Children's Participation [352], specifically the rung "Adult-initiated, shared decisions with young people". Recognising that agency is not fixed but evolves with time, context, and social interactions [352], I therefore propose that AEMs incorporate adjustable settings of level of automation (see section 8.1.2), with C3 as the default. However, I should note that the desire for

agency from children might not always align with their capability for agency. Future work can investigate the latter to better understand how those levels of automation can be adjusted not only based on the desired agency but also based on the capability of children to have agency.

Ethical Implications of Semi-Automation. While AEMs may grant children some level of participation, they also pose risks, including inaccuracies in detecting true negatives (i.e., failing to recognise disruptions or issues when they occur). Additionally, a lack of comprehension from the AEM or children can result in unintended consequences, potentially leading to unintentional harm.

7.2.6 Conclusion

In this study, which explores manual (based on findings from study #3, semi-automated, and automated interventions, I demonstrate the importance of designing tools that provide children with control over their protection while offering assistance from the AEM when needed. This includes giving children the ability to choose how and when they interact with the AEM, such as selecting between manual or automated interventions, or determining the level of assistance provided, thereby promoting a sense of agency and empowerment in social VR when facing group harassment.

Transition to Next Study

I have now explored, from a children's perspectives, how varying levels of child initiation and decision of intervention alongside the AEM in a simulated social VR disruption, impact children's sense of agency, safety and emotional valence. In the next study, I aim to understand real-time context-driven parental involvement in practice. I explored how varying levels of involvement affect parents' reassurance and sense of control, as well as children's sense of safety following the child's experience of a simulated social VR disruption.

7.3 Study #8 - Context-driven Parental Involvement for Children in Social VR via AEM-app: Balancing Safety and Privacy

7.3.1 Motivation

Studies #4, #5, #6 and #7 in Chapter 6, highlight the potential of AEMs as a safeguarding method for enhancing child safety in social VR, particularly given that human moderators cannot be present 24/7 nor detect all incidents [19] and that common social VR platforms, such as VRChat [204] and Rec Room [213], offer limited parental controls. While parents acknowledged the usefulness of AEMs in interviews, they emphasised their desire to be more involved, particularly by receiving real-time notifications about disruptive incidents (study #4). Experts in psychology and online child safety recommend a degree of parental involvement in real-time when an incident occurs, via notifications (studies #5 and #6), acknowledging that parents are often

busy and may not have direct access to the VR environment (also mentioned in prior work section 2.2.4). Consequently, any safety-enhancing companion tool should provide meaningful insight and oversight without resorting to intrusive real-time monitoring, such as continuously streaming the child's social VR activity video feed.

In this study, I designed a four-condition VR study (N=34 parent-child dyads) that varied parental involvement using a custom-built mobile app to receive real-time notifications and child support in a social VR simulated game scenarios based on prior work (see study #4): no support, AEM-only support, parent-written messages spoken by AEM, and parent voice messages delivered via an in-VR avatar.

I aimed to answer the following research questions:

S8-RQ1 How do parents perceive the different levels of involvement via a social VR companion app, considering their availability level and provide support to their child in social VR?

S8-RQ2 How do the varying levels of parental involvement affect children's emotional perceptions, sense of safety and overall experience when supported after simulated social VR harassment?

7.3.2 Related Work

This section synthesises related work on the specific topics addressed in these studies. (A more comprehensive literature review is presented in chapter 2, with specific relevant sections for context: section 2.3.2, section 2.3.3, section 2.2.4; and for methodology: section 2.4.2, section 2.4.3).

7.3.2.1 Challenges of Child Safety and Parental Awareness in Social VR

Unlike traditional online platforms, social VR is experienced through headsets that isolate the user from their physical environment. This occlusion makes it even more difficult for parents to oversee or participate in children's experiences [39]. The invisibility of these interactions creates a unique challenge for parents who are unfamiliar with social VR. Unlike co-watching or sitting beside a child using a tablet, immersive VR offers little to no ambient cues, leaving parents unaware of who their children are interacting with or what they are experiencing. In Study #1, one participant (P13) noted that the headset's immersive nature can hide age-inappropriate situations from parents, allowing children to encounter them without parental awareness [2]. Similarly, in another recent study investigating ethical concerns and considerations of children's VR usage with families, a parent also mentioned the challenges faced in the way to monitor children's VR usage because of the immersive nature *"With a mobile phone we can just glance over his shoulder, but with VR, we can't peek through the headset to see what he's up to."* (P8) [210].

These blind spots lead to gaps in parental mediation and undermine opportunities for support or intervention.

7.3.2.2 From Existing to Innovative Solutions for Parental Involvement in Social VR

Existing Safety Features for Families To address these challenges, social VR platforms have implemented safety tools like blocking, personal space bubbles, muting, and reporting [187, 205, 204], to protect users from disruptive individuals, but are not always adequate for children and are not standardised across platforms [9]. They place responsibility on children, who might lack the necessary understanding or familiarity with the technology [122, 114, 363, 115]. Along with existing mitigation tools, there are human moderators to address incidents in real-time in social VR [31, 20]. However, only 24% of incidents observed in a study in social VR were addressed by human moderators [19], highlighting the lack of scalability for human moderators in a growing VR ecosystem.

VR headsets (e.g., Meta Quest) support casting [209], meaning the VR headset's live view can be mirrored to another device, allowing others to view it in real time. However, relying on casting for parental supervision is cumbersome. A prior survey [210] showed that parents find current safety features insufficient for flexible oversight and control, and although casting exists, maintaining constant observation is difficult in practice. Parents also reported wanting low-effort, non-intrusive ways to stay aware of their child's VR use without needing to put on a headset themselves [210]. This study directly responds to that need.

Some platforms offer standard parental control features and junior accounts. For instance, Meta Horizon allows parents to supervise their child's activity through a Family Centre dashboard, where they can approve follower requests, manage privacy settings, monitor screen time, and restrict communication [212]. Rec Room provides linked junior accounts that enable parents to control spending and disable in-app purchases [213]. Despite these measures, there is a clear lack of consistent and robust parental control tools across platforms. Notably, VRChat [25], one of the most widely used social VR environments [214], offers no parental controls at all, leaving children potentially exposed. Moreover, even when tools are available, their effectiveness heavily relies on active parental engagement, consistent use, and digital literacy, factors that vary significantly between households. This highlights a broader gap in the design of social VR safety systems: the need for more accessible and parent-friendly tools that do not assume a high level of technical skill or availability.

Towards Innovative Solutions to Safeguard Children in Social VR AEMs, embodied AI-driven agents that take action to mitigate harmful interactions, are a promising alternative solution to safeguard children in Social VR (studies #4 [4], studies #5 and #6, [6]). Parents were interviewed regarding how they envisioned the AEMs, and while parents noted its usefulness, they mentioned wanting to be involved and to be notified in real-time about the disruptive situations (see study #4). Another study with parents, children and child safety and child online safety experts, investigated AEMs' potential to safeguard children in social VR (studies #5 and #6, [6]). While participants highlighted AEMs' benefits, including immediate actions, presence and victim support, experts and parents suggested that parents should be involved, have some degree of control (e.g., overriding AI decisions through settings) and receive real-

time notifications about significant events within the environment (studies #5 and #6, [6]). Guardians also noted that AEMs can serve as an intermediary between parents and children, facilitating conversation and strengthening the parent–child relationship (studies #5 and #6, [6]). This desire for a balance of automation with human oversight and involvement reflects a broader parental need for external intervention capacity in immersive environments.

7.3.2.3 Gap and Proposed Research

As children increasingly engage with social VR platforms, where incidents of harassment can occur while users are occluded from their surroundings through VR headsets, parental oversight becomes both more difficult yet more necessary. Unlike traditional 2D screens, VR limits a parent's ability to observe or intervene, and many parents remain unaware of the risks involved (see section 7.3.2.1).

The current challenges of safety features (see section 7.3.2.2) and limited parental awareness of social VR reveal a significant gap: the absence of accessible, effective real-time parental mediation tools that are both parent-friendly and respectful of children's privacy.

In this study, I explored varying degrees of parental involvement (see section 1.2) from both the parent and child perspectives. I propose an innovative companion app that alerts parents to potentially disruptive behaviour while also examining how children perceive such support in terms of agency and emotional safety.

7.3.3 Methodology

Using an AEM to mediate parental interactions, I examined how varying levels of involvement affect parents' reassurance and sense of control, as well as children's sense of safety following the child's experience of a simulated social VR disruption (the same disruption scenario from study #4).

This study was conducted in February-March 2025 with **66 participants, 34 parent-child pairs**.

7.3.3.1 Study Design

This study follows a within-participant design with one independent variable consisting of four conditions, each resembling a level of parental involvement, from the parent perspective (see figure 7.6) and the child perspective (see figure 7.7).

(C1) [**No Support**]
- Child Support: No Support.
- Parent: Only Informed (Sound Alert, Text Notification).

The child experiences a disruption when another player destroys their tower. No support is provided by either the AEM or the parent. The parent receives a text notification: *"Your child has been disrupted by another player."*.

(C2) [**AEM Support Only**]
- Child Support: Supported by AEM.
- Parent: Sound Alert, Text Notification & Video Clip.

The AEM supports the child during the disruption by informing them: "What the other player did is wrong. They should not have done that. I have notified your parent." The parent receives a notification: *"Big Buddy: Your child has been disrupted by another player. I have supported your child after the incident."* A video recording of the event is also made available.

(C3) [C2 + text message: Indirect Parental Support via AEM]

- Child Support: Supported indirectly by the parent through AEM.
- Parent: Informed with Sound Alert, Text Notification & Video Clip; Sends Message.

The AEM is present, and the parent is notified with: *"Big Buddy: Your child has been disrupted by another player. How can I support your child? I will pass on your message."* A video recording of the event is available. The parent can type a message, which the AEM then reads aloud to the child: *"Hi, here is a message from your parent: [message]."*.

(C4) [C2 + audio: Direct Parental Support]

- Child Support: Supported directly by the parent avatar in the VR scene.
- Parent: Informed with Sound Alert, Text Notification & Video Clip; Sends Audio.

The AEM is present, and the parent is notified: *"Big Buddy: Your child has been disrupted by another player. You can directly support your child and enter as your own avatar."* A video recording is available. The parent records a voice message, and in VR, an avatar labelled 'Parent' enters the scene to play the voice message aloud to the child.

Conditions were counter-balanced using a balanced Latin-Square.

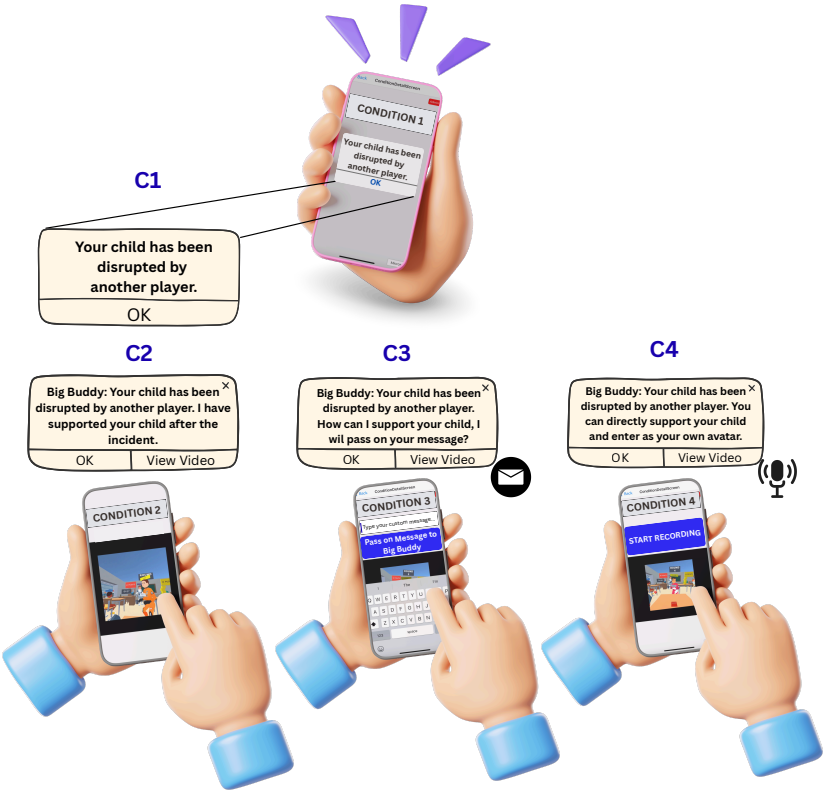


Figure 7.6: Visuals with screenshots of parent perspectives (mobile App view) for each condition.

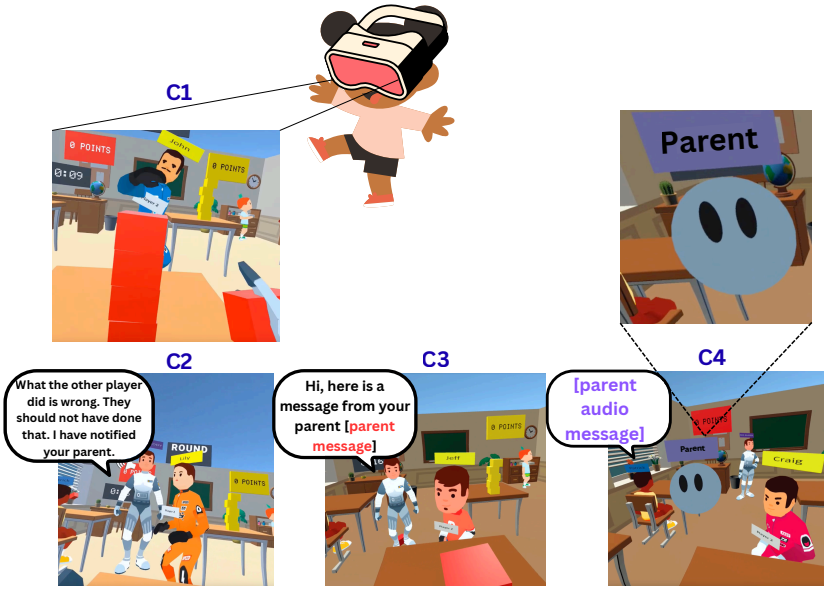


Figure 7.7: Visuals with screenshots of child perspectives (in VR view) for each condition. Note: the text in bubbles is spoken out loud in the VR simulation.

7.3.3.2 Social VR Scenarios and Mobile App Implementation

I employed the same social VR environment and disruption scenario previously used and demonstrated to be effective in prior work (see study #4), suitable to build on for this study. The setting was a virtual classroom populated with non-player characters (NPCs), where participants engaged in a tower-building game using blocks. The disruption consisted of a virtual character intentionally sabotaging the participant's progress by destroying their constructed tower. The AEM was also the same as in prior work "Big Buddy" (see study #4). To represent the parent in **C4** *[Direct Parental Support]*, I used a spherical avatar labelled "parent", designed to be neutral in appearance and suitable for all genders figure 7.7.

I developed an iOS application to deliver condition-specific notifications to parents. While participants were told notifications were real-time (as the VR incident occurred), the mobile app notifications were not in real-time. For example, parents might receive a notification with pre-recorded videos of the disruption, before the child was actually disrupted. However, since parents and children were in separate rooms, this discrepancy went unnoticed. This approach allowed parents in **C3** *[Indirect Parental Support]* to send written messages, which were converted to speech using the Google Text-to-Speech API [380], stored until the disruption actually occurred in VR and relayed to the child via the AEM. (After listening to the recordings, all messages were successfully translated, except for emojis, which were read aloud as-is.) And in **C4** *[Direct Parental Support]*, parents could record a voice note, which was then stored and subsequently played within the VR environment at the time of disruption. Only the most recent message and audio recording were retained and played in the VR environment.

The prototype was pilot-tested with five adult participants: three assumed the role of the child, and two assumed the role of the parent, with the main researcher acting as the counterpart in each case. This resulted in minor adjustments, including increasing the alert volume and extending the duration of each round in VR (45 seconds).

A video of the prototype can be found in the Supplementary Material STUDY 8/VIDEO.

7.3.3.3 Procedure

Parent–child pairs were invited to participate in the study. Prior to the experimental session, parents received an information sheet, consent form, and a demographics questionnaire via email. On the day of the experiment, children also provided written assent before participation.

At the start of the session, both participants were introduced to the social VR environment and the study's context. Key concepts such as bullying, support, and child privacy were defined and explained using a slide presentation (see Supplementary Material STUDY 8/SLIDES). Children were told they would be participating in a virtual classroom, competing with real players from around the world. To ensure children understood the Likert-type scales used in the study, researchers asked simple questions (e.g., "How are you feeling today?").

Children used a Meta Quest 2 headset with an adjustable strap for comfort. They remained standing but were not required to move around the room. Each child began with a practice session in which they could freely build a tower using virtual blocks; they were allowed to continue practising until confident. Researchers ensured the child achieved a minimum of four stacked blocks before proceeding, as the sabotage event primarily targeted tower-building performance. In each condition, children completed 2 to 4 randomly assigned rounds, one of which was disrupted.

While the child participated in the VR session, the parent was seated in another room, separated by a movable partition wall (soundproof). Parents were provided with an iPhone X running the custom-built application and instructed to complete a filler task (either a crossword puzzle or Sudoku) while awaiting notifications. For ethical transparency, the doors of the rooms remained open; however, care was taken to minimise audio transmission between rooms.

Each dyad completed four rounds of gameplay. One round included the scripted disruption event, while the order of rounds was randomised to mitigate habituation. After each round, both the child and the parent completed a questionnaire on a laptop using custom-developed Likert scales.

At the end of the experiment, both participants took part in separate, 5-minute semi-structured interviews to explore their experiences and preferences in more depth.

7.3.3.4 Measures

I used both quantitative and qualitative methods to understand perceptions of the different conditions.

Quantitative **Children** completed seven custom-designed Likert-scale questions for each condition (See Supplementary Material STUDY 8/QUESTIONNAIRES). Three of these scales were adapted from prior work [3], measuring emotional responses: sad to happy, unsafe to safe, and calm to annoyed. All had an "Other" option in addition to the items provided.

1. SAD–HAPPY How did you feel in this round, from sad to happy? [1-6]
2. UNSAFE–SAFE How did you feel in this round, from unsafe to safe? [1-6]
3. CALM–ANNOYED How did you feel in this round, from calm to annoyed? [1-6]

Additional scales assessing perceived support and other dimensions:

1. UNSUPPORTED–SUPPORTED How did you feel in this round? From Not supported to Very supported [1-4]
2. IGNORE–KNOW How much do you think your parent knows about what happened in this round in VR? From Nothing to Everything [1-4]
3. DISCOMFORT–COMFORT How comfortable would you be if your parent had this amount of details? From Not Comfortable at All to Very Comfortable [1-4]
4. DISLIKE–LIKE Did you like how much your parent was involved in this round? From I did not like it all to I really liked it [1-4]

Parents also completed seven 7-point Likert-scale items (ranging from Strongly Disagree to Strongly Agree) after each condition (See Supplementary Material STUDY 8/QUESTIONNAIRES). The items assessed the following:

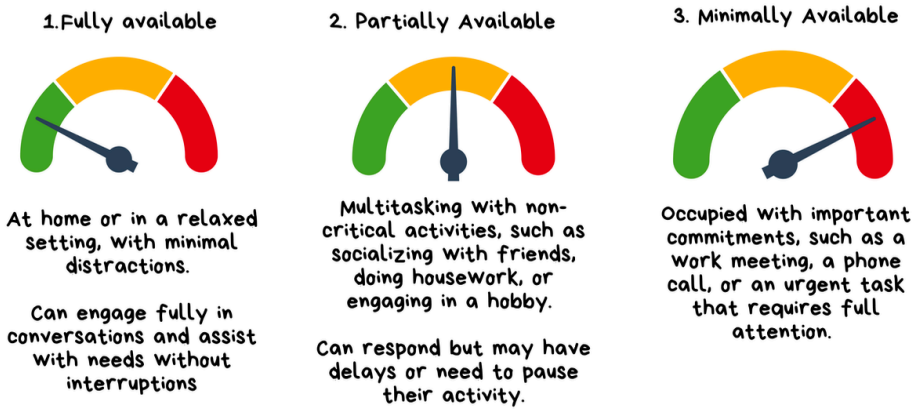


Figure 7.8: Three hypothetical availability scenarios: 1) Fully Available, 2) Partially Available and 3) Minimally Available.

1. BALANCING TASK WITH SUPPORT The round made it easy to balance completing my task and supporting my child. [1-7]
2. INVOLVEMENT My involvement in this scenario was sufficient to support my child. [1-7]
3. CONTROL I felt I had enough control over how I supported my child. [1-7]
4. REASSURANCE This round made me feel reassured about my child's safety. [1-7]
5. COMFORT I felt comfortable with the level of involvement I had in this scenario. [1-7]
6. NOTIFICATION INFORMATIVENESS The notifications provided were informative. [1-7]
7. RESPECTING CHILD PRIVACY I feel my involvement respected my child's privacy. [1-7]

At the end of the survey, parents ranked the four conditions from most to least preferred, under three hypothetical availability scenarios: 1) Fully Available, 2) Partially Available and 3) Minimally Available (see figure 7.8). By asking parents to rank the intervention according to their availability, it better captures the realistic conditions under which digital mediation may occur, acknowledging that parental involvement fluctuates (see section 2.2.4).

Qualitative Each Likert-scale item included a follow-up open-ended question asking participants to explain their rating ("Why?"). For parents, the iPhone screen was recorded to capture typed messages (**C3** *[Indirect Parental Support]*), and video recordings of the VR scene were used to observe spoken interactions with the child (**C4** *[Direct Parental Support]*). At the conclusion of the experiment, I conducted brief (approximately five-minute) semi-structured interviews with children and parents separately to gather deeper insights into their experiences and preferences.

7.3.3.5 Analysis

Quantitative Data was analysed using R. I used ARTool for Aligned ranks transformation (ART) ANOVA [281], a non-parametric approach for multiple independent variables, interactions, and repeated measures [281, 328]. I set the significance level at 0.05, and interpreted effect sizes using partial eta squared η_p^2 with small effects ($<$ 0.01), medium effects ($\approx$ 0.06) and large effects ($\approx$ 0.14 or higher) [286, 287]. Post-hoc comparisons were Tukey-Adjusted. (Note: Responses in which children selected 'Other' on the Likert scales were considered as 'NA' answers and subsequently excluded from the corresponding test). The R Markdown with the code can be found in the Supplementary Material STUDY 8/DATA_ANALYSIS.

Qualitative I used inductive thematic analysis [283] to analyse the qualitative data. After generating transcripts, I listened to recordings ($\sim$5min each) and corrected mistakes in the automatically generated transcripts. A second coder and I familiarised with the data, before creating a coding scheme using *QCAmap* [308], line by line for 20% of the data. Codes generated are words or short phrases that describe any new idea. Coding schemes were compared and readjusted into one final scheme. I then coded all the data with the final scheme. Two other researchers and I then developed a set of higher-level categories by bringing together related codes, discussing, and refining themes to generate a rich description (Supplementary Material STUDY 8/CODING_SCHEMES). *A Miro board gathering themes can be viewed via* `https://miro.com/app/board/uXjVIzd9etM=/?share_link_id=854529644589`.

7.3.3.6 Participants

Recruitment This research was conducted in English and in the UK. The children (8-16 years old) were recruited via their parents on a voluntary basis and required the parent's permission to participate. The study was advertised via university forums and mailing lists, social media, the local library and charity RespectMe [381]. Consent and demographics questions were completed by the children's parent before the session. On the experiment day, children signed a consent form and were told they could take breaks or stop at any time. I assured participants that no wrong answer can be given and assured them of the confidentiality of their responses.

Demographics This study involved 34 child–parent dyads, comprising 34 unique children aged 8–16 years from 31 families. Two parents (P14, P24) participated twice, each completing the full study with two different children, denoted as A (younger) and B (older). When quoting these parents, I indicate the corresponding child with a letter (e.g., P14A for the session with the younger child). Five children were noted to have disabilities (Child4: cystic Fibrosis and dyslexia, Child5: hypermobility, Child12: dyslexia, Child26: ADHD and Autism, Child33: dyslexia)

Table 7.5: Child Participant Demographics and VR/Social VR Experience and Familiarity (N=34). (Responses given by parents)

Question	N	Mean (SD)
Age	34	11.50 (2.43)
Frequency of online gaming (1'never'-6'daily')	34	5.21 (1.04)
Experience with VR (1'none'-5'a great deal')	34	2.56 (0.96)
Familiarity with social VR (1'none'-5'a great deal')	34	2.18 (1.59)
Experience with social VR (1'none'-5'a great deal')	34	1.91 (1.12)

Table 7.6: Parent Participant Demographics and VR/Social VR Experience and Familiarity (N=32)

Question	N	Mean (SD)
Age	32	44.50 (5.54)
Frequency of online gaming (1'never'-6'daily')	32	1.97 (1.62)
Experience with VR (1'none'-5'a great deal')	32	1.97 (0.78)
Familiarity with social VR (1'none'-5'a great deal')	32	1.75 (1.39)
Experience with social VR (1'none'-5'a great deal')	32	1.22 (0.49)
Concern for child's online safety (1'strongly disagree'-7'strongly agree')	32	5.94 (1.14)
Concern for child's safety in social VR (1'strongly disagree'-7'strongly agree')	32	6.09 (1.00)
Belief in child's physical privacy rights (1'strongly disagree'-7'strongly agree')	32	6.00 (1.19)
Belief in child's digital privacy rights (1'strongly disagree'-7'strongly agree')	32	4.59 (1.78)

Among the child participants, 22 were male, and 12 were female. There were 21 White/Caucasian, eight Asian, two Arab, two Black/African and one identified as Other. Regarding parents, nine were male, and 23 were female. There were 20 White/Caucasian, eight Asian, two Arab, one Black/African and one identified as Other. Two mothers [P14A/P14B & P24A/P24B] did the experiment twice with each of their two children [Child14A/Child14B & Child24A/Child24B]. From the same family, the mother P22A did the experiment with their daughter C22A, and the father P22B did the experiment with their son P22B.

There were a total of 34 parent-child pairs across 31 households (32 parents). Of these households, 11 owned a VR headset, while 20 did not.

Ethics and Compensation This study was approved by the ethics committee of the University of Glasgow (College of Science & Engineering) (Approval Number: [300240054]). I was registered with the national Protecting Vulnerable Groups (PVG) scheme to conduct this study with children. A £10 Love2Shop voucher was given to the accompanying guardian, while children received a certificate as a token of appreciation. Documents with the information sheet, pre-survey with consent form and demographics questions can be found in the Supplementary Material.

Current Moderation of Child Online Space and Child Online Safety I present an overview of the strategies parents reported using to moderate their child's online activity, as indicated in the demographics questionnaire before the experiment.

Parents use time limits (N=14) and access child devices (N=5), they block content (N=1), use standard parental controls (N=18). Some parents do not use parental controls (N=8). Parents noted the importance of balance between child agency/privacy and parent knowledge/monitoring, and one does not monitor but trusts their child (N=4). They communicate and have conversations with their child about online usage (N=11). Children can only play with friends (N=4). Parents monitor their child's interactions and content (N=7), with some being in earshot when their child plays online games (N=3). Parents restrict their child from downloading some applications (N=2), and one uses blockers via their broadband. A parent mentioned their lack of knowledge and time to learn parental controls. They consider child online safety important (N=12). Child online safety depends on age (N=1) and who they are interacting with (N=1). Parents have expectations of safety, such as 1) no exposure to malicious people (N=4), 2) no exposure to age-inappropriate content (N=4), 3) no bullying or negative influence (N=5), 4) expects moderation presence (N=1), 5) Low risk of negative harm (N=2) and preventing harms (N=4). They also consider child online safety with the psychological impact on children (N=3). They offered several suggestions to improve safety on social apps, including features like log-off options, protections against addictive content, ad blockers, and tools that promote informed and supported use by children. They emphasised the importance of fostering genuine friendships rather than unregulated open spaces and advocated for mirroring real-world protections in online environments. Encouraging positive peer influence was also seen as a valuable strategy. One parent expressed concern about adults misrepresenting their age in VR. Another suggested that if a child requires constant parental supervision to stay safe, then they may not be ready for social VR platforms.

Selection and Participation of Children Before participation, the details of the studies were explained to the children's parents via an information sheet sent after they expressed interest. A digital consent form was also provided to be signed before coming to the session, noting that the child's participation (aged 8 to 16 years old) was voluntary. At the start of each session, I explained the procedure to the parent and the child using slides (see Supplementary Material STUDY 8/SLIDES). The child was then given a paper consent form to sign, where they could agree or disagree (by ticking boxes), to be audio recorded for the interviews, and were informed that they could take breaks or stop at any point. I assured them that they could ask questions and that there were no wrong answers. Participants were found via university forums and mailing lists (researchers or lecturers with children), social media and the local library. All participants' personal data were stored securely, and all personally identifiable data were removed.

7.3.3.7 Limitations

While repeated exposure to the same harassment scenario in social VR (although by a different second NPC) could lead to habituation, I prioritised consistency across conditions to avoid introducing new variables. To mitigate this, I used a Latin Square design to counterbalance conditions, ensuring that any observed differences were due to the experimental conditions, not repeated exposure. Additionally, to further reduce habituation where participants might anticipate the disruptive event, I had each participant complete between two and four randomised rounds per condition, with only one round containing the disruption.

Although the study was conducted in English, within the UK, and most parents and children were unfamiliar with social VR, the insights can inform the development of interventions tailored for new child users, who may lack experience and have greater needs compared to more experienced users.

Additionally, three parents participated in the study twice, each time with a different child. While this may have introduced some habituation or learning effects, it also provided valuable insights into how a single parent's perspective may vary between children of different ages.

Result-derived Considerations from Experimental Design Impact Some parents experienced a learning effect (N=2), *"Got easier to complete towards the end and be informed of the disruption was clear to see on the phone".* [P31], *"In a routine of doing task and then stopping when I heard the bell."* [P32]. Others noted that the experimental setting influenced their evaluations (N=5). *"So they weren't in order like 1 2 3 4. I suppose coming into it kind of fresh, I think for me it might have been a bit more useful to try level four first, just so that I could gauge the other responses a bit better, because I've maybe agreed with things, but I haven't seen how far the application went. I didn't have that comparison to make."* [P2]. *"As this was my second attempt, I better understood what was going on, so I felt more confident that this might work."*

[P22A]. These observations suggest that repeated exposure may have influenced parent experiences and should be accounted for in future research. However, in this study, I still value their input as relevant, especially since the experimental conditions were balanced using a Latin Square design to minimise order effects.

Although most children reacted negatively to the tower disruptions, supporting the ecological validity of the disruption, some perceived the simulation differently. Two children noted that the classroom setting made them feel safe. Some focused more on game performance than the disruption itself (N=2), while others were unaffected or less bothered as they knew it was not a real danger (N=16). Some felt positive emotional valence either because no attack occurred in some rounds (N=1) or due to a sense of accomplishment in building the tower quickly (N=3). One child noted that they could not trust anyone in the game (N=1), one child thought the second player was a hacker (N=1), and another felt the NPCs were gossiping (N=1).

7.3.4 Results - Quantitative

In this section, I present Likert scales results, statistics and overall ranks of conditions from both parents and children.

7.3.4.1 Likert Scales

Parents figure 7.9 illustrates parental responses across the four conditions for each Likert Scale. Likert scales, except 'Respecting Child Privacy' show a consistent increase across the conditions, from **C1** *[No Support]* to **C4** *[Direct Parental Support]*. Notably, conditions **C1** *[No Support]* and **C2** *[AEM Support Only]* exhibit similar percentages, as do conditions **C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]*, with a clear difference observed between these two pairs. This pattern is further supported by the statistics presented in table 7.7.

Significant differences were found across all parent Likert Scales, except for 'Respecting Child Privacy', and significant post-hoc contrasts were primarily found between **C1** *[No Support]* and the other conditions. (Full table with non-significant results can be found in the Appendices in section D.1.)

Table 7.7: One-way ART ANOVA and Post-hoc Results Across Conditions for All Measures from Parent Participants. I set the significance level at 0.05, and interpreted effect sizes using partial eta squared η_p^2 with small effects (< 0.01), medium effects (≈ 0.06) and large effects (≈ 0.14 or higher) [286, 287].

Analysis	Result (significance level 0.05)
BALANCE TASK WITH SUPPORT	
ART ANOVA BALANCE TASK WITH SUPPORT	$F(3,99) = 9.18$, $p = 2.05 \times 10^{-5}$, $\eta_p^2 = 0.22$ **(large effect)**
Post-hoc Contrasts (Tukey-adjusted)	
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -3.16$, $p = 0.0111$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -4.76$, $p < 0.0001$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(99) = -3.76$, $p = 0.0016$
INVOLVEMENT	
ART ANOVA INVOLVEMENT	$F(3,99) = 34.40$, $p = 2.57 \times 10^{-15}$, $\eta_p^2 = 0.51$ **(large effect)**
Post-hoc Contrasts (Tukey adjusted)	

Analysis	Result (continued)
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(99) = -2.69,\ p = 0.041$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -7.94,\ p < 0.001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -8.58,\ p < 0.001$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -5.25,\ p < 0.001$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(99) = -5.88,\ p < 0.001$
NOTIFICATION INFORMATIVENESS	
ART ANOVA NOTIFICATION INFORMATIVENESS	$F(3,99) = 29.75,\ p = 8.49 \times 10^{-14},\ \eta_p^2 = 0.47$ **(large effect)**
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(99) = -5.41,\ p < 0.001$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -8.08,\ p < 0.001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -8.28,\ p < 0.001$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -2.67,\ p = 0.0428$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(99) = -2.87,\ p = 0.0255$
REASSURANCE	
ART ANOVA REASSURANCE	$F(3,99) = 27.65,\ p = 4.50 \times 10^{-13},\ \eta_p^2 = 0.46$ **(large effect)**
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(99) = -3.50,\ p = 0.0038$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -6.69,\ p < 0.0001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -8.44,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -3.19,\ p = 0.0101$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(99) = -4.94,\ p < 0.0001$
CONTROL	
ART ANOVA CONTROL	$F(3,99) = 31.58,\ p = 2.08 \times 10^{-14},\ \eta_p^2 = 0.49$ **(large effect)**
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -7.40,\ p < 0.0001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -8.36,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -4.86,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(99) = -5.82,\ p < 0.0001$
COMFORT with level of involvement	
ART ANOVA COMFORT with level of involvement	$F(3,99) = 25.67,\ p = 2.28 \times 10^{-12},\ \eta_p^2 = 0.44$ **(large effect)**
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(99) = -3.12,\ p = 0.0126$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -7.41,\ p < 0.0001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -7.31,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -4.29,\ p = 0.0002$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(99) = -4.20,\ p = 0.0003$

Finding 1: Parents rated **C1** *[No Support]* and **C2** *[AEM Support Only]* significantly lower compared to **C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]* in terms of sufficient involvement, notification informativeness, reassurance, control and comfort with level of involvement. Moreover, **C1** *[No Support]* was rated significantly lower compared to **C2** *[AEM Support Only]* in terms of sufficient involvement, notification informativeness, reassurance and comfort with level of involvement. There were no significant differences in terms of respecting child privacy.

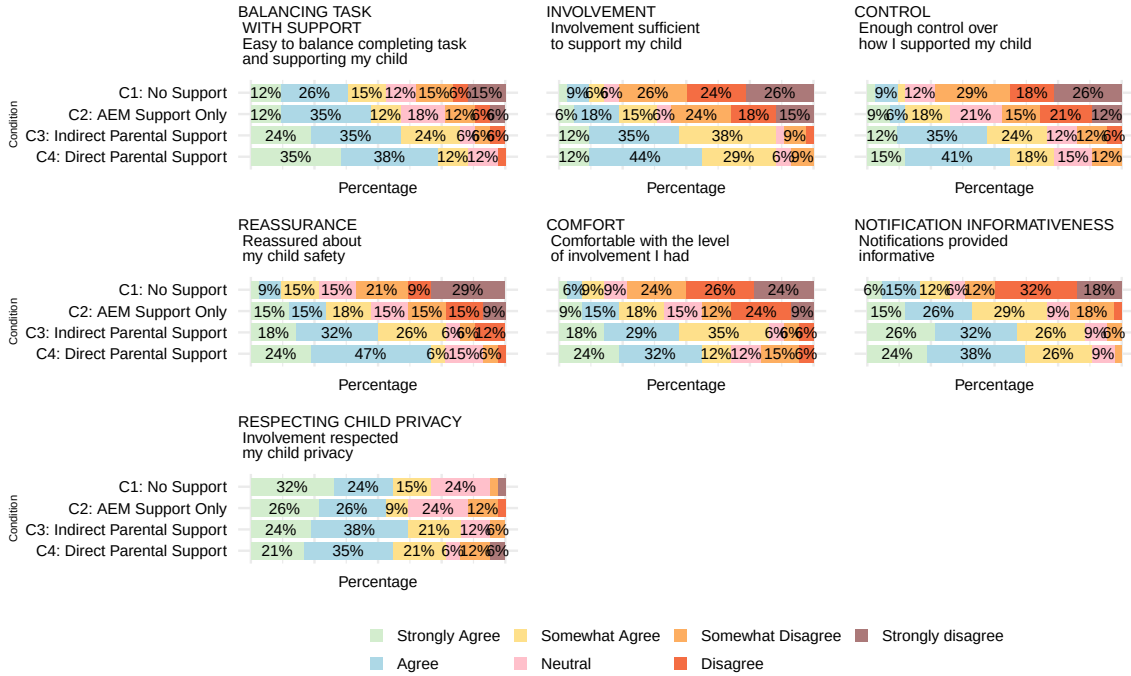


Figure 7.9: Parent Likert Scales (ranging from Strongly Agree to Disagree) as percentages for each condition.

