Beyond technical fixes: Exploring ethical and political education in the training of data scientists

Nicola Jane Weaver

BA (Value & Policy Studies), BA (Hons) Economics, MSc

Submitted in fulfilment of the requirements of the Degree of Doctor of Education (EdD)

School of Education, College of Social Sciences

University of Glasgow

September 2025

Abstract

This dissertation adopts a primarily qualitative approach, grounded in semi-structured interviews with in-service data scientists, to explore how ethical dilemmas are experienced, interpreted and navigated in professional contexts. The research is motivated by the increasing societal scrutiny of data-driven technologies and the recognition that data science, far from being a neutral or purely technical field, is entangled with questions of power, justice and social impact.

The aim is to investigate the everyday ethical challenges data scientists face and their perceptions of the adequacy and relevance of existing (or absent) ethics education and professional guidelines. The interviews examine how data scientists make sense of their responsibilities, the pressures and constraints they encounter in their organisational settings, and the various ways by which they attempt to balance technical objectives with broader social values. In doing so, the dissertation addresses a gap in the literature: the lived realities of data scientists who face the continuous need for ethical reasoning and decision-making in the rapidly evolving landscape of data science.

These empirical insights form the foundation for a secondary, but vital, philosophical dimension of the project: a normative argument that ethics education in data science should be reimagined to move beyond technical compliance and individual responsibility. Drawing on concepts from liberal and critical theories, the dissertation highlights the limitations of current approaches that emphasise personal virtue or adherence to professional codes, arguing that such frameworks are insufficient to address the structural and systemic dimensions of harm that data science can produce.

Instead, this dissertation proposes how ethics training can be reconceptualised to address broader structural concerns, including data and algorithmic biases, power asymmetries and social justice. In this way, my research bridges empirical inquiry and theory, using practitioners' voices to describe the present landscape as well as to inform a more politically engaged vision for the future of ethics education in data science. By examining the individual experiences of data scientists, the study contributes to ongoing efforts to develop more robust, justice-oriented frameworks for both the teaching and practice of data science ethics. In doing so, it aspires to cultivate a data science profession that is not only technically proficient but also socially responsible and attuned to the demands of justice in a digital and data-driven world.

Table of contents

Abstract **1**

List of tables **5**

Acknowledgements **6**

Author's declaration **7**

Definitions/abbreviations **8**

Chapter 1: Introduction **9**

The aims of my study *15*

Drawing on critical and liberal theories *16*

Chapter 2: Outline and structure **19**

Research approach *22*

Chapter 3: Literature review **24**

Key concepts and issues in data science *24*

Ethics education in data science, and data justice *27*

Approaches to moral education *40*

Conclusion *43*

Chapter 4: Research approach, methods and methodology **45**

Research paradigms *46*

From paradigm to methodology and method(s) *55*

Methodology, methods and the collection and analysis of data and writing *56*

Reflexive thematic analysis *58*

Chapter 5: Interviews **62**

5.1 Theme 1: The importance of ethics in data science *62*

5.2 Theme 2: Ethical challenges encountered in practice *66*

5.2.1 Feeling unprepared to grapple with ethical issues 66

5.2.2 The source and nature of the data, and dealing with sensitive data 70

5.2.3 Data labelling and classification 72
5.2.4 The building of the model 75
5.2.5 Outcomes, use and impact of the model 77
5.2.6 Unethical requests from the client 81
5.2.7 Lack of diversity and representation in data science teams 82

5.3 Theme 3: The custodian of ethics 83

5.3.1 The organisation as the custodian of ethics 84
5.3.2 The self as the custodian of ethics 87
5.3.3 Blended custody of ethics 88

5.4 Theme 4: The role of ethical education (including suggestions for ways of teaching data science ethics) 90

5.4.1 Criticism of the current dearth of ethical education for data scientists 90
5.4.2 Suggestions on how best to teach ethics to data scientists 94

5.5 Theme 5: Implications of gen AI and emerging technologies 101

Conclusion 108

Chapter 6: Analysis and discussion 111

6.1 The importance of ethics in data science 111

6.2 Ethical challenges encountered in practice 117

6.3 The custodian of ethics 124

6.4 The role of ethical education and teaching data science ethics 127

6.5: Implications of gen AI and emerging technologies 128

Conclusion 132

Chapter 7: Education in and for ethics and ethical conduct in data science 134

7.1 Recommendations for education in and for ethics and ethical conduct in data science 135

7.2 Transitioning from ethics to political consciousness in data science education 143

Summary 150

Chapter 8: Conclusions 152

8.1 Summary of key findings 152

8.2 Contributions to my professional practice 155

8.3 Broader implications for practice 156

8.4 Limitations and recommendations for future research *159*

8.5 Final reflections *160*

List of references **162**

List of accompanying material **174**

Appendix A: Participant Information Sheet *174*

Appendix B: Privacy Notice *176*

Appendix C: Consent Form *179*

List of tables

Table 1: Ethical frameworks and data science……35
Table 2: Critical theory and data science……45

Acknowledgements

I want to thank Professor Nicki Hedge, the programme convenor while I was doing my doctorate, Dr. Stephen Daniels, my supervisor and each of my tutors who created our study modules and encouraged and stimulated our thinking. Thank you to Denise Porada for all the support.

I am especially indebted to Stephen, who supervised me through my various drafts and revisions, for his regular advice, insightful questions, kindness and thoughtful feedback. I literally could not have done this without his guidance.

I owe special thanks to all the data scientists I interviewed. Their honesty, vulnerability and willingness to share their experiences openly made this research possible and meaningful.

I am deeply grateful to my doctoral cohort, whose camaraderie, sense of fun and steadfast encouragement have been a vital and vibrant support throughout this journey, both in Glasgow and on our ever-active WhatsApp group. Their wisdom, humour and quick responses to all kinds of questions are a continuous delight.

My thanks also go to Aaron and Rafi, the founders of the company where I first became interested in ethics in data science teaching and who supported my efforts to incorporate that interest into our course curricula. Thanks are due, too, to the academic faculty from the University of Virginia who opened my mind to liberal education and its implications for technical and practical fields, in particular Ian Baucom, Jeff Holt, Dudley Doane and Rachel Most.

To my friends who cheered me on and believed in me at every turn: I am grateful to have such a remarkable circle behind me.

Most of all, I thank my beloved family. To my parents Jane, Rob, John and Sue, my siblings Christopher, Peter and Laura, my partner Matthew and my son Max: your love and support have meant everything. I am grateful for the love of my faithful and brave dog Pearl, who started on this journey with me, and to Chester, the corgi, now asleep across my feet, who, of my entire family, would be the happiest to see me do this all over again.

Author's declaration

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Printed Name: Nicola Jane Weaver

Definitions/abbreviations

Algorithm: A set of rules or a step-by-step procedure for solving a problem or performing a task, often used in data science to process data and generate outcomes or predictions.

Artificial intelligence (AI): The capability of machines to imitate intelligent human behaviour, enabling computers or software systems to perform tasks that typically require human intelligence, such as learning, reasoning and decision-making.

Data: Facts or information, such as observations, measurements or records, that are collected and used to discover insights or make decisions and can be either structured (organised in rows and columns) or unstructured (such as text, images or videos).

Data science: The interdisciplinary field combining techniques from computer science and statistics to clean, structure, analyse and visualise data to generate insights or predictions.

Generative AI (gen AI): AI systems designed to create new content, such as text, images or other media, by learning patterns from existing data and producing outputs that are not simply retrieved or recombined, but newly generated.

Chapter 1: Introduction

> Ethics is not a vaccine that can be administered in one dose and have long-lasting effects no matter how often, or in what conditions, the subject is exposed to the disease agent (National Academy of Engineering, 2009:34).

Data science has emerged as a powerful tool for society, businesses and governments in recent years. But as data becomes more ubiquitous, so, too, do the accompanying ethical challenges. Algorithms govern many aspects of our daily lives, including the content that we see when we go online, the universities that accept us, the jobs that hire us, the profiles of potential matches sent to us on dating apps, and the home loans, bank loans, shop credit and various insurance that we get. These systems do not merely reflect existing social inequities, but also actively reshape power dynamics, privileging corporate and institutional interests over individual and collective agency. The metaphor above, which compares ethics to a vaccine that can be administered in one dose for long-lasting effect, is a powerful image. Yet even this metaphor assumes that ethical challenges are purely technical problems to be 'solved' rather than political struggles over power, justice and the distribution of societal harms.

As data science becomes more powerful and pervasive, the urgency for ethical and politically informed decision-making in the field intensifies. Although data science offers great potential to improve both private and public life (Floridi and Taddeo, 2016), such as driving more accurate medical diagnoses or enhancing the delivery of public services, these benefits are inseparable from significant ethical and structural challenges. A growing list of high-profile controversies in data science has raised concerns about the ethical implications of data collection, analysis and use. Without proper ethical and justice-oriented frameworks, the potential for harm is substantial, as recent scandals make clear.

Prominent examples include the Cambridge Analytica scandal, where the political consulting firm Cambridge Analytica obtained personal data of millions of Facebook users without their consent to influence the 2016 presidential election in the United States of America.[1] Or consider the racial bias

[1] For a detailed summary, readers can consult the BBC News article titled 'Meta settles Cambridge Analytica scandal case for $725m' which covers the legal and privacy implications of the scandal: https://www.bbc.com/news/technology-64075067

that is inherent in COMPAS[2], an algorithm used in America by courts and parole boards to forecast future criminal behaviour, which meant that black defendants were twice as likely to be incorrectly labelled as higher risk than white defendants (Angwin & Larson, 2016). Such failures represent more than technical oversights – they reveal systemic misrecognition and reinforce structural inequalities. In another example, thousands of Google employees protested the company's involvement in a Pentagon programme in 2018 and the ethical implications of using artificial intelligence (AI) to interpret video imagery, which could be used to improve the targeting of drone strikes (Shane & Wakabayashi, 2018).[3]

There is ongoing concern about TikTok, a platform that grants excessive data collection permissions by default, including the ability to collect user contact lists, access calendars, scan hard drives and geolocate devices on an hourly basis (Touma, 2022). In this issue, scale matters. As of 2025, TikTok has over 1.6 billion monthly active users, making it the fifth most widely used social platform globally. In the United States alone, TikTok has more than 136 million users, with over half of all users worldwide falling within the 16-34 age range. Concerns abound about the potential impact on privacy and autonomy, which impacts a significant portion of the global population.[4]

TikTok's practices exemplify how data extraction entrenches corporate power over individual autonomy, reframing privacy as a transactional commodity rather than a collective right. This is particularly concerning given the breadth and depth of data TikTok collects, including real-time location, device information, browsing habits and biometric identifiers such as faceprints and voiceprints, which are retained even after account deletion and often shared across a complex network of third parties and advertisers. Individual users have little meaningful control over their personal information in this environment, while corporations and governments gain unprecedented data and power.

---

[2] The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algorithm is a risk assessment tool used in the criminal justice system to evaluate the likelihood of a defendant reoffending. It evaluates individuals based on various factors, which can influence decisions related to bail, sentencing and parole.

[3] This article explains the employee protests, the ethical concerns about using AI for military purposes, and the broader implications for tech companies working with defence agencies: https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html

[4] For more information about TikTok's privacy issues, see here: https://apnews.com/article/tiktok-ireland-european-union-data-privacy-regulation-d386ec74becc716905d7f686d6a448e2 and here: https://pirg.org/articles/demystifying-tiktok-data/

These cases show that the rise of data science brings urgent ethical and political challenges, requiring justice-oriented rather than purely technical 'fixes' for issues like surveillance and systemic injustice. Understanding the ethical and political challenges in data science requires looking beyond isolated incidents to the broader process by which data-driven decisions are made. Ethical and political issues can arise at each stage of the data science process, including the design of the study, the process of data collection, the analysis of the data, and the use of the data and associated outcomes. These concerns are not limited to technical decisions but instead reflect deeper societal questions about how data is used and who benefits, and can be clustered under key themes, including privacy, cybersecurity, power asymmetries, opacity, and bias and discrimination.

As data science grows in scale and influence, it becomes imperative to embed strong professional and ethical standards alongside comprehensive ethics and political education so that data-driven technologies can be developed and used to promote justice, public trust and the collective good. Ethics education in other professional fields – in terms of both content and structure – has been a topic of discussion in many disciplines for years, particularly in law, medicine, business, and engineering (Gille & Nardo, 2020; Fiesler, Garrett & Beard, 2020). The relationship between ethical codes, professionalisation and moral responsibility in any profession is 'often fraught' (Stark & Hoffmann, 2019), but in data science seems more fraught – and to hold more potential for remedy – than others. This is because data science's technical veneer often obscures its role in reproducing structural inequities, requiring practitioners to interrogate individual ethical choices as well as the political status quo that their work sustains.

In addition, the pervasiveness of data science in our daily lives and its significant impact on humanity demand that we address this matter. The need for ethics in data science has been frequently noted within the literature (Baumer et al., 2022; Poirier, 2021; Greene, 2020; Salz et al., 2018; Appel, 2005), and this is not just a theoretical concern but one with real-world consequences. As data-driven systems increasingly determine decisions in areas like finance, healthcare, education and criminal justice, the absence of robust ethical frameworks can reinforce existing power imbalances and marginalise vulnerable communities. However, what is frequently ignored in these discussions is the acknowledgement that ethical frameworks that are grounded in liberal individualism – focused on consent, fairness and transparency – do not address how data science entrenches power imbalances between companies, governments, communities and individuals. For example, when algorithms are designed without considering broader social context, they may inadvertently encode, embed and amplify historical biases, resulting in discriminatory outcomes that disproportionately impact minority groups. In addition, the opacity of many data science

systems makes it difficult for individuals and communities to understand or challenge decisions that impact their lives, which further embeds inequalities and reduces accountability for those in power. There is an urgent need for ethical frameworks that extend beyond individual fairness and transparency to address data science's structural and political dimensions.

Why have there been such apparent gaps in training data scientists in and for ethics, ethical conduct, and political consciousness? O'Neil asserts that the blame lands squarely on academics, who she claims have 'been asleep at the wheel' (2017). Martens is more measured and states that data scientists are 'not inherently unethical, but at the same time not trained to think this through either' (2022, p. 3). This training gap is compounded by a reluctance to confront data science's political nature: the field often frames itself as a neutral technical discipline, which serves to insulate practitioners from accountability for how their tools shape social outcomes. Without a strong foundation in ethical reasoning and political awareness, data scientists may unknowingly contribute to systems that perpetuate injustice or harm.

This gap is exacerbated by the field's tendency to present itself as neutral and technical, which can discourage communities – and even practitioners – from questioning the broader societal impacts of data science. Lundie (2024) contends that it is important to recognise that technology reflects underlying values, and that the ways data are gathered, combined and applied in decision-making can reinforce or worsen social inequalities and risks. Addressing this issue requires integrating ethics more deeply into data science curricula and creating a culture of critical reflection and accountability within the profession. By encouraging data scientists to engage with the political and social implications of their work, data science can move towards more responsible, reflective and equitable practices that truly serve the public good.

Several challenges are associated with teaching ethics for data science. Firstly, data science is a relatively new, rapidly evolving field, with the latest technologies and techniques constantly being developed. This means that ethical issues and training to prepare for these have not yet been embedded in data science teaching:

> The ethical issues a data scientist may encounter have received little attention to date, and ethics training within a data science curriculum has received even less attention (Salz et al., 2018, p. 952).

Secondly, the multidisciplinary nature of data science has made it difficult to codify the teaching of data scientists in and for ethics, ethical conduct and political consciousness. Data Science is a field

that draws on expertise from various disciplines, and people come to be 'data scientists' from a wide variety of undergraduate degrees, including economics, science, mathematics, statistics, computer science and engineering. This means that their ethical teaching, where it existed, has been as varied as their degree disciplines. With little coordination across the diverse backgrounds of data scientists and limited formal ethics training within the field, the implication is that most data scientists are unlikely to have received formal ethics training. As O'Neil (2017) puts it:

> There is essentially no distinct field of academic study that takes seriously the responsibility of understanding and critiquing the role of technology – specifically, the algorithms that are responsible for so many decisions – in our lives (p. 23).

Thirdly, there is an absence of global frameworks and standardisation. There are currently no global ethical guidelines or best practices for data science. There are increasingly exemplars of ethical training (which I will examine later in this paper) in data science. However, nowhere is there an overarching and coherent ethical framework for data science that combines both deep knowledge of data science practices with rigorous ethical argumentation. There is also no framework situating data science within broader struggles for economic redistribution, cultural recognition and political representation. In this dissertation, I will argue that such a framework is urgently needed to connect data science to these broader social and political struggles.

Fourth, the technical complexity of data science means that understanding the field requires advanced quantitative understanding, including of complex algorithms, statistical methods and data structures. The result is that regulation is challenging – both to formulate and enforce – for regulators who lack these skills. Yet technical complexity should not excuse practitioners from grappling with political consequences. The rapid pace of technological change in data science makes it even harder for regulators and educators to keep up, increasing the risk that ethical considerations are sidelined in favour of technical innovation. As algorithms increasingly mediate access to resources and opportunities, data scientists must recognise themselves as political actors shaping normative visions of justice, and whose daily decisions result in outcomes for real humans living in the real and often unequal world. Many data scientists do not receive formal training in ethics, making it easy for them to focus on technical solutions while missing the broader social impacts of their work. This gap can lead to unintentional harm, as technical fixes alone cannot address the complex and value-laden nature of data science.

Fifth, value conflicts make it difficult to achieve an aligned framework. Even within a country, there can be differences of opinion between businesses, government and society on priorities and

values, making decisions on ethical trade-offs difficult. Countries themselves differ in values. For example, some countries prioritise individual freedoms and liberties over community cohesion, while other countries place a higher value on community cohesion. The challenge this poses to the development of a global framework for ethics in data science is clear. For example, one country may reject facial recognition software for policing as a violation of privacy, while another country might reject individual freedom and privacy in favour of effective community policing through the use of facial recognition software. These differences reveal that data science is not a neutral arbiter of 'objective' solutions, but a terrain for negotiating competing political values about security, liberty and equity. The global nature of data science, which allows data and algorithms to cross borders and impact people in vastly different legal and cultural contexts, is an additional challenge. Efforts to regulate or guide data science ethically must grapple with the reality that there are no universally accepted standards, and that any framework must be adaptable to the diverse and evolving values of different communities. This ongoing tension highlights the importance of inclusive, participatory approaches to ethical decision-making in data science, ensuring that diverse perspectives are considered and respected.

A sixth reason is that university data science faculty are unfamiliar with the content of ethics teaching. Many faculty recognise their responsibility to teach students about ethics and responsibilities. However, as Burton, Goldsmith & Mattei (2018, p. 54) acknowledge, this is 'a kind of teaching for which most of us have not been trained, and that faculty and students approach with some trepidation'. This lack of training can lead to uncertainty and inconsistency in how ethics is taught, with some educators feeling ill-equipped to address complex ethical dilemmas that arise in data science. As a result, students may receive only a superficial understanding of ethical issues, missing out on the deeper critical reflection needed to navigate real-world challenges. To address this gap, faculty need dedicated resources and training to help them to integrate ethics into their courses with confidence and hold meaningful discussions about responsibility and justice in data science.

A final – and most dangerous – challenge is that many people believe that because statistics and mathematics underpin data science, it is objective, neutral and unbiased, and as such, there is no need for education in and for ethics and ethical conduct in data science. But this is far from true. The myth of neutrality persists because it legitimises data science's authority while disguising its role in consolidating power for companies and governments. This misconception can dissuade both students and practitioners from questioning the broader societal impacts of their work, allowing harmful biases and inequities to go unchallenged. By treating data science as a purely technical

discipline, we risk ignoring how algorithms and data science can reinforce or even amplify social injustices.

Dismantling this myth requires a fundamental shift in how data science is taught and practised, reconceptualising data science not as a technical discipline but as a site of political struggle and encouraging critical engagement by data scientists with the power structures it shapes. The challenges facing data scientists cannot be divorced from broader questions of political economy. Data infrastructures mediate power, redistribute risk and frequently concentrate decision-making in the hands of a few actors, often without meaningful democratic oversight or community participation.

The aims of my study

As questions about the ethical and political implications of data science intensify, spurred by recent controversies and calls for greater scrutiny (Fiesler et al., 2020), I set out to critically examine how education in and for ethics and ethical conduct should be reimagined for those working in this field. This research is significant to the broader community because increasingly, the decisions that affect our lives are being made not by humans but by data science algorithms. Automated systems systematically produce unfair outcomes that can render even well-intentioned data science products profoundly destructive (Baumer et al., 2022). These harms are not simply technical glitches, but manifestations of structural power imbalances embedded in data infrastructure. Platforms (such as Amazon, Facebook or Uber) invert the relationship between individuals and markets, privatising decision-making power while rendering personal domains public, a dynamic that entrenches corporate authority over democratic accountability (Benthall & Goldenfein, 2020). For example, Amazon decides which products are pushed to which customers, Facebook determines which content is sent to which user and Uber uses algorithms that shape pricing and conditions under which drivers interact with riders. In this way, digital platforms act as intermediaries that connect individuals, like buyers and sellers, privatising decisions that would otherwise be made collectively.

Furthermore, data scientists often frame their work as apolitical basic research. Yet, their technical choices, which range from dataset selection to error metrics, reflect normative judgements about whose safety, dignity or autonomy matters. Ethics-focused critiques alone cannot address these systemic issues, which require interrogating how algorithms encode historical inequities (for example, over-policing in predictive policing tools) and legitimise the status quo. A core aim of this dissertation is to interrogate how data scientists can be equipped to recognise, analyse and contest

the political conditions and structures, both local and global, that both enable and constrain ethical practice in their field.

In this context, my study sought to investigate the overarching question: How do data scientists experience and respond to ethical challenges in their work, and how can these experiences inform the development of ethics education that addresses structural and political dimensions of data science? In my research, I aimed to answer the following questions:

1. How do in-service data scientists experience and respond to ethical challenges in their work?
    a. What role, if any, has formal ethics education played in shaping their responses?
    b. What gaps do data scientists identify in their existing training regarding ethics, and how do they articulate the need for structured ethical and political education within the field?
2. How might concepts from liberal theory – such as critical thinking, civic responsibility and ethical reasoning – inform the design and delivery of ethics education in data science?
3. What are the perceived needs and potential benefits of incorporating ethics and ethical conduct into in-service training for data scientists?
4. What are the perceived challenges and possibilities for incorporating political and social justice considerations into in-service ethics training for data scientists?

Drawing on critical and liberal theories

In addressing data science's ethical and political challenges, I have drawn on both critical and liberal theories. Each offers distinct but complementary perspectives that, when integrated, provide a comprehensive framework for understanding and responding to the complexities of data science practice and education.