Children figure 7.10 illustrates children's responses across the four conditions for each Likert Scale. Likert scales except 'SAFE–UNSAFE' AND 'DISCOMFORT–COMFORT' show a staircase-like increase in ratings across the conditions from **C1** *[No Support]* to **C4** *[Direct Parental Support]*. Children seem to feel safer and more comfortable in **C2** *[AEM Support Only]* and **C4** *[Direct Parental Support]*, although **C3** *[Indirect Parental Support]* is still generally rated more positively compared to **C1** *[No Support]*. This pattern is further supported by the statistics presented in table 7.8.

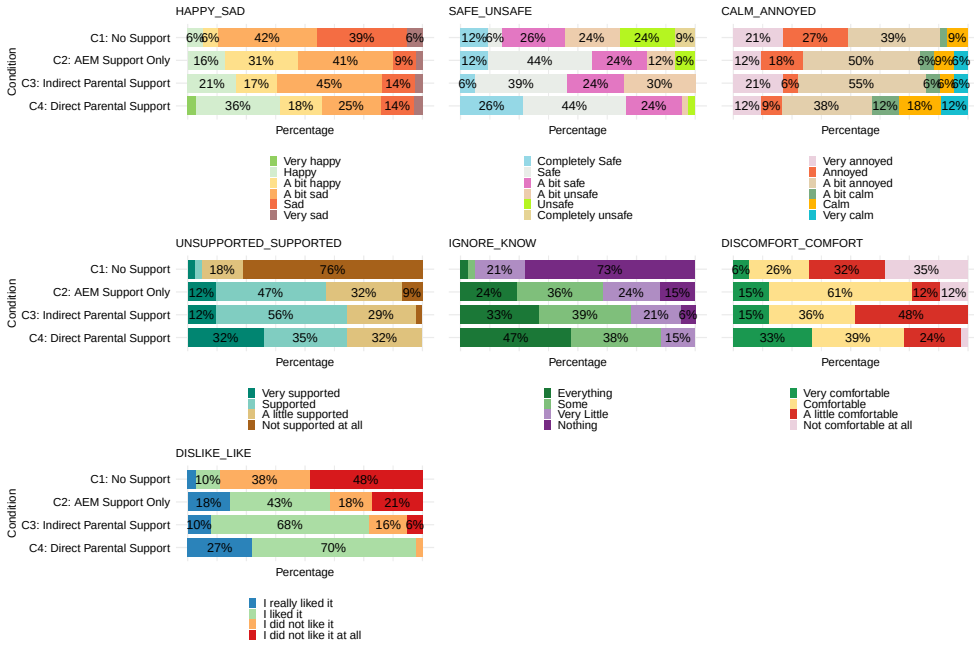


Figure 7.10: Children Likert scales results as percentages for each of the four conditions.

Statistical tests showed significant results in all Likert scales except 'IGNORE–KNOW'. Similar to parents' results, significant post-hoc contrasts were primarily found between **C1** *[No Support]* and the other conditions. (Full table with non-significant results can be found in the Appendices section D.2.)

Table 7.8: One-way ART ANOVA and Post-hoc Test Results Across Likert Scales from Children Participants with significance and effect sizes. I set the significance level at 0.05, and interpreted effect sizes using partial eta squared η_p^2 with small effects ($<$ 0.01), medium effects ($\approx$ 0.06) and large effects ($\approx$ 0.14 or higher) [286, 287].

Analysis	**Result (significance level 0.05)**
SAD-HAPPY	
ART ANOVA SAD–HAPPY	$F(3, 85.76) = 8.39,\ p = 5.92 \times 10^{-5},\ \eta_p^2 = 0.23$ **(large effect)**
Post-hoc Contrasts (Tukey-adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(84.1) = -3.80,\ p = 0.0015$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(85.2) = -3.32,\ p = 0.0071$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(87.0) = -4.64,\ p = 0.0001$
UNSAFE–SAFE	
ART ANOVA UNSAFE–SAFE	$F(3, 98.14) = 18.10,\ p = 2.00 \times 10^{-9},\ \eta_p^2 = 0.36$ **(large effect)**
Post-hoc Contrasts (Tukey-adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(98.0) = -4.74,\ p < 0.0001$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(98.3) = -3.58,\ p = 0.0030$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(98.0) = -7.24,\ p < 0.0001$
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(98.3) = -3.60,\ p = 0.0028$
CALM–ANNOYED	
ART ANOVA CALM–ANNOYED	$F(3, 97.21) = 6.29,\ p = 0.0006,\ \eta_p^2 = 0.16$ **(large effect)**
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(97.2) = 4.33,\ p = 0.0002$
UNSUPPORTED–SUPPORTED	

Table 7.8 (continued): One-way ART ANOVA and Post-hoc Results

Analysis	Result (significance level 0.05)
ART ANOVA UNSUPPORTED–SUPPORTED	$F(3,99) = 41.89,\ p < 2.22 \times 10^{-16},\ \eta_p^2 = 0.56$ **(large effect)**
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(99) = -7.91,\ p < 0.0001$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -8.85,\ p < 0.0001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -10.15,\ p < 0.0001$
IGNORE–KNOW	
ART ANOVA IGNORE–KNOW	$F(3,96.95) = 36.13,\ p < 9.19 \times 10^{-16},\ \eta_p^2 = 0.53$ **(large effect)**
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(97.2) = -6.52,\ p < 0.0001$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(97.2) = -8.02,\ p < 0.0001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(96.7) = -9.75,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(96.7) = -3.18,\ p = 0.0106$
COMFORT with parental knowledge	
ART ANOVA COMFORT with parental knowledge	$F(3,94.69) = 8.82,\ p = 3.27 \times 10^{-5},\ \eta_p^2 = 0.22$ **(large effect)**
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(95.2) = -3.78,\ p = 0.0015$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(95.2) = -2.72,\ p = 0.0384$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(95.2) = -4.93,\ p < 0.0001$
DISLIKE–LIKE	
ART ANOVA DISLIKE–LIKE	$F(3,88.88) = 22.73,\ p = 5.18 \times 10^{-11},\ \eta_p^2 = 0.43$ **(large effect)**
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(89.6) = -4.42,\ p = 0.0002$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(88.8) = -5.59,\ p < 0.0001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(90.3) = -8.11,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(89.5) = -3.47,\ p = 0.0043$

Finding 2: Children rated **C1** *[No Support]* significantly lower than all other conditions on emotional valence (sadder/more annoyed), safety, support, parental knowledge, comfort with parental knowledge, and likeness. Moreover, **C2** *[AEM Support Only]* was rated significantly lower compared to **C4** *[Direct Parental Support]* in terms of parental knowledge and liking. In **C3** *[Indirect Parental Support]* children felt significantly less safe compared to **C4** *[Direct Parental Support]*.

7.3.4.2 Overall Ranks

Parents Parents ranked each condition from most to least preferred across three levels of availability described in figure 7.8. table 7.9, table 7.10 and table 7.11 describe the overall ranks for each availability level.

Finding 3: As parental availability decreased (from Level 1 to Level 3), preferences for support conditions shifted, there was a clear shift in preference away from direct parental involvement (**C4** *[Direct Parental Support]*) toward alternative forms of support, particularly AEM support (**C2** *[AEM Support Only]*). This suggests that parents adjust their preferences based on their availability.

Table 7.9: Availability Level 1 Parents' Weighted Rankings - Rankings were converted to coefficients, with 4 points assigned to the most preferred option, 3 to the second, 2 to the third, and 1 to the least preferred. Cell colour intensity corresponds to the weight, with stronger colours indicating higher preference.

Condition	1st Fav.	2nd Fav.	3rd Fav.	Least Fav.	Total
C4 *[Direct Parental Support]*	80	21	4	5	110
C3 *[Indirect Parental Support]*	44	57	8	0	109
C2 *[AEM Support Only]*	8	21	48	1	78
C1 *[No Support]*	4	3	8	28	43

Table 7.10: Availability Level 2 Parents' Weighted Rankings - Rankings were converted to coefficients, with 4 points assigned to the most preferred option, 3 to the second, 2 to the third, and 1 to the least preferred. Cell colour intensity corresponds to the weight, with stronger colours indicating higher preference.

Condition	1st Fav.	2nd Fav.	3rd Fav.	Least Fav.	Total
C3 *[Indirect Parental Support]*	52	39	14	1	106
C2 *[AEM Support Only]*	40	36	24	0	100
C4 *[Direct Parental Support]*	36	27	12	10	85
C1 *[No Support]*	8	0	18	23	49

Table 7.11: Availability Level 3 Parents' Weighted Rankings - Rankings were converted to coefficients, with 4 points assigned to the most preferred option, 3 to the second, 2 to the third, and 1 to the least preferred. Cell colour intensity corresponds to the weight, with stronger colours indicating higher preference.

Condition	1st Fav.	2nd Fav.	3rd Fav.	Least Fav.	Total
C2 *[AEM Support Only]*	96	12	10	1	119
C3 *[Indirect Parental Support]*	16	51	22	2	91
C1 *[No Support]*	16	24	8	18	66
C4 *[Direct Parental Support]*	8	15	28	13	64

Children Children also ranked conditions from the favourite to least favourite (table 7.12).

Table 7.12: Children's weighted rankings of conditions based on their most and least preferred choices. Rankings were converted into weighted scores: 4 points for the most preferred, 3 for the second, 2 for the third, and 1 for the least preferred option. Cell colour intensity corresponds to the weight, with stronger colours indicating higher preference.

Condition	1st Fav.	2nd Fav.	3rd Fav.	Least Fav.	Total
C4 *[Direct Parental Support]*	88	9	12	3	112
C3 *[Indirect Parental Support]*	20	69	10	1	100
C2 *[AEM Support Only]*	20	18	40	3	81
C1 *[No Support]*	8	6	6	27	47

Finding 4: Overall, the more support, the higher the children rated the conditions.

7.3.5 Results - Qualitative

In this section, I present the results from the qualitative data collected in the questionnaires and interviews, combining **child perspectives** and **parental perspectives**. I refer to the quantitative data when relevant and describe qualitative findings from each condition and interviews. The themes and sub-themes identified are:

THEME 1: Awareness and Involvement (1.1) Parental Involvement (Finding 5), (1.2) Big Buddy Involvement (Finding 6), (1.3) Awareness and Parental Knowledge (Finding 7), (1.4) Parental Availability and Factors Impacting Involvement (Finding 8)

THEME 2: Immersive Parental Presentation (2.1) Parent Avatar Representation in the VR scene (Finding 9), 2.2) Communication through Messages and Audio Recordings (Finding 10)

THEME 3: Parental Role and Child Agency (3.1) Interaction (Finding 11), (3.2) Post-Disruption Real-Time Support, Safety and Reassurance (Finding 12), (3.3) Control (Finding 13), (3.4) Respecting Child Privacy while Navigating Social Boundaries (Finding 14)

THEME 4: Usability and Preferences (4.1) Overall Preferences (Finding 15), (4.2) Benefits of app, Usefulness and Usability (Finding 16), (4.3) Impact of Notifications and Alerts: Disruption, Distraction, and Task Interference (Finding 17), (4.4) Unfamiliarity, Uncertainties and Lack of Clarity (Finding 18)

7.3.5.1 Theme 1: Awareness and Involvement

Parental involvement Parents' Perspectives

Adaptive Involvement Overall, parents want adaptive involvement (N=9), based on child age and emotional maturity, if offence is repeated several times, context (e.g., ages of other players, whether they are on the friend list, which platform, etc.), and situations (e.g, type of harassment).

Lack of Involvement Moreover, where parental involvement was not sufficient, this negatively impacted parents. In **C1** *[No Support]*, parents would supplement the low involvement via the app with other methods (N=8) *"I would have to decide whether to physically attend the location where my child is playing the game or try to phone him without knowing if there were sufficient grounds to do so."* [P2]. Parents noted that knowledge only without involvement led to anxiety (N=12) *"I had no visibility or control, just the knowledge that something was wrong. This would make me panic!"* [P24], *"There wasn't an option to do anything [...] this felt worrying and like I wanted to move and physically go through to the other room and see if they were ok."* [P22A]. P32 noted it was difficult to know something had happened while staying uninvolved. In **C2** *[AEM Support Only]*, parents were also concerned and anxious as they were only more aware of the situation rather than involved (N=6) *"It makes me think that this type of bullying is not just detrimental for the child without any action, it is harming the parent even more."* [P4]. Others considered themselves involved as they received the video clip (N=2).

Sufficiently Involved via Big Buddy In **C3** *[Indirect Parental Support]*, parents discussed being directly involved (N=20). They liked the ability to communicate with their child via Big Buddy (N=5) and the use of their own words (N=1). A parent [P2] felt the level of involvement was appropriate, depending on the child's age and the incident.

Limited Involvement Some parents still found that their involvement was limited (N=5). A parent [P5] wanted stronger actions (e.g., blocking the bully), while others felt uncertain about what to say (N=2), with a parent [P12] wanting suggestions or prompts.

Discretion In **C4** *[Direct Parental Support]*, parents mentioned their involvement was too obvious and could be more discreet (N=2) *"I am not sure that a parent intervening as an avatar would necessarily be supportive of my child's development and sense of reality. But if I could be incognito, perhaps it would be ok."* [P34]. However, others noted it allows them to optimise their involvement (N=2) *"This makes sure that my child receives the message and he is able to consume it."* [P7].

The qualitative findings are supported by the quantitative results (see figure 7.9 and Statistics Parent Involvement). Involvement ratings increased across conditions, with **C1** *[No Support]* and **C2** *[AEM Support Only]* rated significantly lower than both parental support conditions (**C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]*).

Children's Perspectives

Lack of Involvement In **C1** *[No Support]*, children would have wanted their parent involvement (N=8). In **C2** *[AEM Support Only]*, children wished for direct parental involvement (N=11), specifically wishing for more and real-time involvement and presence (N=2) *"I think they should have been there in the moment"* [Child7].

Positive Towards Big Buddy Notifying Parent However, children were positive towards their parents being notified (N=12), with positive/diminished negative emotions from parent's mention (N=8) and children noted feeling safe because their parent was notified about the incident (N=7) *"Since Lily knocked down my tower I didn't feel completely safe but Big Buddy said he was going to notify my mom which made me feel really safe"* [Child24B].

Positive Towards Parent Involvement and Presence In **C3** *[Indirect Parental Support]*, children were positive towards their parent involvement (N=16). Some felt positive emotions, comfort, encouraged by a message from a parent (N=6) and calm (N=3), liking the parent involvement and presence to help and support them (N=7). They found the message effective (N=8). A child mentioned liking their parent's involvement because the child would be able to share/express discomfort instead of keeping it in, *"then my emotions would not be stuck inside me"* [Child14A]. In **C4** *[Direct Parental Support]*, children felt supported (N=31). As a result, most children felt calmed and reassured by the parent, specifically noting it was due to their parent's presence (N=19) or due to their parent's voice (N=20) *"It made me feel warm inside. I recognised her voice instantly".* [Child28]. Children also felt safe (N=22), specifically due to their parents' presence (N=16). Parents could de-escalate the situation (N=4).

Ineffective Message However, children either found it ineffective for Big Buddy to alert them about their parent's message (N=3) or they found their parent's message ineffective (e.g., comforting but irrelevant) (N=3) *"Having them send a message makes you feel comfortable but at the same time because they don't really know much and they just said love you. I didn't get some help or someone say 'someone is mean', or I was not moved to another room"* [Child9]. Children would prefer additional involvement methods (N=10) such as presence (N=8), actions taken to help (N=3) and support via direct voice (N=1).

Parents could de-escalate the situation (N=4). A child mentioned liking their parent's involvement because the child would be able to share/express discomfort instead of keeping it in, *"then my emotions would not be stuck inside me"* [Child14A].

Negative or Neutral Towards Parent Presence However, children felt it was intrusive and did not want their parents' help (N=3) *"My parent was there, though I didn't ask them to be there."* [Child1]. One child [Child27] mentioned liking having their parent involved, even if they did not need it. Others wished for more action from the parent and two-way interaction (N=3).

The qualitative findings are supported by the quantitative results (see figure 7.10 and Statistics Child DISLIKE–LIKE). **C1** *[No Support]* was significantly less liked than other conditions, while **C4** *[Direct Parental Support]* was significantly more liked than **C2** *[AEM Support Only]*.

Finding 5: Both parents and children were positive towards some level of parental involvement (**C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]*) over low or none (**C1** *[No Support]* and **C2** *[AEM Support Only]*), as reflected in both the qualitative and quantitative data. Some parents even expressed a desire for greater

involvement in **C4** *[Direct Parental Support]*. However, important considerations were noted, such as the need for discretion, guidance for uncertain parents, and the possibility that some children may not want their parent present. Some children also found the message from their parent irrelevant or ineffective.

Big Buddy involvement Parents' Perspectives

Big Buddy: A Valuable Source of Support, Reassurance, and Mediation Overall, parents found Big Buddy's presence valuable and reassuring (N=5), especially noting their limited awareness (N=4) *"I'm not experienced with VR at all so the presence of Big Buddy inside the game is reassuring because I understand the Buddy knows the rules and actually is better equipped than I am in knowing how to resolve conflicts."* [P8] . **C2** *[AEM Support Only]*: Parents found Big Buddy helpful (N=14) and supportive (N=5). Some parents found it convenient (N=2), and others felt reassured by its help (N=7). Parents also mentioned trusting Big Buddy (N=3), especially for incidents of lower severity (N=1). Two felt guilty for preferring this option over their direct involvement *"I shouldn't do that if I've got loads of time, maybe I should"* [P34]. **C3** *[Indirect Parental Support]*: Parents found Big Buddy useful as the mediator to support their child (N=5). However, instead of providing their own support, parents relied on Big Buddy's assistance (N=4), *"this relies on Big Buddy being supportive."* [P32], *"I believe I was able to pass on my concern to the helper and hope that the helper would do the right thing and stop this behaviour"* [P4]. One parent [P10] highlighted liking sending a message to their child via Big Buddy's character rather than directly themself.

Wished for Information about Big Buddy and Aftermath However, they wished for more information about the actions taken in VR (N=14) and its capabilities (N=4) *"Big Buddy was a good concept, but it has to learn in what ways it can help the child."* [P20], *"I didn't know whether Big Buddy was also kicking people and whether it had the ability to say that user is gone."* [P29].

Children's Perspectives

Positive Towards Big Buddy Involvement and Presence Overall, children want Big Buddy to be involved (N=13). Big Buddy's presence improved their emotional state and safety (N=10) *"I felt really safe with Big Buddy."* [Child24B] while Big Buddy's absence led to negative emotions (N=2) *"I think the round where Buddy wasn't there made me feel really sad because there was nothing I could do about it."* [Child26] and feeling less safe [Child25] (N=1). Some children noted being continuously aware of Big Buddy in the background while playing (N=2). **C2** *[AEM Support Only]*: Big Buddy's presence was associated with positive emotional responses (N=20), including positive valence (N=16), feeling supported (N=12), comforted (N=4), calm (N=4), and a reduction in negative emotions (N=5). Feelings of safety were also reported (N=11), with children linking this specifically to the actions taken by Big Buddy (N=5). The impact of what Big Buddy said also contributed to these feelings (N=12). However, children expected their parents to appear after Big Buddy's message (N=2), and one child [Child27] expressed not wanting Big Buddy to serve as a mediator. **C3** *[Indirect Parental Support]*: Children were positive towards Big Buddy's presence (N=14) leading to feeling safe (N=12), reducing negative emotions (N=4), feeling supported (N=2) and Big Buddy as mediator (N=11). Many children were positive towards having both

Big Buddy and the parental involvement (N=14), especially appreciating having the presence of two familiar guardians to support them *"I had everybody I knew with me to get that boy away. I knew Big Buddy and my dad."* [Child4]. **C4** *[Direct Parental Support]*: Big Buddy was present even though it did not intervene and was only in the background. Children still felt safe because of its presence, noticing it was there (N=5) *"My parent is here and the robot came to help just in case."* [Child6] and its presence led to positive valence and calmness (N=4). Here too, children appreciated having Big Buddy and the parent as a double presence and support (N=7).

Negative Aspects and Limitations of Big Buddy Children reported not liking Big Buddy's text-to-speech (N=2). They noted that direct help from AI might be wrong (N=1, [Child6]) but would rather have Big Buddy than not (N=2). In a real scenario, relying on a single guardian chatbot may be insufficient (N=1, [Child17]), especially when it cannot provide adequate help *"maybe support from helpers outside of VR and chat bots to stop bullying, but it wouldn't always work."* [Child17]. Additionally, a child [Child31] perceived Big Buddy's ability to stop bullying as limited "[Big Buddy in a real-scenario] to stop bullying is obviously a hard thing." [Child31]. In **C3** *[Indirect Parental Support]*, some noted that the support was ineffective, wanting Big Buddy's direct support rather than a message from their parent (N=3). Children also noted that Big Buddy had more capabilities than their parent (N=5) *"Big Buddy was there and his job is to stop situations [...]. I feel like she [parent] couldn't see what was exactly going on, but it was better than not being able to communicate with me."* [Child9].

Finding 6: Both parents and children viewed Big Buddy as a reassuring and valuable presence, particularly in **C2** *[AEM Support Only]* and **C3** *[Indirect Parental Support]* where it provided indirect support. Children associated Big Buddy with increased emotional safety and calmness, while parents appreciated its mediating role when they felt underprepared or distant. However, limitations were noted: some parents wanted more transparency about Big Buddy's actions, and some children questioned its effectiveness or preferred direct parental involvement. The positive impact of Big Buddy's presence was reflected in emotional and support-related measures across these conditions.

Awareness and Parental Knowledge **Parents' perspectives**

Wished for more Information In all conditions, parents wished for more information regarding 1) the incident and context (N=17) (e.g., severity, age of harasser, repeated, awareness from all perspectives both victim and bully, game information not just incidents), 2) Big Buddy's actions and aftermath (N=14), 3) child reaction and feelings (N=11) and if their child had misbehaved (N=3).

Video Usefulness and Informativeness Parents found the video useful (N=9), but they also wished for the possibility to review the information, re-watch the video (N=2), or view before and after-the-incident clips (N=1):

> *"And maybe it can have like a playback video in short run up, like to see okay what he was playing through what mood he was in and how he went on to the next one."* [P20]

> *"I felt the video had no background. I couldn't see what happened before that. Maybe my child had aggravated the situation. I think it would be useful to see that, but then I get that you might not be able to have such a long video. [...] It would be good to see if they've had interactions with the person who's annoying him before, and it's been negative, so then I can maybe think if we need to block this person."* [P24A]

In **C2** *[AEM Support only]*, parents found the notification informative with the video (N=14), which is enough to drive action (N=2) *"It would enable me to intervene if necessary"* [P2]. One parent [P22B] noted that *"the fact that the message came from 'Big Buddy' felt like they were more informative."*.

The qualitative findings are supported by the quantitative results see figure 7.9 and Statistics Parent Informativeness. There was a significant effect on Notification Informativeness. **C1** *[No Support]* was rated significantly less informative than all other conditions. Moreover, **C2** *[AEM Support Only]* was rated significantly less informative than **C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]*.

Children's Perspectives

Low Involvement Means Low Parental Knowledge **C1** *[No Support]*: Children considered that if the parent was not involved then the parent had little to no knowledge (N=9) *"[Parent doesn't know what happened] because they didn't ask if everything was alright"* [Child32] and similarly if the parent was not present (N=7) *"if my parent knew they would be there as well"* [Child14A]. However, some would just assume they had knowledge (N=2). Similarly, in **C2** *[AEM Support Only]*, children considered that if the parent was not directly present, then the parent had little to no knowledge (N=7).

Big Buddy as Informant Moreover, no Big Buddy means no parental knowledge (N=8), with one child [Child9] noting that the parent would not know much as AI cannot tell them everything. **C2** *[AEM Support Only]*: With Big Buddy present and being told a notification was sent, children considered it provided partial information (N=14) or even all information (N=7). Children were positive towards Big Buddy, providing enough information for parents to be aware (N=9). Children noted feeling comfortable because it was a good balance between parental knowledge and child privacy. Most children expressed a desire for their parent to be informed about the incident (N=19), with one child [Child10] specifically noting that they would feel safe if someone had proof of what happened (N=1).

Partial Parental Involvement Means Partial or All Knowledge **C3** *[Indirect Parental Support]*: Big Buddy's presence with still low parental presence in the VR scene led to children thinking the parent has little knowledge (N=18) while others thought they had full knowledge (N=10).

Content of Message In some cases, the content of the message led to thinking parent had full information (N=3) while in other cases it led to thinking the parent only had partial information (N=3) *"If my mother had known about the destruction of the tower her message to me would not have been as cheerful as it was."* [Child16].

Direct Parental Presence Means More Knowledge **C4** *[Direct Parental Support]*: Children mainly thought their parent had more knowledge due to their in-game presence (N=7), the timing of parental appearance just after the incident (N=1) *"She appeared right when i struggled to build my tower again"* [Child24A], Big Buddy's

presence [Child12] (N=1) and the audio [Child34] (N=1). Some children still thought parents only had partial knowledge as their parent presence was still partial (after the incident or only providing a glimpse of what happened *"[They know some] because they appeared after it happened" [Child1]*) (N=8).

Impact on Comfort **C1** *[No Support]*, children noted wanting their parents to know what happened (N=12), with some feeling discomfort if their parent did not know (N=3) while others mentioned they would feel comfortable even if their parent did not know (N=2). One child [Child34] expressed a preference for their parent to know about trivial situations rather than serious events, and another would want their parent to also know about their accomplishments. In **C3** *[Indirect Parental Support]* Most children felt comfortable if the parent knew what happened (N=14). Children also felt comfortable if their parent knew, as they relied on their parent to take action (N=4). However, one child [Child6] felt it was "weird" that the parent would only know part of what happened. Some children assumed their parents were casting and seeing everything (N=3). In **C4** *[Direct Parental Support]* Children felt comfortable if their parent knew what happened (N=15), while others felt that their parent did not need to know everything (N=7).

The qualitative findings are supported by the quantitative results (see figure 7.10 and Statistics Child Comfort, Statistics Child perspective on parent knowledge). There was a significant effect of support condition on perceptions of being known by parents (IGNORE-–KNOW). **C1** *[No Support]* was rated significantly lower than all other conditions. **C4** *[Direct Parental Support]* received the highest ratings. A similar pattern was found for COMFORT with parental knowledge. Participants felt least comfortable in **C1** *[No Support]*, and most comfortable with **C4** *[Direct Parental Support]*.

Finding 7: Both parents and children emphasised the importance of parental awareness. Parents consistently wanted more detailed information about the incident, Big Buddy's actions, and their child's emotional state, with significantly higher informativeness ratings in **C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]*. Children viewed parental knowledge as limited in **C1** *[No Support]* and **C2** *[AEM Support Only]* but increased and felt more comfortable with regard to their parental knowledge, in **C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]*. Children also appreciated the balance between parental knowledge and maintaining some privacy, especially with Big Buddy serving as an informant. These patterns were reflected in quantitative data: parental informativeness and children's perceived parental knowledge and comfort all significantly increased with higher levels of parental involvement.

Parental Availability and Factors Impacting Involvement **Involvement Not Impacted by Availability** Parents consider that their involvement is based on the severity rather than their availability (N=3), i.e., high severity would lead to high involvement regardless of their availability, and it depends on further information given (N=2):

> *"I think maybe having more of an option to pick at that moment in time, so for example, if it was a more high-level thing, serious bad language or something, then you might want to step in, stop what you're doing, and deal with it separately. It's all about risk and making sure they're safe. So I think if I were in a busy environment doing something intense, and that notification came up and it was serious, then I'd probably want to be able to do the message or voice thing, even though I was engaged in something else, because that's quite important."* [P21].

Parents want options to be involved, regardless of their level of availability (N=7), to interact when something happens (N=4), to be notified (N=3) and to have more information (N=3).

Lower involvement when lower availability Parents would be less involved when less available (N=11). They would prefer typing when less available rather than audio recording, as it does not require their full attention (N=4). A parent [P13] would like the ability to join their child in-game via the app, in their free time. Big Buddy would be useful, especially if very busy (N=12). It would reassure parents in case they cannot support their child (N=4). One parent would always want Big Buddy present regardless of their availability *"I would always want Big Buddy about. I would want a moderator of some sort, either an AI moderator or a human moderator."* [P9]. If the parent is not available (Level 3), the child should not use VR (N=2).

Involvement Impacted by Other Factors than Availability Parents also explained that the level of involvement would depend on several factors (N=9). Some suggested using a rank or colour-coded alert system based on severity (N=4), allowing for a quick visual cue to determine the urgency of a situation. *"If there was a kind of a red alert, I would want to send the voice note."* [P32], *"Green is okay for a go and orange is I'm not that comfortable and red is I'm quite upset."* [P9]. Others emphasised that involvement should be limited to certain situations and context (N=2), or that a clear severity threshold should be established to trigger action (N=2) *"I would have rather have the threshold or the bar for this sort of notifications to be a bit higher in terms of the nature of events that are triggering those notifications".* [P4].

It was also noted that the child's age and level of maturity should influence how and when to intervene (N=2).

> *"The level of action a parent would take towards some situations depends on their child's personality and emotional level. Because for some kids, probably their parents' involvement should be with some words of support. When it comes to older kids or more emotionally stable kids, you're kind of sure that they're going to be fine, but you've just been notified, and you know what to speak about it after with your child to make some notes for the future."* [P1].

Ranks of conditions by parents based on levels of availability are presented in table 7.9, table 7.10 and table 7.11.

Finding 8: Parental involvement is influenced more by the perceived severity of incidents than by availability. While lower availability led some parents to prefer less demanding forms of involvement (e.g., typed messages), many emphasised wanting flexible options to intervene even when busy, particularly for serious events. Parents proposed systems such as severity-based alerts or contextual cues to help guide their response. Factors such as the child's age and emotional maturity also shape parental preferences.

7.3.5.2 Theme 2: Immersive Parental Presence and Communication

Parent Avatar Representation in the VR Scene Parents' Perspectives Parents have diverse and sometimes conflicting preferences for avatar resemblance, visibility, and noticeability.

Resemblance Some parents prefer their avatar to resemble their real appearance (N=3), while others want their voice to be masked or modified (N=2) or their avatar's appearance to be distinctly different (N=2).

Visibility Parents would like to be an invisible manifestation (N=5), communicating only via voice or text, while two prefer to be visible only to their child *"I think just to give my child some privacy as well as Independence and autonomy, I would really much be preferring an invisible manifestation in the space."* [P4]. However, parents want to be fully embodied for a stronger sense of presence (N=2) *"It's probably good to have an avatar there so that maybe a voice isn't a surprise. It doesn't just come out of the blue, so that there is a kind of presence there."* [P2].

Noticeability Even when masked or modified, parents want to remain noticeable to their child (N=3), for example, by having a more colourful avatar than the current design (N=1) *"I would definitely prefer it to be with another colour. That grey is a little bit associated with something bad. I would prefer it to be like a brighter colour."* [P23]. Conversely, one parent expressed a preference to blend into the environment *"I don't want my real voice to get in there and make them feel like it's not a whole game anymore. I would choose to fit in with what she's doing."* [P10].

Child Agency Parents generally believe the child should decide (N=6), with some stating that customisation should be available for the child (N=2). *"I would like the avatar to be something that she would like, prefers, but of course it would be nice also to have my voice to have a degree of something to be closer."* [P27].

Customisation A few parents expressed interest in customising their own avatar (N=4) *"It would be nice to be able to choose your avatar like in Netflix, you can pick and personalise it a little bit."* [P24A].

Abstract Design Some parents preferred an abstract avatar style (N=3).

Impact on the child Parents emphasised that the avatar 1) should not embarrass the child (N=3), 2) should avoid being obtrusive or frightening (N=2), and 3) must respect the child's privacy (N=2) and personal space (N=2) *"I wouldn't want to jump into where they are because it's respecting their personal space, particularly with it as being sort of a parasocial environment. I don't want to be kind of like a cloud wandering in to cause problems. It's not helpful for their development if I'm kind of there on hand at all times to wander in and go 'What's been happening now' ".* [P26].

A one-size-fits-all design would fail to accommodate these diverse preferences. A one-to-one, solution-based personalisation system, enabling parents (and children) to customise visibility, resemblance, style, and presence, would be the most suitable approach.

Children's Perspectives Children's preferences for their parents' avatar representation in VR closely align with the expectations expressed by parents.

Realistic Representation Some children expressed a desire for their parent's avatar to closely resemble their real-life appearance (N=11), reflecting a preference for familiarity. However, one child [Child7] indicated discomfort with a realistic likeness if the avatar were visible to others.

Abstract or Non-Human Representation. Alternatively, several children favoured more abstract or fantastical avatars (N=12), incorporating non-human elements such as the sphere or imaginative objects like a *"guitar with animated strings when the parent speaks"* [Child21]. These choices often included modified voices (N=1) or other fantasy features (N=2). *"She'd look like a fairy, and she'd have a little squeaky voice, and then she'd be saying "Hi, cheer up", and she'd have a little pet unicorn."* [Child14B]. One child mentioned not wanting their parent's avatar to be unrealistic to prevent the parent from being bullied *"I don't want them to be bullied either, so I want them to be just like the sphere."* [Child17].

Avatar Appearance and Behaviour in the Scene **Visibility:** Children desired control over whether their parent's avatar is visible to others (N=9). Some preferred the parent's avatar to be invisible, communicating only through messaging or Big Buddy (N=2), while others wanted the avatar to blend into the environment (N=4) *"Just like any other avatar that you would see inside."* [Child12] or to stand out distinctly (N=3), either with a smaller avatar than the others (N=2) or with colour contrasts to make it pop (N=1).

Entrance in Scene How the avatar's entered into the scene was also discussed by children (N=18); suggestions included magical or attention-grabbing entrances such as popping into view (N=9), walking towards them (N=4), "boing" [Child9] (N=1), "magical spark" [Child33] (N=1) or being introduced by a prompt (N=2), as one was surprised when the avatar entered (N=1).

Floating Object or Avatar Next to Child Some children (N=2) would want the avatar as a floating object or companion beside the child.

Dynamic Representation A child [Child9] wanted the avatar to evolve over time during gameplay *"That [sphere] would be the default setting and then once you played enough rounds you could get other little things like actual people and like little animals".* [Child9].

Personalisation Customisation options were mentioned (N=4), with children expressing interest in adding humorous touches (N=2) or displaying the avatar's name above it (N=1, Child17). Additionally, attributes such as cuteness were mentioned (N=1, Child24A).

Finding 9: Both parents and children emphasised the importance of flexibility and personalisation in the design of the parent avatar within the VR scene. While some parents preferred realism, others favoured abstract, symbolic forms or invisibility to preserve child privacy, autonomy, and to avoid embarrassment or intrusion. Similarly, children were divided between familiar, realistic avatars and imaginative or fantastical representations, seeking control over parent visibility. Preferences converged around respecting child comfort, ensuring presence without disruption, and allowing for customisation and dynamic expression.

Communication through Messages and Audio Recordings Analysis of parents' text messages and audio clips revealed several distinct categories of communication. Firstly, a significant portion of messages were directed specifically **toward the child or victim** (Count=13), often providing emotional reassurance and empathetic check-ins (Count=4), as illustrated by:

> "I hope you aren't getting too upset. People do silly things sometimes. Don't take it to heart." [P26 - message]

Messages were also personalised (Count=3), using the child's name or identifying the sender, such as:

> "*Child name* stay cool and don't get triggered. You can ask him, Is he having a bad day? Is he okay?" [P25 - text message]

Some parents frequently offered direct advice or aimed to de-escalate conflict and reduce the emotional intensity of the situation (Count=4).

> "Just ignore him, he is being a bully." [P2 - text message], "Move away from this situation. Go away from this bully." [P30 - audio], "Just keep rebuilding. Ignore that mad man." [P28 - audio]

A parent encouraged autonomy and independent decision-making (Count=1):

> "Do what you think the best it is." [P20 - audio]

Secondly, several parents **directed their messages to Big Buddy or relied on Big Buddy** as a protective figure (Count=7), seeking intervention or assistance:

> "Can you keep him away from the guy?" [P15 - text message], "Can you be of help to her so that she is safe" [P14A - text message], "Remove the disruptive person/bully from there" [P30 - text message], "What has happened to bully?" [P5 - text message], "Big Buddy is there to rescue you." [P14B - audio]

Thirdly, in some cases, messages were **directed at the bully and other players** (Count=2):

> "Hello! Please stop doing it to my child." [P23 - text message]
>
> "Uh, hey guys! Play nicely!" [P34 - audio]

Finally, **children actively reacted** to their parents' messages or audios (Count=4), demonstrating engagement with the support being offered.