Liberal theory centres on the values of individual autonomy, freedom, equality and justice. It provides a foundation for ethical reflection and the development of professional responsibility among data scientists. Liberal theory emphasises the importance of cultivating ethical judgement, respect for rights and the capacity for critical self-reflection, focusing on human flourishing. In the context of data science education, concepts from liberal theory support the development of practitioners who are both technically competent as well as attuned to the ethical implications of their work. Consider, for example, Nussbaum's liberal theory (1997), which centres on human flourishing, or the development of each person's capabilities to live a life of dignity and purpose.

Her ‘capabilities approach’ identifies a set of fundamental opportunities and freedoms that every individual should be able to access, regardless of nationality or background. These include the ability to reason, to affiliate with others, to express emotions, and to participate in society. In the context of data science education, concepts from this framework might encourage educators to look beyond technical proficiency and economic productivity and instead empower students to use data science in ways that promote well-being, justice and the full development of human potential, both for themselves and others, including those in different countries and cultures.

Nussbaum (1997) contends that a central goal of liberal education is to nurture empathy and narrative imagination, which is the capacity to envision what it is like to be in someone else’s situation, particularly those who are different or distant from us. She argues that this educational aim is best achieved by developing three essential capacities in students. The first capacity is critical self-examination, which involves questioning one’s own assumptions to develop deeper self-awareness and intellectual humility. The second capacity is to see oneself as a world citizen, which means recognising moral obligations that extend beyond one’s own nation and appreciating the interconnectedness of global communities. The third capacity is narrative imagination, or the ability to empathise with the experiences and perspectives of others, especially those in other countries or marginalised groups. These three capacities prepare students to engage thoughtfully and ethically with the broader world. In data science education, these capacities are crucial. Data scientists increasingly work with global datasets and develop technologies that impact people worldwide. By developing empathy and a sense of global citizenship, data scientists can begin to appreciate their work's ethical and social implications, including how data-driven decisions may affect vulnerable populations in other countries.

On the other hand, the tenets of critical theory help interrogate the underlying power structures, systemic inequalities and institutional dynamics that shape data science. This helps researchers and practitioners to look beyond surface-level ethical considerations and examine how data science practices can reinforce existing social dynamics and inequities. By focusing on issues such as bias, discrimination and the social construction of knowledge, critical theory highlights the ways in which data science is embedded within broader political and historical contexts. This deeper perspective can help to identify and address injustices that more individualistic approaches may overlook. Without understanding the historical, economic and legislative forces and power dynamics that shape data science, practitioners are at risk of unintentionally sustaining the very inequalities and structures of domination that ethics education seeks to address.

Drawing on these approaches for contributions to data science education and practice can address the systemic and personal dimensions of ethical decision-making. Neither perspective alone can fully address data science's complex ethical and political challenges. Liberal theory provides a necessary foundation for ethical practice by emphasising the development of individual agency, moral reasoning and the pursuit of fairness and respect for persons. At the same time, concepts from critical theory play a vital role by exposing the structural and institutional forces that shape ethical dilemmas, ensuring that questions of justice and power are brought to the forefront of analysis. By combining these approaches, data science education and practice are better equipped to navigate the personal and the systemic dimensions of ethical decision-making.

Chapter 2: Outline and structure

This chapter outlines the dissertation's structure, summarising each chapter's rationale and content. The structure is designed to develop a coherent argument from foundational concepts through empirical findings to practical recommendations and conclusions. Chapter 1 introduces the context for the study, articulates the research aims and positions the investigation within the broader debates about ethics and political issues in data science. This chapter also outlines the motivation for the research and the significance of examining ethics education for data scientists, including the need for justice-oriented frameworks in data science ethics. It also outlines the overarching research questions and sub-questions guiding the dissertation.

Chapter 2 describes the organisation of the dissertation, providing the reader with a roadmap of the chapters and clarifying how each contributes to the overall research objectives. In addition, Chapter 2 summarises the rationale and content of each chapter, ensuring the reader understands the logical progression of the argument.

In Chapter 3, I investigate data science and current ethical issues in the field, including definitions of what data is, what big data is, what data science is, how data science is a subset of machine learning and AI, and how algorithms work. This is necessary to understand the ethical issues in data science, why they exist and how they are currently being addressed. This chapter also explores the ethical issues that pervade the field, including privacy, cybersecurity, bias and discrimination, data quality and opacity, and it lays the groundwork for understanding why these ethical challenges matter in both technical and societal terms. In this chapter, I also introduce different theories in ethics, focusing particularly on utilitarianism, deontology and virtue ethics, which are among the most widely used ethical theories in technology education. At the most basic level, ethics refers to the perception of something being good or right. However, within the context of this dissertation, ethics is best understood as a multifaceted and dynamic field concerned with the principles, values and frameworks that guide human conduct, both individually and collectively, toward what is considered good, right or just. Rather than being a static set of rules, ethics is more accurately thought of as a process of critical reflection and deliberation about how we ought to act in situations that involve competing interests, values and potential harms. We can speak of an ‘ethical use of data science’ and mean that it is performed in a way that is ‘right, proper, acceptable, or socially appropriate’ (Salz and Dewar, 2019, p. 197).

Beyond ‘ethics’, there are several related concepts, such as values, justice, responsibility, and ongoing conversations situated in diverse philosophical traditions and literature (Fiesler et al., 2020). Ethics in data science is an endeavour that requires ongoing engagement with philosophical principles, contextual realities and the pursuit of justice. This characterisation moves beyond a checklist of rules to embrace ethics as an evolving and adaptive practice essential for responsible professional and societal life. Additionally, this chapter discusses how these theories apply to data science, reviews the limitations of prevailing approaches and explores the intersection of ethics with broader political and social justice concerns.

I also investigate liberal education, and examine whether the concepts in liberal education – freedom, equality, justice and autonomy (Nussbaum, 2010; Zinser, 2004; Bridges, 1997) – could be brought to bear on ethical issues in data science, as according to Zinser, the aims of a strong liberal education include ‘shaping ethical judgement and the capacity for insight and concern for others, our habitats, and the future’ (2004, p. 40). It is the shaping of ethical judgements with which I am most concerned, specifically regarding the ethical judgements that data scientists make daily. Principles of liberal education are already – tentatively – being used in ethics in data science. For example, Ratti and Graves (2021) make use of Nussbaum’s capabilities approach to define a technical act as ethically relevant when:

> It impacts one or more of the basic human capabilities of data subjects. Therefore, rather than ‘applying ethics’ (which can be mindless), data scientists should cultivate ethics as a form of reflection on how technical choices and ethical impacts shape one another (p. 1819).

Further, this chapter examines moral education more broadly. And finally, this chapter also explores the intersection of ethics with political and social justice..

In Chapter 4, I describe and motivate my research approach, including my methods and methodology, and I outline the research design and methodological framework underpinning this dissertation. I describe who I interviewed and how, and how the study was structured to explore the ethical challenges faced by data scientists and the potential for ethics education in the field. I describe how my interviewees’ diverse backgrounds and institutional decision-making logics may influence their ethical views and practices, and how my own positionality and power relations were mitigated. I outline how my research draws on critical theory for its analysis of power and structure, while engaging with liberal theory for its educational and ethical dimensions.

Chapter 4 also explains how the research was designed to explore the lived experiences of data scientists regarding ethical challenges and the adequacy of ethics education in their field. Semi-structured interviews were conducted and analysed using Braun and Clarke's reflexive thematic analysis, allowing for the identification of nuanced themes and patterns in participants' accounts while maintaining a focus on reflexivity and the researcher's positionality throughout the process. The chapter also discusses the rationale for drawing on critical theory for power and structural analysis, highlighting its relevance for interrogating structural inequalities and advocating for emancipatory change in data science education. This chapter also details the process of data collection, analysis and writing.

In Chapter 5, I use Reflexive Thematic Analysis to garner insights from my research interviews and discuss the implications for education in and for ethics and ethical conduct in in-service training for data scientists. In this chapter, I present an analysis of empirical findings from interviews with practising data scientists alongside the existing literature. The chapter is structured around five themes generated through thematic analysis:

- Theme 1: The importance of ethics in data science
- Theme 2: Ethical challenges encountered in practice
- Theme 3: The custodian of ethics
- Theme 4: The role of ethical education
- Theme 5: The implications of generative AI

Chapter 6 bridges the findings and insights from my interviews with data scientists and the literature. In this section, I compare my findings with existing research to contextualise my results within the broader academic discourse and discuss how my findings from the interviews with practising data scientists align with existing research. This chapter offers a deeper understanding of education in and for ethics and ethical conduct in data science through a comparative analysis with the existing body of literature. This analysis bridges practitioner experiences with theoretical frameworks, identifies critical gaps in current education and highlights areas for positive change in data science ethics practices.

Chapter 7 offers practical recommendations for integrating ethics and ethical conduct into training for data scientists. Drawing from interview insights and literature, this chapter advocates for embedding ethics throughout technical training, promoting interdisciplinary and liberal arts approaches, and cultivating political and social consciousness among data scientists. It also

discusses pedagogical strategies such as case-based learning, experiential education and collaboration with humanities disciplines. This chapter advocates for a focus on a politics of justice in data science education.

In Chapter 8, I provide a summary of the main findings, reflecting on the implications for professional learning and practice, and I outline the limitations of the study, referring the reader to the relevant chapters for detailed discussion. This chapter also reflects on the contributions to professional practice and the broader implications for data science education and the profession. It provides recommendations for future research, including the need for interdisciplinary and justice-oriented approaches. The chapter concludes with a strong argument for the urgent transformation of ethics education in data science, highlighting the stakes for society and the field.

The structure of my dissertation moves from foundational definitions and theoretical frameworks to empirical analysis, practical recommendations and conclusions. This progression ensures that each chapter builds upon the chapter before, giving the reader an understanding of the ethical and political dimensions of data science and the urgent need for reform in ethics education:

- **Literature review (Chapters 3–4):** These chapters collectively explore the key concepts, ethical theories and current practices that inform the research questions.
- **Empirical analysis (Chapter 5):** The analysis chapter grounds the theoretical discussion in real-world experiences, offering nuanced insights from practising data scientists.
- **Recommendations and conclusions (Chapters 7–8):** The final chapters translate analytical insights into actionable recommendations and reflect on the broader implications for the field.

Research approach

My study focused on a group of data science teachers and their experiences with ethical issues in the practice of data science. An organisation I work with teaches technology skills, including data science. Our data science teachers are also data science professionals and are, therefore, accustomed to dealing with ethical issues in their daily work. However, in our courses, instruction remains focused almost exclusively on the technical aspects of data science. This overwhelmingly technical emphasis reflects the approach of most data science degrees, diplomas and training programmes worldwide. My research examined the experiences of our data science teachers regarding ethical issues in their field, how prepared they feel to address these issues, and – if they feel unprepared – how this might be addressed in future data science education.

My research is an empirical study. For my research paradigm, I drew on critical theory for power and structural analysis while engaging liberal theory for educational and ethical dimensions. I used semi-structured interviews, which I recorded and transcribed, and then used Braun and Clarke's six-phase process for reflexive thematic analysis to analyse the data from the interviews (Braun & Clarke, 2019). Reflexive thematic analysis is particularly well-suited to this research because it allows for a systematic yet flexible approach to identifying patterns and themes within qualitative data, making it ideal for exploring complex, nuanced experiences and meanings.

Because reflexive thematic analysis helps address research questions related to people's experiences, understanding and representation, social processes, rules and norms, people's practices and behaviours and the construction of meaning (Braun and Clarke, n.d.), it aligns with my research intention of finding out about the experiences that data scientists have of ethical issues in their field, and how they think this might be addressed in ethical education. This approach allowed me to go beyond superficial answers to reveal insights into how participants experience and navigate ethical challenges in their professional practice, and how they make sense of these experiences within broader social and institutional contexts. The flexibility of reflexive thematic analysis made it possible to explore both explicit statements and underlying assumptions, providing a rich, nuanced understanding of the field. This depth of analysis was essential for developing practical, context-sensitive recommendations for improving ethics education in data science, discussed in Chapters 8 and 9.

Chapter 3: Literature review

This chapter brings together the conceptual, ethical and educational foundations that inform my research. It first clarifies key terms and processes in data science and then examines the ethical and justice-related concerns that arise from data science processes. It goes on to review current approaches to ethics education in data science alongside broader work on moral and professional education. The chapter closes by sketching the theoretical frameworks that inform my analysis, indicating why I draw on liberal and critical traditions, and how these perspectives relate to the empirical chapters.

To understand why ethical questions arise in data science, it is helpful to understand what we are talking about when we are talking about data science. Data has been hailed as 'the world's most valuable resource' (Economist, 2017) and 'the new oil' (Stark & Hoffman, 2019, p.6), and data science has been labelled 'the sexiest job of the twenty-first century' (Davenport & Patil, 2012, p. 70). In part because of the hype surrounding data science, some key concepts – and how these key concepts fit together – are not clearly understood. These key concepts include 'data', 'big data', 'data science', 'algorithm' and 'artificial intelligence (AI)'.

Key concepts and issues in data science

The key building block of all data science is data. Data includes 'facts or information, especially when examined and used to find out things or to make decisions' (Martens, 2022, p. 7). In other words, data is any information – facts, observations, measurements or records – collected or stored for further analysis, processing or use. Data generally falls into two types (Hale, 2017): structured and unstructured. Structured data is data that we can imagine having rows and columns, that fits into spreadsheets or relational databases. Unstructured data, on the other hand, includes free text, images and videos (Hale, 2017). Often, data science is used to extract some structure from, or impose structure on, unstructured data to answer questions of social or other importance.

Increasingly, our online and offline interactions generate large amounts of data. Credit card transactions, loyalty programmes, internet searches, social media platforms and user-generated content platforms (where individuals upload posts, share opinions and post photographs) are all examples of our online and offline interactions and behaviour that leave a data trace behind. 'Big data' is a term used to refer to 'large sets of data compiled from various sources (e.g., existing administrative data, online interactions, data collected by devices) and stored in a digital form to be analysed with computers' (Hosseini et al, 2022, p. 2). The 'three v's' are often used to characterise

big data: volume, velocity and variety. Volume refers to the large size of the datasets, ranging from terabytes to petabytes or even exabytes.[5] Velocity refers to the speed at which the data is generated, processed and analysed, which can be in real-time or near real-time. Variety refers to the diversity of data types, formats and sources, including structured, semi-structured and unstructured data from different sources, such as social media, sensors and devices.

To understand the ethical and educational challenges discussed in this chapter, it is necessary to clarify what is meant by 'data science', as this term underpins much of the debate about technology's role in society and the professional practices I examine. 'Data science' includes cleaning the data and extracting structure from the data if the data is unstructured. It also involves the analysis of the data, and often the visualisation of the data, too. Data science is, therefore, a combination of techniques from computer science and statistics.

Data science relies on computational processes called algorithms. Algorithms are the building blocks of data analysis and play a central role in shaping the outcomes and ethical considerations of data-driven decision-making. An 'algorithm' is 'a procedure for solving a mathematical problem in a finite number of steps that frequently involves repeating an operation' (Merriam-Webster, n.d.). In data science, algorithms follow rules or procedures to perform a specific task on a dataset. The particular steps involved in an algorithm depend on the problem it is designed to solve and the type of data it is analysing. For example, a predictive algorithm takes historical data and outcomes and predicts future outcomes.

Because many of the ethical, legal and societal issues addressed in this work arise from the increasing use of artificial intelligence (AI) in data science and related fields, it is crucial to define what is meant by 'artificial intelligence' and to distinguish it from, yet relate it to, data science. 'Artificial intelligence' is, as the term implies, intelligence that is artificial. In other words, it is the 'capability of a machine to imitate intelligent human behaviour' (Merriam-Webster, n.d.). It is a subfield of computer science that deals with creating intelligent machines that can perform tasks that typically require human intelligence, such as perception, reasoning, learning and decision-making. AI encompasses various approaches, including rule-based systems, expert systems,

[5] A terabyte is 1,000 gigabytes; a petabyte is 1,000 terabytes and an exabyte is 1,000 petabytes. Each unit is a thousand times larger than the previous. The world produces enough data daily to fill billions of smartphones, with individuals generating about 147GB per day, which is more than a typical smartphone's 128GB capacity. See, for example, https://edgedelta.com/company/blog/how-much-data-is-created-per-day.

machine learning, deep learning and neural networks. Although the terms ‘data science’ and ‘AI’ are sometimes used interchangeably by the layperson, they are distinct fields. Data science is often used as a tool in developing AI models, but it is possible to train a machine in AI without using data science. The field of AI has a rich history, dating back to the mid-20th century. It has evolved through cycles of optimism and scepticism, with recent advances in machine learning and neural networks driving a new wave of applications across society. AI has sparked ethical, legal and philosophical debates about responsible innovation, fairness and the societal impact of technology.

Ethical issues can arise at each step of the practice of data science. For example, at the outset of all data science processes is the collection of data. Ethical issues can arise regarding how the data is collected and whether informed consent is given by the person whose data is being collected. Both ‘informed’ and ‘consent’ can be problematic, ethically. For example, if someone is told how their data will be used, but they do not fully understand, is this truly ‘informed’ consent? What if they give consent and fully understand, but the data is then used for a different purpose years later? Is that informed ‘consent’? Further issues arise regarding how securely the data is stored and whether it is anonymised. There have been several cases where researchers or companies collected data for a purpose different from that for which it was used, or where data was used without informed consent (O’Neil, 2016; Stark & Hoffman, 2019; Van Noorden, 2020; Martens, 2022).

There are five main clusters of ethical issues in data science. The first area is privacy, since the collection and analysis of data for large data sets can raise serious concerns around privacy. The second area is cybersecurity, since data science requires the collection, storage and transmission of large amounts of sensitive data. The protection of this data from hacking or theft is a key ethical issue in data science. The third area is bias and discrimination, as data science can reproduce – or even amplify – structural inequalities, biases and discrimination, including on race, gender, age and socioeconomic status (O’Neil, 2016; Angwin & Larson, 2016; Dastin, 2018; Martens, 2022). The fourth area is data quality, since data science is highly dependent on the quality and accuracy of the data being used, which is provided and collected by humans and therefore prone to human error.

The fifth area is opacity. Data science is a complex field, and data science algorithms can be complex and difficult to understand by the layperson, and importantly, by regulators. Algorithm-driven decisions can lead to better health outcomes for people living in richer suburbs, lower university acceptance rates for lower-income applicants or perpetuating discrimination in the criminal justice system. It is nonsensical to talk of an unethical algorithm, as algorithms are simply

mathematical rules. However, the choice of algorithm – always made by humans – can be laden with ethical issues. Methodological biases and personal prejudices (Hosseini et al, 2022) can creep into algorithm selection and application. Inappropriately chosen or unsupervised algorithms can have compounding and negative impacts, creating a reinforcing nature, which is why O'Neil (2016) talks of 'weapons of math destruction'.

In summary, the use to which data science is put, either within an artificial intelligence application or simply in analysis and insights, is not just a matter of academic debate or theoretical concern. Data science is already causing serious harm through biased algorithms in hiring, policing, lending and misinformation amplification. These incidents have deepened inequalities, eroded privacy and undermined trust, proving that the risks are urgent and widespread. Ethics education and training for data scientists must address these issues directly. Without decisive reform, the social costs will grow, making it imperative that we investigate how to embed robust ethical standards and critical consciousness at every level of the data science profession.

Ethics education in data science, and data justice

Three key approaches central to technology education are utilitarianism, deontological ethics and virtue ethics (Boddington, 2023; Fiesler et al., 2020; Shapiro et al., 2020; Tractenberg, 2020; Dignum, 2019; Burton et al., 2018). I discuss how each approach's core ethical commitments and concepts could be applied to data science teaching. At the most basic level, ethics refers to the perception of something being good or right. Ethics are 'a set of moral principles: a theory or system of moral values' and 'the principles of conduct governing an individual or a group', 'a consciousness of moral importance', 'a guiding philosophy', and 'a set of moral issues or aspects (such as rightness)' (Merriam-Webster, n.d.). This chapter also explores moral education, as well as concepts in liberal education, such as critical thinking, civic responsibility and ethical reasoning. Finally, I look at political and social justice. I critique the false separation of ethics from politics in data science and emphasise structural critiques of power, inequality and justice.

Ethics is typically understood to be normative, in that it is aimed at establishing norms of thought, values or conduct. According to Burton et al (2018), this assumption is particularly present in many professional ethics courses that are 'typically used as a means to steer students' future behaviour toward a set of professionally agreed-upon values, such as professionalism and honesty'. But it is important also, argue Burton et al, to acknowledge that ethics can also serve as a tool for description, giving decision-makers a helpful framework to understand 'what is happening in a

given situation and what is at stake in any action they might take' (p. 57). Ethics includes thought, a structured and intentional reflection on morality, and practice, the effort to make decisions and actions in good, just and right ways.

We can speak of an 'ethical use of data science' and mean that it is performed in a way that is 'right, proper, acceptable, or socially appropriate' (Salz and Dewar, 2019, p. 197). Wylie (2020) makes a compelling argument for all people to be involved in the defining of ethical data science:

> Who decides what good data science looks like? And who gets to decide what 'data ethics' means? The answer is all of us. Good data science should incorporate the perspectives of people who create and work with data, people who study the interactions between science and society, and people whose lives are affected by data science (p. 1).

This call for democratic participation reflects a broader recognition that ethics in data science cannot be developed in isolation from those whose lives are most affected by these technologies.

Utilitarianism is an ethical theory that focuses on maximising overall happiness or well-being to achieve the greatest good for the greatest number of people. In utilitarianism, therefore, the right action is the action that is expected to produce the greatest good. In this way, moral decisions are made simple by supplying a single measure of rightness: maximisation of utility. Utilitarian thinking can be traced back to Mozi, a Chinese philosopher who lived 490–403 BC, according to Lazari-Radek and Singer (2017). However, Jeremy Bentham (1748–1832), an English moral philosopher and legal reformer, founded the doctrine of utilitarianism. The central tenet is simply stated, easily understood, and as a result, appealing to many. The main idea is that the highest principle of morality is 'to maximise happiness, the overall balance of pleasure over pain' (Sandel, 2010, p. 20). In other words, according to Bentham, the right thing to do is whatever will maximise 'utility' – by which he means whatever produces pleasure or happiness and prevents pain or suffering.

The implication is that, as far as it is within our power, we should bring about a world where every individual has the highest possible level of well-being (Lazari-Radek & Singer, 2017, p. 2). According to Sandel (2010), 'citizens and legislators should ask themselves this question: If we add up all of the benefits of this policy, and subtract all the costs, will it produce more happiness than the alternative?' (p. 20). John Stuart Mill, a follower of Bentham and an advocate of utilitarianism, outlines:

> The creed which accepts as the foundation of morals, Utility, or the Greatest Happiness Principle, holds that actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness (1879, p. 14).