> "Hi. Hope you're safe" [P14B] Mommy, I'm not okay. Craig is being mean." [Child14B]

It also allowed a conversation starter for after the fact.

> "Hi *Child name*. Just try to ignore that guy, he looks really annoying. But I'm sure it'll be okay. Let's chat about it later" [P22B - audio] "yeah, he's called Craig, and I don't like him" [Child22B]

Finding 10: Parents used text and audio communication to offer emotional reassurance, practical advice, and conflict de-escalation, primarily directed at their child or Big Buddy. Many messages reflected empathy and personalisation, with some encouraging autonomy or directly addressing other players. Children responded actively to these messages, indicating engagement. The communication also creates opportunities for follow-up conversations. These interactions highlight the value of timely, supportive messaging as both a coping and bonding mechanism.

7.3.5.3 Theme 3: Parental Role and Child Agency

Interaction Parents' Perspectives

Two-Way Interaction Many parents wished for two-way interactions either with the child (N=14) or via Big Buddy (N=2). They wanted a response from the child. Some additional ways of interaction were also suggested, such as the integration of continuous chat (N=3) or video messages/calls (N=3).

Non-Disruptive and Quick Responses Parents suggested a quick response from the child with an emoji that would be less intrusive and disruptive to their game (N=2) *"Even thumbs up thumbs down - just something to respond to 'are you okay': yes nothing major so or good."* [P26], *"A reaction from my child would be I think that would probably be the best way to do it. Just an emoji smiley face"* [P16] or two-way interaction that is for a limited period of time allowing a quick check-in with the child without over-interrupting (N=2) *"In my view it would be for a limited period of time anyway it's not as though you would want to ask a question about something completely unrelated during the game. I think you would just accept it, there's maybe half an hour or an hour of that going on, so that would be mark time."* [P2]. In **C1** *[No Support]*, parents wished for more involvement and communication options (N=13).

Children's Perspectives

Two-Way Interaction Children also want two-way interaction (N=18) with the parent (N=16) or with Big Buddy (N=1), expressing a desire to respond with minimal and easy-quick responses (N=2) *"There could be like a little keyboard that would pop up and then there would be two buttons which would say send and you don't want to reply to them and then you could just do like the normal thing where you just press the keys and then send it."* [Child9].

One-Way Interaction Preferred However, some children would want to keep it as a one-way interaction, where the parent speaks but the child does not respond (N=2) *"I'd probably not want to answer my parent. I just want support and then continue the game."* [Child31].

Finding 11: Both parents and children expressed wanting two-way interactions. Quick, low-disruption responses such as emojis or brief messages were commonly suggested to maintain immersion while enabling check-ins. Parents emphasised wanting reassurance through minimal feedback, while children appreciated the option to reply but also valued autonomy, with some preferring one-way support without needing to respond. This highlights the importance of designing communication features that balance responsiveness with non-intrusiveness.

Post-Disruption Real-Time Support, Safety and Reassurance Parents' Perspectives

Insufficient Support and Reassurance In **C1** *[No Support]*, parents were uncertain if children were given enough support (N=3) and wished for more reassurance about their child's safety (N=5). One parent [P31] assumed the game would safeguard the child. In **C2** *[AEM Support Only]*, parents were negatively impacted by not having options to support their child (N=8) and felt like the support given was not sufficient (N=3), while others were unsure (N=3).

Reassured by Big Buddy However, parents were reassured by Big Buddy's help (N=7). In **C3** *[Indirect Parental Support]*, parents found the support useful, especially with Big Buddy (N=5). They felt reassured (N=15) with 1) the ability to pass on message (N=7) 2) to view what happened (N=6) 3) by Big Buddy (N=5) and 4) the ability to de-escalate the situation (N=1) *"It is a good tool for intervening with reassurance that the bully should be ignored"* [P2].

Questioning Child Reactions to Support Others were uncertain if their support would be well-received by the child (N=3) *"He might not have liked the fact that I asked Big Buddy to help him."* [P31].

Presence and Voice Reassure and Support **C4** *[Direct Parental Support]*: Parents felt reassured by the benefits of visual and voice presence (N=11). The ability to send a voice note not only comforted the parents into balancing the task and supporting their child (N=4) but also allowed them to support and reassure their child using their own voice (N=8) *"Having a voice message would provide more reassurance rather than the text message in the first round"* [P31], *"Expressing myself to her and to tell her that I love her when she was bullied will be of great support to her"* [P14A].

The qualitative findings are supported by the quantitative results (see figure 7.9 and Statistics Parent Reassurance). Parents felt significantly more reassured in **C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]* compared to **C1** *[No Support]* and **C2** *[AEM Support Only]*. Moreover, parents felt significantly more reassured in **C2** *[AEM Support Only]* compared to **C1** *[No Support]*.

Children's Perspectives

Disruption Impact Children generally reacted negatively to the tower disruptions. Many felt annoyed (N=23), particularly due to repeated disruptions (N=12) or due to the destruction of their efforts (N=14). The sense of injustice and lack of consequences further fueled their annoyance (N=10). Some described feeling helpless (N=2), surprised (N=1), or emotionally jolted by the attacks (N=4).

Negative Impact of Low Support In **C1** *[No Support]*, children noted low to no support as nobody helped them in VR (N=11), with some feeling annoyed by the lack of support (N=5) and others mentioned negative emotions and loneliness due to no support (N=13). The classroom environment was also leading to expected help/support (N=1) *"Nothing could be done about this, even though the room looked like a classroom setting."* [Child16]. In the interviews, children mentioned that no support had a negative impact (N=17). It led to negative emotions (N=6), and the absence of Big Buddy was disliked (N=3). A child mentioned that no guardian means no protection *"The support was good, although the last round there wasn't any support. Nobody was there for protection. It took a minute to notice nobody was there until they destroyed the tower."* [Child17]. Some felt hopeless when no support was given (N=3) *"The fact that no one's there, I just felt kind of just hopeless".* [Child34], and some felt unsettled or unsafe (N=3).

Parent Notified Increases Safety but Insufficient In **C2** *[AEM Support Only]*, children felt safe because the parent was notified by the incident (N=7). However, children felt support was not sufficient (N=4), some trusted that their parent would help and support (N=2), and one felt unsafe.

Support and Reassurance from Big Buddy In **C3** *[Indirect Parental Support]*, some children felt safe anticipating Big Buddy's support and help as they noticed Big Buddy from the start (N=2) and were continuously aware of Big Buddy in the background (N=2) *"I saw Big Buddy. When Big Buddy talked to me, I felt safe."* [Child20]. Children also wished for support from NPCs and bystanders (N=2) *"I felt unsupported since there were many people in the class and they didn't help me"* [Child7].

Parental Presence and Voice Increase Safety In **C4** *[Direct Parental Support]*, children felt safe with parent presence (N=16), *"Feeling safe because my parent was there. I recognised her voice directly."* [Child15]. The increased support helped care less about disruption (N=1), feeling safe (N=1) and diminishing negative emotions (N=2). Children found it surprising to hear their parent (N=2) and were amused by unexpected VR interaction with their parent (N=1) *"It was kind of funny because we're in different rooms just to hear him from like a VR headset"* [Child29].

The qualitative findings are supported by the quantitative results (see figure 7.10, Statistics Child Safety, Statistics Child Support and Statistics Child Annoyance). Children felt significantly more unsafe in **C1** *[No Support]* compared to all other conditions. Moreover, children felt significantly more unsafe in **C3** *[Indirect Parental Support]* compared to **C4** *[Direct Parental Support]*. Children felt significantly calmer in **C4** *[Direct Parental Support]* compared to **C1** *[No Support]*.

Finding 12: Parental and child experiences of support and reassurance varied significantly across conditions. In **C1** *[No Support]* and **C2** *[AEM Support Only]*, both groups expressed concern over insufficient support, leading to uncertainty and negative emotions, particularly among children who felt unsafe and alone. As support increased in **C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]*, reassurance and safety improved. Parents were comforted by the ability to view a video clip, send messages, or use their own voice, while children felt more protected, especially when parents were visually or audibly present. These findings align with the quantitative data.

Control Parents' Perspectives

More Control and Actions Wanted Parents wished for more control and options in **C1** *[No Support]* (N=11). In **C2** *[AEM Support Only]*, parents also wished for more control (N=18), with possibilities of communication (N=13) or to intervene remotely (N=3). Although they did feel they had more control and agency in **C3** *[Indirect Parental Support]* (N=7) and **C4** *[Direct Parental Support]* (N=9), some wanted to have both options to communicate either via message or audio (N=3) as well as being able to take direct actions via the app such as blocking the other player (N=2) *"More choices on what to do (i.e., send a custom message to my child)"* [P34] or freezing/pausing game (N=1) *"I would also like to opt for option to pause/ freeze while communicating"* [P9].

Excessive Control In **C4** *[Direct Parental Support]*, parents felt that appearing in VR with an avatar was too controlling (N=2) *"Also, immediately appearing as an avatar might be a bit controlling?"* [P29].

The qualitative findings are supported by the quantitative results, see figure 7.9 and Statistics Parent Control. Parents rated control significantly lower in **C1** *[No Support]* and **C2** *[AEM Support Only]* compared to **C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]*.

Children's Perspectives

Retaliation In all conditions, children wished for retaliation (in **C1** *[No Support]* N=7, in **C2** *[AEM Support Only]* N=1, in **C3** *[Indirect Parental Support]* N=2 and in **C4** *[Direct Parental Support]* N=3. In **C4** *[Direct Parental Support]*, one child [Child12] noted feeling positive emotions as the harasser got caught.

Want More Actions Taken Overall, in interviews, children mentioned they wished for more actions taken (N=6) *"I really liked that it was helpful but what I didn't like is it couldn't be dealt with right away it had to be like everything paused and then dealt with."* [Child26] and retaliation (N=7) *"I wanted to go over and knock his tower down."* [Child33].

Want to be Part of the Decision-Making They also wanted to be part of the decision-making, such as wanting to decide to inform the parent (N=2) or to call Big Buddy (N=1) *"maybe you can click something to inform the parent instead of just they informed automatically. Calling Big Buddy."* [Child22A].

Finding 13: Parents consistently desired more control over how they could support their child, especially in **C1** *[No Support]* and **C2** *[AEM Support Only]* seeking options like direct messaging, freezing gameplay, or blocking players. While feelings of control improved in **C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]*, some still wished for more flexible intervention tools, and others found the avatar's presence in VR too controlling. Children also expressed a desire for more control, particularly through retaliatory actions or being included in decisions like when to alert a parent or call Big Buddy. This highlights the importance of balancing adult oversight with child agency in VR safety interventions.

Respecting Child Privacy while Navigating Social Boundaries Parents' Perspectives

Balancing Responsibility of Child Safety and Privacy Parents consider privacy important (N=14) and they consider privacy age dependent (N=6). There is tension between responsibility, safety and privacy (N=9) as parents consider their child's well-being and safety more important than their privacy (N=4) *"We are talking about a child. Someone who I am 100% responsible for as a parent and whose, e.g., agency of consent is still developing. Therefore, my child's social world needs to be completely accessible to me as a parent."* [P8], *"I feel personally he is a child, so I should know what he is doing. In that he still lives with me, so I am still looking after him."* [P16].

Involvement via a Mobile App and Little Information Uphold Child Privacy They consider that no information and involvement preserves child privacy (N=15) in **C1** *[No Support]*. In **C2** *[AEM Support Only]*, only minimal knowledge and no direct involvement also mean it respects the child's privacy (N=10), and it is not intrusive (N=4). In **C3** *[Indirect Parental Support]*, most parents found that it is a good balance between parental knowledge and involvement, and respecting the child's privacy (N=11), with low to no invasion of privacy (N=3) *"Being able to pass a message to my child via Big Buddy felt like a good balance between my child's privacy and my involvement."* [P22B]. In **C4** *[Direct Parental Support]*, parents considered they did not invade their child's privacy (N=7), and some parents thought it was a good balance between involvement and child privacy (N=3).

Parental Involvement Mean Privacy Invasion and Child Embarrassment In **C3** *[Indirect Parental Support]*, one parent felt it invaded the child's privacy *"His mum having to be involved to sort out a dispute does not really respect his privacy, but this does not really matter, children do not/should not have ultimate privacy from their parents."* [P34]. A few parents were worried about distracting their child's game (N=3) *"Hopefully he is able to carry on his game now without disruption"* [P31], and one [P28] felt it would cause their child "to cringe". In **C4** *[Direct Parental Support]*, some parents felt they did invade their child's privacy (N=4), especially if other players would see them (N=2). Parents also felt it would be embarrassing to have a parent in the VR scene (N=3), *"He would probably be annoyed/embarrassed if I entered as an*

avatar" [P12]. One would consider incognito involvement more acceptable *"I am not sure that a parent intervening as an avatar would necessarily be supportive of my child's development and sense of reality. But if I could be incognito, perhaps it would be ok."* [P34]. In the interviews, parents mentioned feeling "awkward" intervening directly in the gameplay (N=2) *"Occasionally it made me feel a little bit awkward if I would have to intervene in a game."* [P34] and would not want to embarrass their child (N=2) *"I think it could be reassuring to send a voice message to *Child name*, but could also be perceived as embarrassing?"* [P30].

Invasion of Privacy Depends on Child Age In **C3** *[Indirect Parental Support]*, a parent mentioned it would depend on the child's age. In **C4** *[Direct Parental Support]*, parents also mentioned it depends on the child's age and experience (N=3), with one [P28] mentioning they would feel it is like "spying" for older children.

Briefing Before and Obtaining Child Consent However, overall, parents would want to brief before and ensure child consent and awareness, and would want to be mindful of what their child would feel comfortable with (N=3) *"I chickened out of telling them they could come off of the game as I didn't know if he knew I could watch what he was doing or not. I did feel like I should have told him beforehand that I would be watching anything bad that happened."* [P16].

Importance of Child Independence Other parents (N=2) want their child to be independent, emotional, resilient to manage their online issues and talk about it, building a trusting relationship *"I do like to raise them to be quite strong and independent. I think they are quite confident, and we try to create an atmosphere of you can tell us anything but nobody should really feel sorry for themselves or get too upset about small things. If it's big things, then it's different."* [P34].

The qualitative findings are supported by the quantitative results (see figure 7.9, Statistics Parent Comfort and Statistics Child Privacy). Parents rated comfort with level of involvement significantly lower in **C1** *[No Support]* compared to all other conditions. They also rated **C2** *[AEM Support Only]* significantly lower compared to **C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]*.

Children's Perspectives

Self-Reliance and Child Independence Some children preferred relying on themselves and be independent. Children said they acted as their own support and moral compass in **C1** *[No Support]* (N=2) *"I was supported by myself. I told myself that it was rude if I were you, I wouldn't do that."* [Child6], one liked having some space [Child21], and one was fine handling by themselves [Child10]. In **C3** *[Indirect Parental Support]*, children mentioned they would stand up for themselves (N=2). In **C4** *[Direct Parental Support]*, one child mentioned wanting to take responsibility independently. *"It felt like no big deal, and to avoid these situations, you have to take responsibility for your own actions and predict what will happen. If you say something or do so, the parents won't have to worry or get involved"* [Child1]. In the interview, children mentioned that in a real scenario, they would rather stand up for themselves (N=8). One would *"leave the game and join another."* [Child24B].

Intrusion and Embarrassment of Parental Presence Parents entering VR was viewed as an intrusion of privacy and social interaction (N=3). Even when only the child can see the parent, concerns over parental judgment were noted (N=1) *"Me and my friends we can sometimes be silly and then my dad might get annoyed because*

we're being a bit silly." [Child29]. A child would prefer low distraction and disruption, such as text messages that can be discreet (N=1, Child27). They also reported feeling embarrassed if a parent was present (N=4) or if the Big Buddy came to help while others could see (N=1). In terms of parent visibility in a real scenario, children would not want others to see the parent (N=6) *"That's just kind of embarrassing for a parent to come in."* [Child27] and it would be a new reason to get bullied (N=3)*"I'd prefer if only I could see the parent and only I could send messages because if your parent comes in the VR scene, then somebody might have another reason to bully you because your parents are interfering."* [Child9]. Others would want their parent visible to deter the situation (N=4) or if it is a serious event (N=1, Child12).

Tension between Privacy and Safety **C1** *[No Support]*: Another child [Child9] indicated feeling uncomfortable if they were closely monitored, but would still prefer their parent to be informed for safety reasons. **C2** *[AEM Support Only]* was considered a good balance between parent knowledge and child privacy (N=1, Child9). One child noted feeling comfortable if their parent knows little in terms of privacy, but not in terms of safety (N=1, Child7). Some children felt that **C3** *[Indirect Parental Support]* respected their privacy (N=2), and one [Child12] added that it also respected the parent's privacy since it was communicated via the Big Buddy voice.

Privacy Is Invaded, Requiring Awareness and Consent **C4** *[Direct Parental Support]*: Some children felt their privacy was invaded or that the parent was intrusive (N=2) *"It felt like a little bit too personal coming, getting a bit too involved"* [Child34]. Children also mentioned wanting their parent to ask for consent before involvement (N=2). Overall, in the interviews, children expressed a desire for awareness: one [Child29] found it surprising (N=1) and unexpected (N=1), and another [Child5] wanted to know exactly what their parent sees and knows. Children noted that direct parental presence can be disruptive and intrusive (N=3), and one child felt that both **C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]* were excessive.

The qualitative findings are supported by the quantitative results (see figure 7.10 and Statistics Child Comfort with parent knowledge). As parental availability decreased (from Level 1 to Level 3 figure 7.8), preferences for support conditions shifted, there was a clear shift in preference away from direct parental involvement (**C4** *[Direct Parental Support]*) toward alternative forms of support, particularly AEM support (**C2** *[AEM Support Only]*).

Finding 14: Both parents and children navigated a delicate balance between safety, involvement, and privacy. Parents emphasised their responsibility to protect their child, often placing safety above privacy, especially for younger children. However, many also valued non-intrusive forms of involvement, such as indirect messaging. Children, on the other hand, frequently expressed a strong desire for independence, low disruption, and control over whether and how their parents intervened. Parental presence in VR, particularly visible to others, was often seen as embarrassing, socially risky or invasive. Preferences leaned toward discreet, private support and prior consent. These tensions were reflected in quantitative results.

7.3.5.4 Theme 4: Usability and Preferences

Overall Preferences Parents' Perspectives

Direct Involvement Direct involvement was the most preferred option overall (N=10). One parent even reported feeling powerless in (**C1** *[No Support]*). Among the communication formats, text messaging was preferred (N=6), with some describing it as more discreet (N=1) or feeling more natural (N=1). Others preferred sending audio messages (N=7), with some specifically noting that the tone of voice could convey greater emotional support (N=3). One parent expressed a preference for synchronous rather than asynchronous communication, while another emphasised the importance of having the option to choose between text and audio formats.

Indirect Involvement For parents, indirect involvement through Big Buddy was the top choice (N=3). They appreciated the hands-off, agent-based style (N=2). The type of information provided also influenced preferences: conditions that included videos were favoured (N=2), whereas **C1** *[No Support]* was ranked as their least favourite because of the lack of information (N=1).

Ranks align with these findings presented in table 7.9, table 7.10 and table 7.11.

Children's Perspectives

Low Intrusion Some children expressed a preference for support that is non-intrusive (N=5), low in interruption, and low in involvement, allowing for autonomy during gameplay, for example, with text messages (N=5).

Parent-Big Buddy Team There was a preference for combined human and Big Buddy support (N=7), where Big Buddy played the mediator (N=7), or both were present (N=1). Some children still wanted real human intervention (N=3), and others mentioned that the more parental involvement, the better (N=7).

Wanting Big Buddy Some children preferred Big Buddy's support (N=6). Children noted that Big Buddy improved their emotional state and sense of safety (N=10), preferring its presence over its absence (N=2). In real scenarios, they would want Big Buddy support when needed (N=2).

Wanting Parent or Familiar Figures Some preferred hearing their parent's voice (N=18) as it is familiar (N=3) and increases parental presence (N=2) while others would either want it for a limited time (N=1), for severe scenarios (N=1) or with written text (N=1). In real scenarios, children would want direct parental help (N=6). However, some would want parental involvement only in specific cases (N=4): depending on the game played (N=1), scenario severity (N=1), whether the child noticed something was wrong (N=1) or if settings were set beforehand (N=1). One child expressed a desire for familiar figures to be present: *"I want my dad, my mother and my sisters as well and my friends."* [Child15].

Ranks are presented table 7.12.

Finding 15: While preferences varied, parents tended to favour direct involvement. However, they also appreciated having flexible options, including indirect, agent-mediated approaches like Big Buddy when less hands-on involvement was desired. In contrast, children expressed more diverse needs: many preferred low-intrusion

support that preserved autonomy, such as text messages or indirect involvement through Big Buddy. Yet some still valued hearing a familiar voice or having direct parental help in severe or pre-agreed situations. Preference rankings supported these patterns, with both parent and AEM-based supports rated higher than no support.

Benefits of app, Usefulness and Usability **Mobile App increases Parental Agency** Parents noted positive aspects of the app and saw potential (N=15). The app increased their feelings of agency and control (N=3) *"I feel like I had more agency or like I could do something about it. At least I know that something is happening there."* [P30].

Convenient External to VR and Low Interruption It was noted as easy and convenient (N=1) *"It was quite easy, just press the button."* [P1] and useful especially as a conversation starter for a later discussion with their child (N=1, P1). They like the idea of supporting their child with little to no interruption (N=1) *"I actually thought that was quite good because it didn't mean I had to go and stop her doing what she was doing."* [P33]. One mentioned the benefit of circumventing the requirement to wear a VR headset as they do not like wearing one: *"App is useful as I don't like wearing the VR headset, it makes me dizzy and makes me sick."*[P24].

Usefulness based on Different Factors They also found it useful for younger children (N=1) but would be more useful for more serious events (N=1):

> *I felt it was a really good tool to use. I mean, obviously *Child name* is older, 15, so he probably knows a bit more about what to ignore when things are online and maybe doesn't quite take things to heart like a younger child would, but I think it's a really good tool to be able to intervene."* [P2].

Finding 16: Parents generally viewed the mobile app as a valuable tool for increasing their sense of agency, enabling supportive involvement without disrupting their child's gameplay or requiring them to wear a VR headset.

Impact of Notifications and Alerts: Disruption, Distraction, and Task Interference **Positive Aspects** Parents were positive towards visuals and sounds as it drew attention (N=7) *"The sound makes it easy to monitor my child's gaming."* [P25], *"I can concentrate on what I am doing as the loud notification will alert me to a problem."* [P22A]. However, a parent [P9] would want to silence them or be notified only for severe scenarios, with others wishing to set a severity threshold for when to be notified to avoid distraction when it is a minor incident (N=2). In sum, parents described the alerts: valuable in C1 (N=7), quick and easy to process in C1 (N=1) and C2 (N=1), convenient without further action in C1 (N=8), straightforward (N=5) in C3, clear and concise in C3 (N=5), short interruption in C2 (N=3), timely, understandable and accurate in C3 (N=3), informative enough in C4 (N=7).

Negative Aspects Some found it distracting to focus on task C2 (N=1), and could be worrisome in C1 (N=2) if it lacked information: *"With respect C1 and C2, I think that it's just a buzzer where you are informed that something is happening but you feel a little bit isolated because you cannot do anything. Someone is telling*

you that something bad might be happening, but you cannot do anything, so I felt a little bit, let's say, trapped." [P27], but, overall, the notification can drive action for after-the-fact (N=2). In **C3** *[Indirect Parental Support]*, parents felt it was distracting to type a message (N=2) and less instant than a voice message (N=1) *"It was quick and easy to send a message, but maybe not as instant as being able to record a voice message."* [P21]. Some considered the interruptions as causing potential issues (N=2), others found the interruption low (N=3). In **C4** *[Direct Parental Support]*, one parent [P22B] mentioned it requires more work, and another [P34] mentioned the lack of game context was distracting. Some parents were more worried about distracting their child (N=2).

Finding 17: Notifications and alerts were appreciated for their clarity, timeliness, and ability to prompt awareness or later discussion, especially in serious scenarios. However, parents expressed mixed views on distraction: while many praised the app's convenience and low-interruption design, some found certain conditions distracting, especially when typing messages or lacking sufficient context.

Unfamiliarity, Uncertainties and Lack of Clarity **Parents' Perspectives**

Impact of Unfamiliarity A few parents were unsure how to support their child due to limited knowledge of VR (N=10), *"With our age gap and my lack of familiarity using newer technologies is something I'm very concerned with."* [P12], *"Do I really care so much about just a game? Because for me it's still hard to decide if it is just a game or if it's more reality already. Probably I'm too old."* [P23].

Hesitations in Supporting Children and Understanding Their Needs They were also unsure about what their child would want (N=1, P31) and how the app should support based on the individual (age, disabilities, etc.) (N=1, P14B). In **C3** *[Indirect Parental Support]*, parents were not sure if the message they sent was helpful (N=4), and some hesitated on what to say when writing the message (N=2) *"it was a bit unexpected, not knowing what advice to provide"* [P27]. In **C4** *[Direct Parental Support]*, similarly, parents were unsure if their audio was sufficient and beneficial (N=4) *"Without fully knowing what is going on in the game it is difficult to know if my voice message was beneficial"* [P31] and one wondered if the child ignored them *"I understand I can speak to my child in the game but what if she doesn't listen to me and ignore me?"* [P10].

Children's Perspectives

Confusion, Misunderstandings and Irrelevance In **C2** *[AEM Support Only]*, children were confused (N=4), they did not know exactly what was sent (N=2), and one child [Child2] did not understand what "parent was notified" meant. In **C3** *[Indirect Parental Support]*, children felt confused or frustrated (N=3). One [Child20] did not expect a message from their parent via Big Buddy (N=1), another [Child22A] did not realise it came from their parent (N=1), one [Child26] found it confusing (N=1), and another [Child9] comforting, but the parent's message was irrelevant (N=1). One child was annoyed due to a lack of awareness [Child20], and another [Child13] expressed frustration.

Finding 18: Both parents and children experienced moments of uncertainty and confusion across support conditions, particularly when the interaction was indirect or lacked sufficient context. Parents often felt unsure about how to support their child effectively, especially due to unfamiliarity with VR, uncertainty about their child's preferences, or not knowing whether their message text or audio was helpful. Children, in turn, were sometimes confused about who the message came from or its relevance, particularly in indirect support conditions.

7.3.6 Discussion

Reflection around the AEM and future work are provided in chapter 8 to situate this study's findings within the wider contributions of the thesis. Here we present a summary of the findings and references to corresponding guidelines in chapter 9.

7.3.6.1 Summary of Findings

This study explored how different levels of parental involvement via a mobile app influence both parent and child experiences in the context of social VR support following simulated harassment. I focused on four levels of involvement, ranging from **C1** *[No Support]* to **C4** *[Direct Parental Support]* in VR, with Big Buddy as an intermediary.

S8-RQ1: Parental Perceptions of Involvement via the App Overall, parents appreciated having a means of involvement, especially in **C3** *[Indirect Parental Support]* and **C4** *[Direct Parental Support]*, where they felt greater agency and control. The mobile app was viewed as usable and helpful, particularly because it enabled remote support without needing to wear a VR headset. However, the level of informativeness and degree of control were important: while some parents desired more detailed information about incidents and greater intervention capabilities (e.g., pausing the game or blocking users) others expressed concerns about overly intrusive control and invading child privacy, particularly in **C4** *[Direct Parental Support]*, where a parent's presence in VR was perceived as potentially invasive and embarrassing. Additionally, parents' unfamiliarity with VR, uncertainty about what to say, doubts regarding the effectiveness of their messages and reliance on Big Buddy highlighted the need for supportive guidance, including prompts, enhanced information, and increased awareness of Big Buddy's capabilities. Parents also adjust their involvement based on availability, relying more on the AEM's support during busier periods, but opting for direct or indirect support when they are available or when tasks are less important, still allowing them to respond.

S8-RQ2: Impact on Children's Emotional Perceptions and Experience. Children generally appreciated parental support, especially when it was timely, emotionally resonant, and minimally disruptive to gameplay. Many expressed emotional comfort and a heightened sense of safety when hearing their parent's voice (C4), although some preferred lower-involvement options (e.g., text in C3 or support via Big Buddy in C2). Importantly, children's preferences were nuanced: while some wanted

greater parental presence, others preferred autonomy and felt irritated or confused by irrelevant or unclear messages. The presence of Big Buddy was largely seen as emotionally reassuring, with children often preferring hybrid approaches where both the agent and the parent played complementary roles.

Based on these findings, I gather practical design guidelines in section 9.2, specifically guidelines 8, 11, 13 and 14.

7.3.6.2 From 2D to 3D Digital Parental Mediation

The findings not only align with and extend prior research on parental mediation and digital safety but also provide novel empirical insights into the underexplored social VR environments within emerging immersive platforms.

Consistent with foundational frameworks [112, 113], parents noted factors that influence parental involvement, including personal/family preferences, child age and context of situations, underscoring the importance of layered and configurable parental mediation strategies. Echoing [118], results highlight that parents prefer mediation tools that balance control with child autonomy and facilitate transparent communication rather than purely controlling mechanisms.

The challenges parents reported around unfamiliarity with VR and uncertainties about how to intervene are in line with prior studies emphasising gaps in parental digital skills [121, 117, 122, 123]. These knowledge gaps often lead to hesitation and reduced confidence in mediation, pointing to the need for improved onboarding and guidance, as I suggest. From the child's perspective, the findings reinforce prior research emphasising the importance of agency and self-regulation, as outlined in the TOSS framework [115, 117], described in section 2.2.4. Children expressed a strong desire to have control over when and how parental support is provided, highlighting the need for mediation strategies that respect their autonomy.

While TOSS was proposed as a complement to traditional parental control approaches, aiming to balance external restrictions with internal self-regulation [115, 117], results show that parents often relied on the AEM to support their child, and children may still want support based on their own decisions. This highlights the need to expand digital frameworks with two new pillars (the three existing pillars are described in Section 2.1). The fourth would aim to encourage and empower the child to request support based on personal preferences (e.g., child interaction with parent or AEM perceptible only to them). The fifth would introduce AI-supported parental mediation to enable timely interventions and bridge gaps in digital literacy and availability.

Additionally, the emotional impact of a parent's presence, whether through voice or avatar, aligns with earlier work on maintaining emotional connection while supporting independence [113, 53]. This study adds nuance by revealing the tension around avatar visibility in VR spaces (i.e., only the child can see their parent's avatar), where privacy concerns and potential intrusiveness call for flexible, child-controlled visibility options, an area not yet extensively explored.

Overall, this work provides novel empirical insights by examining family dynamics in social VR with a prototype system and paired parent-child interaction, advancing the design of parental mediation beyond traditional screen-based contexts toward immersive, embodied social spaces.

7.3.7 Conclusion

In this study, I conducted a four-condition VR study involving 34 parent-child dyads to investigate how varying levels of parental involvement, facilitated by a custom mobile app and the AEM (Big Buddy), impact parents' perspectives of their involvement and ability to support their child remotely without having to wear a VR headset. I also explored how these varying involvement levels influence children's emotional responses, sense of safety, and overall experience within a simulated social VR environment.

The findings showed that parents and children both preferred some degree of parental involvement in social VR safety, with stronger support conditions (Indirect and Direct Parental Support) rated significantly higher than no support. Parents generally favoured direct involvement and valued concise but informative notifications with video clips of incidents and control options, especially in serious incidents. However, they would adjust their preferred level of involvement based on their availability, often opting for the AEM's support when hands-on participation is not possible, and some consider their child's needs and consent. Some parents acknowledged their own limited knowledge of the VR environment and expressed a reliance on Big Buddy, trusting the AEM's protection capabilities, even sometimes with a degree of blind trust. At the same time, others showed self-awareness about the need to be more actively involved to ensure their child's safety. Children, meanwhile, felt safer and more supported with higher levels of parental involvement and parental knowledge, but often preferred less intrusive forms such as indirect support or messages through Big Buddy, which respect their autonomy and privacy. Both groups wished for customisable options for parental presence and in-VR parent avatar design to avoid embarrassment or disruption while maintaining connection and safety. Parents desired communication features that would enable two-way interaction (including brief, low-disruption feedback from the child, like emojis), greater transparency and context information about the incident and AEM's actions, as well as more control over interventions, while children would want agency over when and how parents are involved.

Overall, effective VR safety support requires balancing parental awareness and involvement with children's safety and the need for independence and privacy. While the instinct for many parents is to increase oversight to protect their child, the ultimate goal should be to equip children with the skills to protect themselves and navigate the world confidently. Systems should provide context-sensitive options that adapt to incident severity, parental availability, and family preferences, ensuring serious risks are addressed without micromanaging everyday interactions. Without such safeguards, tools risk enabling overbearing involvement that undermines a child's growth and resilience. This nuanced approach promotes safety, reassurance, and comfort for both parents and children within immersive social VR experiences.

7.4 Chapter Conclusion

This chapter aimed to investigate what the requirements of the AEMs are to safeguard children in social VR while still empowering them **(RQ4: Informed by RQ3, what are the requirements of the Automated Embodied Moderators to safeguard children in social VR while still empowering them?**). Building on insights from Chapter 6, I studied two levers: the level of child participation in AEM's moderation (Study #7) and the degree of parental involvement supported by the AEM and a custom-built app (Study #8). Guided by theories of child participation and parental mediation [352, 112, 117, 378] and prior stakeholder input (studies #5 and #6), the results in this chapter converge on a simple principle: safety should be adjustable.

From these two studies (along with the studies from the previous chapter), I produced design considerations towards eventual realisation of AEMs to safeguard children in social VR, around the lifecycle of a moderation incident (see chapter 8).

7.4.1 Contribution to Research Question 4

Study #7 showed that children prefer tools that keep them in the loop. When the AEM helps but still invites children into the decision process, for example, suggesting actions, asking for confirmation, letting them choose manual or automated responses, children report stronger agency and safety. Study #7 showed that parent-child dyads value parental involvement, but in ways that respect the child's context and privacy. Both parents and children rated stronger support conditions higher than no support; parents favoured concise notifications with context, such as short video clips with low disruption on both ends (child and parent). Children tended to prefer less intrusive, indirect parental presence mediated by the AEM. Across both studies, participants asked for customisability: escalation that matches incident severity, parental availability, and the child's preferences.

In summary, by giving children options to be part of the decision-making and by offering parents right-sized, context-driven ways to help, **AEMs should allow immediate protection without normalising constant surveillance or removing the child's say.**

In the next chapter, I propose an architecture of how the AEM can work in practice to safeguard children in social VR.

Chapter 8

Summary & Reflection of Part III

8.1 Design Guidelines of Automated Embodied Moderators to Safeguard Children in Social VR

In Part III, I introduced AEMs and conducted foundational user studies to identify design requirements that balance children's privacy and agency with parents' needs for reassurance and involvement, ensuring that AEMs are both useful and effective from the perspectives of children and parents alike. Operationally, these requirements suggest an **AEM architecture centred on a configurable triad of: child, AEM, and parent**, with explicit consent and low-disruption feedback channels for both parents and children. Considerations emerged in terms of 1) the System and 2) the Embodiment of the AEM. To consider individual preferences, social VR platforms should make it simple to change AEMs' automation levels, set thresholds for notifications and AEM interventions, personalise embodiment features and review or override AEM actions.

8.1.1 A Triadic Human-in-the-Loop AEM Framework to Safeguard Children in Social VR Pre/During/Post-Event

Considering the lifecycle of a moderation incident (a child harassed in social VR), and informed by the findings, the following is how I envision an AEM to work in practice. I highlight key considerations regarding the role of AEMs, child users and parental settings; what actions the AEM should take during the incident, how parents can be involved; and after-the-fact, what impact AEMs have on children, and how their role may evolve.

The following diagram presents the various factors considered from the setup to when an incident occurs (see figure 8.1), and the triadic actor interaction is represented in figure 8.2.

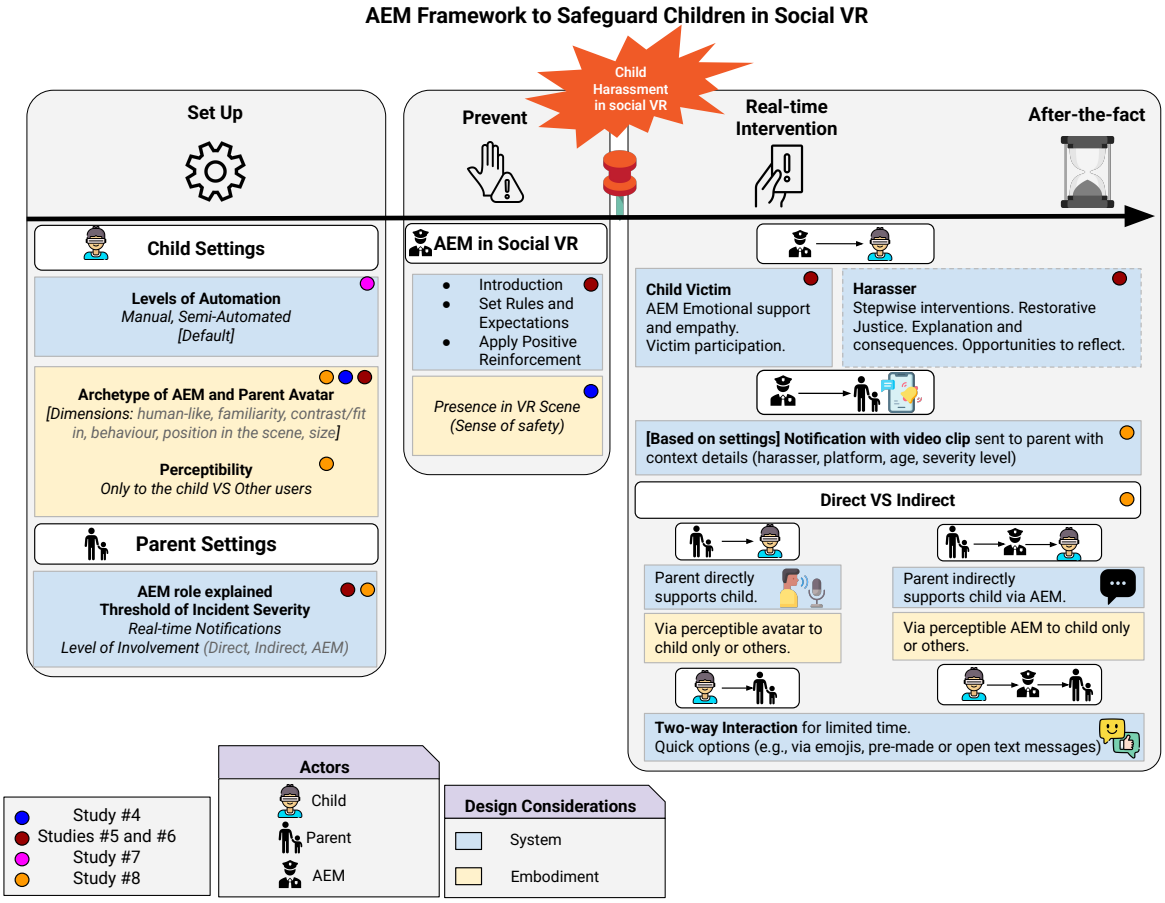


Figure 8.1: A Triadic Human-in-the-Loop AEM Framework to Safeguard Children in Social VR

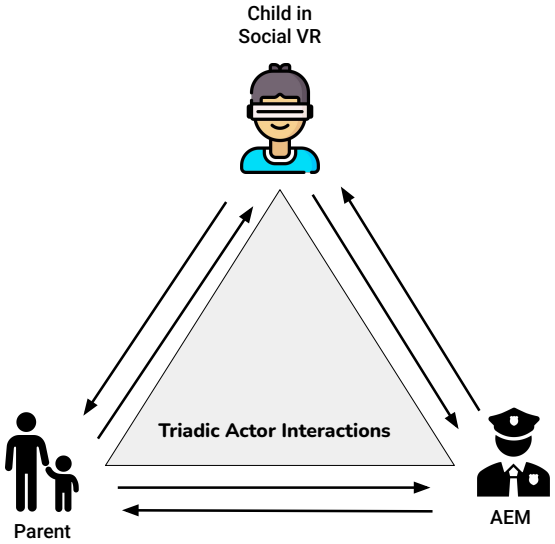


Figure 8.2: Diagram representing the three actors (Child in social VR, parent and AEM) in a triadic model.

Setup Phase **Parental settings** At the setup, parents should familiarise themselves with the AEM's role through a provided tutorial, for example, that explains the code of conduct, the AEM's functions, consequences for actions, and how parents can be involved with the AEM. Parents would configure how and when they wish to receive notifications and debriefs (e.g., via a threshold of severity of offence), and set personalised intervention methods of the AEM, such as helpful phrases, behavioural cues, or strategies tailored to their child's needs and preferences.

Child-Centred Solution Children should be able to customise their AEM in terms of embodiment and the system, thereby enhancing their sense of control, and potentially also trust. Indeed, the experts emphasised the AEM should avoid reinforcing stereotypes and consider cultural factors, promoting social cohesion and avoiding labels by suggesting neutral appearances or letting the child design its moderator's appearance features (from non-human to human-like to existing fictional characters) and communication features (from language to accents and tones) (Studies #5 and #6 Finding 7). They should also be able to choose the perceptibility of the AEM and parent interventions to others (i.e., if they want other users to see their parent and AEM in the scene when they provide support to the victim, for example) (Study #8 Finding14). Based on the clusters from study #4 (section 6.2.4.4), I would propose three default archetypes that can be readjusted and personalised: *a Coach, a Mentor and a Referee* (see table 8.1). The *Coach* would be an authoritative leader who is actively present, offering support, teaching, and clear direction for families who want expertise and a confident guiding hand. The *Referee* would be a light-touch authority, ensuring fairness but keeping out of the spotlight for families who value independence with minimal direct intervention. The *Mentor* would be a supportive and approachable figure, present and encouraging but not overly strict, striking a balance between friendliness and authority, visible enough to guide without being overbearing. Future research would need to validate these archetypes (see section 8.2).

Table 8.1: Examples of AEM Embodiment Archetypes based on Clusters from Study #4

Archetype	Children's Cluster Preference	Parents' Cluster Preference	Synthesised Attributes
Coach	Authoritarian, visible, humanised, teacher-like	Humanised, expert, formal, authoritarian, visible, assertive	Confident leader: actively present and directive. Seen as an expert teacher and authority figure, guides firmly and clearly.
Referee	Non-visible, background presence	Indulgent, non-humanised, amateur, informal, not too assertive	Light-touch authority: sets rules but stays in the background. Provides structure with minimal intervention and in an informal manner.
Mentor	Friendly, indulgent, visible, humanised supervisor	Visible, indulgent but formal and assertive, non-humanised	Supportive and approachable guide: visible and encouraging, balancing warmth with structure. Formal and assertive enough to maintain authority, but still indulgent.

The child can also set the level of automation, deciding how they would like help and if they would want to be given options if an incident is detected, having a say in the decision-making or if they would prefer to use manual interactive tools provided with the possibility of calling the AEM for help. These settings would be useful as they would allow crucial adaptation for children of different ages, developmental skills and personal preferences.

Prevention Phase The AEM's role When children join social VR, the AEM should introduce itself and explain its role and scope. When present in a room, it should transparently notify users. According to the Information Commissioner's Office (ICO) [294], children have the right to know if they are being monitored, with an "obvious sign to the child" [294]. AEMs offer an embodied, relatable route to providing such transparency. Children suggested a verbal announcement and introduction of the AEM if it is present in the social VR space.

The simple presence of the AEM within the VR scene was found to enhance participants' sense of safety (studies #4, #7 and #8). More broadly, research has established guardianship as a key strategy for crime prevention and community safety [382]. Within the routine activity framework, Cohen and Felson (1979) define a capable guardian as any person or object that discourages crime [383]. Authors further noted that the mere presence of guardians can deter offending, whereas their absence heightens the risk of crime. As Felson and Boba [384] put it, *"a guardian is not usually someone who brandishes a gun or threatens an harasser with quick punishment,*

but rather someone who mere presence serves as a gentle reminder that someone is looking.". Findings in studies #4, #7 and #8 extend this line of research by demonstrating that even the continuous presence in the background of a virtual guardian can enhance the sense of safety. Future research may further explore whether such virtual guardianship can not only increase perceived safety but also reduce actual instances of harassment.

Instead of relying solely on negative punishments when incidents occur that are similar to punishments in schools (Study #4), the AEM should establish clear expectations about appropriate behaviour and consequences for actions from the start and most notably and importantly, employ positive reinforcement and restorative justice. The latter is an approach highly recommended by experts as it may be more effective for long-term behavioural change [385]. Considerations regarding privacy and child data protection should also be taken into account, including what data can be accessed for the AEMs to initiate actions, ensuring transparency for users, for example.

During the Incident - Intervention and Support Automated Detection In case of an AEM that automatically detects incidents, it is important to note that incidents can be subjective. While the goal is to achieve high accuracy in detection, procedures must be in place to address errors, which could be false positives resulting in unintended consequences by misinterpreting a situation, or false negatives where the system fails to recognise an incident because of ambiguous or subjective harassment. This procedure may involve a human (either the user or a trusted third party, like a human moderator or parents) to verify outcomes, assist with decision-making, and explore the possibility of reconciliation. I propose the following workflow as default: the AEMs detect events, then discreetly ask the victim for confirmation; if confirmed, the intervention occurs. This approach aligns with adult user preferences for Human-User-AI collaboration in social VR settings [40].

Automated detection of harassment in social VR will require further multidisciplinary research, from AI to psychology, to HCI, for accurate harm detection and interpretation, and for developing strategies to address mistakes and system failures.

Taking Action Once the incident is detected, tailored interventions to harassment type, the environment, and interaction types (e.g., one-on-one vs group interactions) should take place. Responses should be proportionate to severity and frequency, ranging from warnings to immediate actions. In contrast to experts, there was a particularly high prevalence of guardians and children recommending immediate actions (sanctions) before reflection and repair. Some immediate proposed intervention approaches also seemed to align with the concept of altering the physical environment, such as creating greater separation between avatars or relocating to entirely different virtual spaces.

Informing Parents and Optional Support Based on the settings (notification threshold), child preferences in terms of parental involvement and perceptibility, the AEMs can be programmed to send parents real-time notifications about incidents, including contextual explanations of what happened with video clips and guidance

for talking with their child. Depending on their preference settings and availability, parents may rely on the AEMs' support, or they may participate in supporting their child indirectly or directly in the VR scene, with the child's consent, either perceptible to the child only or to other users as well.

Two-Way Interaction Parent-Child The child may respond to the parental support with quick, low-disruption messages or emojis to reassure the parent, as findings from study #8 showed the desire for two-way interaction with the child for reassurance and checking they are okay. The child may also accept having a limited synchronous conversation with their parent via text message or a live voice call (perceptible or imperceptible to other users).

After the Incident - Long Term Impact ***Psychological Impact*** Immediate actions can stop the harassment, but the AEM's intervention and embodiment must be carefully considered for their impact on children, whether harasser or victim. Empathy and understanding the emotional impact of actions are important for both victims and harassers, as harassers could be victims of the past and victims may become harassers [386, 387]. Therefore, interventions must support the mental well-being of child users. As agreed by the experts and prior work [388], harsh tactics like negative punishments without explanation, reflection and support, may lead to self-harm or extreme reactions from children. The actions the AEM takes must avoid causing damage and toxicity and instead, focus on positive reinforcement and facilitating conversations and open discussions. They should also involve the children in the decision-making process, with discretion if preferred, to empower them. From the findings, it is also suggested that the AEM serves as a link, connecting the child users to guardians or other resources like child helplines.

The AEM's Role VS Parent's Role By imbuing AEMs with a customisable, human-like personality, it can potentially become akin to a tutor or a friend. However, this ambiguity of scope raises ethical concerns, necessitating the clear definition of an AEM's role and continuous evolution within the system, with adaptation of rules over time and ensuring transparent explanations of system functionality and limitations to child users.

Parents may come to rely on the AEM, which could lead to improving the parent-child relationship in the case of children becoming indifferent to the rules set by the parents. However, it might also hinder the responsibility of the parents, relying solely on the AEM and neglecting parenting duties. Children may also miss out on valuable opportunities for self-development and building social resilience, which can diminish their ability to stand up for themselves. Therefore, control, agency, reflection and learning should be accounted for as part of the system's design.

8.1.2 Journeys of Families Navigating AEM

This section is only a suggestion of potential journeys based on the findings. The aim is to accommodate both parental and child needs. Future work is needed to validate and further research these suggestions (see section 8.2).

I present four exhaustive journeys of triad parent-child-AEM based on the level of involvement a parent may want and the level of participation a child may want. I have illustrated the first two journeys into a diagram figure 8.3. In these scenarios, I consider a child desiring high agency as one who wants a high level of participation and a child desiring low agency as one who wants a low level of participation. However, the child's capability of agency and their desirability of agency may vary and may not be aligned. As mentioned previously, agency is a complex process that can change depending on the situation [66]. For example, a child who may desire high agency may also "freeze" in a disruptive situation, which is why I do not suggest fully manual solutions for those children. Settings may also be re-adjusted by the children and parents over time, and a child may transition from one journey to another.

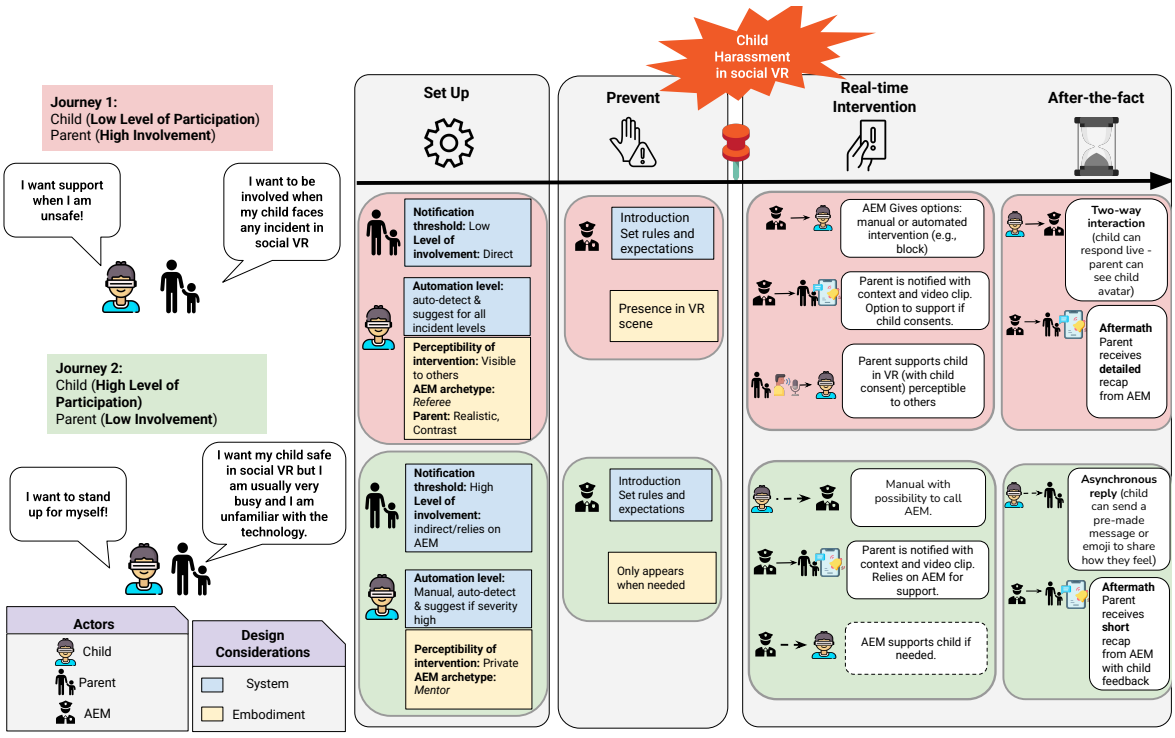


Figure 8.3: Schema representing two journeys as examples to see how the system may work with different preferences from the child and the parent.

Journey 1: Child desires low agency, Parent desires high involvement In this journey, notifications sent to parents are set to a low threshold, meaning parents will be alerted even for incidents classified as "low severity". The level of parental involvement set is direct (i.e., the parent may want to enter the VR scene to directly support their child). The AEM archetype is *Referee*. It enables the parent to be engaged in their child's VR experiences while intervening when needed. The level of automation is semi-automated, where the AEM detects the events and gives options to the child. During prevention, rules are set, and the AEM is present in the VR scene throughout the child's play. In the intervention phase, the AEM suggests options for manual or automated actions, while notifying the parent with context and a video. If the child consents, the parent can enter the VR scene in a way that is perceptible to others. Afterwards, the child and parent can interact directly and synchronously, and the parent receives a detailed recap from the AEM.

Journey 2: Child desires high agency, Parent desires low involvement Here, the AEM archetype is *Mentor*. It prioritises child autonomy while keeping parents informed only when necessary. Notifications are set at a high threshold, with parental involvement being indirect and reliant on AEM. The level of automation is manual, except for high-severity events where the AEM may suggest options to the child if no action has been taken. Prevention focuses on establishing rules, with the AEM appearing only when needed. During intervention, the child can call for AEM support, while the parent receives context and video but relies on AEM for active assistance if needed. Afterwards, the child can send a check-in asynchronously with simple messages or emojis to the parent, and the parent receives a brief recap with the child's feedback.

Journey 3: Child desires low agency, Parent desires low involvement In this case, the AEM takes on a *Coach* archetype and manages most of the oversight while parents remain in the background. Notification thresholds are high, with reliance on automated detection and suggestions for all incidents. Prevention includes rules and expectations, with the AEM present in VR as a steady guide. Interventions can be automated or semi-automated, with parents receiving real-time updates but leaving support primarily to the AEM. The visibility of the intervention depends on severity. After incidents, the child can interact directly with the AEM to express their feelings, while parents receive concise recaps paired with tailored support suggestions.

Journey 4: Child desires high agency, Parent desires high involvement This journey balances child independence with parent engagement, supported by a *minimalist Mentor-style* AEM. Notifications are set low, and parents remain directly involved but without intruding on the child's autonomy. The level of automation is manual, with suggestions triggered only for severe cases. Prevention includes setting clear expectations, but the AEM appears only when necessary. During interventions, parents receive real-time notifications with context and a video, but child consent determines the mode of support, either allowing parental presence in VR or limiting it to private messages mediated by the AEM. Afterwards, the child can reassure their parent asynchronously with simple responses, while the parent receives a detailed recap from the AEM to reassure them.

8.2 Limitations and Future Work

Despite limitations, chapters 6 and 7 highlight a rich space for future exploration. Several promising directions emerged.

Practical and Technical Challenges Implementing effective interventions from the AEM and the parents in social VR environments involves overcoming several practical and technical challenges. It requires developing the "backend" of the AEMs, identifying suitable AI technologies capable of detecting harassment, exploring which LLMs or similar models can enable meaningful and child-friendly conversations, real-time monitoring, supported by VR headsets with sufficient processing power to run these AI models either locally or via a cloud. Throughout Part II, I assumed that automated detection would be feasible and therefore designed Wizard-of-Oz prototypes, focusing on how AEMs could be presented, the level of automation and how parents could be involved while balancing child safety, privacy, and agency in social VR. Future research should explore how AI can detect harassment using, for example, biometrics (e.g., heart rate, eye gaze), voice chat analysis, behavioural changes, proxemics, and non-verbal cues [7] while minimising data collection and adhering to international regulations, such as the GDPR and the DSA. Referring to my published work ([6, 4, 5]), automated detection has begun to be explored through hybrid pipelines that incorporate appeal mechanisms to address potential false positives, cases where harassment in social VR may be incorrectly detected [389]. More recently, a study developed a system, HardenVR, which leverages pose information in social VR as contextual data to detect harassment, achieving a detection accuracy of up to 98.26% in experimental evaluations [208], demonstrating the potential of automation for harassment detection in social VR environments.

Moreover, standardised APIs must be established to enable parental control applications to access and manage data consistently across diverse social VR platforms. With safety features varying across platforms [9], standardising these tools would enhance child safety and well-being while increasing parental reassurance. To maintain uniform safety standards, integrated and enforceable moderation controls should be accessible through these APIs. Additionally, a standalone mobile or web-based interface is essential to provide parents with configurable, extensible oversight of multiple social VR applications across multiple children from a single dashboard. This approach could also be applied to other emerging social immersive environments. Realising this vision demands close collaboration among platform developers, device manufacturers, and app creators to build an interoperable ecosystem that facilitates seamless and effective parental mediation.

Exploring Additional Child and Parent Perspectives to Enhance AEMs for Safeguarding Children in Social VR This part's user studies revealed aspects of child desirability in terms of agency and level of participation through qualitative data, for example, children expressing a desire to "take action instead of something just swooping in" [Child18] or "wanting to stand up for myself and have control." [Child2] (study #7). Future research should aim to investigate child agency capability and how it may align with or differ from child agency desirability, to better inform the design of settings and customisations that support it. Similarly, although parental

involvement was conceptualised as a continuum in section 1.2, it was measured using a discrete Likert scale (study #8). Future work is needed to more accurately map system settings to parents' sense of involvement and to understand how both child agency and parental involvement evolve based on different interactions and situations in social VR.

The sentences and prompts used by the AEM to convey messages in the simulations were intentionally kept simple and concise for the purpose of this thesis. However, it would be interesting to investigate and validate the appropriateness, child-friendliness, and effectiveness of these messages.

Research should also assess long-term impact, looking at how prolonged interactions with AEMs may influence children's social skills, resilience to online harassment, and trust in digital systems. This would help refine AEMs to support children's emotional and psychological development effectively.

Personalisation and Customisation of AEM Embodiment, Automation and Parental Involvement Personalisation of AEM Embodiment, Automation and Parental Involvement Even though, some findings showed prevalence in aspects of embodiment (e.g., apparent clusters of attributes in study #4), level of automation and level of parental involvement preferred via ranks and quantitative data (studies #7 and #8), the qualitative data also showed the importance of customisation and personalisation of aspects as each participant may have individual preferences and needs. Therefore, while I propose settings, AEM archetypes and journeys, future work is needed to better understand what settings may be needed in terms of embodiment of AEM, the system and parental involvement, considering long-term changes, age, developmental needs and family preferences. Exploring embodiment in AEMs, such as how different forms of embodiment in practice affect children's perception of safety and support in different situations, could guide the design of AEM interfaces.

While parents and children may wish to control how perceptible the parent's presence is to others, future work should also investigate how the child appears to other users during child–parent interactions, and how these interactions can be made more discreet (e.g., through messaging, as participants suggested, or by masking them with animations and other visual cues).

Considering Additional Perspectives Although Studies #5 and #6 focused on children, parents, and grandparents in the UK, future research could broaden the scope to include other trusted figures, such as siblings, teachers, or close friends, and explore diverse cultural contexts to gain a more comprehensive understanding of support networks.

Moreover, I mainly focused on the "pure" victims' perspectives as described in Chapter 2 (see section 2.2.3.1). Future work should explore the experiences and perspectives of the different profiles of bullying and victimisation, including "pure" bullies and "bully-victims", to design interventions that are better tailored to their specific needs and psychological characteristics (e.g., low self-control, anger as described in prior work [49]).

PART IV
Conclusion

Chapter 9

Contributions & Final Reflection

In addition to the reflections presented in chapter 5 and chapter 8, this chapter offers broader reflections and conclusions on the thesis as a whole. First, I outline the overall contributions. Second, I present practical design guidelines derived from the findings of the studies that can be used to develop safety features for children in social VR. Third, I discuss the methodological approaches employed to evaluate safety tools for children in social VR. Finally, I synthesise the answers to the thesis' research questions, propose directions for future work, and conclude the thesis.

9.1 Contributions Overview

9.1.1 Empirical Contributions

This thesis makes a number of empirical contributions that help us understand children's experiences in social VR and contributes empirical findings on children's and parents' experiences using developed safety tools in simulated social VR across eight studies. In total, **70** social-VR-aware adults, **79** social-VR-aware parents, **16** experts, **164** children aged 8-16 years old, **53** parents and **3** grandparents participated in the studies of this thesis.

In part I, through three empirical studies (a mixed-methods questionnaire with 79 parents and 70 non-parent social-VR aware adult users, interviews with 16 experts and a user study with 40 children 8-16 years old) the research 1) explores and better understands issues and concerns associated with child use of social VR and 2) informs the design of child safety features in social VR.

In part II, through five empirical studies with children aged 8-16 years old, parents and experts (user studies with 43 children and 17 parents, 34 children and 34 parent-child dyads respectively, workshops with 13 children 8-16 years old, 5 parents, 3 grandparents and interviews with 16 experts), the research 1) explores both key factors of AEMs from stakeholder perspectives and 2) identifies the requirements for AEMs to effectively engage both children and their parents, while ensuring children remain empowered and supported in the process.

Based on findings, I propose 14 guidelines to safeguard children in social VR (see section 9.2.

9.1.2 Methodological Contributions

In terms of methodology, as described in section 9.3, this thesis presents ethical ways of evaluating child safety tools. The evaluation can either be from a first-party perspective with provocative scenarios to create an emotional response or a two-step procedure, from third-party to first-party perspective, where the child does not directly face the harassment situation themselves but acts and uses the tools as if they were the victim.

9.1.3 Conceptual Contributions

Finally, I propose the concept of AEMs, AI-driven embodied safety agents that respond to harassment in social VR in real-time, designed with parents, children and experts. I developed "Big Buddy", a Wizard-o AEM prototype, investigated how children and parents can participate with AEMs, and I present an initial framework of how it would work in practice in chapter 8.

9.2 Design Guidelines Derived from Findings to Safeguard Children in Social VR

Based on the findings of this thesis' studies, I provide practical guidelines for designers to develop child-friendly safety features in social VR. While these guidelines can also support the design and development of AEMs, AEM-specific design principles are discussed in chapter 8. These guidelines are grounded in the findings of the studies and open promising directions for further research.

9.2.1 Preventive, Proactive and Reactive Harassment Interventions

Guideline 1 - Tackling intentional harassment: enforce rules and community guidelines, consequences and stakeholders' responsibility

[Based on study #2, Finding 4 and Finding 10]. While most safety features in social VR rely on the child to enact reactive protection measures, preventive and educational solutions are crucial. Research would be required to better understand the root causes of harassment in social VR environments, i.e., whether harassers themselves were subject to abuse, if they have insecurities or other reasons for their behaviour. This would enable designers and developers to create safer social VR experiences for children and adults, potentially reducing harassment. The responsibility also lies with the industry in establishing standards and social norms in social VR, supporting meaningful parental involvement.

As preventive solutions, tips and in-game training/tutorials to use the safety features and abide by community guidelines could be shared via, e.g., in-game nudges, to dissuade users from harassing and set appropriate expectations for behaviour within these communities. Such efforts could, for example, draw from Instagram's approach to

protecting users [390] via warnings to discourage harassment or prevent unwanted interactions that would automatically hide users that are not in the child's list of friends. In light of the recent OSA in the UK section 2.2.5.2 and age verification requirement, age-matching would also help reduce unwanted interactions between minors and adults, but would require robust age-verification, which may use biometrics (itself a notable challenge and open to age ambiguity). Even if effectively applied, the latter may lead to privacy issues and identity leakage, whilst age-grouping may lead to the children not learning how to interact with adults or children of other ages. Challenges are faced by the industry in balancing engagement, costs, and safety. Therefore, there needs to be regulations to enforce safety (e.g., OSA [139]) and a collaboration between multiple stakeholders, including the platforms, user communities, and families, as shown in prior work investigating approaches to address online harm [371].

Guideline 2 - Tackling unintentional harassment: encourage harasser reflection, discuss with moderators, opportunities to explain and apologise and positive reinforcement

[Based on study #2, Finding 2, Finding 6 and Finding 10].

Experts also recommended proactive and reactive strategies that, instead of focusing solely on sanctioning harassers, foster a positive environment by rewarding respectful behaviour and offering opportunities for change.

Harassers should be allowed to reflect, learn from their actions and apologise, especially in mild cases. This could be done via a human moderator in a safe space. Social skills training in VR could also be put in place to help children or even adult users understand the impact of their actions on others and develop empathy [263]. A prior study investigated ways of designing social media platforms for people with intellectual disabilities [391]. One of the features that was included is education strategies for cyberbullying awareness, integrating structured teaching and video modelling into platforms [391].

Users could also be allowed to change their behaviour and receive positive reinforcement in the form of rewards. This approach could not only set a standard for desirable conduct but also motivate users to adopt positive behaviours and social cues. This follows Skinner's Reinforcement Theory [392, 393] whereby behaviour is strengthened by rewarding. Other forms of positive reinforcement, such as the Social Learning Theory [394] that focuses on imitation via attention, retention, reproduction and motivation, may be further researched as a form of improving behaviour in social VR.

Recently, according to OSA in the UK [139], companies are mandated to have the opportunity to appeal before content undergoes removal or moderation. This provision enables harassers to offer explanations and engage in reflection.

9.2.2 Safety Tools

Guideline 3 - Implement quick, easy, familiar, fun and visible-grabbing-attention tools to be used when under pressure

[Based on study #2, Finding 1, Finding 3, Finding 6, Finding 8 and Finding 9]. The safety features should be quick, easy, familiar (e.g., derived from existing social media tools or real-world metaphors) and available and able to be used by a child when under pressure. This could be done via a visible Panic Button that the child could quickly identify and use without having to go through multiple menus. Pre-prompts may also be available, such as "I don't feel comfortable" to help the child initiate confrontation or alert a moderator. Social VR platforms are constantly implementing new tools, such as Meta, adding a 'Pause' in Horizon Worlds social VR platform (June 2023) that lets the user take a break from a situation or space [395].

[Based on study #3, Finding 3]. Children want quick, fun and interactive tools. They appreciated the enjoyable aspects of safety tools as it provided relief and decreased stress. Moreover, the "Magic Wand" *[PointBlock]* complements the fantasy of the virtual environment, without interrupting immersion and presence, while allowing space for imagination and personal interpretation of what happens to harassers once they disappear. Nevertheless, safety and practicality must remain the primary focus. Overly engaging tools may lead to potential gamified misuse and enact abuse.

Guideline 4 - Customisation of safety features based on child's individual needs

[Based on study #2, Finding 7 and Finding 8]. At the setup, a questionnaire may be filled in by the child and/or the parent, which could adapt settings based on the responses. The questionnaire could consider age, emotional maturity, personality traits, or needs, for example. The language, the visibility and accessibility of the safety tools, as well as the UX/UI interface, may also need to be adapted depending on those preferences.

Recently, Rec Room has offered Junior Accounts that users can set up initially and alternatively (October 2023), if their moderation system detects users under the age of 12, their account is automatically converted to a Junior Account (no information is given on how these detection systems work) [396]. Junior accounts limit communication features such as voice audio, in-game text messages, as well as attending events or joining clubs [396].

Guideline 5 - Tools taking into account the number of harassers

Scenario: In a virtual world like VRChat, a child enters a public space and is quickly surrounded by a small group of users who begin shouting slurs and making inappropriate comments.

[Based on study #2, Finding 3 and Finding 7]. Existing safety features should enable the capacity to quickly take simultaneous actions on multiple users at once. If the design of the tool is geared towards "one-user-at-a-time" moderation, the delays experienced will expose children to further harm. The implementation of automated moderation can also contribute to addressing situations efficiently, considering both scalability and the potential volume of cases.

[Based on study #3, Finding 2]. For example, the designer could implement a tool such as *[PointBlock]* for children to quickly and precisely block harassers, allowing them to regain control of their space and stop the harassment from specific users quickly. However, if the child enters a scenario where a larger group of harassers (e.g., a "flash mob" of 10 or more users) targets them with coordinated harassment like noise bombardment or surrounding the child physically in the virtual space, *[MassBlock]* would be a helpful tool for blocking. This feature could allow for blocking all of the harassers simultaneously, preventing the situation from escalating and offering a moment of safety to reassess and decide how to proceed.

Guideline 6 - Using perspective manipulation for enhanced safety

[Based on study #3, Finding 3]. Scenario: A child is playing a social VR world when a group of users begins to swarm and uses their virtual avatars to crowd the child's personal space, making them feel increasingly trapped and vulnerable. To empower the child and reduce the overwhelming sensation of being physically cornered, the designer could implement a perspective manipulation feature. This would adjust the child's viewpoint, such as placing them above the group, allowing them to see over the crowd and perceive the harassers as smaller and less threatening. This not only helps the child regain emotional control but also minimises the physical intimidation caused by the harassment, creating a sense of detachment and safety.

9.2.3 Child Participation and Agency

Guideline 7 - Children need a sense of control while receiving assistance

[Based on study #7, Finding 1, Finding 2 and Finding 3.] Children place a high value on having control over their protection. This underscores the importance of providing opportunities for children to exercise agency while interacting with the AEM's interventions. Providing children with choices within VR interventions is also essential for promoting a sense of agency and empowerment. In practice, this means allowing them to decide how and when they interact with AMs. Whether it is choosing between manual or automated interventions, or selecting how much involvement the

AEM should have, offering these options ensures that children feel in control of their virtual experience. Settings for different age groups could be implemented as defaults or suggestions, where the level of AEM's assistance is stronger for younger children under 13 and decreased for teenagers.

Scenario: In a VR space, a group of users decides to harass a child by making obscene gestures and blocking their path. The harassment escalates over time, and the child feels unsure of how to handle the situation, particularly when there are so many harassers involved. The system could automatically detect this harassment, recognising the abusive behaviour and offering the child a range of response options, either tools like *[PointBlock]* or allowing the system to handle the blocking. The child maintains control over how they respond, with the system providing the necessary support to help them manage the situation without feeling overwhelmed.

Guideline 8 - Keeping the child in the loop for parental and AEM support

[Based on study #8, Finding 8, Finding 11, Finding 13]. When designing parental involvement with a companion app, it is important to acknowledge that preferences for involvement, from both parents and children, can vary depending on factors such as the severity of the situation, the child's age, parental availability and individual family dynamics (see Finding 8). Effective systems should therefore provide families with the ability to predefine clear thresholds that determine when and how interventions occur, ensure both parents and children are aware of what actions parents and children can do and the role of AEM, as well as establish preferred communication channels. This customisation at the setup ensures that the system adapts to the unique needs and comfort levels of each family. It also secures the child's consent and provides reassurance to parents who want to discuss matters with their child beforehand, while ensuring both parties understand what information and actions are visible and accessible on each side.

To maintain a balance between oversight and respect for the child's experience, it can be valuable to give the child quick and subtle response options, such as emoji reactions or brief text messages (see Finding 11). These responses could allow children to communicate their feelings and perspectives without breaking immersion or causing significant disruption to their activity. At the same time, this would enable parents to stay informed and engaged with their child's emotional state in a non-intrusive manner (see future work recommendation section 8.2). It is also critical that notifications and support mechanisms are designed to minimise interruptions and avoid imposing burdensome tasks on either children or parents. The system should be as unobtrusive as possible, respecting the flow of interaction while still providing meaningful support.

Importantly, children want to maintain a sense of agency over their experience, including deciding when to involve their parents or Big Buddy (see Finding 13). Designers should therefore consider incorporating child-led triggers and adjustable autonomy settings that allow children to decide when and how assistance is sought. This approach can better align the system with children's developmental needs and preferences, for

example, mature teenagers who may prefer self-regulation, underscoring the importance of agency as outlined in the TOSS framework [117]. Such a setting is predicted to foster empowerment, as demonstrated in study #7, where children participated in initiating and deciding on interventions alongside the AEM in social VR [3].

Ultimately, successful design in this area ensures that children feel heard and actively involved in decisions, providing flexible options for AEM and parental support that can adapt and evolve as they grow and mature.

9.2.4 Human Oversight and Involvement

Guideline 9 - Involving human moderators, buddies and parents

[Based on study #2, Finding 2 and Finding 5]. While there are rapid technological advancements allowing potentially fully automated moderation in social VR [40], experts underscored the importance of human-based moderation, specifically for moderators' ability to incorporate nuance and flexibility. Additionally, the inclusion of supportive buddies and parental involvement was recommended to assist children in navigating harassment situations in social VR.

Guideline 10 - Aim for a balanced parental involvement in social VR

[Based on study #2, Finding 11]. Rather than relying on extreme monitoring or rigid restrictions, parents are encouraged by experts to support their child's autonomy while maintaining safety through open communication and trust (section 4.2.4.5).

At the setup, parents should configure parental controls and safety settings relevant to their child's age and experience. Restrictions may be negotiated with the child. Setting up parental controls should be simplified with pre-suggestions for ease of use.

Parents should be responsible for their child's well-being. However, platforms must share responsibility by providing accessible controls, safety defaults, and user-friendly educational resources. Designers and policymakers should support families through shared responsibility frameworks. Parents should engage proactively through open conversations, setting expectations and maintaining open dialogue about experiences in VR. This includes creating a safe space for the child, regular check-ins, and co-exploring social VR spaces.

Parents play a key role in educating children about online safety, especially in immersive environments. This involves explaining risks using real-world metaphors and preparing children to respond to harmful behaviour. Just as importantly, parents themselves need education. This includes learning about social VR features, understanding potential risks, and knowing how to use parental tools. Training videos, onboarding sessions, and self-learning are critical in this process

As VR occludes from reality, it is difficult for parents to be aware of what is happening in social VR. Tools like mobile real-time notifications about incidents or child-triggered support features can help parents stay involved while respecting child agency. There is no one-size-fits-all approach. Involvement should be tailored to the child's age, maturity, and individual context. Younger children generally require closer guidance, while teens seek and deserve more autonomy. Adjust your role as your child grows, maintaining support while allowing for independence.

Guideline 11 - Addressing parental knowledge gaps and awareness through AEM-based support

[Based on study #8, Finding 6, Finding 7, Finding 10, section 6 and section 7.3.5.1]. When designing a parental companion app with an AEM-based moderation system, it is essential to recognise that parents often have varying levels of knowledge and confidence in managing their child's online or digital experiences. Many children and parents appreciate the presence of an AEM like Big Buddy, turning to it as a trusted guide for assistance. Some parents preferred to support their child directly, while others encouraged their children to independently seek help from Big Buddy when necessary (see Finding 6 and Finding 10). This behaviour highlights an important insight: **parents often recognise their own limitations in addressing certain situations effectively and thus would rely on the AEM as a knowledgeable authority**.