While Mill was a follower of Bentham, their versions of utilitarianism differ considerably. Bentham's utilitarianism is often described as 'act utilitarianism' and argues that each action should be evaluated solely by its consequences for overall pleasure or pain, with all pleasures and pains considered equal in value. In contrast, Mill's utilitarianism distinguishes between higher and lower pleasures, arguing that intellectual and moral pleasures are superior to mere physical pleasure. Mill's version, sometimes called 'rule utilitarianism', also allows for considering general rules that promote the greatest happiness, rather than evaluating each act in isolation. The utilitarian philosophy recognises that we all like pleasure and dislike pain and makes it the basis of moral and political life. It is an appealing philosophy, but there are several objections.

The first objection is that utilitarianism fails to respect individual rights. By caring only about the overall level of happiness of the overall population ('the greatest happiness of the greatest number'), the state of each person is ignored, which can violate fundamental norms. Sandel (2010) uses the example of throwing Christians to the lions to illustrate this point. Even if the violent spectacle amuses thousands of Roman spectators, there is a violation of the rights of individual Christians that cannot be ignored. This objection is highly relevant in data science: algorithms designed to maximise efficiency or accuracy for the majority can ignore the rights of minorities or vulnerable groups. For example, predictive policing tools may reduce crime rates overall but disproportionately target marginalised communities, thereby violating their rights in pursuit of aggregate benefit. A linked criticism is that utilitarianism overlooks whether things are distributed fairly or justly. The outcome that generates the greatest good overall may differ from the outcome whose distribution of goodness comes closest to being just or fair (Savulescu et al., 2020). In data science, this manifests in algorithmic resource allocation, such as distributing healthcare, public services or educational opportunities, where optimising for overall benefit may systematically disadvantage already marginalised groups. The result is often the reinforcement of structural inequalities, justified as acceptable trade-offs under utilitarian logic.

A second objection arises over the 'currency' of value, since, to aggregate preferences, it is necessary to measure them on a single scale (Sandel, 2010, p. 22). If the best outcome is the most happiness over misery (or pleasure over pain), then we need to be able to measure these states. We can't even do this individually, let alone for populations. To illustrate this, say that going for a run

and eating out in a restaurant are both things that provide me with pleasure, but I can't say which provides more. Social policy decisions raise similar problems (Lazari-Radek & Singer, 2017). How can we begin to weigh the pleasure of one half of the population against the displeasure or pain of the other half? This is a major issue in data science, where algorithms often rely on quantifying complex human experiences, like well-being or privacy, into simple metrics. This reduction can lead to decisions that overlook or undervalue certain individuals or groups' unique harms or needs, as seen in credit scoring or healthcare algorithms that may not account for social context or lived experience.

A third objection arises when one considers the principle of autonomy – an individual's freedom to choose and determine how to live their own life. Individual freedoms could conflict with the overall good if, for example, a person decides they want to do something different from what is best for the overall good. In data science, this is seen in the widespread use of automated decision-making systems that make choices on behalf of individuals, such as targeted advertising, personalised news feeds or algorithmic hiring, often without meaningful consent or transparency. This can erode personal autonomy and reduce individuals to mere data points in a larger system, stripping them of individual agency and voice.

Some of the central objections in utilitarianism speak directly to the major ethical questions raised by data science. Utilitarianism is a form of consequentialism: the right action is understood entirely in terms of the consequences produced. So, a utilitarian approach to data science might argue that the best thing to do, morally, is whatever will bring about the best overall outcomes, or as much happiness and as little unhappiness as possible, overall. This goal-oriented focus suits the problems of data science well, as data science is always conducted with an end objective in mind. However, problems could include precisely specifying the ends and objectives, the measurement of the objectives, the value of individuals, and the challenges of tackling ethical questions involving agency. While utilitarianism's focus on outcomes aligns with data science's problem-solving orientation, its application risks reinforcing technical solutionism.

Deontological ethics focuses on rules and duties guiding what we ought to do, in contrast to those theories that guide and assess what kind of person we are and should be (Alexander & Moore, 2021), such as virtue ethics. In other words, deontology centres ethical evaluation on actions themselves and the principles that govern them. Deontology derives from the Greek 'deon', which essentially means duty or obligation – so it is sometimes called duty ethics (Bartneck et al., 2021, p.

19). Unlike utilitarian or consequentialist theories, deontological theories judge the morality of choices by moral norms, not outcomes. Thus, an action is right or wrong based on the motives of the person who carries it out, not its consequences (Dignum, 2019).

Immanuel Kant (1724-1804) is a philosopher central to deontological moral theories and is responsible for developing one of the most frequently cited deontological ethics (Bartneck et al., 2021). Kant argues that an action is only obligatory if it satisfies the 'categorical imperative', which broadly states that one should act only in a way that could be made into a universal law. For Kant, respecting human dignity means treating people as ends in themselves, not as a means to an end. Further, for Kant, an action's moral worth can be judged not from the consequences that flow from it, but from the intention that caused it. Doing the right thing because it's the right thing is what matters morally for Kant. For example, if a person donates to someone begging because it feels good, the act isn't inherently moral; it is moral only if done out of empathy and out of duty, and because it's the right thing to do. The justice of actions depends on intent rather than consequence.

Deontological theories are often categorised into agent-centred, patient-centred and contractualist deontological theories. In agent-centred theories, we have both permissions and obligations to act in a certain way. Morality is personal, and we are each obliged and permitted to 'keep our own agency free of moral taint' (Alexander & Moore, 2021). Patient-centred theories are rights-based rather than duty-based, meaning we all have the right to not be 'used' only as means for producing good consequences, without our consent. Contractualist deontological theories focus on the principles and rules that individuals would voluntarily agree upon under fair and reasonable conditions. Morally wrong acts would be forbidden by principles that people in a suitably described social contract would accept or forbidden only by principles that such people could not 'reasonably reject' (Alexander & Moore, 2021).

A major example is John Rawls' theory of justice as fairness, which asks: what principles of justice would we agree to under a 'veil of ignorance'? In this hypothetical situation, where we do not know our social position, we would select principles ensuring equal basic liberties and only those inequalities that benefit even the least advantaged (Sandel, 2010). In his view, the way to think about justice is to ask what principles we would agree to if we did not know our place in society, class, race, gender, or religion. If – in this hypothetical situation – we think about justice without knowing whether we would be rich or poor, healthy or frail, a banker or a bus driver, we would

adopt a system of equal basic liberties for all citizens and accept only those inequalities in income and wealth that work to the advantage of the least well-off members of society.

Deontology contains strengths and limitations that merit thoughtful consideration. A strength is that it provides a universal moral framework that can apply to all individuals and situations, unlike Utilitarianism's focus on outcomes. Further, it does not override the individual's rights and freedoms, highlighting the importance of respecting the value and dignity of individuals and treating others as ends in themselves, rather than as mere means to an end. However, a limitation is that deontology can be rigid, sometimes offers conflicting duties with little guidance for resolution, and often disregards the consequences of actions.

These limitations are acute in data science. For example, deontology emphasises the importance of fulfilling moral duties and respecting individual rights. How, then, would a data scientist deal with a trade-off between the duty to protect individual privacy and the duty to advance scientific knowledge or provide societal benefits through her data science analysis? Deontological theories typically prioritise adherence to moral rules and duties without significant consideration of consequences and may be of little to no use in weighing up complex ethical trade-offs.

Virtue ethics is another of the three major approaches in normative ethics. Virtue ethics emphasises the virtues or moral character, in contrast to consequentialism, which emphasises the consequences of actions, or deontology, the approach that emphasises duties or rules. All three theories make room for consequences, rules, and virtues, but the distinguishing feature of virtue ethics is the centrality of virtue within the theory. Virtue ethics emphasises the development of virtuous character traits, such as integrity, fairness and responsibility. Applied to data science, cultivating these virtues can help professionals make ethical decisions when collecting, analysing, and using data.

In the West, virtue ethics' founding fathers are Plato and Aristotle, and in the East, it can be traced back to Mencius and Confucius (Hursthouse & Pettigrove, 2022). Three key concepts are *arête* (excellence or virtue), *phronesis* (practical or moral wisdom) and *eudaimonia* (usually translated as happiness or flourishing). Central to all forms of virtue ethics are virtue and practical wisdom. A virtue is an excellent character trait; to possess a virtue is to be a certain sort of person with a certain complex mindset. For example:

> An honest person's reasons and choices concerning honest and dishonest actions reflect her views about honesty, truth, and deception… Valuing honesty as she does,

> she chooses, where possible, to work with honest people, to have honest friends, to bring up her children to be honest (Hursthouse & Pettigrove, 2022).

The example above makes it apparent that while some might be paragons of virtue, others might strive for virtue but fall short. Therefore, possessing a virtue is a matter of degree. Another way in which one can easily fall short of full virtue is through lacking phronesis – moral or practical wisdom. For example, a person might take virtues too far if they lack practical wisdom. People are practically wise when they understand what is truly worthwhile, truly important, and advantageous in life. They know how to 'live well'.

Objections against virtue ethics include the a) application, b) adequacy, c) relativism, d) conflict, e) self-effacement, f) justification, g) egoism and h) situationist problems. To outline these in turn: in a) application, a challenge is that virtue ethics may struggle to provide specific guidance on how to apply virtues to concrete moral dilemmas; and in b) adequacy, focusing on virtues alone may not adequately address other important moral considerations such as consequences or duties. Virtue ethics may be susceptible to c) relativism, as virtues and moral character can vary across cultures and individuals, and different virtues may come into conflict with one another, making it difficult to determine which virtue should take precedence in each situation. Virtue ethics may downplay, or e) self-efface, the importance of rules and principles, and critics claim that virtue ethics does not provide a satisfactory account of how virtues are; f) justified or identified. Critics also argue that virtue ethics may promote self-centredness; or g) egoism, if virtues are pursued solely for personal excellence and flourishing. Finally, empirical studies suggest that people's behaviour is often influenced more by h) situational factors than by stable character traits, challenging the emphasis on virtues in ethical decision-making.

While traditional ethics education often emphasises acquiring knowledge of codes, principles and frameworks, empirical work in medical ethics education suggests the need for more transformative learning approaches (Gille & Nardo, 2020). Transformative learning centres on the change of perspective that takes place when learners critically examine their assumptions in light of 'disorienting dilemmas' or complex, unresolved problems. Experiences that initially generate confusion or discomfort – which Gille and Nardo call 'negative experiences' – can be fertile ground for deeper moral and political understanding. Embedding such perspectives into data science ethics curricula could move training beyond the recitation of abstract principles, towards confronting the socio-political realities of data use, including structural injustice and power asymmetries.

We can consider an oversimplified but useful framework for normative approaches to ethics that divides into utilitarian (or consequentialist), deontological and virtue ethics approaches. Consequentialist approaches focus upon the outcome of actions; deontological approaches focus on the rules that should be followed; and virtue ethics focuses on the character of agents (Boddington, 2023, p. 231). This framework is represented below in Table 1.