Designers should be aware that this reliance can range from healthy trust to a degree of blind trust, where parents may use the system as a substitute for their own active involvement. While the system's notifications are often perceived as more credible and informative because they come from an automated, impartial source (see section 6), this can sometimes lead parents to disengage from direct communication or oversight. It is crucial to strike a balance where the system supports parental involvement without encouraging passivity.

In addition, parents frequently expressed a need for clearer, contextual feedback regarding incidents flagged by the app. Videos were considered very useful; however, clips showed only the incident itself, and some parents expressed wanting more context (see Finding 7). Including a short window of events before and after the incident may make the videos more effective. Parents want to understand the nature of the issue, the actions recommended or taken, and whether their own responses have been effective. Without this, parents may feel uncertain about their role and hesitant to intervene.

To address this, designers should consider integrating onboarding tutorials that clarify the AEM's functions and best practices for parental engagement. A parent requested message templates and AEM-supported communication suggestions (see section 7.3.5.1). Future work could explore how to better support parents and whether such suggestions empower them to engage more confidently, reduce hesitation in initiating difficult conversations, and still support the child effectively. Overall, successful design should aim not only to fill parental knowledge gaps but also to foster an ongoing collaborative relationship between parents, children, and the AEM.

9.2.5 Child Comfort and Privacy

Guideline 12 - Tackling unwanted consequences: discreet ways of using safety features and enacting outcomes

[Based on study #2, Finding 3, Finding 4, Finding 5, Finding 6and Finding 9]. Discreet solutions should be developed, both in the way the child user would take action, but also in the resulting outcome. While social VR is synchronous, what the other users see or perceive around the use of safety features could be invisible or be portrayed differently between victims and harassers. For instance, employing a gesture-based blocking feature could remain effective without revealing the hand movements of the user facing harassment to the other users, and consequently exposing the use of this feature.

Guideline 13 - Managing parental avatar perceptibility and familiarity to balance presence, privacy, and intrusiveness

[Based on study #8, Finding 9, Finding 12, Finding 14, Finding 15 and section 13]. When designing how parents are perceived by the child and other users in immersive environments, it is crucial to balance the emotional impact of parental presence with considerations of privacy and intrusiveness.

The familiarity of a parent's voice and emotional cues, such as tone, can provide significant reassurance to children (Finding 12 and Finding 15). So systems should enable the transmission of these cues when appropriate. It should also provide alternative communication methods or allow users to mask such cues for greater privacy or when voice is unsuitable, for instance, when a parent is unavailable or a more discreet option is needed.

Designers should also recognise that the perceptibility of a parental avatar within the VR space directly influences children's comfort and their sense of safety (see Finding 9 and Finding 14). While many children and parents appreciate a visible parental avatar as a comforting, reassuring presence, there is a strong preference for this perceptibility to be limited to the child alone, shielding it from other users. This restriction helps maintain privacy, reduce embarrassment, and prevent any potential exacerbation of harassment. At the same time, some children may perceive a perceptible parent-child interaction by others as a protective presence to deter bullies (see section 13). To address this, the system should provide flexible controls that allow children to selectively toggle the perceptibility of their interaction with the parent by others in social VR. This can empower children to feel supported when needed while preserving their autonomy and privacy, without feeling embarrassed in front of others, within the immersive experience.

9.2.6 AEM Design in practice

Guideline 14 - Designing the AEM based on the triadic human-in-the-loop AEM framework to safeguard children in social VR pre/during/post-event

[See chapter 8.] I recommend designers and developers of social VR platforms to consider the framework proposed in chapter 8 when designing an AEM (see chapter 8, considering the lifecycle of a moderation incident (a child harassed in social VR), and key factors around the role of AEMs, child users and parental settings; what actions the AEM should take during the incident, how parents can be involved; and after-the-fact what impact AEMs have on children, and how their role may evolve. Overall, AEMs should be envisaged as a balance between moderators that enforce standards and rules, and safety companions that support people in the decisions they make and the options they have.

9.3 Methodological Approaches to Evaluating Safety Tools for Children in Social VR

Social VR is an immersive technology that can increase the risk and impact of harm towards children. Therefore, to assess safety tools with child participants, I had to follow strict ethical procedures to ensure there were no psychological effects on the children, while also maintaining ecological validity and standards.

I needed to find a way to create an emotional response in the child participants without exposing them to actual harmful harassment. This was important in order to evaluate the tools and understand the children's genuine responses in such contexts.

Drawing from prior works' methodological approaches section 2.4.3, I designed methods to evaluate safety tools in social VR under disruptive scenarios, incorporating both third-party (observer imagining themselves as the victim) and first-party (victim) perspectives. Ecological validity was inferred from participants' emotional responses; however, it is important to note that the methods themselves were not validated to determine which approach was most effective.

9.3.1 From Third-Party to First-Person Engagement

Based on research showing that empathy develops in early childhood [270, 271], I designed the experiments (in studies #3 and #7) so that the child participant acted as a third-party observer during the harassment scenario, but became a first-party user when interacting with the safety tool. This approach allowed me to gather meaningful insights into the tool's effectiveness without placing the child directly in a harassment situation.

In designing the VR simulation, several important considerations were made. The environment was child-friendly and engaging to reduce the potential emotional impact of the bullying scene. The characters were cartoon-like and positioned in a semi-circle in front of the participant, to avoid the impression that avatars were behind them,

which might otherwise cause the child to feel personally targeted. Striking a balance between realism and abstraction was also a challenge. I did not want the child to believe the scenario was real, but I still wanted them to understand it as a plausible situation. Therefore, the animations and audio were intentionally repetitive, mimicking a robotic and slightly annoying interaction that hinted at bullying but still felt more animated than real.

Engagement was also a key part of the study design, starting from the moment the child entered the experiment room. I introduced the study using interactive slides before starting the VR simulation. I used simple phrasing and showed images from the scene, telling the children:

> "Imagine you are the child being bullied in the centre. Think about how you would react, how you would feel, and what you think about the tool, whether it is helpful or not."

This approach helped the children to imagine themselves in the victim's position rather than as solely observers of the situation. Before starting the session, I also carried out a short quiz to check their understanding, such as: *"How many rounds will there be, how do you access the blocking tools, and what will you do in each round?"*.

The findings showed that the simulated group harassment scenarios had a clear emotional impact (see section 4.3.4.1). Both quantitative and qualitative results revealed negative emotional responses, with most children imagining themselves as the victim. This emotional engagement highlights the effectiveness of the scenarios in eliciting emotional responses while maintaining ethical research practices.

9.3.2 First-Person Perspective in Social VR: Designed Disruptions for Ethical Insight

While third-person perspective user studies allowed me to evaluate the safety tools in studies #3 and #7, I wanted to further investigate how a child may react when they are first-person victims. Rather than exposing children to simulated harassment scenarios, which would be unethical, I designed alternative situations capable of eliciting emotional responses without causing harm. I developed a simulated social VR environment game with disruptive situations based on prior research [44], that used "Object Provocation" where another NPC player ruined their game.

The findings of study #4 indicated that the simulated harassment scenarios successfully elicited emotional responses from children (including anger and sadness), suggesting that the design sufficiently provoked children without compromising ethical standards. Therefore, this approach enabled to assess the interventions of the AEM ("Big Buddy") in simulated social VR from direct child experiences.

9.4 Research Questions

Research questions 1 and 2 were answered in chapter 3 and chapter 4, with a reflection provided in chapter 5.

Research Questions 3 and 4 were answered in chapter 6 and chapter 7, with a reflection provided in chapter 8.

9.4.1 Research Question 1

RQ1

What are the issues and concerns associated with children's use of social VR?

This initial question aimed at establishing context and understanding the issues and concerns from non-parent adult and parent perspectives via a mixed-methods questionnaire.

The following quote illustrates the results well: *"With the added obfuscation a headset provides in seeing these age-inappropriate situations, it is very easy for a child to get into them without parental knowledge." - [Study #1 Non-parent Adult Social VR user (P13)].*

Results showed that despite age restrictions, a large proportion of children under 13 (43% in the data) regularly use social VR, which raises several concerns related to safety. One of the main issues is children's exposure to age-inappropriate behaviour and content. Both parents and non-parent adults report that children often face abuse or inappropriate interactions with adults. Some children even use social VR to seek moral support or to disclose personal issues to adults, which creates ethical and safety risks. A lack of supervision and regulation is another major concern. Responses showed that many children navigate these spaces freely, leading to worries about excessive independence and insufficient platform moderation. Parents and non-parent adults expressed concerns about the potential psychological effects of social VR, such as addiction, emotional strain, and the possibility that children may lose real-world social skills and abilities. Differences also exist between how parents and non-parent adults perceive and respond to these issues. Parents tend to prioritise trust, open communication, and guidance, while non-parent adults would favour stricter rule enforcement and stronger controls. Both groups agree, however, on the need for improved safety measures. Parents want more robust authentication systems, time limits, and better tools for supervision. Non-parent adults emphasise the importance of age verification, clearer age restrictions, and enhanced moderation by developers. There is a shared expectation that both parents and platform providers should take active responsibility for protecting children in social VR.

Overall, the key concerns centre on children's unsupervised access to social VR, their exposure to harmful behaviours, and the absence of effective, developmentally informed safeguards.

9.4.2 Research Question 2

RQ2

What design considerations can help shape safety features to address child harassment in social VR?

Moving beyond the problems and concerns of child usage of social VR from adult and parent perspectives, I then aimed to explore solutions and design considerations to safeguard children effectively, integrating the perspectives of 16 experts through interviews (study #2) and 40 children through a user study with three proposed blocking tools (study #3). Most of the design guidelines can be referred to: guidelines 1-6, 9, 12.

Familiarity Experts agree that safety in social VR for children is most effective when it builds on their existing social media experience and uses real-world metaphors to make safety concepts familiar and relatable or for parents to explain risks using real-world metaphors. Major design concerns of existing safety features include inconsistency across platforms.

Give Control and Empower Victims Effective safety tools should empower children to take immediate control in harmful situations, such as muting or blocking multiple users during group harassment. As the responsibility is usually placed on children to enact reactive protection measures, who may have limited ability to manage overwhelming situations, experts recommend safety features that are simple to use, discreet, supportive, and empowering, with options like panic buttons, preset prompts, safety bubbles, and reflection tools such as pause features.

In a more practical case with children evaluating blocking tools in a simulated social VR scenario of group harassment, children responded positively to the three proposed blocking tools, especially *[PointBlock]* and *[MassBlock]*, which they found fast, effective, and easy to use. In overall rankings, *[PointBlock]* was the most preferred tool for both small and large group scenarios, followed by *[MassBlock]*, with *[SelectBlock]* rated lowest. The *[PointBlock]* tool was described as fun and engaging, while *[MassBlock]* gave a sense of psychological relief during large-scale harassment. However, the *[SelectBlock]* tool was criticised for being slow and awkward to use.

Children also suggested new and improved safeguarding tools, proposing interactive and VR-specific actions such as throwing objects, using gestures, or pulling items from virtual pockets to block others. They recommended adding visual indicators like ghost avatars for blocked users and tools to mute bullies, pause the environment, or block everyone except trusted friends.

Manipulating Perspective of Events Viewing harassment scenarios from above made children feel calmer and safer, as the elevated perspective helped them emotionally distance themselves from the situation and see bullies as less threatening. They appreciated the better visibility, control, and discretion this view offered, allowing them more time to think and respond. Many reported feeling more empowered from this perspective, though a few disliked it due to fear of heights or disorientation, preferring instead to stay grounded but at a comfortable distance. Quantitative data supported these perceptions, showing a clear increase in calmness and perceived safety from the above view.

Adapt to Context and Personalise Safety systems should be continually updated for accuracy, privacy, and inclusivity while remaining resistant to misuse. These should adapt to children's age, social and cognitive abilities, and the severity of incidents.

Human VS AI Human moderators can play an essential role by providing authority, flexibility, and understanding of social VR norms, while automated moderation can offer scale and speed.

Less Punitive Approaches, More Encouragement of Good Behaviour According to experts, solutions should also encourage harassers to reflect on their actions and understand consequences, whether their behaviour was intentional or not. To discourage harassment, children suggested isolating harassers or using sensory deterrents while promoting empathy.

Consider Parental Involvement in a Balanced Approach Parental involvement is considered vital, but should form part of a shared responsibility for safety. A balanced approach is encouraged, focusing on parent–child communication, informed setup, and real-time awareness during incidents, tailored to each family's needs.

9.4.3 Research Question 3

As described in part II, existing real-time safety features in social VR have significant limitations. Moreover, AI-based moderation has demonstrated promise (see section 5.4.1) and embodiment in VR has proven to be highly influential, particularly for children (see section 5.4.2), and could be leveraged in safety-focused applications.

Therefore, I have proposed a solution: *Automated Embodied Moderators* (AEMs): *AI-driven, embodied safety agents that combine automated moderation with the affordances of VR embodiment.* AEMs are designed to detect and respond to harassment in real-time, visibly protecting via an embodied avatar within social VR.

RQ3

How are Automated Embodied Moderators (AEMs) perceived as a solution to safeguard children from harassment in social VR and what factors are considered important to the prospective usage of AEMs?

To answer this, I developed and tested *"Big Buddy"*, a Wizard-of-Oz AEM system, to evaluate children's real-time reactions to a disruptive scenario while their parents simultaneously viewed the disruption on an iPad. This was followed by a study involving key stakeholders (13 children, 5 parents, 3 grandparents and 16 experts) to further unpack the requirements, expectations, and concerns related to the feasibility and acceptability of AEMs in terms of safeguarding children in social VR via qualitative methodologies, interviews and workshops.

Results from the user study with 43 children aged 8 to 16 years old showed that Big Buddy's presence and interventions have significantly impacted children's reactions to the disruption. Its absence induced negative feelings (e.g., sadness, frustration), whilst applying the full set of punishments (C4) significantly reduced sadness compared to no intervention (C1). Interestingly, C2 has also been perceived as substantially fairer than C1, which led me to think that perceived fairness is a potential determinant of

children's post-harassment reactions. Big Buddy's perceived role is congruent with these results, as children compared him with a referee to keep people safe, punish bad behaviours and ensure game fairness. Its presence is associated with active punishments and increased perceived safety. Regarding punishments, some children underlined the unfairness of applying different punishments for the same action. They also criticised the ban and notifying parents as too harsh. However, others thought it was appropriate and preferred the most severe interventions. Children nonetheless raised concerns about Big Buddy's appearance and shared their design preferences. They can be clustered in three trends: 1) a more realistic authority figure, 2) a more friendly, indulgent moderator, and 3) a non-embodied moderator. Parents showed empathy to their children experiencing the attacks and said they felt reassured by the presence of an AEM such as Big Buddy. Parents provided similar feedback to children's and added that notifying parents may not be ideal. Its effectiveness relies on others' parents, and bullies may not be sensitive to it. Some proposed graduated interventions and an initial warning before intervening. They expressed similar design preferences as their children but mentioned that its appearance and type of intervention would depend on various factors (e.g., gaming context, children's preferences, bad behaviour characteristics). They were not completely convinced by letting an AI moderating system ensure the safety of their children in Social VR and expressed their desire to better understand the process and remain in the loop.

Then, as I captured broader perspectives of the AEMs crossing stakeholders' perspectives, findings showed that when designing and developing AEMs to safeguard children in social VR, designers should take into account both system and embodiment features. Summary figures of the features proposed by participants are provided in figure 6.14 and figure 6.15, reflecting on convergence and divergence of perspectives in chapter 5. The following synthesis draws on examples proposed by children, parents, and experts in studies #4, #5 & #6; however, it is not exhaustive and warrants further research and evaluation.

System features concern the operational design of the AEM, focusing on the rules, logic and interventions that ensure fairness, psychological comfort and transparency. Clear and consistently enforced rules should be established, with age-appropriate consequences and opportunities for children to participate in rule-setting. The system should support context-responsive actions, tailoring moderation to the severity, frequency, and nature of behaviours, and distinguishing between individual and group harassment. AEMs should prioritise positive support and reflection rather than purely punitive responses. Interventions should encourage rehabilitation, dialogue, and emotional understanding. Human-in-the-loop remains essential to ensure accountability, allowing AEMs to review automated decisions and users to appeal outcomes where appropriate.

Psychological support mechanisms should be integrated to provide empathy and reassurance. Discreet engagement options and shared decision-making should be available for victims. For harassers, interventions should follow a stepwise model, beginning with private warnings and escalating only when necessary. Educational and reflective activities, such as gamified learning moments, should be incorporated to encourage behavioural change.

Embodiment features address how the AEM appears, behaves, and interacts within the VR environment. According to experts, AEM should function as a buffer between victims and harassers. Its presence should be reassuring to children and convey both protection and trustworthiness.

Customisation is also a key strength of embodied moderation. AEMs can be designed to appear differently to different child users, for example, taking on a more authoritative form when interacting with harassers while adopting a comforting, supportive appearance for victims. Such personalisation should remain age-appropriate, culturally sensitive, and inclusive. Children expressed preferences for creative embodiments, such as animals, familiar characters, or friendly avatars, and valued the ability to name or personalise the AEM to suit their comfort level. The AEM's level of visibility should be adaptable: some children prefer it to be constantly present as a background protector, while others favour a subtler form that appears only when needed.

9.4.4 Research Question 4

RQ4

> Informed by RQ3, what are the requirements of the Automated Embodied Moderators to safeguard children in social VR while still empowering them?

While a number of features were given by stakeholders as described above, I further investigated two levers with empirical user studies: 1) automation levels of the AEM (34 children), 2) parental involvement levels, while considering child privacy and agency (34 parent-child pairs). I then propose a triadic framework of how the AEM may work in practice with examples of parent-child pairs in chapter 8.

Children want AEMs to include them, with a preference for automated detection with a proposed list of actions for the child to decide.

Parental involvement should respect child consent and should be context-driven when incidents occur, offering a spectrum from AEM only to direct embodied parent presence, with defaults set to less intrusive options for minor issues and more direct involvement for serious incidents (these can be changed by parents in settings). When parents are unavailable, they should be able to rely on the AEMs, but should still be given options to support via messaging. Notifications should be concise, providing summaries but sufficient context with a video clip, while avoiding continuous streaming and practising data minimisation for privacy. Communication between parent and child should be low disruption and two-way, enabling quick, lightweight exchanges such as emojis, or "I'm okay" and "I need help" prompts that do not distract, embarrass the child or disrupt their activity. Transparency is essential, with clear explanations of what the AEM did and why. Embodiment should be customisable, enabling families to tailor the parents' avatar perceptibility to other users.

9.5 Future Work

While section 5.3 and section 8.2 discussed future work directions that are related to part II and part III, respectively, the thesis as a whole can be seen as a stepping stone to address issues related to child safety and parental involvement, in social VR but also beyond social VR.

Evaluating Safety Tools In Situ and with Longitudinal Studies New safety features proposed, such as *[MassBlock]* or *[PointBlock]*, as well as the AEM, could be evaluated in Situ, collaborating with existing social VR platforms to measure the impact of user behaviours over time and on a larger scale. I also call for companies to evaluate any new safety tools empirically for children, similarly done in this thesis, while transparently documenting the rationale behind tool development, providing evidence of their effectiveness, and actively involving stakeholders throughout the process.

Putting AEM Personalisation into Practice This thesis highlights the importance of embodiment in AEMs, with different clusters of attributes emerging and the need for customisation of AEMs in terms of appearance, communication and perceptibility that may enhance children's trust, sense of agency and safety. It would be interesting to examine these findings in practice, how different customisation options of the AEMs may influence child users' sense of comfort, trust, and safety in a social VR harassment simulation. Future systems could allow users to adjust the AEM's communication style, tone of voice, appearance, and behaviour according to their preferences. Another avenue of exploration involves creating archetypes or default AEM configurations based on a questionnaire and individual user profiles. With advances in AI, it may also be possible to design an AEM that learns from a child's gameplay, social interactions, and behavioural signals, adapting over time in a dynamic and personalised way. However, this may create privacy and ethical concerns.

Investigating Proposed Interventions While chapter 7 focused on investigating requirements of AEMs based on two main levers (automation levels and guardian involvement), future research should also examine the effectiveness of other proposed interventions from studies #5 and #6. In particular, preventive and proactive approaches, such as positive reinforcement via a reward-based system rather than a sanction-reward-based system.

Beyond the child victim's perspective This thesis also primarily focuses on the victim's short-term experience; future studies could consider the harasser's perspective and explore how interventions may affect long-term behaviour change. Study #1 indicated that adults also engage in harassment, and marginalised groups (e.g., women, LGBTQ+) have been reported as particularly vulnerable to such harassment [397, 398]. Although the interventions in this research were designed to protect children, similar systems could be researched for other groups of the population. For example, recent work has started exploring AI companions to protect women and other marginalised users in social VR [399]. While this thesis offers a framework for AEMs in chapter 8, this must be further researched and tested through collaboration with developers, designers, children and parents.

Diverse Sample Given that this thesis' research was conducted in the UK, future studies could explore the experiences of users from more diverse cultural and socioeconomic backgrounds. Such cross-cultural research would provide a more comprehensive understanding of how safety mechanisms are perceived and how they can be tailored to different communities.

Emerging Technologies Beyond social VR, Augmented Reality (AR) technologies and generative AI are developing rapidly. For instance, the global AR market size is estimated at $83.65 billion in 2024 and is predicted to reach $599.59 billion by 2030 [400]. Meta has already released AI smartglasses with Ray Ban [401], not to mention Orion AI glasses planned to be released in the next two years, as well as Snapchat SPECS AR glasses in 2026 [402], featuring integration of AI. As these technologies evolve, new forms of safety risk will emerge. Imagine a friendly virtual companion encouraging a child to influence their parents' purchasing decisions, or a malicious adult exploiting generative AI to manipulate a child's emotions. Children already face difficulties distinguishing between the virtual and real worlds [38]; differentiating between AI-driven avatars and human-controlled ones will be even more complex. Future research involving various stakeholders will be essential to identify effective interventions. The design guidelines and the concept of AEMs proposed in this work can serve as a foundation for these interventions, though they will need to be tested and refined in the context of other emerging technologies.

The directions outlined here are not exhaustive but illustrate the many opportunities for expanding upon this work.

9.6 Conclusion

This thesis aimed to explore the intervention space, including existing safety tools, as well as new possible tools that can effectively safeguard children in social VR, through eight studies addressing four research questions. In total, the research involved 203 children, 53 parents, 3 grandparents, and 16 experts, employing a combination of quantitative and qualitative methodologies, including surveys, interviews, workshops, and user studies. It demonstrates that the existing moderation solutions in place for safeguarding children from harassment in social VR are inadequate, highlighting the need for supervision and child-friendly tools (RQ1, RQ2). Instead, it proposes a novel approach to safeguarding these spaces using AEMs, designed with parents, children and experts (RQ3, RQ4).

The research makes empirical, methodological, and conceptual contributions and proposes design guidelines for creating child-friendly safety features in social VR. It also introduces and demonstrates the potential of AEMs. By considering the full lifecycle of a moderation incident, before, during, and after a child experiences harassment in social VR, the thesis proposes a practical vision for how AEMs could work in practice. It highlights considerations related to the role of AEMs, child users, and parental settings, as well as ways to involve both children and parents in the design and decision-making of AEMs.

Technology is moving fast, but safety needs to catch up! This work is a stepping stone towards child safety in social VR but as predictions show, the next generation of VR is coming (the global VR market is projected to grow from 20.83 billion in 2025 to USD 123.06 billion US dollars by 2032 [403]) with the convergence of generative AI, sensors and new accessories, but also other emerging technologies such as AR as described above (with an estimate of $83.65 billion in 2024 and is predicted to reach $599.59 billion by 2030 [400]), new issues and harms towards children may arise. If there is one takeaway from this thesis is: we should start putting the children at the centre of safety design, increase parental awareness and constantly and regularly update regulations, codes of conduct and safety features to adapt to the new harms and protect children, it should be a continuous loop, because even if the virtual world is not "real" it still can "feel real" and can have negative impact on children! Our children deserve better than this. Cristina Fiani

Appendices

Appendix A

Study #1

A.1 Intervention Categories Presented to Participants

In Block 3.3 of the questionnaire, participants were presented with the following parental interventions and asked to indicate the age range for which they would apply each, and to leave blank the intervention if they would not use it.

- **Pre-planned interventions**
  - I would explain to them the risks and benefits of social VR. *[Active Mediation]*
  - I would choose the platforms they can use. *[Restrictive Mediation]*
  - I would set up rules on usage (e.g., time limits). *[Restrictive Mediation]*
  - I would restrict functions such as audio or chat. *[Restrictive Mediation]*
  - I would allow negotiation about any rules I set up. *[Active Mediation]*
- **Knowledge (of what happens as it happens)**
  - I would supervise them by joining them in VR. *[Active Mediation]*
  - I would supervise them outside of VR (e.g., using a smartphone to view what they are doing) **with** their knowledge. *[Monitoring]*
  - I would supervise them outside of VR (e.g., using a smartphone to view what they are doing) **without** their knowledge. *[Monitoring]*
  - I would allow automatic recognition of misuses or problematic events (verbal and non-verbal). *[Monitoring]*
  - I would not observe what they are doing while playing but I would like to receive notifications in real time **in case of problematic events.** *[Monitoring]*
  - I would not observe what they are doing while playing but I would like to receive notifications in real time **in case of new social interactions.** *[Monitoring]*
  - I would not observe what they are doing while playing but I would like to receive notifications in real time **when they're starting a new activity.** *[Monitoring]*

- **Real-time interventions**
  - Based on notifications received, I would like to intervene remotely **by blocking or restricting access to activity.** *[Technical Mediation]*
  - Based on notifications received, I would like to intervene remotely **by disabling voice and text chat.** *[Technical Mediation]*
  - Based on notifications received, I would like to intervene remotely **by quitting the virtual environment my child is currently in.** *[Technical Mediation]*
  - I would want a computer-based guardian to supervise my child and intervene automatically **when necessary.** *[Technical Mediation]*
  - I would want a computer-based guardian to supervise my child and intervene automatically **based on my settings.** *[Technical Mediation]*
- **After-the-fact interventions**
  - I would observe what they did after the fact (e.g., reviewing video clips, chat logs). *[Monitoring]*

Appendix B

Study #3

B.1 Quantitative Summary Tables

Table B.1: Summary statistics by condition and likert scale Study #3.

Condition	Mean	Median	SD	Likert Scales
C1: *[SelectBlock]*	2.8	3.0	1.2	SAD-HAPPY [1-6]
C2: *[MassBlock]*	2.8	3.0	1.3	
C3: *[PointBlock]*	3.0	3.0	1.4	
C4: Top View *[SelectBlock]*	3.1	3.0	1.5	
C1: *[SelectBlock]*	3.2	3.0	1.4	UNSAFE-SAFE [1-6]
C2: *[MassBlock]*	3.2	3.0	1.5	
C3: *[PointBlock]*	3.5	3.0	1.5	
C4: Top View *[SelectBlock]*	3.7	4.0	1.7	
C1: *[SelectBlock]*	4.4	5.0	1.4	CALM-ANNOYED [1-6]
C2: *[MassBlock]*	4.4	5.0	1.4	
C3: *[PointBlock]*	4.2	5.0	1.5	
C4: Top View *[SelectBlock]*	4.0	4.0	1.6	
C1: *[SelectBlock]*	4.2	4.0	0.8	UNHELPFUL-HELPFUL [1-5]
C2: *[MassBlock]*	4.2	4.0	0.9	
C3: *[PointBlock]*	4.3	5.0	0.9	
C4: Top View *[SelectBlock]*	4.0	4.0	0.8	
C1: *[SelectBlock]*	4.5	5.0	1.2	BORED-ENJOYED [1-6]
C2: *[MassBlock]*	4.5	5.0	1.2	
C3: *[PointBlock]*	4.8	5.0	1.2	
C4: Top View *[SelectBlock]*	4.4	4.5	1.2	
C1: *[SelectBlock]*	4.8	5.0	1.2	DIFFICULT-EASY [1-6]
C2: *[MassBlock]*	4.6	5.0	1.4	
C3: *[PointBlock]*	5.0	5.0	1.1	
C4: Top View *[SelectBlock]*	4.7	5.0	1.2	

B.2 Non-Significant Statistical Results

Table B.2: Non-Significant Results from Two-Way ART ANOVAs for Group Size and Tools/Intervention Perspective (IP) Study #3.

Dependent Variable	**Independent Variable and Interactions**	**F-statistic**	**p-value**
SAD-HAPPY (1-6)	Condition (Tools)	$F(2,196) = 0.91$	$p = 0.404$
SAD-HAPPY (1-6)	Group:Condition (Tools)	$F(2,196) = 0.31$	$p = 0.732$
SAD-HAPPY (1-6)	Group (IP)	$F(1,118) = 0.87$	$p = 0.352$
SAD-HAPPY (1-6)	Condition (IP)	$F(1,118) = 3.38$	$p = 0.069$
SAD-HAPPY (1-6)	Group:Condition (IP)	$F(1,118) = 0.36$	$p = 0.548$
UNSAFE-SAFE (1-6)	Condition (Tools)	$F(2,196) = 3.02$	$p = 0.051$
UNSAFE-SAFE (1-6)	Group:Condition (Tools)	$F(2,196) = 0.09$	$p = 0.914$
UNSAFE-SAFE (1-6)	Group:Condition (IP)	$F(1,118) = 0.91$	$p = 0.341$
CALM-ANNOYED (1-6)	Condition (Tools)	$F(2,196) = 1.35$	$p = 0.261$
CALM-ANNOYED (1-6)	Group:Condition (Tools)	$F(2,196) = 0.32$	$p = 0.723$
CALM-ANNOYED (1-6)	Group:Condition (IP)	$F(1,118) = 0.37$	$p = 0.543$
UNHELPFUL-HELPFUL (1-5)	Condition (Tools)	$F(2,196) = 0.69$	$p = 0.501$
UNHELPFUL-HELPFUL (1-5)	Group:Condition (Tools)	$F(2,196) = 0.57$	$p = 0.565$
UNHELPFUL-HELPFUL (1-5)	Group:Condition (IP)	$F(1,118) = 0.88$	$p = 0.351$
BORED-ENJOYED (1-6)	Group:Condition (Tools)	$F(2,196) = 1.17$	$p = 0.311$
DIFFICULT-EASY (1-6)	Condition (Tools)	$F(2,196) = 1.23$	$p = 0.294$
DIFFICULT-EASY (1-6)	Group:Condition (Tools)	$F(2,196) = 1.95$	$p = 0.145$
DIFFICULT-EASY (1-6)	Group:Condition (IP)	$F(1,118) = 0.57$	$p = 0.451$

Appendix C

Study #7

C.1 Quantitative Summary Tables

Table C.1: Summary Data for Emotional Perceptions

Condition	Mean	Median	SD	Likert Scales
C1: manual	2.6	3	0.7	BEFORE SAD-HAPPY [1-6]
C2: Call AEM	2.5	2	0.7	
C3: AEM gives options	2.5	2	1.0	
C4: AEM fully automated	2.4	2	0.8	
C1: manual	5.0	5	0.7	AFTER SAD-HAPPY [1-6]
C2: Call AEM	5.1	5	0.6	
C3: AEM gives options	5.1	5	0.9	
C4: AEM fully automated	4.9	5	1.0	
C1: manual	3.1	3.0	1.4	BEFORE UNSAFE-SAFE [1-6]
C2: Call AEM	2.9	2.0	1.4	
C3: AEM gives options	2.9	2.0	1.4	
C4: AEM fully automated	3.0	2.5	1.3	
C1: manual	5.1	5.0	1.4	AFTER UNSAFE-SAFE [1-6]
C2: Call AEM	5.2	5.0	1.4	
C3: AEM gives options	5.3	5.0	1.4	
C4: AEM fully automated	5.1	5.0	1.3	
C1: manual	4.6	5.0	1.1	BEFORE CALM-ANNOYED [1-6]
C2: Call AEM	4.7	5.0	0.9	
C3: AEM gives options	4.7	5.0	1.1	
C4: AEM fully automated	4.7	5.0	1.0	
C1: manual	2.1	2.0	0.8	AFTER CALM-ANNOYED [1-6]
C2: Call AEM	2.1	2.0	0.7	
C3: AEM gives options	2.0	2.0	0.8	
C4: AEM fully automated	2.0	2.0	0.8	

Table C.2: Summary Data for Usability Perceptions

Condition	Mean	Median	SD	Likert Scales
C1: manual	3.5	5	0.5	UNHELPFUL-HELPFUL [1-4]
C2: Call AEM	3.5	5	0.5	
C3: AEM gives options	3.5	5	0.5	
C4: AEM fully automated	3.5	5	0.7	
C1: manual	5.1	5	0.8	BORED-ENJOYED [1-6]
C2: Call AEM	4.3	5	0.8	
C3: AEM gives options	4.6	5	0.9	
C4: AEM fully automated	3.9	5	1.2	
C1: manual	5.3	5	0.8	DIFFICULT-EASY [1-6]
C2: Call AEM	5.6	5	0.7	
C3: AEM gives options	5.5	5	0.7	
C4: AEM fully automated	5.7	5	0.8	

C.2 Non-Significant Statistical Results

Table C.3: Non-Significant Results from One-Way ART ANOVAs for Conditions Study #7.

Non-Statistically Significant of Conditions on Various Dependent Variables			
Dependent Variable	**Independent Variable**	**F-statistic**	**p-value**
SAD-HAPPY (Before, 1-6)	Conditions	$F(3,93.38) = 1.28$	$p = 0.284$
SAD-HAPPY (After, 1-6)	Conditions	$F(3,97.09) = 0.74$	$p = 0.533$
UNSAFE-SAFE (Before, 1-6)	Conditions	$F(3,96.32) = 0.58$	$p = 0.626$
UNSAFE-SAFE (After, 1-6)	Conditions	$F(3,98.37) = 1.36$	$p = 0.260$
CALM-ANNOYED (Before, 1-6)	Conditions	$F(3,95.43) = 0.03$	$p = 0.993$
CALM-ANNOYED (After, 1-6)	Conditions	$F(3,98.29) = 0.28$	$p = 0.841$
UNHELPFUL-HELPFUL (1-4)	Conditions	$F(3,95.59) = 0.28$	$p = 0.842$

Appendix D

Study #8

D.1 Parents Results

Parent Statistics Results Summary: One-Way ART ANOVA by Conditions			
Dependent Variable	**Effect Size (η_p^2)**	**F-statistic**	**p-value**
BALANCING TASK WITH SUPPORT	Large (0.22)	$F(3,99) = 9.18$	$p = 2.05 \times 10^{-5} < 0.05$
INVOLVEMENT	Large (0.51)	$F(3,99) = 34.40$	$p = 2.57 \times 10^{-15} < 0.05$
CONTROL	Large (0.49)	$F(3,99) = 31.58$	$p = 2.08 \times 10^{-14} < 0.05$
REASSURED	Large (0.46)	$F(3,99) = 27.65$	$p = 4.50 \times 10^{-13} < 0.05$
COMFORT	Large (0.44)	$F(3,99) = 25.67$	$p = 2.28 \times 10^{-12} < 0.05$
NOTIFICATION INFORMATIVENESS	Large (0.47)	$F(3,99) = 29.75$	$p = 8.49 \times 10^{-14} < 0.05$
CHILD PRIVACY	NA	$F(3,99) = 0.73$	Not signifcant $p = 0.535$

Table D.1: One-Way ART ANOVA results Parents.

Table D.2: ART ANOVA and Post-hoc Results Across Conditions for All Measures from Parent Participants

Analysis	**Result (significance level 0.05)**
BALANCE TASK WITH SUPPORT	
ART ANOVA BALANCE TASK WITH SUPPORT	$F(3,99) = 9.18,\ p = 2.05 \times 10^{-5},\ \eta_p^2 = 0.22$
Post-hoc Contrasts (Tukey-adjusted)	
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -3.16,\ p = 0.0111$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -4.76,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(99) = -3.76,\ p = 0.0016$
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(99) = -1.00,\ p = 0.7488$ (not significant)
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -2.16,\ p = 0.1423$ (not significant)
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -1.60,\ p = 0.3811$ (not significant)
INVOLVEMENT	
ART ANOVA INVOLVEMENT	$F(3,99) = 34.40,\ p = 2.57 \times 10^{-15},\ \eta_p^2 = 0.51$
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(99) = -2.69,\ p = 0.041$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -7.94,\ p < 0.001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -8.58,\ p < 0.001$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -5.25,\ p < 0.001$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(99) = -5.88,\ p < 0.001$
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -0.637,\ p = 0.9196$ (not significant)
NOTIFICATION INFORMATIVENESS	
ART ANOVA NOTIFICATION INFORMATIVENESS	$F(3,99) = 29.75,\ p = 8.49 \times 10^{-14},\ \eta_p^2 = 0.47$
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(99) = -5.41,\ p < 0.001$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -8.08,\ p < 0.001$

Analysis	Result (continued)
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -8.28,\ p < 0.001$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -2.67,\ p = 0.0428$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(99) = -2.87,\ p = 0.0255$
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -0.195,\ p = 0.9974$ (not significant)
REASSURANCE	
ART ANOVA REASSURANCE	$F(3,99) = 27.65,\ p = 4.50 \times 10^{-13},\ \eta_p^2 = 0.46$
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(99) = -3.50,\ p = 0.0038$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -6.69,\ p < 0.0001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -8.44,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -3.19,\ p = 0.0101$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(99) = -4.94,\ p < 0.0001$
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -1.747,\ p = 0.3051$ (not significant)
CONTROL	
ART ANOVA CONTROL	$F(3,99) = 31.58,\ p = 2.08 \times 10^{-14},\ \eta_p^2 = 0.49$
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -7.40,\ p < 0.0001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -8.36,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -4.86,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(99) = -5.82,\ p < 0.0001$
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(99) = -2.544,\ p = 0.0594$ (not significant)
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -0.963,\ p = 0.7710$ (not significant)
COMFORT with level of involvement	
ART ANOVA COMFORT with level of involvement	$F(3,99) = 25.67,\ p = 2.28 \times 10^{-12},\ \eta_p^2 = 0.44$
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(99) = -3.12,\ p = 0.0126$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -7.41,\ p < 0.0001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -7.31,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -4.29,\ p = 0.0002$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(99) = -4.20,\ p = 0.0003$
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = 0.096,\ p = 0.9997$ (not significant)
RESPECTING CHILD PRIVACY	
ART ANOVA RESPECTING CHILD PRIVACY	$F(3,99) = 0.73,\ p = 0.535$

D.2 Children Results

Children Statistical Results Summary - One-Way ART ANOVA by Conditions

Dependent Variable	Effect Size (η_p^2)	F-statistic	p-value
SAD-HAPPY	Large (0.23)	$F(3,85.76) = 8.39$	$p = 5.92 \times 10^{-5} < 0.05$
UNSAFE-SAFE	Large (0.36)	$F(3,85.76) = 8.39$	$p = 5.92 \times 10^{-5} < 0.05$
CALM-ANNOYED	Large (0.16)	$F(3,97.21) = 6.29$	$p = 0.0006 < 0.05$
UNSUPPORTED-SUPPORTED	Large (0.56)	$F(3,99) = 41.89$	$p < 2.22 \times 10^{-16} < 0.05$
IGNORE-KNOW	Large (0.53)	$F(3,96.95) = 36.13$	$p = 9.19 \times 10^{-16} < 0.05$
DISCOMFORT-COMFORT	Large (0.22)	$F(3,94.69) = 8.82$	$p = 3.27 \times 10^{-5} < 0.05$
DISLIKE-LIKE	Large (0.43)	$F(3,88.88) = 22.73$	$p = 5.18 \times 10^{-11} < 0.05$

Table D.3: One-Way ART ANOVA results for Children with effect sizes and significance.

Appendix D: Study #8

Table D.4: ART ANOVA and Post-hoc Test Results Across Likert Scales from Children Participants

Analysis	Result (significance level 0.05)
SAD-HAPPY	
ART ANOVA SAD–HAPPY	$F(3,85.76) = 8.39,\ p = 5.92 \times 10^{-5},\ \eta_p^2 = 0.23$
Post-hoc Contrasts (Tukey-adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(84.1) = -3.80,\ p = 0.0015$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(85.2) = -3.32,\ p = 0.0071$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(87.0) = -4.64,\ p = 0.0001$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(84.9) = 0.36,\ p = 0.9840$ (not significant)
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(86.7) = -1.04,\ p = 0.7267$ (not significant)
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(86.9) = -1.36,\ p = 0.5297$ (not significant)
UNSAFE–SAFE	
ART ANOVA UNSAFE–SAFE	$F(3,98.14) = 18.10,\ p = 2.00 \times 10^{-9},\ \eta_p^2 = 0.36$
Post-hoc Contrasts (Tukey-adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(98.0) = -4.74,\ p < 0.0001$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(98.3) = -3.58,\ p = 0.0030$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(98.0) = -7.24,\ p < 0.0001$
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(98.3) = -3.60,\ p = 0.0028$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(98.3) = 1.12,\ p = 0.6782$ (not significant)
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(98.0) = -2.50,\ p = 0.0664$ (not significant)
CALM–ANNOYED	
ART ANOVA CALM–ANNOYED	$F(3,97.21) = 6.29,\ p = 0.0006,\ \eta_p^2 = 0.16$
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(97.2) = 4.33,\ p = 0.0002$
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(97.2) = 2.13,\ p = 0.1496$ (not significant)
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(97.4) = 1.87,\ p = 0.2486$ (not significant)
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(97.2) = -0.25,\ p = 0.9946$ (not significant)
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(97.0) = 2.22,\ p = 0.1263$ (no significance)
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(97.2) = 2.44,\ p = 0.0759$ (not significant)
UNSUPPORTED–SUPPORTED	
ART ANOVA UNSUPPORTED–SUPPORTED	$F(3,99) = 41.89,\ p < 2.22 \times 10^{-16},\ \eta_p^2 = 0.56$
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(99) = -7.91,\ p < 0.0001$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -8.85,\ p < 0.0001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -10.15,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(99) = -0.94,\ p = 0.7836$ (not significant)
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(99) = -2.24,\ p = 0.1204$ (not significant)
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(99) = -1.30,\ p = 0.5667$ (not significant)
IGNORE–KNOW	
ART ANOVA IGNORE–KNOW	$F(3,96.95) = 36.13,\ p < 9.19 \times 10^{-16},\ \eta_p^2 = 0.53$
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(97.2) = -6.52,\ p < 0.0001$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(97.2) = -8.02,\ p < 0.0001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(96.7) = -9.75,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(96.7) = -3.18,\ p = 0.0106$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(97.2) = -1.50,\ p = 0.4392$ (not significant)
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(96.7) = -1.66,\ p = 0.3509$ (not significant)
COMFORT with parental knowledge	
ART ANOVA COMFORT with parental knowledge	$F(3,94.69) = 8.82,\ p = 3.27 \times 10^{-5},\ \eta_p^2 = 0.22$
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(95.2) = -3.78,\ p = 0.0015$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(95.2) = -2.72,\ p = 0.0384$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(95.2) = -4.93,\ p < 0.0001$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(94.2) = 1.08,\ p = 0.7006$ (not significant)
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(94.2) = -1.17,\ p = 0.6495$ (not significant)
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(94.2) = -2.25,\ p = 0.1176$ (not significant)
DISLIKE–LIKE	
ART ANOVA DISLIKE–LIKE	$F(3,88.88) = 22.73,\ p = 5.18 \times 10^{-11},\ \eta_p^2 = 0.43$
Post-hoc Contrasts (Tukey adjusted)	
C1 *[No Support]* vs. **C2** *[AEM Support Only]*	$t(89.6) = -4.42,\ p = 0.0002$
C1 *[No Support]* vs. **C3** *[Indirect Parental Support]*	$t(88.8) = -5.59,\ p < 0.0001$
C1 *[No Support]* vs. **C4** *[Direct Parental Support]*	$t(90.3) = -8.11,\ p < 0.0001$

Table D.4 (continued): ART ANOVA and Post-hoc Results

Analysis	**Result (significance level 0.05)**
C2 *[AEM Support Only]* vs. **C4** *[Direct Parental Support]*	$t(89.5) = -3.47,\ p = 0.0043$
C2 *[AEM Support Only]* vs. **C3** *[Indirect Parental Support]*	$t(88.0) = -1.04,\ p = 0.7275$ (not significant)
C3 *[Indirect Parental Support]* vs. **C4** *[Direct Parental Support]*	$t(87.4) = -2.49,\ p = 0.0681$ (not significant)

References

[1] *Protecting Vulnerable Groups (PVG) scheme - mygov.scot*. URL: https://www.mygov.scot/pvg-scheme.

[2] Cristina Fiani et al. 'Exploring the Perspectives of Social VR-Aware Non-Parent Adults and Parents on Children's Use of Social Virtual Reality'. In: *Proc. ACM Hum.-Comput. Interact.* 8.CSCW1 (Apr. 2024), p. 25. DOI: 10.1145/3652867.

[3] Cristina Fiani et al. '"Abracadabra, Block the Bullies!": Child Perceptions of Manual, Semi-Automated and Automated Interventions against Group Harassment in Social VR'. In: *Proceedings of the 24th Interaction Design and Children*. Association for Computing Machinery, 2025, pp. 314–339. ISBN: 9798400714733. DOI: 10.1145/3713043.3728850.

[4] Cristina Fiani et al. 'Big Buddy: Exploring Child Reactions and Parental Perceptions towards a Simulated Embodied Moderating System for Social Virtual Reality'. In: *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*. IDC '23. Chicago, IL, USA: Association for Computing Machinery, 2023, pp. 1–13. ISBN: 9798400701313. DOI: 10.1145/3585088.3589374.

[5] Cristina Fiani et al. 'Big Buddy: A Simulated Embodied Moderating System to Mitigate Children's Reaction to Provocative Situations within Social Virtual Reality'. In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Hamburg, Germany: ACM, Apr. 2023, p. 7. DOI: 10.1145/3544549.3585840.

[6] Cristina Fiani et al. '"Pikachu would electrocute people who are misbehaving": Expert, Guardian and Child Perspectives on Automated Embodied Moderators for Safeguarding Children in Social Virtual Reality'. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. CHI '24. Honolulu, HI, USA: Association for Computing Machinery, 2024. ISBN: 9798400703300. DOI: 10.1145/3613904.3642144.

[7] Cristina Fiani and Stacy Marsella. 'Investigating the Non-verbal Behavior Features of Bullying for the Development of an Automatic Recognition System in Social Virtual Reality'. In: *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*. AVI 2022. Frascati, Rome, Italy: Association for Computing Machinery, 2022. ISBN: 9781450397193. DOI: 10.1145/3531073.3534492.

[8] Cristina Fiani, Mark Mcgill and Mohamed Khamis. *Ensuring Child Safety in Social VR: Navigating Different Perspectives and Merging Viewpoints*. CHI'23 Workshop. 2023. URL: http://mkhamis.com/data/papers/fiani2023chiworkshop.pdf.

[9] Maheshya Weerasinghe et al. 'Beyond Mute and Block: Adoption and Effectiveness of Safety Tools in Social VR, from Ubiquitous Harassment to Social Sculpting'. In: *IEEE Transactions on Visualization and Computer Graphics* 31.5 (2025), pp. 3275–3284. DOI: 10.1109/TVCG.2025.3549860.

[10] Guo Freeman and Divine Maloney. 'Body, Avatar, and Me'. In: *Proceedings of the ACM on Human-Computer Interaction* 4 (CSCW3 Jan. 2021). ISSN: 25730142. DOI: 10.1145/3432938.

[11] Guo Freeman et al. 'My Body, My Avatar: How People Perceive Their Avatars in Social Virtual Reality'. In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. Honolulu, HI, USA: Association for Computing Machinery, 2021, pp. 1–8. ISBN: 9781450368193. DOI: 10.1145/3334480.3382923.

[12] Qijia Chen, Michiel M. Spapé and Giulio Jacucci. 'Understanding Phantom Tactile Sensation on Commercially Available Social Virtual Reality Platforms'. In: *Proc. ACM Hum.-Comput. Interact.* 8.CSCW1 (Apr. 2024). DOI: 10.1145/3637418.

[13] *Rec Room Studio - Rec Room*. URL: https://recroom.com/studio.

[14] *Become a VRChat Creator | VRChat Creation*. URL: https://creators.vrchat.com/.

[15] Christopher Ball, Kuo-Ting Huang and Jess Francis. 'Virtual reality adoption during the COVID-19 pandemic: A uses and gratifications perspective'. In: *Telematics and Informatics* 65 (2021), p. 101728. ISSN: 0736-5853. DOI: 10.1016/j.tele.2021.101728.

[16] Miguel Barreda-Ángeles and Tilo Hartmann. 'Psychological benefits of using social virtual reality platforms during the covid-19 pandemic: The role of social and spatial presence'. In: *Computers in Human Behavior* 127 (2022), p. 107047. ISSN: 0747-5632. DOI: 10.1016/j.chb.2021.107047.

[17] *Meta Quest Safety Center*. Last Accessed: 03-08-2025. 2024. URL: https://www.meta.com/gb/quest/parent-info/.

[18] *Generation VR - The Institution of Engineering and Technology*. Last Accessed: 06-10-2025. 2022. URL: https://www.theiet.org/media/press-releases/press-releases-2022/press-releases-2022-april-june/19-april-2022-generation-vr/.

[19] Nazanin Sabri et al. 'Challenges of Moderating Social Virtual Reality'. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. Hamburg, Germany: Association for Computing Machinery, 2023. ISBN: 9781450394215. DOI: 10.1145/3544548.3581329.

[20] Guo Freeman et al. 'Disturbing the Peace: Experiencing and Mitigating Emerging Harassment in Social Virtual Reality'. In: *Proc. ACM Hum.-Comput. Interact.* 6.CSCW1 (Apr. 2022). DOI: 10.1145/3512932.

[21] Divine Maloney, Guo Freeman and Andrew Robb. 'Stay Connected in An Immersive World: Why Teenagers Engage in Social Virtual Reality'. In: *Proceedings of Interaction Design and Children, IDC 2021* (June 2021), pp. 69–79. DOI: 10.1145/3459990.3460703.

[22] Divine Maloney, Guo Freeman and Andrew Robb. 'It Is Complicated: Interacting with Children in Social Virtual Reality'. In: *Proceedings - 2020 IEEE Conference on Virtual Reality and 3D User Interfaces, VRW 2020* (Mar. 2020), pp. 343–347. DOI: 10.1109/VRW50115.2020.00075.

[23] Divine Maloney, Guo Freeman and Andrew Robb. 'A Virtual Space for All: Exploring Children's Experience in Social Virtual Reality'. In: *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. CHI PLAY '20. Virtual Event, Canada: Association for Computing Machinery, 2020, pp. 472–483. ISBN: 9781450380744. DOI: 10.1145/3410404.3414268.

[24] SteamDB. *Steam Charts by SteamDB | Most played VR games*. Last Accessed: 14-10-2025. URL: https://steamdb.info/charts/?tagid=21978.

[25] *VRChat*. Last Accessed: 21-01-2024. 2014. URL: https://hello.vrchat.com/.

[26] *VRChat on Steam | PlayTracker*. Last Accessed: 14-10-2025. URL: https://playtracker.net/insight/game/2704.

[27] *Rec Room*. Last Accessed: 21-01-2024. 2016. URL: https://recroom.com/.

[28] *Rec Room on Steam | PlayTracker*. Last Accessed: 14-10-2025. URL: https://playtracker.net/insight/game/1442.

[29] Lauren B. McInroy and Faye Mishna. 'Cyberbullying on Online Gaming Platforms for Children and Youth'. In: *Child and Adolescent Social Work Journal* 34 (6 Dec. 2017), pp. 597–607. ISSN: 07380151. DOI: 10.1007/S10560-017-0498-0.

[30] Vivian H Wright et al. 'Cyberbullying'. In: *Journal of Computing in Teacher Education* 26 (1 2009), pp. 35–42. ISSN: 2332-7421. DOI: 10.1080/10402454.2009.10784629.

[31] Lindsay Blackwell et al. 'Harassment in Social Virtual Reality: Challenges for Platform Governance'. In: *Proc. ACM Hum.-Comput. Interact.* 3.CSCW (2019). DOI: 10.1145/3359202.

[32] Clare McGlynn and Carlotta Rigotti. 'From Virtual Rape to Meta-rape: Sexual Violence, Criminal Law and the Metaverse'. In: *Oxford Journal of Legal Studies* (Apr. 2025), gqaf009. ISSN: 0143-6503. DOI: 10.1093/ojls/gqaf009.

[33] Arthi Nachiappan. *Police 'investigate sexual abuse of young girl's avatar in the metaverse' - prompting NSPCC warning | Science, Climate & Tech News | Sky News*. Last Accessed: 29-07-2025. 2024. URL: https://news.sky.com/story/police-investigate-sexual-abuse-of-young-girls-avatar-in-the-metaverse-prompting-nspcc-warning-13041003.

[34] *Metaverse app allows kids into virtual strip clubs - BBC News*. Last Accessed: 29-07-2025. 2022. URL: https://www.bbc.co.uk/news/technology-60415317.

[35] Joshua McVeigh-Schultz et al. 'What's It Mean to "Be Social" in VR? Mapping the Social VR Design Ecology'. In: *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*. DIS '18 Companion. Hong Kong, China: Association for Computing Machinery, 2018, pp. 289–294. ISBN: 9781450356312. DOI: 10.1145/3197391.3205451.

[36] Jakki O. Bailey et al. 'Virtual reality's effect on children's inhibitory control, social compliance, and sharing'. In: *Journal of Applied Developmental Psychology* 64 (2019), p. 101052. ISSN: 0193-3973. DOI: 10.1016/j.appdev.2019.101052.

[37] Jakki O. Bailey and Jeremy N. Bailenson. 'Considering virtual reality in children's lives'. In: *Journal of Children and Media* 11 (1 Jan. 2017), pp. 107–113. ISSN: 17482801. DOI: 10.1080/17482798.2016.1268779.

[38] Jakki O. Bailey and Jeremy N. Bailenson. 'Chapter 9 - Immersive Virtual Reality and the Developing Child'. In: *Cognitive Development in Digital Contexts*. Ed. by Fran C. Blumberg and Patricia J. Brooks. San Diego: Academic Press, 2017, pp. 181–200. ISBN: 978-0-12-809481-5. DOI: 10.1016/B978-0-12-809481-5.00009-2.

[39] Joseph O'Hagan et al. 'Safety, Power Imbalances, Ethics and Proxy Sex: Surveying In-The-Wild Interactions Between VR Users and Bystanders'. In: *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. Bari, Italy: IEEE, 2021, pp. 211–220. DOI: 10.1109/ISMAR52148.2021.00036.

[40] Kelsea Schulenberg et al. 'Towards Leveraging AI-based Moderation to Address Emergent Harassment in Social Virtual Reality'. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. Hamburg, Germany: ACM, 2023, p. 17. DOI: 10.1145/3544548.3581090.

[41] Shenghui Cheng. 'Metaverse'. In: *Metaverse: Concept, Content and Context*. Cham: Springer Nature Switzerland, 2023, pp. 1–23. ISBN: 978-3-031-24359-2. DOI: 10.1007/978-3-031-24359-2_1.

[42] Sevtap Gurdal and Emma Sorbring. 'Children's agency in parent–child, teacher–pupil and peer relationship contexts'. In: *International Journal of Qualitative Studies on Health and Well-being* 13 (sup1 June 2019). ISSN: 17482631. DOI: 10.1080/17482631.2019.1565239.

[43] Dhanasekhar Kesavelu, K Sheela and Ponsekar Abraham. 'Stages of Psychological Development of Child-An Overview'. In: *Int J Cur Res Rev* | 13 (2021). ISSN: 0975-5241. DOI: 10.31782/IJCRR.2021.131320.

[44] Rogier E.J. Verhoef et al. 'Interactive virtual reality assessment of aggressive social information processing in boys with behaviour problems: A pilot study'. In: *Clinical Psychology and Psychotherapy* 28 (3 May 2021), pp. 489–499. ISSN: 10990879. DOI: 10.1002/cpp.2620.

[45] Grace Icenogle et al. 'Adolescents' cognitive capacity reaches adult levels prior to their psychosocial maturity: Evidence for a "maturity gap" in a multinational, cross-sectional sample'. In: *Law and Human Behavior* 43.1 (2019), pp. 69–85. DOI: 10.1037/lhb0000315.

[46] Laurence Steinberg. 'A dual systems model of adolescent risk-taking'. In: *Developmental Psychobiology* 52.3 (2010), pp. 216–224. DOI: 10.1002/dev.20445.

[47] Danielle A. Robson, Mark S. Allen and Steven J. Howard. 'Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review'. In: *Psychological Bulletin* 146.4 (2020), pp. 324–354. DOI: 10.1037/bul0000227.

[48] Sarah E. Moore et al. 'Consequences of bullying victimization in childhood and adolescence: A systematic review and meta-analysis'. In: *World Journal of Psychiatry* 7.1 (2017), pp. 60–76. DOI: 10.5498/wjp.v7.i1.60.

[49] James D. Unnever. 'Bullies, aggressive victims, and victims: Are they distinct groups?' In: *Aggressive Behavior* 31 (2 Apr. 2005), pp. 153–171. ISSN: 1098-2337. DOI: 10.1002/AB.20083.

[50] Qingxiao Zheng et al. 'Understanding Safety Risks and Safety Design in Social VR Environments'. In: *Proc. ACM Hum.-Comput. Interact.* 7.CSCW1 (Apr. 2023). DOI: 10.1145/3579630.

[51] John Zimmerman, Jodi Forlizzi and Shelley Evenson. 'Research through design as a method for interaction design research in HCI'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '07. San Jose, California, USA: Association for Computing Machinery, 2007, pp. 493–502. ISBN: 9781595935939. DOI: 10.1145/1240624.1240704.

[52] Jodi Zimmerman John and Forlizzi. 'Research Through Design in HCI'. In: *Ways of Knowing in HCI*. Ed. by Wendy A Olson Judith S. and Kellogg. Springer New York, 2014, pp. 167–189. ISBN: 978-1-4939-0378-8. DOI: 10.1007/978-1-4939-0378-8_8.

[53] K. Jayasankara Reddy and G. Balasubramanian. 'Misuse of Internet Among School Children: Risk Factors and Preventative Measures'. In: *Child Safety, Welfare and Well-being: Issues and Challenges, Second Edition* (Jan. 2022), pp. 321–334. DOI: 10.1007/978-981-16-9820-0_19.

[54] *A window into young children's online worlds - Ofcom*. Last Accessed: 16-07-2025. 2024. URL: https://www.ofcom.org.uk/media-use-and-attitudes/media-habits-children/a-window-into-young-childrens-online-worlds.

[55] Ofcom. *Children and Parents: Media Use and Attitudes*. 2023. URL: https://www.ofcom.org.uk/media-use-and-attitudes/media-habits-children/children-and-parents-media-use-and-attitudes-report-2023.

[56] *The United Nations Convention on the Rights of the Child*. Tech. rep. PDF Last Accessed: 13-09-2025. United Nations Children's Fund (Unicef), 1989. URL: https://www.unicef.org.uk/what-we-do/un-convention-child-rights/.

[57] *The Global Strategy for Women's, Children's and Adolescents' Health (2016-2030)*. Tech. rep. World Health Organization (WHO), 2018. URL: https://www.who.int/publications/i/item/the-global-strategy-for-women-s-children-s-and-adolescents-health-(2016-2030)-early-childhood-development-report-by-the-director-general.

[58] Susan M. Sawyer et al. 'The age of adolescence'. In: *The Lancet Child & Adolescent Health* 2 (3 Mar. 2018), pp. 223–228. ISSN: 2352-4642. DOI: 10.1016/S2352-4642(18)30022-1.

[59] Erik H. Erikson. 'Eight Ages of Man'. In: *International Journal of Psychiatry*. Vol. 2. 3. 1966, pp. 281–300.

[60] Nicki R. Crick and Kenneth A. Dodge. 'A review and reformulation of social information-processing mechanisms in children's social adjustment'. In: *Psychological Bulletin* (1994). DOI: 10.1037/0033-2909.115.1.74.

[61] Andy Miller, Eamonn Ferguson and Rachel Simpson. 'The Perceived Effectiveness of Rewards and Sanctions in Primary Schools: adding in the parental perspective'. In: *Educational Psychology* 18.1 (1998), pp. 55–64. DOI: 10.1080/0144341980180104.

[62] Jie He et al. 'Anger and selective attention to reward and punishment in children'. In: *Journal of experimental child psychology* 115 (3 July 2013), pp. 389–404. ISSN: 1096-0457. DOI: 10.1016/J.JECP.2013.03.004.

[63] Julie A Hubbard. 'Emotion expression processes in children's peer interaction: The role of peer rejection, aggression, and gender'. In: *Child development* 72.5 (2001), pp. 1426–1438.

[64] James A. Russell. 'A circumplex model of affect'. In: *Journal of Personality and Social Psychology* 39 (6 Dec. 1980), pp. 1161–1178. ISSN: 00223514. DOI: 10.1037/H0077714.

[65] Peter J Lang, Margaret M Bradley, Bruce N Cuthbert et al. *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. NIMH, Center for the Study of Emotion & Attention Gainesville, FL, 2005.

[66] Anna Pauliina Rainio. 'From resistance to involvement: Examining agency and control in a playworld activity'. In: *Mind, Culture, and Activity* 15 (2 Apr. 2008), pp. 115–140. ISSN: 10749039. DOI: 10.1080/10749030801970494.

[67] Rebecca N.H. de Leeuw and Christa A. van der Laan. 'Helping behavior in Disney animated movies and children's helping behavior in the Netherlands'. In: *Journal of Children and Media* 12 (2 Apr. 2017), pp. 159–174. ISSN: 17482801. DOI: 10.1080/17482798.2017.1409245.

[68] Roser Cañigueral and Antonia F.de C. Hamilton. 'Being watched: Effects of an audience on eye gaze and prosocial behaviour'. In: *Acta Psychologica* 195 (Apr. 2019), pp. 50–63. ISSN: 0001-6918. DOI: 10.1016/J.ACTPSY.2019.02.002.

[69] Alex Bradley, Claire Lawrence and Eamonn Ferguson. 'Does observability affect prosociality?' In: *Proceedings of the Royal Society B: Biological Sciences* 285 (1875 Mar. 2018). ISSN: 14712954. DOI: 10.1098/RSPB.2018.0116.

[70] Marie S. Tisak et al. 'Mothers' and Teachers' Home and School Rules: Young Children's Conceptions of Authority in Context'. In: *Merrill-Palmer Quarterly* 46.1 (2000), pp. 168–187. ISSN: 0272930X, 15350266. URL: http://www.jstor.org/stable/23093347.

[71] Lila Ghent Braine et al. 'Conflicts With Authority: Children's Feelings, Actions, and Justifications'. In: *Developmental Psychology* 27 (5 1991), pp. 829–840. ISSN: 00121649. DOI: 10.1037/0012-1649.27.5.829.

[72] *Protection from Harassment Act 1997*. 2011. URL: https://www.legislation.gov.uk/ukpga/1997/40/section/8.

[73] Dan Olweus. *Bullying at school: What we know and what we can do*. Blackwell Publishers, 1993.

[74] Anthony A. Volk, Andrew V. Dane and Zopito A. Marini. 'What is bullying? A theoretical redefinition'. In: *Developmental Review* 34 (4 Dec. 2014), pp. 327–343. ISSN: 02732297. DOI: 10.1016/j.dr.2014.09.001.

[75] Robert Slonje and Peter K. Smith. 'Cyberbullying: Another main type of bullying?' In: *Scandinavian Journal of Psychology* 49 (2 Apr. 2008), pp. 147–154. ISSN: 1467-9450. DOI: 10.1111/J.1467-9450.2007.00611.X.

[76] Dan Olweus. 'School Bullying: Development and Some Important Challenges'. In: 34 (2013), p. 758. DOI: 10.1146/annurev-clinpsy-050212-185516.

[77] Maeve Duggan. *Online Harassment 2017 | Pew Research Center*. Last Accessed: 14-10-2025. 2017. URL: https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/.

[78] Justin W. Patchin and Sameer Hinduja. 'Bullies Move Beyond the Schoolyard: A Preliminary Look at Cyberbullying'. In: *Youth Violence and Juvenile Justice* 4 (2 2006), pp. 148–169. ISSN: 15412040. DOI: 10.1177/1541204006286288.

[79] Alice E. Marwick. 'Morally Motivated Networked Harassment as Normative Reinforcement'. In: *Social Media and Society* 7 (2 June 2021). ISSN: 20563051. DOI: 10.1177/20563051211021378.

[80] Jacob M. Rabbie and H. Lodewijkx. 'Individual and Group Aggression'. In: *Current Research on Peace and Violence* 10.2/3 (1987). PSYCHOLOGICAL ISSUES OF WAR AND PEACE, pp. 91–101. URL: https://www.jstor.org/stable/40725065.

[81] KIRSTI M.J. LAGERSPETZ et al. 'Group aggression among school children in three schools'. In: *Scandinavian Journal of Psychology* 23 (1 Sept. 1982), pp. 45–52. ISSN: 1467-9450. DOI: 10.1111/J.1467-9450.1982.TB00412.X.

[82] Robert Faris and Susan Ennett. 'Adolescent aggression: The role of peer group status motives, peer aggression, and group characteristics'. In: *Social Networks* 34 (4 Oct. 2012), pp. 371–378. ISSN: 0378-8733. DOI: 10.1016/J.SOCNET.2010.06.003.

[83] Dorothy L Espelage and Susan M Swearer. 'Research on School Bullying and Victimization: What Have We Learned and Where Do We Go From Here?' In: *School Psychology Review* 32 (3 2003), pp. 365–383. DOI: 10.1080/02796015.2003.12086206.

[84] Michele L Ybarra and Kimberly J Mitchell. 'Online aggressor/targets, aggressors, and targets: a comparison of associated youth characteristics'. In: (2004). DOI: 10.1111/j.1469-7610.2004.00328.x.

[85] Robert S. Tokunaga. 'Following you home from school: A critical review and synthesis of research on cyberbullying victimization'. In: *Computers in Human Behavior* 26 (3 May 2010), pp. 277–287. ISSN: 0747-5632. DOI: 10.1016/J.CHB.2009.11.014.

[86] Dieter Wolke and Suzet Tanya Lereya. 'Long-term effects of bullying'. In: (2015). DOI: 10.1136/archdischild-2014-306667.

[87] Sonja Perren et al. 'Bullying in school and cyberspace: Associations with depressive symptoms in Swiss and Australian adolescents'. In: *Child and Adolescent Psychiatry and Mental Health* 4 (1 Nov. 2010), pp. 1–10. ISSN: 17532000. DOI: 10.1186/1753-2000-4-28.

[88] Christoph Burger, Dagmar Strohmeier and Lenka Kollerová. 'Teachers Can Make a Difference in Bullying: Effects of Teacher Interventions on Students' Adoption of Bully, Victim, Bully-Victim or Defender Roles across Time'. In: *Journal of Youth and Adolescence 2022* (Sept. 2022), pp. 1–16. ISSN: 1573-6601. DOI: 10.1007/S10964-022-01674-6.

[89] L. Christian Elledge et al. 'School-based mentoring as selective prevention for bullied children: A preliminary test'. In: *Journal of Primary Prevention* 31 (3 June 2010), pp. 171–187. ISSN: 0278095X. DOI: 10.1007/S10935-010-0215-7.

[90] Giulio D'Urso et al. 'Bullies, victims, and meanies: the role of child and classmate social and emotional competencies'. In: *Social Psychology of Education* 25 (1 Feb. 2022), pp. 293–312. ISSN: 15731928. DOI: 10.1007/S11218-021-09684-1.

[91] Giulio D'Urso and Ugo Pace. 'Homophobic bullying among adolescents: The role of insecure-dismissing attachment and peer support'. In: *Journal of LGBT Youth* 16 (2 Apr. 2019), pp. 173–191. ISSN: 19361661. DOI: 10.1080/19361653.2018.1552225.

[92] Gary W. Giumetti and Robin M. Kowalski. 'Cyberbullying via social media and well-being'. In: *Current Opinion in Psychology* 45 (June 2022), p. 101314. ISSN: 2352-250X. DOI: 10.1016/J.COPSYC.2022.101314.

[93] Kathleen Van Royen et al. '"Thinking before posting?" Reducing cyber harassment on social networking sites through a reflective message'. In: *Computers in Human Behavior* 66 (Jan. 2017), pp. 345–352. ISSN: 0747-5632. DOI: 10.1016/J.CHB.2016.09.040.

[94] Richard S. Lazarus and Susan Folkman. *Stress, Appraisal, and Coping - Richard S. Lazarus, Susan Folkman - Google Books*. Springer Publishing Company New York, 1984. URL: https://books.google.co.uk/books?hl=en&lr=&id=i-ySQQuUpr8C&oi=fnd&pg=PR5&ots=DhHUhpkdUd&sig=pR-rcjBL0hLbD8IBaP5ekZtrLoM&redir_esc=y#v=onepage&q&f=false.

[95] Grace Skrzypiec et al. 'School bullying by one or more ways: Does it matter and how do students cope?' In: *School Psychology International* 32 (3 June 2011), pp. 288–311. ISSN: 01430343. DOI: 10.1177/0143034311402308.

[96] Richard S. Lazarus and Susan Folkman. 'Transactional Theory and Research on Emotions and Coping'. In: *European Journal of Personality* (Sept. 1987).

[97] Juliana Raskauskas and Amanda Huynh. 'The process of coping with cyberbullying: A systematic review'. In: *Aggression and Violent Behavior* 23 (July 2015), pp. 118–125. ISSN: 1359-1789. DOI: 10.1016/J.AVB.2015.05.019.

[98] Kimberley R. Allison and Kay Bussey. 'Individual and collective moral influences on intervention in cyberbullying'. In: *Computers in Human Behavior* 74 (Sept. 2017), pp. 7–15. ISSN: 0747-5632. DOI: 10.1016/J.CHB.2017.04.019.

[99] Sara Bastiaensens et al. ''Can I afford to help?' How affordances of communication modalities guide bystanders' helping intentions towards harassment on social network sites'. In: *Behaviour &Information Technology* 34 (4 Apr. 2014), pp. 425–435. ISSN: 13623001. DOI: 10.1080/0144929X.2014.983979.

[100] Vaishali U. Gongane, Mousami V. Munot and Alwin D. Anuse. 'Detection and moderation of detrimental content on social media platforms: current status and future directions'. In: *Social Network Analysis and Mining 2022 12:1* 12 (1 Sept. 2022), pp. 1–41. ISSN: 1869-5469. DOI: 10.1007/S13278-022-00951-3.

[101] Tarleton Gillespie. 'Content moderation, AI, and the question of scale'. In: *Big Data & Society* 7.2 (2020), p. 2053951720943234. DOI: 10.1177/2053951720943234.

[102] Andreas Veglis. 'Moderation techniques for social media content'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8531 LNCS (2014), pp. 137–148. ISSN: 16113349. DOI: 10.1007/978-3-319-07632-4_13.

[103] H. Rosa et al. 'Automatic cyberbullying detection: A systematic review'. In: *Computers in Human Behavior* 93 (Apr. 2019), pp. 333–345. ISSN: 0747-5632. DOI: 10.1016/J.CHB.2018.12.021.

[104] Nasiru I Aliyu et al. 'Analysis of Cyber Bullying on Facebook Using Text Mining'. In: 1 (2020), pp. 1–12. ISSN: 2709-5908. DOI: 10.48185/jaai.v1i1.30.

[105] Thabo Mahlangu, Chunling Tu and Pius Owolawi. 'A review of automated detection methods for cyberbullying'. In: *2018 International Conference on Intelligent and Innovative Computing Applications, ICONIC 2018* (Jan. 2019). DOI: 10.1109/ICONIC.2018.8601278.

[106] Robert Gorwa, Reuben Binns and Christian Katzenbach. 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance'. In: *Big Data & Society* 7.1 (2020), p. 2053951719897945. DOI: 10.1177/2053951719897945.

[107] Marie Ozanne et al. 'Shall AI moderators be made visible? Perception of accountability and trust in moderation systems on social media platforms'. In: *Big Data & Society* 9.2 (2022), p. 20539517221115666. DOI: 10.1177/20539517221115666.

[108] Robyn Caplan and Tarleton Gillespie. 'Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy'. In: *Social Media + Society* 6.2 (2020), p. 2056305120936636. DOI: 10.1177/2056305120936636.

[109] Motahhare Eslami et al. 'First I "like" it, then I hide it: Folk Theories of Social Feeds'. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: Association for Computing Machinery, 2016, pp. 2371–2382. ISBN: 9781450333627. DOI: 10.1145/2858036.2858494.

[110] Tijana Milosevic et al. 'Effectiveness of Artificial Intelligence–Based Cyberbullying Interventions From Youth Perspective'. In: *Social Media + Society* 9 (1 Jan. 2023). ISSN: 20563051. DOI: 10.1177/20563051221147325.

[111] Jenny S. Radesky et al. 'Parent Perspectives on Their Mobile Technology Use: The Excitement and Exhaustion of Parenting while Connected'. In: *Journal of Developmental and Behavioral Pediatrics* 37 (9 2016), pp. 694–701. ISSN: 15367312. DOI: 10.1097/DBP.0000000000000357.

[112] Sonia Livingstone Ph.D. and Ellen J. Helsper Ph.D. 'Parental Mediation of Children's Internet Use'. In: *Journal of Broadcasting & Electronic Media* 52.4 (2008), pp. 581–599. DOI: 10.1080/08838150802437396.

[113] Patti M. Valkenburg et al. 'Developing a scale to assess three styles of television mediation: "Instructive mediation," "restrictive mediation," and "social coviewing"'. In: *Journal of Broadcasting & Electronic Media* 43 (1 1999), pp. 52–66. ISSN: 15506878. DOI: 10.1080/08838159909364474.