Table 1: Ethical frameworks and data science

	Utilitarianism / Consequentialism	Deontological ethics	Virtue Ethics
Summarised description	An action is right if it promotes the best consequences (the most happiness for the most people).	An action is right if it is in accordance with a moral rule or principle.	An action is right if it is what a virtuous person would do in the circumstances.
Central concern	What matters is the outcome of the actions.	What matters is that there are rules that should be followed, and that people must be seen as ends and may never be used as means.	What matters is the virtues and practical wisdom of the person acting.
Guiding value	Maximum happiness for the maximum number of people.	Right, and doing one's moral duty.	Virtue (leading to the attainment of eudaimonia).
Implications for data science	Individual decisions taken by data scientists do not matter in and of themselves – what matters is that the outcome achieves the most overall happiness.	The data scientist should adhere to moral duties and principles when making decisions. Individual rights and freedoms should not be sacrificed in the search for overall levels of happiness for the population.	The data scientist should cultivate her own moral virtues and apply these along with practical wisdom when making decisions.

The three approaches focus on outcomes, actions, and agents. The utilitarian focus on outcomes is clearly important, but many questions in data science ethics concern issues of agency, such as when humans outsource their agency to algorithmic decision-making. Further, the assessment of outcomes across 'the population' (the entire world?) is immensely challenging. Utilitarianism, focusing on maximising the most happiness, or utility, for the most people, prioritises outcomes and

consequences. This is helpful for optimising algorithms for user engagement, satisfaction and overall societal benefit, but raises concerns about individual rights, privacy, and potential trade-offs in pursuing the greater good.

The rule-based focus of deontology is appealing, but rules must always be interpreted, which leaves data science open to human error. Further, strict adherence to rules when making data science decisions might result in ludicrous and undesirable outcomes. For example, an absolute rule on data privacy would have prevented most data analysis undertaken during the COVID-19 pandemic. Deontology, emphasising ethical principles, duties and rights, provides a strong foundation for ensuring user privacy, informed consent and fairness. Still, it may not provide clear guidance in cases where conflicting duties or principles arise.

Virtue ethics and the cultivation of virtues by each data scientist is an appealing proposition, yet it places a great individual burden of responsibility on each data scientist's shoulders and leaves little scope for universal principles. Virtue ethics can guide data scientists in cultivating virtues like honesty, integrity and empathy, which can positively influence algorithm design and user experiences, but does not provide specific decision-making guidelines. Further, relying solely on individual virtue is insufficient in a context where institutional norms and structures shape the daily realities of data science practice. Institutions, whether companies, universities or government agencies, play a critical role in establishing ethical cultures. Placing all the responsibility on individual data scientists' shoulders means that institutions evade responsibility.

It is helpful to use an example to illustrate how these different theories of ethics might impact a data science decision. Consider, for example, a hypothetical scenario where a data scientist needs to design an algorithm that allocates treatment to patients with a life-saving drug. Using a utilitarian approach, the data scientist would focus on maximising overall utility, which in this case, means saving as many lives as possible. Using a vast database that correlates patient characteristics (for example, age, medical history, co-existing conditions, and similar factors), she develops an algorithm that allocates the drug to those patients who statistically have the highest probability of survival. In this case, the drug might be allocated to younger patients with a higher likelihood of survival, or those with no comorbidities. For example, suppose there are 100 doses and 200 patients. In that case, the algorithm might recommend giving all doses to those with the highest predicted survival rates, even if older or sicker patients are excluded entirely. This approach

maximises the total number of lives saved but can result in difficult moral trade-offs and the systematic exclusion of certain groups.

If she uses a deontological approach, she will prioritise moral principles and duties and develop an algorithm that follows predetermined ethical guidelines and ensures fairness, justice, and equal treatment of patients. The algorithm might consider factors such as equitable distribution. It may then allocate the drug to patients based on objective criteria that prioritise those in dire conditions or with the greatest medical necessity, regardless of age or other demographic factors. For instance, the algorithm could use a lottery system or prioritise patients in the most urgent need, ensuring that no group is categorically excluded. This might mean that some doses go to patients with lower chances of survival, but the process is guided by justice and equal treatment, not just the maximisation of saved lives.

Using a virtue ethics approach, the data scientist would make sure she cultivates in herself the virtues of honesty and fairness, compassion and empathy, and the like. In this case, the algorithm might balance decisions between saving lives and treating patients with dignity and respect. It might incorporate values like patient autonomy, shared decision-making, and a comprehensive assessment of patients' circumstances, including medical factors and personal and social considerations, and it may not solely focus on maximising the number of lives saved. For example, the data scientist could convene a multidisciplinary team, including ethicists, clinicians and patient advocates, to consider each case. Decisions might consider whether a patient is a primary caregiver, their expressed wishes regarding treatment, or other social and personal factors, aiming for a compassionate and context-sensitive allocation process.

The hypothetical example I have chosen mimics the decision-making that many leaders worldwide had to make during the COVID-19 pandemic, in this case, often for life-saving and scarce ventilators, instead of my hypothetical example of life-saving and scarce drugs. Savulescu et al (2020) make a strong argument for the use of utilitarianism in such a situation:

> There are no egalitarians in a pandemic. The scale of the challenge for health systems and public policy means that there is an ineluctable need to prioritize the needs of the many. It is impossible to treat all citizens equally, and a failure to carefully consider the consequences of actions could lead to massive preventable loss of life. In a pandemic there is a strong ethical need to consider how to do most good overall (p. 620).

Each of the normative ethical frameworks discussed above has advantages and disadvantages. However, as this example shows, each ethical framework leads to different outcomes and trade-offs in specific situations, emphasising the importance of critically examining which approach is most appropriate in real-world data science dilemmas. A good general strategy would be to consider all three. Drawing on utilitarianism, deontology and virtue ethics, data scientists can more comprehensively identify potential harms, conflicts and blind spots that any single approach might miss, leading to more balanced and justifiable decisions. However, combining these frameworks is not without risk. Because each is grounded in fundamentally different principles, drawing on all three simultaneously could result in inconsistent or even conflicting ethical guidance. This potential for conflict underscores the need for careful reflection when integrating multiple ethical perspectives. Further, as Lundie (2024) contends, while ethical frameworks can be useful for guiding practice, they can become overly complicated and even obscure harms if the underlying structures and motivations within Big Data practices are not carefully examined.

It is still all too common for data scientists to have limited engagement with these ethical frameworks and considerations. This lack of engagement could be due to gaps in training, the pressures of fast-paced professional environments, or an unwillingness, unease or even fear of engaging with unknown ethical frameworks. But given the high stakes and far-reaching consequences of data-driven decisions, when data scientists fail to engage with ethical perspectives, the result can be the perpetuation of injustice. Considering multiple ethical frameworks is not just an academic exercise but a practical necessity for responsible and trustworthy data science.

There is, of course, a challenge in focusing on training individual data scientists, when data science happens in teams in organisations. The reasons that I chose deontology, utilitarianism and virtue ethics as the starting point for this dissertation is because they represent three dominant and contrasting approaches to moral reasoning, focusing respectively on duties, consequences and character. This makes them well-suited, although admittedly imperfect, as analytical lenses for identifying the implicit patterns in how the data scientists I interviewed reasoned about ethical dilemmas in their work. Their prominence in professional ethics education also supports their relevance here, since these are the frameworks that data science students are most likely to encounter, if they encounter any ethical frameworks at all. At the same time, their shared focus on the individual moral agent makes them a useful starting point for a critical examination of their limitations in addressing the structurally embedded harms that data justice approaches bring into view.

Although deontology, utilitarianism and virtue ethics differ in how they evaluate moral action, they share a broadly liberal-individualist assumption that locates ethics in the autonomous, rational agent. However, this focus risks obscuring how the choices data scientists face are shaped and constrained by structural inequalities, organisational pressures and power relations, limiting these frameworks' capacity to address the collective harms that are central to data justice. As I will go on to discuss in Chapter 6, the data scientists I interviewed often described ethical dilemmas that were not really individual choices at all, but the product of organisational priorities, client demands and inherited patterns in the data they were given to work with.

Data science is not a politically neutral discipline. Technical choices, from dataset selection to error metrics, can reflect normative judgements about whose safety, dignity, or well-being matters. As Dencik et al. (2019) put it, data must be understood as 'situated and necessarily understood in relation to other social practices' (p. 873). Recent scholarship has interrogated data science's political and social justice dimensions, often under the banner of data justice, which asks how data practices produce or contest structural inequalities and what a just datafied society would require. This approach questions both the field's self-conception as neutral, as well as the adequacy of existing ethical frameworks, paying attention to how data practices make people and communities visible or not visible, how they redistribute risks and resources and how they reinforce or contest existing power relations. Rather than treating problems as isolated 'biases' that can be fixed through technical adjustment, data justice scholars show how histories of oppression and structural inequalities can show up in data science. This focus is important for my study because it highlights the limitations of narrow, liberal-individualist ethics and points towards more structural and political questions that ethics education for data scientists needs to address. Taylor (2017) frames the field's animating concern: 'an idea of data justice — fairness in the way people are made visible, represented and treated as a result of their production of digital data — is necessary to determine ethical paths through a datafying world' (p. 1).

Green (2021) argues that data science is a political activity, not merely a technical or neutral endeavour, arguing that data scientists who design algorithms that shape decisions in domains such as policing, welfare and employment are actively participating in the distribution of rights, resources and power. Attempts by data scientists to claim neutrality or frame their work as 'just engineering' are themselves political acts, often reinforcing the status quo and existing social

hierarchies. Green suggests that striving for neutrality is not, in fact, apolitical but is instead a conservative stance that privileges dominant perspectives and legitimises existing injustices.

Benthall and Goldenfein (2020) contend that liberal approaches to law and ethics, grounded in notions of individual autonomy, privacy and property, are increasingly incompatible with the techno-political realities of data science. They explain how the rise of data platforms has inverted the traditional relationship between individuals and markets, eroding individual capacity for meaningful autonomy and self-determination. Both Benthall and Goldenfein (2020) and Green (2021) critique prevailing ethics training and liberal legal frameworks as insufficient for addressing the social harms and injustices perpetuated by data science. Arguing for an 'ethics of care' approach, Fotopoulou (2019) suggests that framing data practices as 'matters of care' (p. 239) helps reveal their emotional, physical and material aspects, along with the often-overlooked work of people who use, create and are affected by data. Green points to 'ethics-washing', a practice where ethical principles are invoked in abstract terms but do not allow for substantive change, often serving corporate interests rather than challenging structural sources of harm.

Green (2021) proposes a four-stage framework for moving data science towards a politics of justice: interest in social issues; critical reflection on the political nature of data science; application of methods to address injustice; and the development of participatory, community-centred practices. Green emphasises the importance of moving beyond incremental, 'reformist' reforms that simply optimise existing systems, advocating instead for 'non-reformist' reforms that challenge and reimagine social structures in search of substantive equality and anti-oppression. In summary, Green calls for the field of data science to:

> Abandon its self-conception of being neutral to recognize how, despite not being engaged in what is typically seen as political activity, data science logics, methods, and technologies shape society (p. 261).

He argues that it is not enough to have good intentions. Instead, data scientists must ground their efforts in detailed knowledge about how their work will be used, and in clear political commitments. He argues, 'By deliberating about political goals and strategies and by developing new methods and norms, data scientists can more rigorously contribute to social justice' (p. 261).

Approaches to moral education

To understand how ethics might be taught more effectively to data scientists, it is useful to look beyond the specific context of data science and draw on broader scholarship on moral education. This section reviews influential approaches to moral development and ethics education. In this section, I distinguish between descriptive accounts of how moral judgement develops, normative accounts of what ethical education should aim for and pedagogical questions about how ethics is taught in practice. For example, Kohlberg and Haidt describe how individuals reason about right and wrong. Curren, Carr and Freire outline the aims of ethical education and what kinds of character, judgement and responsibility education should cultivate. Finally, questions of pedagogical approaches discuss case-based methods, the design of standalone versus integrated modules and the possibility of apprenticeship under the guidance of mentors.