[114] Heidi Hartikainen, Netta Iivari and Marianne Kinnula. 'Should We Design for Control, Trust or Involvement? A Discourses Survey about Children's Online Safety'. In: *Proceedings of the The 15th International Conference on Interaction Design and Children*. IDC '16. Manchester, United Kingdom: Association for Computing Machinery, 2016, pp. 367–378. ISBN: 9781450343138. DOI: 10.1145/2930674.2930680.

[115] Lindsay Blackwell, Emma Gardiner and Sarita Schoenebeck. 'Managing expectations: Technology tensions among parents and teens'. In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW* 27 (Feb. 2016), pp. 1390–1401. DOI: 10.1145/2818048.2819928.

[116] Kristin L. Moilanen, Katie E. Rasmussen and Laura M. Padilla-Walker. 'Bidirectional Associations Between Self-Regulation and Parenting Styles in Early Adolescence'. In: *Journal of Research on Adolescence* 25 (2 June 2015), pp. 246–262. ISSN: 15327795. DOI: 10.1111/JORA.12125.

[117] Pamela Wisniewski et al. 'Parental control vs. teen self-regulation: Is there a middle ground for mobile online safety?' In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW* (Feb. 2017), pp. 51–69. DOI: 10.1145/2998181.2998352.

[118] Ge Wang et al. 'Protection or Punishment? Relating the Design Space of Parental Control Apps and Perceptions about Them to Support Parenting for Online Safety'. In: *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW2 Oct. 2021). ISSN: 25730142. DOI: 10.1145/3476084.

[119] Lynn Schofield Clark. 'Parenting in a Digital Age: The Mediatization of Family Life and the Need to Act'. In: *The Parent App: Understanding Families in the Digital Age*. Oxford University Press, Nov. 2012, pp. 201–226. ISBN: 9780199899616. DOI: 10.1093/acprof:oso/9780199899616.003.0009.

[120] Lorrie Faith Cranor et al. 'Parents' and Teens' Perspectives on Privacy In a Technology-Filled World'. In: (2014).

[121] Pamela Wisniewski et al. '"Preventative" vs."Reactive:" How Parental Mediation Influences Teens' Social Media Privacy Behaviors'. In: (Jan. 2014).

[122] Patricia M. Greenfield. 'Developmental considerations for determining appropriate Internet use guidelines for children and adolescents'. In: *Journal of Applied Developmental Psychology* 25.6 (2004). Developing Children, Developing Media - Research from Television to the Internet from the Children's Digital Media Center: A Special Issue Dedicated to the Memory of Rodney R. Cocking, pp. 751–762. ISSN: 0193-3973. DOI: 10.1016/j.appdev.2004.09.008.

[123] Justine Humphry, Jonathon Hutchinson and Olga Boichak. 'Social digital dilemmas: Young people's and parents' negotiation of emerging online safety issues'. In: *New Media &Society* (Jan. 2025). ISSN: 1461-4448. DOI: 10.1177/14614448241310247.

[124] Benoît Rapoport and Céline Le Bourdais. 'Parental time and working schedules'. In: *Journal of Population Economics* 21 (4 Oct. 2008), pp. 903–932. ISSN: 09331433. DOI: 10.1007/S00148-007-0147-6.

[125] Lyn Craig and Judith Brown. 'The Multitasking Parent: Time Penalties, Dimensions, and Gender Differences'. In: *The Economics of Multitasking* (2016), pp. 33–59. DOI: 10.1057/9781137381446_4.

[126] Phyllis Moen and Donna I. Dempster-McClain. 'Employed Parents: Role Strain, Work Time, and Preferences for Working Less'. In: *Journal of Marriage and the Family* 49 (3 Aug. 1987), p. 579. ISSN: 00222445. DOI: 10.2307/352203.

[127] Chris A. Higgins, Linda E. Duxbury and Sean T. Lyons. 'Coping with overload and stress: Men and women in dual-earner families'. In: *Journal of Marriage and Family* 72 (4 Aug. 2010), pp. 847–859. ISSN: 00222445. DOI: 10.1111/J.1741-3737.2010.00734.X.

[128] Lynn Schofield Clark. 'Risk, Media, and Parenting in a Digital Age'. In: *The Parent App: Understanding Families in the Digital Age*. Oxford University Press, Nov. 2012, pp. 3–27. ISBN: 9780199899616. DOI: 10.1093/acprof:oso/9780199899616.003.0001.

[129] Mizuko Ito et al. 'Gaming'. In: *Hanging Out, Messing Around, and Geeking Out*. The MIT Press, Dec. 2018. Chap. 5. DOI: 10.7551/MITPRESS/8402.001.0001. URL: https://mitpress.mit.edu/9780262518543/hanging-out-messing-around-and-geeking-out/.

[130] Danah Boyd and Eszter Hargittai. 'Connected and concerned: Variation in parents' online safety concerns'. In: *Policy & Internet* 5 (3 Sept. 2013), pp. 245–269. ISSN: 1944-2866. DOI: 10.1002/1944-2866.POI332.

[131] Marie Danet. 'Parental Concerns about their School-aged Children's Use of Digital Devices'. In: *Journal of Child and Family Studies* 29 (Oct. 2020). DOI: 10.1007/s10826-020-01760-y.

[132] Candice L. Odgers, Stephen M. Schueller and Mimi Ito. 'Screen Time, Social Media Use, and Adolescent Development'. In: *Annual Review of Developmental Psychology* 2.1 (2020), pp. 485–502. DOI: 10.1146/annurev-devpsych-121318-084815.

[133] Constance Steinkuehler. 'Parenting and Video Games'. In: *Journal of Adolescent & Adult Literacy* 59.4 (2016), pp. 357–361. DOI: 10.1002/jaal.455.

[134] Meghan Hollis, Marcus Felson and Brandon Welsh. 'The capable guardian in routine activities theory: A theoretical and conceptual reappraisal'. In: *Crime Prevention and Community Safety* 15 (2013), pp. 65–79. DOI: 10.1057/cpcs.2012.14.

[135] Meghan E. Hollis-Peel et al. 'Guardianship for crime prevention: a critical review of the literature'. In: *Crime, Law and Social Change* 56 (2011), pp. 53–70. DOI: 10.1007/s10611-011-9309-2.

[136] Zarina I. Vakhitova, Alisa Go and Clair L. Alston-Knox. 'Guardians Against Cyber Abuse: Who are They and Why do They Intervene?' In: *American Journal of Criminal Justice* 48 (1 Feb. 2023), pp. 96–122. ISSN: 19361351. DOI: 10.1007/S12103-021-09641-W.

[137] Mariana Olaizola Rosenblat, Ayushi Agrawal and Isaac Yap. *Online Safety Regulations Around The World: The State of Play and The Way Forward-A Resource Guide*. Tech. rep. 2025. URL: https://bhr.stern.nyu.edu/wp-content/uploads/2025/04/NYU-CBHR-Online-Regulations_Updated-Jun-17-1.pdf.

[138] Directorate-General for Communication. *Legal documents on The Digital Services Act - European Commission*. 2022. URL: https://commission.europa.eu/publications/legal-documents-digital-services-act_en.

[139] *Online Safety Act 2023*. 2023. URL: https://www.legislation.gov.uk/ukpga/2023/50.

[140] *Online Safety (Miscellaneous Amendments) Act 2022*. 2022. URL: https://sso.agc.gov.sg/Acts-Supp/38-2022/Published/20221221.

[141] *Online Safety Act 2021*. 2021. URL: https://www.legislation.gov.au/C2021A00076/latest/text.

[142] *Online Safety Amendment (Social Media Minimum Age) Bill 2024*. 2024. URL: https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=r7284.

[143] *Guidance - Online Safety Act: explainer*. 2025. URL: https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer.

[144] Hedvig Schmidt. 'The online safety act 2023'. In: *Journal of Media Law* 16 (2 July 2024), pp. 202–210. ISSN: 1757-7632. DOI: 10.1080/17577632.2025.2459440.

[145] General Assembly resolution 44/25. *Convention on the Rights of the Child | OHCHR*. Tech. rep. 1989. URL: https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child.

[146] *Children and the law | NSPCC Learning*. 2025. URL: https://learning.nspcc.org.uk/child-protection-system/children-the-law.

[147] *General comment No. 25 (2021) on children's rights in relation to the digital environment | OHCHR*. 2021. URL: https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation.

[148] *5Rights*. Last Accessed: 09-09-2025. 2016. URL: https://5rightsfoundation.com/wp-content/uploads/2024/09/5rights_Framework-.pdf.

[149] Andrea Gaggioli. 'Virtually Social'. In: *CYBERPSYCHOLOGY, BEHAVIOR AND SOCIAL NETWORKING* 21 (5 May 2018), pp. 338–339. ISSN: 2152-2715. DOI: 10.1089/CYBER.2018.29112.CSI.

[150] *VRChat stats, graphs and player estimates | PlayerTracker estimates*. Last Accessed: 25-11-2025. URL: https://playtracker.net/insight/game/2704?utm_source=SteamDB.

[151] *VRChat (App) · Steam Charts · SteamDB*. Last Accessed: 19-11-2022. URL: https://steamdb.info/app/438100/graphs/.

[152] Ben Delaney and F Biocca. *Immersive virtual reality technology*. Lawrence Erlbaum Associates Hillsdale, NJ, US, 1995, pp. 57–124.

[153] Bob G. Witmer and Michael J. Singer. 'Measuring Presence in Virtual Environments: A Presence Questionnaire'. In: *Presence: Teleoperators and Virtual Environments* 7 (3 June 1998), pp. 225–240. ISSN: 10547460. DOI: 10.1162/105474698565686.

[154] Mel Slater. 'Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments'. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1535 Dec. 2009), p. 3549. ISSN: 14712970. DOI: 10.1098/RSTB.2009.0138.

[155] Mel Slater and Sylvia Wilbur. 'A framework for immersive virtual environments five'. In: *Presence: Teleoperators and Virtual Environments* 6 (6 Dec. 1997), pp. 603–616. ISSN: 10547460. DOI: 10.1162/PRES.1997.6.6.603.

[156] Konstantina Kilteni, Raphaela Groten and Mel Slater. 'The Sense of Embodiment in Virtual Reality'. In: *Presence Teleoperators & Virtual Environments* 21 (Nov. 2012). DOI: 10.1162/PRES_a_00124.

[157] Guo Freeman and Dane Acena. 'Hugging from A Distance: Building Interpersonal Relationships in Social Virtual Reality'. In: *Proceedings of the 2021 ACM International Conference on Interactive Media Experiences*. IMX '21. Virtual Event, USA: Association for Computing Machinery, 2021, pp. 84–95. ISBN: 9781450383899. DOI: 10.1145/3452918.3458805.

[158] Divine Maloney, Guo Freeman and Donghee Yvette Wohn. '"Talking without a Voice": Understanding Non-Verbal Communication in Social Virtual Reality'. In: *Proc. ACM Hum.-Comput. Interact.* 4.CSCW2 (Oct. 2020). DOI: 10.1145/3415246.

[159] Jessica Lindblom. 'Embodiment and social interaction'. In: *Cognitive Systems Monographs* 26 (2015), pp. 115–159. ISSN: 18674933. DOI: 10.1007/978-3-319-20315-7_4.

[160] Lawrence W Barsalou et al. 'Social Embodiment'. In: *THE PSYCHOLOGY OF LEARNING AND MOTIVATION Advances in Research and Theory* (2003).

[161] Jakki O. Bailey and Isabella Schloss. '"Awesomely Freaky!" The Impact of Type on Children's Social-Emotional Perceptions of Virtual Reality Characters'. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. Hamburg, Germany: Association for Computing Machinery, 2023. ISBN: 9781450394215. DOI: 10.1145/3544548.3581501.

[162] Ruizhi Cheng et al. 'Are We Ready for Metaverse? A Measurement Study of Social Virtual Reality Platforms'. In: *Proceedings of the 22nd ACM Internet Measurement Conference*. IMC '22. Nice, France: Association for Computing Machinery, 2022, pp. 504–518. ISBN: 9781450392594. DOI: 10.1145/3517745.3561417.

[163] Artur Pilacinski, Marita Metzler and Christian Klaes. 'Phantom touch illusion, an unexpected phenomenological effect of tactile gating in the absence of tactile stimulation'. In: *Scientific Reports* 13 (2023), p. 15453. DOI: 10.1038/s41598-023-42683-0.

[164] *Virtual Reality Titles*. URL: https://store.steampowered.com/vr/?flavor=contenthub_popular&offset=36.

[165] *Gorilla Tag*. 2021. URL: https://www.gorillatagvr.com/.

[166] *Meta Horizon Worlds*. Last Accessed: 16-09-2025. 2019. URL: https://horizon.meta.com/.

[167] Anya Kolesnichenko, Joshua McVeigh-Schultz and Katherine Isbister. 'Understanding emerging design practices for avatar systems in the commercial social VR ecology'. In: *DIS 2019 - Proceedings of the 2019 ACM Designing Interactive Systems Conference* (June 2019), pp. 241–252. DOI: 10.1145/3322276.3322352.

[168] Tony Manninen and Tomi Kujanpää. 'Non-Verbal Communication Forms in Multiplayer Game Session'. In: *People and Computers XVI - Memorable Yet Invisible* (2002), pp. 383–401. DOI: 10.1007/978-1-4471-0105-5_23.

[169] Nick Yee and Jeremy Bailenson. 'The proteus effect: The effect of transformed self-representation on behavior'. In: *Human Communication Research* 33 (3 July 2007), pp. 271–290. ISSN: 03603989. DOI: 10.1111/J.1468-2958.2007.00299.X.

[170] Julie R. Williamson et al. 'Digital Proxemics: Designing Social and Collaborative Interaction in Virtual Environments'. In: *Conference on Human Factors in Computing Systems - Proceedings* (Apr. 2022). DOI: 10.1145/3491102.3517594.

[171] *Maker AI - Rec Room*. 2025. URL: https://blog.recroom.com/posts/maker-ai.

[172] Divine Maloney and Guo Freeman. 'Falling Asleep Together: What Makes Activities in Social Virtual Reality Meaningful to Users'. In: *CHI PLAY 2020 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Nov. 2020), pp. 510–521. DOI: 10.1145/3410404.3414266.

[173] Samaneh Zamanifard and Guo Freeman. '"The Togetherness that We Crave": Experiencing Social VR in Long Distance Relationships'. In: (2019). DOI: 10.1145/3311957.3359453.

[174] *Weddings In The Metaverse: Would You Get Married In VR And AR?* Last Accessed: 10-01-2025. 2022. URL: https://www.forbes.com/sites/bernardmarr/2022/03/07/weddings-in-the-metaverse-would-you-get-married-in-vr-and-ar/.

[175] Qijia Chen, Andrea Bellucci and Giulio Jacucci. '"I'd rather drink in VRChat": Understanding Drinking in Social Virtual Reality'. In: *Conference on Human Factors in Computing Systems - Proceedings* 16 (May 2024). DOI: 10.1145/3613904.3642405/SUPPL_FILE/3613904.3642405-TALK-VIDEO.VTT.

[176] *AR/VR Trends and Predictions for 2025 & Beyond: Charting the Future of Digital Realities*. Last Accessed: 03-08-2025. 2024. URL: https://sti-corporate.com/news/ar-vr-trends-and-predictions-for-2025-beyond-charting-the-future-of-digital-reali.

[177] Steven Vosloo et al. *The Metaverse, Extended Reality and Children | Innocenti Global Office of Research and Foresight*. Last Accessed: 30-07-2025. 2023. URL: https://www.unicef.org/innocenti/reports/metaverse-extended-reality-and-children.

[178] Minyue Zhang et al. 'Virtual Reality Technology as an Educational and Intervention Tool for Children with Autism Spectrum Disorder: Current Perspectives and Future Directions'. In: *Behavioral Sciences 2022, Vol. 12, Page 138* 12 (5 May 2022), p. 138. ISSN: 2076-328X. DOI: 10.3390/BS12050138.

[179] Nyaz Didehbani et al. 'Virtual Reality Social Cognition Training for children with high functioning autism'. In: *Computers in Human Behavior* 62 (Sept. 2016), pp. 703–711. ISSN: 0747-5632. DOI: 10.1016/J.CHB.2016.04.033.

[180] Marijn Mado et al. 'Accessibility of Educational Virtual Reality for Children During the COVID-19 Pandemic'. In: *Technology, Mind, and Behavior* 3 (1 Mar. 2022). DOI: 10.1037/TMB0000066.

[181] Arik Kurniawati, Ari Kusumaningsih and Imam Hasan. 'Class VR: Learning Class Environment for Special Educational Needs using Virtual Reality Games'. In: *2019 International Conference on Computer Engineering, Network, and Intelligent Multimedia, CENIM 2019 - Proceeding* 2019-November (Nov. 2019). DOI: 10.1109/CENIM48368.2019.8973353.

[182] Nigel Newbutt, Ryan Bradley and Iian Conley. 'Using Virtual Reality Head-Mounted Displays in Schools with Autistic Children: Views, Experiences, and Future Directions'. In: *https://home.liebertpub.com/cyber* 23 (1 Jan. 2020), pp. 23–33. ISSN: 21522723. DOI: 10.1089/CYBER.2019.0206.

[183] Azadeh Bashiri, Marjan Ghazisaeedi and Leila Shahmorasdi. 'The opportunities of virtual reality in the rehabilitation of children with attention deficit hyperactivity disorder: a literature review'. In: *Korean Journal of Pediatrics* 60 (11 Nov. 2017), p. 337. ISSN: 20927258. DOI: 10.3345/KJP.2017.60.11.337.

[184] *Opening Horizon Worlds to Everyone 18+ in the US and Canada*. Last Accessed: 03-08-2025. 2021. URL: https://about.fb.com/news/2021/12/horizon-worlds-open-in-us-and-canada/.

[185] *Parent and Guardian Guide to Meta Horizon Worlds - ConnectSafely*. Last Accessed: 03-08-2025. 2024. URL: https://connectsafely.org/horizon/.

[186] *VRChat Community Guidelines*. Last Accessed: 21-01-2024. 2024. URL: https://hello.vrchat.com/community-guidelines.

[187] *Comfort and Safety - Rec Room*. Last Accessed: 10-09-2024. 2024. URL: https://recroom.com/safety.

[188] *Junior Accounts & Age-Verification – Rec Room Help Center*. Last Accessed: 03-08-2025. 2025. URL: https://recroom.zendesk.com/hc/en-us/articles/4426900227735-Junior-Accounts-Age-Verification.

[189] *PlayStation VR2*. Last Accessed: 21-01-2024. 2024. URL: https://www.playstation.com/en-us/ps-vr2/.

[190] Heidi Hartikainen, Netta Iivari and Marianne Kinnula. 'Should We Design for Control, Trust or Involvement? A Discourses Survey about Children's Online Safety'. In: *Proceedings of the The 15th International Conference on Interaction Design and Children*. IDC '16. Manchester, United Kingdom: Association for Computing Machinery, 2016, pp. 367–378. ISBN: 9781450343138. DOI: 10.1145/2930674.2930680.

[191] Joseph O'Hagan, Florian Mathis and Mark McGill. 'User Reviews as a Reporting Mechanism for Emergent Issues Within Social VR Communities'. In: *ACM International Conference Proceeding Series* (Dec. 2023), pp. 230–237. DOI: 10.1145/3626705.3627780.

[192] Philipp Sykownik et al. 'Something Personal from the Metaverse: Goals, Topics, and Contextual Factors of Self-Disclosure in Commercial Social VR'. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22. New Orleans, LA, USA: Association for Computing Machinery, 2022. ISBN: 9781450391573. DOI: 10.1145/3491102.3502008.

[193] Elmira Deldari et al. *An Investigation of Teenager Experiences in Social Virtual Reality from Teenagers', Parents', and Bystanders' Perspectives*. 2023, pp. 1–17. ISBN: 978-1-939133-36-6. URL: https://www.usenix.org/conference/soups2023/presentation/deldari.

[194] *Sex offenders using virtual reality to abuse children as victims describe feeling 'phantom touch' | Science & Tech News | Sky News*. Last Accessed: 10-09-2024. 2023. URL: https://news.sky.com/story/sex-offenders-using-virtual-reality-to-abuse-children-as-victims-describe-feeling-phantom-touch-12954906.

[195] Nichola Copson and Shane D. Johnson. 'A scoping review of metaverse-facilitated child sexual abuse'. In: *Social Sciences & Humanities Open* 12 (Jan. 2025), p. 101667. ISSN: 2590-2911. DOI: 10.1016/J.SSAHO.2025.101667.

[196] Sam R. Sharar et al. 'Factors Influencing the Efficacy of Virtual Reality Distraction Analgesia During Postburn Physical Therapy: Preliminary Results from 3 Ongoing Studies'. In: *Archives of Physical Medicine and Rehabilitation* 88.12, Supplement 2 (2007). The NIDRR Burn Injury Rehabilitation Model System Program: Selected Findings, S43–S49. ISSN: 0003-9993. DOI: 10.1016/j.apmr.2007.09.004.

[197] Guo Freeman, Lingyuan Li and Kelsea Schulenberg. '"I Have Abused Someone Who Abused Me": Understanding People Who Have Experienced Both Sides of Harassment Accusations in Social VR'. In: *Proceedings of the ACM on Human-Computer Interaction* 9 (2 May 2025). ISSN: 25730142. DOI: 10.1145/3711005.

[198] Bernard Marr. *The Metaverse And Its Dark Side: Confronting The Reality Of Virtual Rape*. Last Accessed: 10-09-2024. 2024. URL: https://www.forbes.com/sites/bernardmarr/2024/01/16/the-metaverse-and-its-dark-side-confronting-the-reality-of-virtual-rape/.

[199] Nina Jane Patel. *Reality or Fiction?* Last Accessed: 10-09-2024. 2021. URL: https://medium.com/kabuni/fiction-vs-non-fiction-98aa0098f3b0.

[200] kentbye. *1057: What Parents Should Know about Social VR, Understanding Social VR Harassment, & Parental Guidance for the Metaverse with Lance G. Powell, Jr. – Voices of VR Podcast*. Last Accessed: 25-11-2024. Feb. 2022. URL: https://voicesofvr.com/1057-what-parents-should-know-about-social-vr-understanding-social-vr-harassment-parental-guidance-for-the-metaverse-with-lance-g-powell-jr/.

[201] Anne Hobson. 'Phantoms, Crashers, and Harassers: Emergent Governance of Social Spaces in Virtual Reality'. Last Accessed: 25-11-2024. 2020. URL: https://www.thecgo.org/wp-content/uploads/2020/09/Phantoms-Crashers-and-Harassers-Emergent-Governance-of-Social-Spaces-in-Virtual-Reality.pdf.

[202] *Code of Conduct*. Last Accessed: 07-05-2024. URL: https://recroom.com/code-of-conduct.

[203] *Code of Conduct for Virtual Experiences*. Last Accessed: 07-05-2024. URL: https://www.meta.com/en-gb/help/quest/articles/accounts/privacy-information-and-settings/code-of-conduct-for-virtual-experiences/.

[204] *VRChat Safety and Trust System*. Last Accessed: 16-11-2025. 2025. URL: https://docs.vrchat.com/docs/vrchat-safety-and-trust-system.

[205] *Safety and privacy tools available for your teen in Meta Horizon Worlds - Hotizon World*. Last Accessed: 10-09-2024. 2024. URL: https://www.meta.com/en-gb/help/quest/articles/horizon/safety-and-privacy-in-horizon-worlds/safety-tools-for-your-teen-horizon/.

[206] *Safety with PS VR2 (UK)*. Last Accessed: 03-08-2025. 2024. URL: https://www.playstation.com/en-gb/support/hardware/ps-vr2-safety/.

[207] Xuetong Wang et al. 'Avatar Appearance and Behavior of Potential Harassers Affect Users' Perceptions and Response Strategies in Social Virtual Reality (VR): A Mixed-Methods Study'. In: *Proceedings of the ACM on Human-Computer Interaction* 8 (CSCW2 Nov. 2024). ISSN: 25730142. DOI: 10.1145/3686934.

[208] Na Wang et al. 'HardenVR: Harassment Detection in Social Virtual Reality'. In: *Proceedings - 2024 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2024* (2024), pp. 94–104. DOI: 10.1109/VR58804.2024.00033.

[209] *Cast to a Screen*. Last Accessed: 16-07-2025. 2025. URL: https://www.meta.com/en-gb/help/quest/192719842695017/.

[210] Qiao Jin et al. 'Is Your Family Ready for VR? Ethical Concerns and Considerations in Children's VR Usage'. In: *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*. IDC '24. Delft, Netherlands: Association for Computing Machinery, 2024, pp. 436–454. ISBN: 9798400704420. DOI: 10.1145/3628516.3655804.

[211] *Introducing VR Parental Supervision Tools on Quest*. Last Accessed: 13-11-2025. Mar. 2022. URL: https://about.fb.com/news/2022/03/vr-parental-supervision-tools-on-quest/.

[212] *Get parents started with supervision for Meta Horizon*. Last Accessed: 16-07-2025. 2025. URL: https://www.meta.com/en-gb/help/quest/529890378674396/.

[213] *A Parent's Guide to Rec Room - Rec Room*. Last Accessed: 16-07-2025. URL: https://recroom.com/parents-guide.

[214] *Stream Stats*. Last Accessed: 16-07-2025. URL: https://streamscharts.com/channels/vrchat.

[215] Jiaxun Cao et al. 'Understanding Parents' Perceptions and Practices Toward Children's Security and Privacy in Virtual Reality'. In: *Proceedings - IEEE Symposium on Security and Privacy* (2024), pp. 1554–1572. ISSN: 10816011. DOI: 10.1109/SP54263.2024.00222.

[216] Juan Pablo Hourcade. 'Interaction design and children'. In: *Foundations and Trends in Human-Computer Interaction* 1 (4 2007). Cited by: 194, pp. 277 – 392. DOI: 10.1561/1100000006.

[217] Juan Pablo Hourcade et al. 'Conducting Ethical Research on Emerging Technologies for Children'. In: *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. CHI EA '25. Association for Computing Machinery, 2025. ISBN: 9798400713958. DOI: 10.1145/3706599.3716397.

[218] Karla Badillo-Urquiola et al. 'Stranger Danger! Social Media App Features Co-designed with Children to Keep Them Safe Online'. In: *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. IDC '19. Boise, ID, USA: Association for Computing Machinery, 2019, pp. 394–406. ISBN: 9781450366908. DOI: 10.1145/3311927.3323133.

[219] Naulsberry Jean Baptiste et al. 'Teens on Tech: Using an Asynchronous Remote Community to Explore Adolescents' Online Safety Perspectives'. In: *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. CSCW '23 Companion. Minneapolis, MN, USA: Association for Computing Machinery, 2023, pp. 45–49. ISBN: 9798400701290. DOI: 10.1145/3584931.3606967.

[220] Brenna McNally et al. 'Co-designing Mobile Online Safety Applications with Children'. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–9. ISBN: 9781450356206. DOI: 10.1145/3173574.3174097.

[221] Zainab Agha, Karla Badillo-Urquiola and Pamela J. Wisniewski. '"Strike at the Root": Co-designing Real-Time Social Media Interventions for Adolescent Online Risk Prevention'. In: *Proc. ACM Hum.-Comput. Interact.* 7.CSCW1 (2023). DOI: 10.1145/3579625.

[222] Zahra Ashktorab and Jessica Vitak. 'Designing Cyberbullying Mitigation and Prevention Solutions through Participatory Design With Teenagers'. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: Association for Computing Machinery, 2016, pp. 3895–3905. ISBN: 9781450333627. DOI: 10.1145/2858036.2858548.

[223] Melissa Mazmanian and Simone Lanette. '"Okay, One More Episode": An Ethnography of Parenting in the Digital Age'. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. CSCW '17. Portland, Oregon, USA: Association for Computing Machinery, 2017, pp. 2273–2286. ISBN: 9781450343350. DOI: 10.1145/2998181.2998218.

[224] Ji Youn Shin et al. 'Designing Technologies to Support Parent-Child Relationships: A Review of Current Findings and Suggestions for Future Directions'. In: *Proc. ACM Hum.-Comput. Interact.* 5.CSCW2 (Oct. 2021). DOI: 10.1145/3479585.

[225] Martina Smahelova et al. 'Mediation of young children's digital technology use: The parents' perspective'. In: *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 11.3 (Nov. 2017), Article 4. DOI: 10.5817/CP2017-3-4. URL: https://cyberpsychology.eu/article/view/8561.

[226] Radhika Garg and Subhasree Sengupta. 'Conversational Technologies for In-home Learning: Using Co-Design to Understand Children's and Parents' Perspectives'. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. Honolulu, Hawaii: Association for Computing Machinery, 2020, pp. 1–13. ISBN: 9781450367080. DOI: 10.1145/3313831.3376631.

[227] Kerryann Walsh et al. 'Best Practice Framework for Online Safety Education: Results from a rapid review of the international literature, expert review, and stakeholder consultation'. In: *International Journal of Child-Computer Interaction* 33 (2022), p. 100474. ISSN: 2212-8689. DOI: 10.1016/j.ijcci.2022.100474.

[228] Patrícia Dias and Rita Brito. 'Criteria for selecting apps: Debating the perceptions of young children, parents and industry stakeholders'. In: *Computers & Education* 165 (2021), p. 104134. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2021.104134.

[229] John W Creswell and Cheryl N Poth. *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications, 2016.

[230] Maarten Van Mechelen et al. '18 Years of ethics in child-computer interaction research: a systematic literature review'. In: *Proceedings of the Interaction Design and Children Conference*. IDC '20. London, United Kingdom: Association for Computing Machinery, 2020, pp. 161–183. ISBN: 9781450379816. DOI: 10.1145/3392063.3394407.

[231] Elizabeth Foss, Allison Druin and Mona Leigh Guha. 'Recruiting and retaining young participants: Strategies from five years of field research'. In: *ACM International Conference Proceeding Series* (2013), pp. 313–316. DOI: 10.1145/2485760.2485798.

[232] Nina Heinrichs et al. 'Parent recruitment and retention in a universal prevention program for child behavior and emotional problems: Barriers to research and program participation'. In: *Prevention Science* 6 (4 Dec. 2005), pp. 275–286. ISSN: 13894986. DOI: 10.1007/S11121-005-0006-1.

[233] Jóhanna Einarsdóttir. 'Research with children: methodological and ethical challenges'. In: *European Early Childhood Education Research Journal* 15 (2 2007), pp. 197–211. DOI: 10.1080/13502930701321477.

[234] Priscilla Alderson and Virginia Morrow. 'Consent'. In: *The Ethics of Research with Children and Young People: A Practical Handbook* (Jan. 2020). DOI: 10.4135/9781529682694.

[235] Allison Druin. 'The role of children in the design of new technology'. In: *Behaviour & Information Technology* 21 (1 Jan. 2002), pp. 1–25. ISSN: 13623001. DOI: 10.1080/01449290110108659.

[236] Anja Stedjeberg Hansen. 'How to best communicate with and encourage children during a design process'. In: *Education, Psychology* (2017). URL: https://www.semanticscholar.org/paper/How-to-best-communicate-with-and-encourage-children-Hansen/ae0e89081be616bfaea307b6de73c3660d9c75af.

[237] Alison Clark. 'Listening to and involving young children: A review of research and practice'. In: *Early Child Development and Care* 175 (6 2005), pp. 489–505. ISSN: 03004430. DOI: 10.1080/03004430500131288.

[238] TTC Labs. *How we've used co-design to develop parental supervision tools at Meta*. Accessed on 2023-04-09. URL: https://www.ttclabs.net/news/how-weve-used-co-design-to-develop-parental-supervision-tools-at-meta.

[239] Ananta Chowdhury and Andrea Bunt. 'Co-Designing with Early Adolescents: Understanding Perceptions of and Design Considerations for Tech-Based Mediation Strategies that Promote Technology Disengagement'. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. Hamburg, Germany: Association for Computing Machinery, 2023. ISBN: 9781450394215. DOI: 10.1145/3544548.3581134.

[240] Jessica Van Brummelen et al. 'What Do Children and Parents Want and Perceive in Conversational Agents? Towards Transparent, Trustworthy, Democratized Agents'. In: *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*. IDC '23. Chicago, IL, USA: Association for Computing Machinery, 2023, pp. 187–197. ISBN: 9798400701313. DOI: 10.1145/3585088.3589353.

[241] Elaheh Sanoubari et al. 'Robots, Bullies and Stories: A Remote Co-design Study with Children'. In: *Proceedings of the 20th Annual ACM Interaction Design and Children Conference*. IDC '21. Athens, Greece: Association for Computing Machinery, 2021, pp. 171–182. ISBN: 9781450384520. DOI: 10.1145/3459990.3460725.

[242] Samiullah Paracha et al. 'Co-design with Children: Using Participatory Design for Design Thinking and Social and Emotional Learning'. In: *Open Education Studies* 1 (1 Jan. 2019), pp. 267–280. ISSN: 25447831. DOI: 10.1515/EDU-2019-0021.

[243] Ross Flom, Douglas A. Gentile and Anne D. Pick. 'Infants' discrimination of happy and sad music'. In: *Infant Behavior and Development* 31 (4 Dec. 2008), pp. 716–728. ISSN: 01636383. DOI: 10.1016/j.infbeh.2008.04.004.

[244] Antony S. R. Manstead. 'Children's Understanding of Emotion'. In: *Everyday Conceptions of Emotion* (1995), pp. 315–331. DOI: 10.1007/978-94-015-8484-5_18.

[245] Albert J Caron, Robert F Caron and David J MacLean. 'Infant discrimination of naturalistic emotional expressions: the role of face and voice'. In: *Child Development* 59.3 (1988), pp. 604–616. DOI: 10.2307/1130370.

[246] Margaret M. Bradley and Peter J. Lang. 'Measuring emotion: The self-assessment manikin and the semantic differential'. In: *Journal of Behavior Therapy and Experimental Psychiatry* 25 (1 Mar. 1994), pp. 49–59. ISSN: 0005-7916. DOI: 10.1016/0005-7916(94)90063-9.

[247] Elaine C.S. Hayashi et al. 'Exploring new formats of the Self-Assessment Manikin in the design with children'. In: *ACM International Conference Proceeding Series* Part F128046 (Oct. 2016). DOI: 10.1145/3033701.3033728.

[248] Brenda J. Shields et al. 'Predictors of a child's ability to use a visual analogue scale'. In: *Child: Care, Health and Development* 29 (4 July 2003), pp. 281–290. ISSN: 1365-2214. DOI: 10.1046/J.1365-2214.2003.00343.X.

[249] Lynn M. Walsh et al. 'Color Preference and Food Choice Among Children'. In: *The Journal of Psychology* 124 (6 1990), pp. 645–653. ISSN: 19401019. DOI: 10.1080/00223980.1990.10543258.

[250] Natacha Borgers, Joop Hox and Dirk Sikkel. 'Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point'. In: *Quality and Quantity* 38 (1 Feb. 2004), pp. 17–33. ISSN: 00335177. DOI: 10.1023/B:QUQU.0000013236.29205.A6.

[251] Frederik Funke. 'A Web Experiment Showing Negative Effects of Slider Scales Compared to Visual Analogue Scales and Radio Button Scales'. In: *Social Science Computer Review* 34 (2 Apr. 2016), pp. 244–254. ISSN: 15528286. DOI: 10.1177/0894439315575477.

[252] Charlotte Bexson, Geralyn Oldham and Jo Wray. 'Safety of virtual reality use in children: a systematic review'. In: *European Journal of Pediatrics* 183 (123), pp. 2071–2090. DOI: 10.1007/s00431-024-05488-5.

[253] Polyxeni Kaimara, Andreas Oikonomou and Ioannis Deliyannis. 'Could virtual reality applications pose real risks to children and adolescents? A systematic review of ethical issues and concerns'. In: *Virtual Reality* (2021). ISSN: 14349957. DOI: 10.1007/s10055-021-00563-w.

[254] Jenifer Miehlbradt et al. 'Immersive virtual reality interferes with default head–trunk coordination strategies in young children'. In: *Scientific Reports* 11 (1 Dec. 2021), p. 17959. ISSN: 20452322. DOI: 10.1038/s41598-021-96866-8.

[255] *Virtual Reality Headsets | NSPCC*. Last Accessed: 14-09-2025. URL: https://www.nspcc.org.uk/keeping-children-safe/online-safety/virtual-reality-headsets/.

[256] *The safety of domestic virtual reality systems A literature review BEIS Research Paper Number 2020/038 RPN 4527 Safe Use of Domestic Virtual Reality Systems v3.3 3*. Tech. rep. Last Accessed: 14-09-2025. 2019, pp. 5–6. URL: https://assets.publishing.service.gov.uk/media/5f763502d3bf7f7c2bcf9eb9/safety-domestic-vr-systems.pdf.

[257] Isabella Schloss, Jakki O. Bailey and Sanjana Tripathi. '"I'm in his belly!": Children's Responses to Different Types of Characters in Virtual Reality'. In: *Proceedings of the 20th Annual ACM Interaction Design and Children Conference*. IDC '21. Athens, Greece: Association for Computing Machinery, 2021, pp. 43–48. ISBN: 9781450384520. DOI: 10.1145/3459990.3460723.