Kohlberg (1958) explains that:

> Moral judgements are judgements about the good and the right of action. Not all judgements of "good" or "right" are moral judgments, however; many are judgements of aesthetic, technological, or prudential goodness or rightness. Unlike judgements of prudence or aesthetics, moral judgements tend to be universal, inclusive, consistent, and to be grounded on objective, impersonal, or ideal grounds (p. 383).

It is this latter category of moral judgement that concerns ethics in data science, rather than questions of technical excellence. Kohlberg's research examines common school case incidents (for example, cheating or stealing) as shared case studies through which students deliberate about rules, fairness and responsibility. Power, Higgins and Kohlberg (1989) show how cluster meetings at the A-School revolved around concrete episodes, allowing students to articulate different perspectives, revise norms and gradually develop a student culture that discourages cheating. Rather than teaching principles in the abstract, moral growth occurs as learners return repeatedly to these lived cases, questioning excuses, clarifying obligations and linking 'what is right' to 'what I must do' in specific situations.

By contrast, Haidt's work on moral dumbfounding, and his broader social-intuitionist model, challenges the assumption that moral judgements are usually the outcome of deliberate reasoning. In classic studies using taboo cases such as consensual incest, participants often reached rapid, confident condemnations, struggled to produce coherent reasons and yet 'stubbornly' maintained their verdicts in a pattern Haidt labelled moral dumbfounding.

Haidt argues that moral judgements typically arise from quick, emotion-laded intuitions, with reasoning playing a largely post hoc, justificatory role and serving to persuade others rather than to generate the initial judgement. As Haidt explains:

> Moral reasoning does not cause moral judgment; rather, moral reasoning is usually a post hoc construction, generated after a judgment has been reached (2001, p. 814).

For ethics education, including in data science, this implies that curricula must engage learners' intuitive responses and social contexts, not only their capacity to articulate arguments, if they are to influence how practitioners decide and act.

Stanley, Yin and Sinnott-Armstrong (2019) complicate this picture by showing that, in many of these cases, participants' wrongness judgements track their beliefs about the likelihood that the action could have caused harm. The authors accept that people often struggle to articulate reasons in Haidt-style cases but contend that this does not mean judgements are irrational or reason-free. Across several experiments on 'harmless' taboo violations, they show that participants' moral 'wrongness' ratings are predicted by their judgements on the probability that the action could have caused harm. In so doing, their work supports ethics education that allows for intuitive response while at the same time helps people to describe the reasons that guide their judgements.

Behavioural ethics research takes this concern with real world judgement further by focusing on why well-intentioned people so often fail to act on their own standards. In their book 'Blind Spots', Bazerman and Tenbrunsel (2011) argue that many ethical failures are best explained not by deliberate immorality but by 'bounded ethicality', the systematic limits on our ability to recognise the ethical dimensions of our own decisions. The authors show how 'ethical fading' allows people to reframe problems in purely business or technical terms, thereby allowing the moral concerns to drop out a view, creating the 'blind spots' of the title. Writing about legal education, Nicolson (2008) makes a parallel point, saying 'knowing what is morally right by no means guarantees moral behaviour' (p. 151) and therefore effective ethical education must address moral motivation and moral courage as well as moral judgement.

Regarding the aims of ethical education, Carr (2007) argues that good teaching depends on the cultivation of teachers' moral character and *phronesis*, not just on technical skills:

> Teaching seems to be the sort of occupation in which professional effectiveness is greatly enhanced by the possession and exercise of personal qualities and practical

> dispositions that are not entirely (if at all) reducible to academic knowledge or technical skills (p. 369).

In so saying, Carr highlights the role of personal qualities and practical judgement in good teaching, in a similar vein to Aristotle's virtue ethics approach, which is 'primarily an ethics of character more than of action guidance' (p. 378). Carr's account of professional ethics therefore challenges approaches that treat ethics as the application of abstract rules. This is relevant in data science, as principle-based reasoning alone cannot solve the complex and uncertain situations data scientists face.

Similarly, Curren and Metzger (2017) argue successful teachers are those 'who are themselves practiced in norms and patterns of ethical reflection and have earned moral authority with their students, in part through exemplifying the requisite virtues, both reflective and interpersonal' (p. 177) and that just educational institutions must equip students with understanding, capabilities and virtues. As they explain:

> Moral virtue and the motivation it involves are acquired through experiencing a nurturing and just social environment and through guided practice. The supervision and coaching of practice would call learners' attention to factors that are relevant to decisions, provide a related moral vocabulary and explanations, and guide them in exercising the forms of discernment, imagination, reasoning, and judgment on which good decisions are based (p. 97).

So saying, they highlight that the aims of moral education should include the development of students' practical judgement in social and institutional situations, instead of just the application of abstract principles. This matters in data science, where ethical decision-making is part of everyday professional practice within organisational settings. Prinsloo and Slade (2016), drawing on Noddings, argue that even 'when a just decision has been reached, there is still much ethical work to be done' (p. 120), highlighting that rule-based ethics, however well-applied, does not exhaust what ethical practice requires.

Finally, case-based approaches to moral reflection offer a pedagogical bridge between abstract theory and practitioners' lived experience. Husu and Tirri (2003) present a qualitative case study approach to investigating and supporting teachers' moral reflection, analysing a single narrative of a classroom dilemma through three 'ethical reference points': the ethic of purpose, the ethic of rules and principles and the ethic of probability. As they explain:

> Our goal is to show how abstract philosophical theories can be translated into real-world ethics in education and how these reference points can help teachers in their practical ethical reflection (2003, p. 345).

The authors show how one case study can lead to ethical tensions in a teacher's reasoning, and how working through the case from different philosophical perspectives helps connect abstract ethical theories to the lived realities of educational practice. They argue that case reports provide rich forums for practising moral reasoning and dialogue, where the aim is not to apply a correct rule but to deepen understanding of values, relationships and consequences in context. This approach offers a helpful template for data science ethics education, where practitioners can analyse concrete project stories from several ethical standpoints and reflect on how their own choices might enact or challenge institutional norms.

In a similar vein, Bazerman and Tenbrunsel (2011) argue that traditional ethics education that assumes conscious recognition of dilemmas and focuses on rules or values is therefore insufficient. Instead, they call for interventions that redesign decision environments by making ethical issues clear and hard to miss. For data science ethics education, their work suggests using realistic organisational cases not only to ask 'what is the right thing to do?' but also 'why might a smart, well-meaning person fail to see this as an ethical problem in the first place?' and to design discussion around spotting and countering these blind spots. As they explain:

> Ethics interventions have failed and will continue to fail because they are predicated on a false assumption: that individuals recognize an ethical dilemma when it is presented to them (2011, p. 1).

Conclusion

There is a key question about what ethical education for data scientists is actually trying to achieve. Is it trying to build conceptual, abstract knowledge about ethical theories, or is it rather trying to activate moral perception and judgement? The Aristotelian tradition in moral education (Carr, 2007; Curren and Metzger, 2017) treats these as distinct educational aims and suggests that the harder educational problem is often recognising that a routine decision has ethical significance in the first place. Most data science ethics curricula presuppose that the ethical dimension has already been identified, and the student's task is to apply a framework, but in practice this recognition is itself something that needs to be cultivated. Gray and Witt (2021) describe this as 'a change in consciousness of participants in the machine learning economy that empowers them to be agents of action' (p. 1). In support of this point, one of the data scientists I interviewed describes a 'switch'

that has to be turned on, suggesting that the capacity to recognise that a technical decision raises ethical questions may matter more than the capacity to apply a framework once the ethical dimension has already been identified.

There is a further dimension worth flagging here: the difference between knowing that something is ethically problematic and being disposed to act on that knowledge, particularly when acting is costly or inconvenient. As Nicolson (2008) puts it, ‘knowing what is morally right by no means guarantees moral behaviour’ (p. 151). The virtue ethics tradition takes this gap seriously, and it raises the question of whether data science ethics education currently addresses only the epistemic dimension (knowing what is right) while neglecting the motivational one. This has implications for how ethics might be taught to data scientists, which I address in Chapter 7.

The next chapter sets out the research approach and methodology I used to investigate how practising data scientists experience these ethical challenges, and the adequacy (or not) of the education that prepared them for such work.

Chapter 4: Research approach, methods and methodology

This chapter outlines the research design and methodological framework underpinning this dissertation. Building on the theoretical foundations and ethical challenges identified in earlier chapters, it explains how the study was structured to explore data scientists' real-world experiences of ethical dilemmas and to assess the adequacy of current ethics education. Recognising the complex, value-laden nature of data science, the chapter details the integration of critical and liberal theoretical perspectives, justifies the selection of qualitative methods and describes the empirical process from the design of semi-structured interviews to reflexive thematic analysis.

Denzin and Lincoln (2018) assert that the design of a qualitative research project always begins with a socially situated researcher who moves from a research question to a paradigm or perspective, to the empirical world:

> The gendered, multiculturally situated researcher approaches the world with a set of ideas, a framework (theory, ontology) that specifies a set of questions (epistemology), which are then examined (methodology, analysis) in specific ways (2018, p. 52).

Following Denzin and Lincoln, I will summarise my research project: an empirical study, drawing on concepts from critical theory for power and structural analysis, and engaging concepts from liberal theory to address ethical considerations and educational objectives. Instead of adopting a single theoretical paradigm, my research integrates insights from both traditions to provide a more nuanced understanding of the ethical and political dimensions of data science education and practice. For example, critical theory offers valuable approaches for examining how data science can reinforce or challenge existing social hierarchies and institutional arrangements, highlighting issues of justice and the social construction of knowledge. At the same time, liberal theory focuses on individual agency, moral reasoning, and the cultivation of ethical judgement, which are necessary for preparing data scientists to act responsibly within diverse professional contexts. This dual engagement allows the research to address both the structural and the personal dimensions of ethical practice in data science.

I used semi-structured interviews, which I recorded and transcribed, and then used Braun and Clarke's six-phase process for reflexive thematic analysis to analyse the interview data. Empirical research involves collecting and analysing data to address a research question rather than purely

theoretical or speculative concepts. Empirical research often employs quantitative, qualitative, or a combination of both to collect and analyse data. Since my research involves collecting data for analysis, my research is empirical.

This empirical approach is crucial because it grounds the study in data scientists' lived experiences and perspectives. Using semi-structured interviews, I was able to capture detailed accounts of data scientists' experiences of ethical challenges in real-world settings, providing insights into the ways in which they encounter and navigate ethical issues in practice. Using reflexive thematic analysis enabled me to systematically identify and interpret patterns and themes within the data, while remaining sensitive to context and the complexities of the interviewees' experiences. This methodological choice strengthens the credibility and depth of the findings and also aligns with my research paradigm by foregrounding issues of power, justice and the social construction of meaning within data science practice. Ultimately, this approach enables the research to contribute both practical and theoretical insights into how ethics education can be improved to better address the realities faced by data scientists today.

Research paradigms

Paradigms are 'sets of beliefs and practices, shared by communities of researchers, which regulate inquiry within disciplines' (Weaver & Olson, 2006, p. 459). They are 'characterised by ontological, epistemological and methodological differences in their approaches to conceptualising and conducting research' (Weaver & Olson, 2006, p. 459). A research paradigm is a broad framework that shapes a researcher's worldview, assumptions and methodological choices, influencing how research problems are defined and addressed.