[258] Heidi Singleton et al. 'Virtual reality used to distract children and young people with long-term conditions from pain or pruritus: A scoping review using PAGER'. In: *Journal of Clinical Nursing* 33 (2 Feb. 2024), pp. 469–480. ISSN: 13652702. DOI: 10.1111/JOCN.16928.

[259] *Meta Quest 2 Elite Strap | VR Head Strap | Meta Quest | Meta Store*. Last Accessed: 29-09-2025. URL: https://www.meta.com/jp/en/quest/accessories/quest-2-elite-strap/.

[260] *Virtual Reality User Study with Children and Seniors in a VR Experience - Equal Entry*. Last Accessed: 29-09-2025. 2023. URL: https://equalentry.com/virtual-reality-user-study-with-children-and-seniors/?utm_source=chatgpt.com.

[261] Melanie Jane Ackers. 'Cyberbullying: through the eyes of children and young people'. In: *Educational Psychology in Practice* 28 (2 June 2012), pp. 141–157. ISSN: 02667363. DOI: 10.1080/02667363.2012.665356.

[262] JoLynn V. Carney. 'Perceptions of bullying and associated trauma during adolescence'. In: *Professional School Counseling* 11.3 (2008), pp. 179–187.

[263] Paula C. Ferreira et al. 'Exploring empathy in cyberbullying with serious games'. In: *Computers & Education* 166 (June 2021), p. 104155. ISSN: 0360-1315. DOI: 10.1016/J.COMPEDU.2021.104155.

[264] Antonio Calvo-Morata et al. 'Serious games to prevent and detect bullying and cyberbullying: A systematic serious games and literature review'. In: *Computers & Education* 157 (Nov. 2020), p. 103958. ISSN: 0360-1315. DOI: 10.1016/J.COMPEDU.2020.103958.

[265] Ana Paiva et al. 'Learning by feeling: Evoking empathy with synthetic characters'. In: *Applied Artificial Intelligence* 19 (3-4 Mar. 2005), pp. 235–266. ISSN: 08839514. DOI: 10.1080/08839510590910165.

[266] Maria Sapouna et al. 'Virtual learning intervention to reduce bullying victimization in primary school: a controlled trial'. In: *Journal of Child Psychology and Psychiatry* 51 (1 Jan. 2010), pp. 104–112. ISSN: 1469-7610. DOI: 10.1111/J.1469-7610.2009.02137.X.

[267] A. Martinelli et al. 'Hostile attribution bias and aggression in children and adolescents: A systematic literature review on the influence of aggression subtype and gender'. In: *Aggression and Violent Behavior* 39 (Mar. 2018), pp. 25–32. ISSN: 1359-1789. DOI: 10.1016/J.AVB.2018.01.005.

[268] R. S. Aylett et al. 'FearNot! – An Experiment in Emergent Narrative'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 3661 LNAI (2005), pp. 305–316. ISSN: 1611-3349. DOI: 10.1007/11550617_26.

[269] Natalie Vannini et al. '"FearNot!": A computer-based anti-bullying-programme designed to foster peer intervention'. In: *European Journal of Psychology of Education* 26 (1 2011), pp. 21–44. ISSN: 18785174. DOI: 10.1007/S10212-010-0035-4.

[270] Carolien Rieffe, Lizet Ketelaar and Carin H. Wiefferink. 'Assessing empathy in young children: Construction and validation of an Empathy Questionnaire (EmQue)'. In: *Personality and Individual Differences* 49.5 (2010), pp. 362–367. ISSN: 0191-8869. DOI: 10.1016/j.paid.2010.03.046.

[271] Janet Strayer and William Roberts. 'Children's empathy and role taking: Child and parental factors, and relations to prosocial behavior'. In: *Journal of Applied Developmental Psychology* 10.2 (1989), pp. 227–239. ISSN: 0193-3973. DOI: 10.1016/0193-3973(89)90006-3.

[272] Divine Maloney, Guo Freeman and Donghee Yvette Wohn. '"Talking without a Voice": Understanding Non-verbal Communication in Social Virtual Reality'. In: *Proc. ACM Hum.-Comput. Interact.* 4.CSCW2 (Oct. 2020). DOI: 10.1145/3415246.

[273] Lawrence Tychsen and Paul Foeller. 'Effects of Immersive Virtual Reality Headset Viewing on Young Children: Visuomotor Function, Postural Stability, and Motion Sickness'. In: *American Journal of Ophthalmology* 209 (Jan. 2020), pp. 151–159. ISSN: 0002-9394. DOI: 10.1016/J.AJO.2019.07.020.

[274] Gülçin Özalp Gerçeker et al. 'Effects of virtual reality on pain, fear and anxiety during blood draw in children aged 5–12 years old: A randomised controlled study'. In: *Journal of Clinical Nursing* 29 (7-8 Apr. 2020), pp. 1151–1161. ISSN: 13652702. DOI: 10.1111/JOCN.15173;CTYPE:STRING:JOURNAL.

[275] Junior Vitorino Fandim et al. 'Effectiveness of virtual reality in children and young adults with cerebral palsy: a systematic review of randomized controlled trial'. In: *Brazilian Journal of Physical Therapy* 25 (4 July 2021), pp. 369–386. ISSN: 1413-3555. DOI: 10.1016/J.BJPT.2020.11.003.

[276] Andrea Stevenson Won et al. 'Immersive virtual reality for pediatric pain'. In: *Children* 4 (7 July 2017). ISSN: 22279067. DOI: 10.3390/CHILDREN4070052.

[277] Veronica Lambert et al. 'Virtual reality distraction for acute pain in children'. In: *Cochrane Database of Systematic Reviews* (10 Oct. 2020). ISSN: 14651858. DOI: 10.1002/14651858.CD010686.pub2.

[278] John C. Flanagan. 'The critical incident technique'. In: *Psychological Bulletin* 51 (4 July 1954), pp. 327–358. ISSN: 00332909. DOI: 10.1037/H0061470.

[279] Brittany Huber, Kate Highfield and Jordy Kaufman. 'Detailing the digital experience: Parent reports of children's media use in the home learning environment'. In: *British Journal of Educational Technology* 49 (5 Sept. 2018), pp. 821–833. ISSN: 1467-8535. DOI: 10.1111/BJET.12667.

[280] Qualtrics. *Qualtrics Survey Platform*. Last Accessed: 21-07-2025. URL: https://uofg.qualtrics.com/.

[281] Lisa A Elkin et al. 'An Aligned Rank Transform Procedure for Multifactor Contrast Tests; An Aligned Rank Transform Procedure for Multifactor Contrast Tests'. In: 15 (21 2021). DOI: 10.1145/3472749.3474784.

[282] Lorelli S. Nowell et al. 'Thematic Analysis: Striving to Meet the Trustworthiness Criteria'. In: *International Journal of Qualitative Methods* 16 (1 Oct. 2017). ISSN: 16094069. DOI: 10.1177/1609406917733847.

[283] Virginia Braun and Victoria Clarke. 'Using thematic analysis in psychology'. In: *Qualitative Research in Psychology* 3 (2 2006), pp. 77–101. ISSN: 14780887. DOI: 10.1191/1478088706QP063OA.

[284] *Health and Safety | HTC United States*. Last Accessed: 22-07-2022. URL: https://www.htc.com/us/sunrise-safety-guide/.

[285] Danah Boyd and Eszter Hargittai. 'Facebook privacy settings: Who cares?' In: *First Monday* 15.8 (July 2010). DOI: 10.5210/fm.v15i8.3086.

[286] John T.E. Richardson. 'Eta squared and partial eta squared as measures of effect size in educational research'. In: *Educational Research Review* 6.2 (2011), pp. 135–147. ISSN: 1747-938X. DOI: 10.1016/j.edurev.2010.12.001.

[287] Peter E. Morris and Catherine O. Fritz. 'Effect sizes in memory research'. In: *Memory* 21 (7 Oct. 2013), pp. 832–842. ISSN: 09658211. DOI: 10.1080/09658211.2013.763984.

[288] Barry Glassner. *The culture of fear: Why Americans are afraid of the wrong things.* Basic Books, 1999, pp. xix–xxviii.

[289] Liliana Pasquale et al. 'Digital Age of Consent and Age Verification: Can They Protect Children?' In: *IEEE Software* 39.3 (2022), pp. 50–57. DOI: 10.1109/MS.2020.3044872.

[290] Andy Phippen. 'Protecting children in the metaverse: it's easy to blame big tech, but we all have a role to play'. In: (Mar. 2022). Last Accessed: 02-10-2023. URL: https://blogs.lse.ac.uk/parenting4digitalfuture/.

[291] Hinduja Sameer. *Child Grooming and the Metaverse – Issues and Solutions - Cyberbullying Research Center*. Last Accessed: 02-10-2023. n.d. URL: https://cyberbullying.org/child-grooming-metaverse.

[292] *I Sold Too Many Copies of GTA V To Parents Who Didn't Give a Damn - Kotaku*. 2013. URL: https://kotaku.com/i-sold-too-many-copies-of-gta-v-to-parents-who-didnt-g-1371011511.

[293] Louise Egan. *Video Game Content Ratings: Does Anyone Care Anymore? | The Artifice*. 2014. URL: https://the-artifice.com/video-game-content-ratings-anyone-care-anymore/.

[294] *Use of parental controls | ICO*. Last Accessed: 09-08-2022. URL: https://ico.org.uk/for-organisations/childrens-code-hub/how-to-use-our-guidance-for-standard-one-best-interests-of-the-child/children-s-code-best-interests-framework/use-of-parental-controls/\#recommendations.

[295] Leonie Maria Tanczer, Isabel López-Neira and Simon Parkin. ‘‘I feel like we’re really behind the game’: perspectives of the United Kingdom’s intimate partner violence support sector on the rise of technology-facilitated abuse’. In: *Journal of Gender-Based Violence* 5 (3 Oct. 2021), pp. 431–450. ISSN: 2398-6808. DOI: 10.1332/239868021X16290304343529.

[296] Julia Slupska and Leonie Maria Tanczer. ‘Threat Modeling Intimate Partner Violence: Tech Abuse as a Cybersecurity Challenge in the Internet of Things’. In: *The Emerald International Handbook of Technology-Facilitated Violence and Abuse* (June 2021), pp. 663–688. DOI: 10.1108/978-1-83982-848-520211049.

[297] Lynn Schofield Clark. ‘75Identity 2.0: Young People and Digital and Mobile Media’. In: *The Parent App: Understanding Families in the Digital Age*. Oxford University Press, Nov. 2012. ISBN: 9780199899616. DOI: 10.1093/acprof:oso/9780199899616.003.0004.

[298] Alexander Bogner, Beate Littig and Wolfgang Menz. ‘Introduction: Expert Interviews - An Introduction to a New Methodological Debate’. In: *Interviewing Experts*. Research Methods Series. Palgrave Macmillan, London, 2009, pp. 1–13. DOI: 10.1057/9780230244276_1.

[299] Christian Von Soest. ‘Why Do We Speak to Experts? Reviving the Strength of the Expert Interview Method’. In: *Perspectives on Politics* 21.1 (2023), pp. 277–287. DOI: 10.1017/S1537592722001116.

[300] *Most played VR Games Steam Charts · SteamDB*. Last Accessed: 17-01-2024. URL: https://steamdb.info/charts/?tagid=21978.

[301] *14 Best Virtual Reality Social Games 2024*. Last Accessed: 17-01-2024. 2023. URL: https://rigorousthemes.com/blog/best-virtual-reality-social-games/.

[302] Marcel Jonas et al. ‘Towards a taxonomy of social VR application design’. In: *CHI PLAY 2019 - Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play* 19 (Oct. 2019), pp. 437–444. DOI: 10.1145/3341215.3356271.

[303] *Basic Controls | Rec Room Wiki | Fandom*. URL: https://rec-room.fandom.com/wiki/Basic_Controls.

[304] *Watch Menu | Rec Room Wiki | Fandom*. Last Accessed: 29-11-2025. URL: https://rec-room.fandom.com/wiki/Watch_Menu.

[305] *Settings - VRChat Wiki*. Last Accessed: 29-11-2025. URL: https://wiki.vrchat.com/wiki/Settings.

[306] *Applying for Volunteer Moderator – Rec Room Help Center*. Last Accessed: 29-11-2025. URL: https://recroom.zendesk.com/hc/en-us/articles/4419902650135-Applying-for-Volunteer-Moderator.

[307] *Little Kid Shows Me the Dark Side of VR Chat - YouTube*. Last Accessed 28-08-2023. 2019. URL: https://www.youtube.com/watch?v=DBV1pTzsevM.

[308] *QCAmap*. Last Accessed 05-09-2023. 2022. URL: https://www.qcamap.org/ui/en/home.

[309] *Free Online Appointment Scheduling Software | Calendly*. URL: https://calendly.com/.

[310] Qijia Chen, Jie Cai and Giulio Jacucci. ‘”People are Way too Obsessed with Rank”: Trust System in Social Virtual Reality’. In: *Computer Supported Cooperative Work (CSCW)* (2024). DOI: 10.1007/s10606-024-09498-7.

[311] *State of Voice Moderation - Rec Room Developer Blog*. Last Accessed: 10-10-2024. 2024. URL: https://blog.recroom.com/posts/voice-moderation-updates.

[312] Robert Thornberg and Sven Knutsen. 'Teenagers' Explanations of Bullying'. In: *Child and Youth Care Forum* 40 (3 Nov. 2010), pp. 177–192. ISSN: 1053-1890. DOI: 10.1007/s10566-010-9129-z.

[313] *Police 'investigate sexual abuse of young girl's avatar in the metaverse' - prompting NSPCC warning | Science & Tech News | Sky News*. Last Accessed: 10-09-2024. 2024. URL: https://news.sky.com/story/amp/police-investigate-sexual-abuse-of-young-girls-avatar-in-the-metaverse-prompting-nspcc-warning-13041003.

[314] Haesoo Kim et al. 'ReSPect: Enabling Active and Scalable Responses to Networked Online Harassment'. In: *Proceedings of the ACM on Human-Computer Interaction* 8 (CSCW1 Apr. 2024). ISSN: 25730142. DOI: 10.1145/3637394.

[315] Joanne E. Gray, Marcus Carter and Ben Egliston. 'Governing Social Virtual Reality: Preparing for the Content, Conduct and Design Challenges of Immersive Social Media'. In: *Governing Social Virtual Reality: Preparing for the Content, Conduct and Design Challenges of Immersive Social Media* (Jan. 2024), pp. 1–94. DOI: 10.1007/978-3-031-61831-4/COVER.

[316] Qijia Chen, Michiel M. Spapé and Giulio Jacucci. 'Understanding Phantom Tactile Sensation on Commercially Available Social Virtual Reality Platforms'. In: *Proceedings of the ACM on Human-Computer Interaction* 8 (CSCW1 Apr. 2024). ISSN: 25730142. DOI: 10.1145/3637418.

[317] Lawrence E. Williams and John A. Bargh. 'Keeping One's Distance: The Influence of Spatial Distance Cues on Affect and Evaluation'. In: *Psychological Science* 19.3 (2008), pp. 302–308. DOI: 10.1111/j.1467-9280.2008.02084.x.

[318] Yaacov Trope, Nira Liberman and Cheryl Wakslak. 'Construal Levels and Psychological Distance: Effects on Representation, Prediction, Evaluation, and Behavior'. In: *Journal of consumer psychology : the official journal of the Society for Consumer Psychology* 17 (2 2007), p. 83. ISSN: 10577408. DOI: 10.1016/S1057-7408(07)70013-X.

[319] Yaacov Trope and Nira Liberman. 'Construal-Level Theory of Psychological Distance'. In: *Psychological Review* 117 (2 Apr. 2010), pp. 440–463. ISSN: 0033295X. DOI: 10.1037/A0018963.

[320] *Safety Features: Chat, Privacy & Filtering – Roblox Support*. Last Accessed: 10-09-2024. 2024. URL: https://en.help.roblox.com/hc/en-us/articles/203313120-Safety-Features-Chat-Privacy-Filtering.

[321] *Pause your experience in Meta Horizon Worlds | Quest help | Meta Store*. Last Accessed: 10-01-2025. 2024. URL: https://www.meta.com/en-gb/help/quest/articles/horizon/safety-and-privacy-in-horizon-worlds/pause-in-horizon-worlds/.

[322] Unity Technologies. *Unity Real-Time Development Platform 3D, 2D, VR And AR Engine*. Last Accessed: 10-09-2024. URL: https://unity.com/.

[323] *VR Beginner: The Escape Room | Tutorials | Unity Asset Store*. Last Accessed: 10-09-2024. 2021. URL: https://assetstore.unity.com/packages/templates/tutorials/vr-beginner-the-escape-room-163264.

[324] *UCL-VR/ubiq*. Last Accessed: 29-11-2022. 2022. URL: https://github.com/UCL-VR/ubiq.

[325] Epidemic Sound Effects. *Small Crowd Talking Ambience*. Last Accessed: 10-09-2024. 2019. URL: https://www.youtube.com/watch?v=IJQdnHT3MCA&t=33s.

[326] David Morgan. 'Follow-up qualitative extensions to core quantitative research projects'. In: *Integrating Qualitative and Quantitative Methods: A Pragmatic Approach*. SAGE Publications, Inc., 2014, pp. 153–178. DOI: 10.4135/9781544304533.

[327] Elaine C. S. Hayashi et al. 'Exploring new formats of the Self-Assessment Manikin in the design with children'. In: *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems*. IHC '16. São Paulo, Brazil: Association for Computing Machinery, 2016. ISBN: 9781450352352. DOI: 10.1145/3033701.3033728.

[328] Jacob O Wobbrock et al. 'The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures'. In: (2011). URL: https://depts.washington.edu/acelab/proj/art/index.html.

[329] *Fortnite - PlayStation*. Last Accessed: 05-12-2024. 2017. URL: https://www.playstation.com/en-us/games/fortnite/.

[330] *Hogwarts Legacy - PlayStation*. Last Accessed: 05-12-2024. 2023. URL: https://www.playstation.com/en-gb/games/hogwarts-legacy/.

[331] *Splatoon - Nintendo*. Last Accessed: 05-12-2024. 2015. URL: https://splatoon.nintendo.com/.

[332] *Wands - Meta Quest*. Last Accessed: 05-12-2024. 2019. URL: https://www.meta.com/en-gb/experiences/wands/2096339947076151/.

[333] Małgorzata Gawlik-Kobylińska. 'The Proteus effect on security and safety exergamers'. In: *Interactive Learning Environments* 31 (5 July 2023), pp. 2707–2721. ISSN: 17445191. DOI: 10.1080/10494820.2021.1898988;CTYPE:STRING:JOURNAL.

[334] *Meta moves to tackle creepy behaviour in virtual reality - BBC News*. Last Accessed: 29-07-2025. 2022. URL: https://www.bbc.co.uk/news/technology-60247542.

[335] Tom Acres. *Sex offenders using virtual reality to abuse children as victims describe feeling 'phantom touch' | Science, Climate & Tech News | Sky News*. Last Accessed: 29-07-2025. 2023. URL: https://news.sky.com/story/sex-offenders-using-virtual-reality-to-abuse-children-as-victims-describe-feeling-phantom-touch-12954906.

[336] Rachael Venables. *Paedophiles are starting to use virtual reality headsets to view child abuse images, data shows | UK News | Sky News*. Last Accessed: 29-07-2025. 2023. URL: https://news.sky.com/story/paedophiles-are-starting-to-use-virtual-reality-headsets-to-view-child-abuse-images-data-shows-12816889.

[337] Susan Hanley Duncan and Louis D. Brandeis. 'Cyberbullying and Restorative Justice'. In: *Cyberbullying Across the Globe: Gender, Family, and Mental Health* (Jan. 2016), pp. 239–257. DOI: 10.1007/978-3-319-25552-1_12.

[338] Elnaz Moghimi et al. 'Using a co-design approach to develop a Preventative Online Mental Health Program for Youth (POMHPY): a quality improvement project'. In: *BMC Health Services Research* 25 (1 Dec. 2025), pp. 1–12. ISSN: 14726963. DOI: 10.1186/S12913-024-12101-W.

[339] Leanne Bowler et al. 'The Meaning of "Participation" in Co-Design with Children and Youth: Relationships, Roles, and Interactions'. In: *Proceedings of the Association for Information Science and Technology* 58 (1 Oct. 2021), pp. 13–24. ISSN: 2373-9231. DOI: 10.1002/PRA2.432.

[340] Leelo Andren et al. 'Co-design As Collaborative Research'. In: *Eesti Teadusliku Seltsi Rootsis : Aastaraamat. Annales Societatis Litterarum Estonicae in Svecia* 11 (Sept. 2018), pp. 216;ill. URL: https://connected-communities.org/wp-content/uploads/2018/07/Co-Design_SP.pdf.

[341] Guo Freeman, Lingyuan Li and Kelsea Schulenberg. '"I Have Abused Someone Who Abused Me": Understanding People Who Have Experienced Both Sides of Harassment Accusations in Social VR'. In: *Proc. ACM Hum.-Comput. Interact.* 9.2 (May 2025). DOI: 10.1145/3711005.

[342] Ming Li Hsieh, Shun Yung Kevin Wang and Yusheng Lin. 'Perceptions of Punishment Risks Among Youth: Can Cyberbullying Be Deterred?' In: *Journal of School Violence* 22 (3 July 2023), pp. 307–321. ISSN: 15388239. DOI: 10.1080/15388220.2023.2183865.

[343] Anne Linda Camerini et al. 'Cyberbullying perpetration and victimization among children and adolescents: A systematic review of longitudinal studies'. In: *Telematics and Informatics* 49 (June 2020), p. 101362. ISSN: 0736-5853. DOI: 10.1016/J.TELE.2020.101362.

[344] Kostas A. Fanti, Andreas G. Demetriou and Veronica V. Hawa. 'A longitudinal study of cyberbullying: Examining risk and protective factors'. In: *European Journal of Developmental Psychology* 9 (2 Mar. 2012), pp. 168–181. ISSN: 17405629. DOI: 10.1080/17405629.2011.643169.

[345] Michael D. Maltz. *Recidivism*. Quantitative Studies in Social Relations. Orlando, FL: Academic Press, 1984, pp. xiv + 240. ISBN: 0-12-468980-9. URL: https://books.google.co.uk/books?hl=en&lr=&id=VxLb2pHetakC&oi=fnd&pg=PR13&dq=recidivism+&ots=b6Y5izdsQj&sig=nuKSBpDQk751sFXNyzKAqtCSuIo&redir_esc=y#v=onepage&q=recidivism&f=false.

[346] *Data protection: The UK's data protection legislation - GOV.UK*. Last Accessed: 02-09-2025. URL: https://www.gov.uk/data-protection.

[347] *Online Safety Act 2023*. Last Accessed: 07-05-2024. 2023. URL: https://www.legislation.gov.uk/ukpga/2023/50/contents/enacted.

[348] Zhendong Chu et al. 'LLM Agents for Education: Advances and Applications'. In: *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)* 1 (Mar. 2025). URL: https://arxiv.org/pdf/2503.11733.

[349] Yongxiang Zhang and Weiwei Pan. 'A scoping review of embodied conversational agents in education: trends and innovations from 2014 to 2024'. In: *Interactive Learning Environments* (Feb. 2025). ISSN: 17445191. DOI: 10.1080/10494820.2025.2468972.

[350] Sandra Wiktor, Mohsen Dorodchi and Nicole Wiktor. 'AI Can Help Instructors Help Students: An LLM-Supported Approach to Generating Customized Student Reflection Responses'. In: *Proceedings - Frontiers in Education Conference, FIE* (2024). ISSN: 15394565. DOI: 10.1109/FIE61694.2024.10893324.

[351] Sameer Hinduja and Justin W. Patchin. 'Cultivating youth resilience to prevent bullying and cyberbullying victimization'. In: *Child Abuse & Neglect* 73 (2017), pp. 51–62. ISSN: 0145-2134. DOI: 0.1016/j.chiabu.2017.09.010.

[352] Alan Reid et al. 'Participation and learning: Perspectives on education and the environment, health and sustainability'. In: *Participation and Learning: Perspectives on Education and the Environment, Health and Sustainability* (2008), pp. 1–345. DOI: 10.1007/978-1-4020-6416-6.

[353] Hannah Bloch-Wehba. 'Automation in Moderation'. In: *Cornell International Law Journal* 53 (1 Mar. 2020). URL: https://scholarship.law.tamu.edu/facscholar/1448.

[354] Maral Dadvar and Franciska de Jong. 'Cyberbullying Detection: A Step toward a Safer Internet Yard'. In: *Proceedings of the 21st International Conference on World Wide Web*. WWW '12 Companion. Lyon, France: Association for Computing Machinery, 2012, pp. 121–126. ISBN: 9781450312301. DOI: 10.1145/2187980.2187995.

[355] Yubo Kou and Xinning Gui. 'Flag and Flaggability in Automated Moderation: The Case of Reporting Toxic Behavior in an Online Game Community'. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: Association for Computing Machinery, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445279.

[356] J Brooke. *Usability evaluation in industry, chap. SUS: a "quick and dirty" usability scale*. 1996.

[357] Sandra G. Hart and Lowell E. Staveland. 'Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research'. In: *Human Mental Workload*. Ed. by Peter A. Hancock and Najmedin Meshkati. Vol. 52. Advances in Psychology. North-Holland, 1988, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9.

[358] Aaron Powers and Sara Kiesler. 'The Advisor Robot: Tracing People's Mental Model from a Robot's Physical Attributes'. In: vol. 2006. Jan. 2006, pp. 218–225. DOI: 10.1145/1121241.1121280.

[359] *K-means Cluster Analysis · UC Business Analytics R Programming Guide*. Last Accessed: 08-01-2023. URL: https://uc-r.github.io/kmeans_clustering.

[360] Shaun Alexander MacDonald et al. 'User Preferences for Calming Affective Haptic Stimuli in Social Settings'. In: *ICMI 2021 - Proceedings of the 2021 International Conference on Multimodal Interaction* (Oct. 2021), pp. 387–396. DOI: 10.1145/3462244.3479903.

[361] Evropi Stefanidi et al. 'When Children Program Intelligent Environments: Lessons Learned from a Serious AR Game'. In: *Proceedings of Interaction Design and Children, IDC 2021* (June 2021), pp. 375–386. DOI: 10.1145/3459990.3462463.

[362] Cristina Fiani and Stacy Marsella. 'Investigating the Non-Verbal Behavior Features of Bullying for the Development of an Automatic Recognition System in Social Virtual Reality'. In: *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*. AVI 2022. Frascati, Rome, Italy: Association for Computing Machinery, 2022. ISBN: 9781450397193. DOI: 10.1145/3531073.3534492.

[363] Alexis Hiniker, Sarita Y. Schoenebeck and Julie A. Kientz. 'Not at the dinner table: Parents' and children's perspectives on family technology rules'. In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW* 27 (Feb. 2016), pp. 1376–1389. DOI: 10.1145/2818048.2819940.

[364] Masahiro Mori, Karl F. MacDorman and Norri Kageki. 'The uncanny valley'. In: *IEEE Robotics and Automation Magazine* 19 (2 2012), pp. 98–100. ISSN: 10709932. DOI: 10.1109/MRA.2012.2192811.

[365] Mairi Therese Deighan, Amid Ayobi and Aisling Ann O'Kane. 'Social Virtual Reality as a Mental Health Tool: How People Use VRChat to Support Social Connectedness and Wellbeing'. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. Hamburg, Germany: Association for Computing Machinery, 2023. ISBN: 9781450394215. DOI: 10.1145/3544548.3581103.

[366] Sasha Alexdottir and Xiaosong Yang. 'Phantom Touch phenomenon as a manifestation of the Visual-Auditory-Tactile Synaesthesia and its impact on the users in virtual reality'. In: *Proceedings - 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct, ISMAR-Adjunct 2022* (2022), pp. 727–732. DOI: 10.1109/ISMAR-ADJUNCT57072.2022.00218.

[367] Melvin Abraham et al. 'Implications of XR on Privacy, Security and Behaviour: Insights from Experts'. In: *Nordic Human-Computer Interaction Conference*. NordiCHI '22. Aarhus, Denmark: Association for Computing Machinery, 2022. ISBN: 9781450396998. DOI: 10.1145/3546155.3546691.

[368] Jeffrey H. Reiman. 'Driving to the panopticon: A philosophical exploration of the risks to Privacy Posed by the Highway Technology of the future'. In: *Privacy*. London, UK: Taylor and Francis, Sept. 2017, pp. 159–176. ISBN: 9781315246024. URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315246024-8/driving-panopticon-philosophical-exploration-risks-privacy-posed-highway-technology-future-jeffrey-reiman.

[369] *Community Safety and Moderation | Discord*. Last Accessed: 14-11-2025. URL: https://discord.com/community-moderation-safety.

[370] Karla A. Badillo-Urquiola et al. 'Stranger Danger!: Social Media App Features Co-designed with Children to Keep Them Safe Online'. In: *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (2019). URL: https://api.semanticscholar.org/CorpusID:174802028.

[371] Sijia Xiao, Coye Cheshire and Niloufar Salehi. 'Sensemaking, Support, Safety, Retribution, Transformation: A Restorative Justice Approach to Understanding Adolescents' Needs for Addressing Online Harm'. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22. New Orleans, LA: Association for Computing Machinery, 2022. ISBN: 9781450391573. DOI: 10.1145/3491102.3517614.

[372] Roy A. Willems et al. 'Internet Use and Perceived Parental Involvement among Adolescents from Lower Socioeconomic Groups in Europe: An Exploration'. In: *Children* 10 (11 Nov. 2023). ISSN: 22279067. DOI: doi.org/10.3390/children10111780.

[373] Michaela Geržičáková, Lenka Dedkova and Vojtěch Mýlek. 'What do parents know about children's risky online experiences? The role of parental mediation strategies'. In: *Computers in Human Behavior* 141 (Apr. 2023), p. 107626. ISSN: 0747-5632. DOI: 10.1016/J.CHB.2022.107626.

[374] Alena Cerna, Hana Machackova and Lenka Dedkova. 'Whom to Trust: The Role of Mediation and Perceived Harm in Support Seeking by Cyberbullying Victims'. In: *Children & Society* 30 (4 July 2016), pp. 265–277. ISSN: 1099-0860. DOI: 10.1111/CHSO.12136.

[375] Pamela J. Wisniewski, Jessica Vitak and Heidi Hartikainen. 'Privacy in adolescence'. In: *Modern Socio-Technical Perspectives on Privacy* (Feb. 2022), pp. 315–336. DOI: 10.1007/978-3-030-82786-1_14.

[376] Craig McDonald-Brown, Kumar Laxman and John Hope. 'Sources of support and mediation online for 9–12-year-old children'. In: *E-Learning and Digital Media* 14 (1-2 Mar. 2017), pp. 52–71. ISSN: 20427530. DOI: 10.1177/2042753017692430.

[377] J. Maya Hernandez et al. 'Parental Monitoring of Early Adolescent Social Technology Use in the US: A Mixed-Method Study'. In: *Journal of Child and Family Studies* 33 (3 Mar. 2024), pp. 759–776. ISSN: 15732843. DOI: 10.1007/S10826-023-02734-6.

[378] Janet Goodall and Caroline Montgomery. 'Parental involvement to parental engagement: a continuum'. In: *Mapping the Field* (Oct. 2023), pp. 158–169. DOI: 10.4324/9781003403722-13.

[379] *Gender Neutral Voice Generator, TTS for Non-Binary content*. Last Accessed: 10-09-2024. URL: https://speechgen.io/en/neutral/.

[380] *Text-to-Speech AI: Lifelike Speech Synthesis | Google Cloud*. Last Accessed: 03-08-2025. 2025. URL: https://cloud.google.com/text-to-speech.

[381] *respectme | Scotland's anti-bullying service*. Last Accessed: 14-11-2025. URL: https://www.respectme.org.uk/.

[382] Danielle M. Reynald. 'Chapter 2: Guarding Against Crime : Measuring Guardianship within Routine Activity Theory'. In: *Guarding Against Crime*. Routledge, Apr. 2016, pp. 13–31. ISBN: 9781315586007. DOI: 10.4324/9781315586007.

[383] Lawrence E. Cohen and Marcus Felson. 'Social Change and Crime Rate Trends: A Routine Activity Approach'. In: *American Sociological Review* 44 (4 Aug. 1979), p. 588. ISSN: 00031224. DOI: 10.2307/2094589.

[384] Marcus Felson and Rachel Boba. *Crime and Everyday Life*. Fourth. SAGE publications, 2010, p. 28. URL: https://books.google.co.uk/books?hl=en&lr=&id=TI6xdKDLwtcC&oi=fnd&pg=PR1&ots=ro8uLXPule&sig=9e59Ny-40QlgLvI3WcYK3EI1A-o&redir_esc=y#v=onepage&q&f=false.

[385] Howard Zehr. *The little book of restorative justice: Revised and updated*. Simon and Schuster, 2015.

[386] Daniel Falla et al. 'Why do Victims become Perpetrators of Peer Bullying? Moral Disengagement in the Cycle of Violence'. In: *Youth and Society* 54 (3 Apr. 2022), pp. 397–418. ISSN: 15528499. DOI: 10.1177/0044118X20973702.

[387] Catherine N. Dulmus, Karen M. Sowers and Matthew T. Theriot. 'Prevalence and Bullying Experiences of Victims and Victims Who Become Bullies (Bully-Victims) at Rural Schools'. In: *Victims & Offenders* 1 (1 Apr. 2006), pp. 15–31. ISSN: 15564991. DOI: 10.1080/15564880500498945.

[388] Petr Slovak et al. 'Mediating Conflicts in Minecraft: Empowering Learning in Online Multiplayer Games'. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–13. ISBN: 9781450356206. DOI: 10.1145/3173574.3174169.

[389] Mirko Franco et al. 'Shielding the Metaverse: How Automated Content Moderation Can Mitigate Harassment in Social VR'. In: ().

[390] Adam Mosseri. *Introducing New Ways to Protect Our Community from Abuse | Instagram Blog*. Last Accessed: 23-01-2024. 2021. URL: https://about.instagram.com/blog/announcements/introducing-new-ways-to-protect-our-community-from-abuse.

[391] Ya-Wen Avon Chang, Laurianne Sitbon and Leonie Simpson. 'Towards a Secured and Safe Online Social Media Design Framework for People with Intellectual Disability'. In: (2021). DOI: 10.1145/3441852.3476540.

[392] B. F. Skinner. *The Behavior of Organisms*. B. F. Skinner Foundation, 1938. URL: https://books.google.co.uk/books?hl=en&lr=&id=S9WNCwAAQBAJ&oi=fnd&pg=PT20&ots=LnwqamwBH-&sig=SGAPVzfIx4ZZIo522nyl55uOP1A&redir_esc=y#v=onepage&q&f=false.

[393] Mark A. Adams. 'Reinforcement theory and behavior analysis.' In: *Behavioral Development Bulletin* 9 (1 2000), pp. 3–6. DOI: 10.1037/H0100529.

[394] Albert Bandura and Richard H Walters. *Social learning theory*. Vol. 1. Prentice hall Englewood Cliffs, NJ, 1977.

[395] *Safety and privacy tools available for your teen in Meta Horizon Worlds*. Last Accessed: 07-11-2025. 2023. URL: https://www.meta.com/en-gb/help/quest/195981279862866/.

[396] *Junior Accounts & Age Verification*. Last Accessed: 07-05-2024. URL: https://recroom.zendesk.com/hc/en-us/articles/4426900227735-Junior-Accounts-Age-Verification.

[397] Kelsea Schulenberg et al. '"Creepy Towards My Avatar Body, Creepy Towards My Body": How Women Experience and Manage Harassment Risks in Social Virtual Reality'. In: *Proc. ACM Hum.-Comput. Interact.* 7.CSCW2 (Oct. 2023). DOI: 10.1145/3610027.

[398] Kelsea Schulenberg et al. 'Does Who You Are or Appear to Be Matter?: Understanding Identity-Based Harassment in Social VR Through the Lens of (Mis)Perceived Identity Revelation'. In: *Proceedings of the ACM on Human-Computer Interaction* 8 (CSCW2 Nov. 2024). ISSN: 25730142. DOI: 10.1145/3686923.

[399] Guo Freeman et al. '"Comforting and Small Like a House Cat, Big and Intimidating Like a Bodyguard": How Women Perceive and Envision AI Companions as a New Harassment Mitigation Approach in Social VR'. In: *Conference on Human Factors in Computing Systems - Proceedings* (Apr. 2025). DOI: 10.1145/3706598.3713473.

[400] *Augmented Reality Market Size, Share & Trends Report 2030*. Last Accessed: 25-10-2025. URL: https://www.grandviewresearch.com/industry-analysis/augmented-reality-market.

[401] *Ray-Ban | Meta AI glasses Gen 2 | Ray-Ban® GB*. URL: https://www.ray-ban.com/uk/ray-ban-meta-ai-glasses.

[402] *Snap to Launch New Lightweight, Immersive Specs in 2026*. Last Accessed: 25-10-2025. URL: https://newsroom.snap.com/launch-specs-2026.

[403] *Virtual Reality (VR) Market Size, Growth, Share | Report, 2032*. Oct. 2025. URL: https://www.fortunebusinessinsights.com/industry-reports/virtual-reality-market-101378.