Among the most common research paradigms is positivism, which assumes knowledge is derived from objective observations and measurable phenomena. This paradigm is grounded in discovering causal relationships and establishing general laws with principles such as objectivity, replicability and control. Post-positivism builds on the empirical rigour of positivism but also recognises and explicitly acknowledges the complexities and interpretive nature of human experience. While post-positivism values systematic observation, it also acknowledges that all knowledge is, to some extent, shaped by context and subject to revision.

The interpretivist paradigm, on the other hand, focuses on understanding individuals' subjective meanings and experiences within their social and cultural contexts. It views reality as socially constructed, with meaning shaped by participants' interactions and interpretations. According to Weaver and Olson (2006, p. 460), the interpretive paradigm highlights understanding of the 'meaning individuals ascribe to their actions and the reactions of others'.

Critical social theory, inspired by the writings of Marx, Habermas and Freire, includes feminist, grassroots and emancipatory movements. Critical theory has both a narrow and a broad meaning in philosophy and the history of the social sciences (Bohman, 2021). In the narrow sense, critical theory is a product of the Frankfurt School, which encompasses several generations of German philosophers and social theorists in the Western European Marxist tradition. It seeks the emancipation of humanity and works to address the needs and powers of human beings. In the broader sense, critical theory relates to the many social movements that identify different dimensions of the domination of human beings today. In both the broad and the narrow senses, however, critical theory is aimed at all forms of decreasing domination and increasing freedom.

Critical social theory is concerned with the study of 'social institutions, issues of power and alienation, and envisioning new opportunities' (Weaver & Olson, 2006, p. 460). In other words, it is an approach to research and analysis that focuses on uncovering and critiquing power structures, inequalities, and social injustices within societies. Critical theory aims to promote social change and emancipation. A critical theory perspective assumes that truth exists as 'taken for granted' realities shaped by social, political, cultural, gender and economic factors (Weaver & Olson, 2006, p. 461). According to Kincheloe et al. (2018), critical theory encompasses:

> An ever-evolving criticality that engages the current crisis of humanity, all life forms, and the Earth that sustains us—a criticality that through its various theories and research approaches maintains its focus on a critique *for* social justice (p. 418).

This means that social justice, and the achievement thereof – or at least, the striving towards achievement– is the aim of research. Interestingly, Hammersley (2005) calls for 'bounded' (my term) criticism, and argues that there ought to be 'proper limits' to criticism. Like anything else, he says, 'criticism is not always a good thing' (p. 175). He believes that those engaging in criticism in research must consider the soundness of their arguments and the likely consequences of their criticism, since these consequences cannot always be assumed to be beneficial (Hammersley, 2005). I heed Hammersley's warning but also believe that in the case of my research, an enhanced

awareness of the potential misuse of data science to perpetuate systemic inequalities is unlikely to be detrimental to emancipation.

Tripp (1992) provides a helpful definition for socially-critical research in education:

> Socially-critical research in education is informed by principles of social justice, both in terms of its ways of working and in terms of its outcomes in and orientation to the community. It involves strategic pedagogic action on the part of classroom teachers, aimed at emancipation from overt and covert forms of domination. In practical terms, it is not simply a matter of challenging the existing practices of the system, but of seeking to understand what makes the system be the way it is, and challenging that, whilst remaining conscious that one's own sense of justice and equality are themselves open to question (p. 13).

Tripp's definition is entirely appropriate to my research project and my reflexivity process as a researcher. In reflecting on the aims of my study, it gradually dawned upon me that what I wanted to research was how the teaching of data science could be shifted from a pure focus on the technical aspects of 'doing' data science (which requires knowledge of statistics and computer programming) to incorporate broader issues that addressed the potential misuse of data science (whether unwittingly or not) to perpetuate the existing practices of the system. This is aimed at 'emancipation from overt and covert forms of domination', including but not limited to the powerful and permeating belief that data are just numbers and therefore devoid of bias. Shifting data science teaching in this manner demands 'strategic pedagogic action' from both the data science teachers and me.

Kincheloe et al. (2018) outline a set of foundational assumptions for the critical scholar, pedagogue, or activist. Firstly, that all thought is fundamentally mediated by power relations that are social and historically constituted. This challenges the notion of objective, context-free knowledge and emphasises that our ways of knowing, interpreting, and questioning the world are shaped by the social structures and power dynamics in which we are embedded. This perspective compels critical scholars to interrogate whose interests are served by dominant ways of thinking and to recognise that knowledge is never neutral because it is always situated within particular historical and social contexts.

Secondly, facts can never be isolated from the realm of values or removed from some form of ideological impressions. The idea that facts cannot be separated from values or ideologies directly challenges positivist traditions that claim objectivity and value-neutrality in knowledge production. In practice, this means that what is counted as a 'fact' is always filtered through personal, cultural,

political and institutional lenses. For critical scholars, this recognition calls for a reflexive approach that acknowledges the values and assumptions underlying research questions and data interpretations.

The third foundational assumption is that the relationship between concept and object is not stable or fixed and is often mediated by the social relations of capitalist production and consumption. This implies that meanings are not fixed but are constantly negotiated and contested within the realms of economic and social power. For example, the meaning of ‘success’ or ‘progress’ in education or technology is often defined by market logics, which can obscure alternative values or priorities. To illustrate this, consider the language that some universities use when claiming job placement statistics of alumni as a measure of success.

Fourth, that language is central to forming subjectivity (conscious and unconscious awareness). This means that the words and narratives available to us shape how we describe the world and how we understand ourselves and our place within it. Critical scholars, therefore, pay close attention to discourse, rhetoric, and how language can empower and constrain individuals and groups.

Fifth, that certain groups in any society and particular societies are privileged over others and although the reasons for this privilege may vary widely, the oppression that characterises contemporary societies is most forcefully reproduced when subordinates accept their social status as natural, necessary or inevitable. This perspective insists that social hierarchies are not inevitable or natural but are actively constructed and maintained through cultural, economic and political mechanisms. Critical scholarship seeks to expose these mechanisms and challenge the normalisation of inequality.

Sixth, that oppression has many faces and focusing on only one at the expense of others (for example, on class oppression and not racism) often ignores the connections between them. Oppression often operates through interconnected systems, such as class, race, gender and sexuality. The insight that one form of oppression may be maintained at the expense of another highlights the importance of intersectionality in critical analysis. Addressing one axis of inequality without recognising its links to others risks reinforcing rather than dismantling systems of domination.

Finally, mainstream research practices are generally, although most often unwittingly, implicated in the maintenance of capitalist production and reproduction of systems of oppression. This critique

calls for a re-examination of research agendas, methodologies and the purposes they serve. Critical scholars are urged to move beyond technical or descriptive accounts and engage with their work's ethical and political dimensions, striving to produce knowledge that challenges rather than perpetuates injustice.

A parallel set of assumptions can be identified for researchers who approach data science with a critical perspective, particularly those attentive to power and institutional responsibility issues. The table below clarifies how some of the tenets of critical theory, especially its focus on power relations, structural dynamics and the social construction of knowledge, could directly inform a critical approach to data science. By highlighting these connections, the table demonstrates why drawing on critical theory is helpful for understanding and addressing the ethical and political dimensions of data science practice. This framing also supports the subsequent analysis and recommendations, highlighting that data science is never value-neutral but is always shaped by broader social forces, institutional contexts and power dynamics.

Table 2: Critical theory and data science

	Critical theory, per Kincheloe et al. (2018)	Data science
Power	All thought is fundamentally mediated by power relations that are social and historically constituted.	'Raw data' is an impossibility; data is inherently mediated by power relations that are social and historically constituted.
Fact	Facts can never be isolated from the domain of values or removed from some form of ideological inscription.	Data is never fact; data can never be isolated from the domain of values or removed from some form of ideological inscription.
Relationships	The relationship between concept and object and between signifier and signified is never stable or fixed and is often mediated by the social relations of capitalist production and consumption.	The way in which data is obtained is often mediated by the social relations of capitalist production and consumption.
Language	Language is central to the formation of subjectivity (conscious and unconscious awareness).	The categorisation of data for analysis takes place through language labels, which are therefore always subjective.
Privilege	Certain groups in any society and particular societies are privileged over others and, although the reasons for this privileging may vary widely, the	Certain groups in any society and particular societies are privileged over others and, although the reasons for this privileging may vary widely, and

	oppression that characterises contemporary societies is most forcefully reproduced when subordinates accept their social status as natural, necessary, or inevitable.	the use of data science can – consciously or most often unconsciously – reinforce patterns of privilege and oppression, through opaque algorithms that are difficult to challenge.
Oppression	Oppression has many faces and focusing on only one at the expense of others (e.g., class oppression vs. racism) often elides the interconnections among them.	The same – oppression has many faces and focusing on only one at the expense of others (e.g., class oppression vs. racism) often elides the interconnections among them.
Reproduction of systems	Mainstream research practices are generally, although most often unwittingly, implicated in the maintenance of capitalist production and in the reproduction of systems of oppression, including poverty, racism, sexism, heteronormativity, religious oppression, ableism, and others.	Mainstream data science practices can be – although most often unwittingly – implicated in the maintenance of capitalist production and in the reproduction of systems of oppression, including poverty, racism, sexism, heteronormativity, religious oppression, ableism, and others.
	Text above from Kincheloe et al. (2018, p. 420-421).	Text above from my own analysis

(Column on left, own analysis; middle column Kincheloe et al., 2018, p. 420-421; column on right, own analysis).

This table helps to bridge theory and practice by providing a concrete link between abstract theoretical concepts and the lived realities of data science. The side-by-side comparison highlights structural parallels to reveal how issues such as power, privilege, language and relationships are not just philosophical concerns but embedded in how data is produced, analysed and used in contemporary society. By showing that data science is always shaped by social, historical, and power relations, the table invites researchers and practitioners to reflect critically on their assumptions, methodologies, and the broader impact of their work. And finally, the table underpins my central claim that ethical and political challenges in data science cannot be addressed through technical solutions alone, but require a broader, justice-oriented framework rooted in critical theory.

For these reasons, I have chosen to draw from critical social theory for my research. Further, critical researchers and teachers understand that praxis involves both theory and action and that each informs the other (Kincheloe et al., 2018). Data science is a field in which praxis is critically important, since the practice of data science itself can perpetuate social injustice.

In summary, in designing this research, I drew on insights from critical theory, interpretivism and post-positivism to inform the methodological approach and the interpretation of findings. Each paradigm offers a distinct perspective on the nature of knowledge, the role of the researcher and the aims of inquiry, and understanding their differences is essential for justifying the choices made in this study.

The critical theory perspective is particularly valuable for interrogating how power relations, structural inequalities and institutional dynamics shape knowledge production and professional practice in data science. Critical theory assumes that knowledge is socially constructed and deeply influenced by historical, political and social contexts. It seeks not only to understand the world but also to critique and transform it by exposing and challenging systems of domination and injustice. This orientation is especially relevant for research that addresses ethical and political dimensions of data science, as it highlights questions of justice, emancipation and the social construction of meaning.

The interpretivist paradigm focuses on understanding individuals' subjective meanings and lived experiences within their social and cultural contexts. Interpretivism views reality as constructed through interaction and interpretation, emphasising the importance of capturing research participants' perspectives and sense-making processes. While interpretivism is well-suited to exploring how data scientists experience and navigate ethical challenges, it does not necessarily highlight the broader structural and power dynamics that critical theory brings to the fore.

Postpositivism builds on positivism's empirical rigour but departs from strict objectivity by recognising that all knowledge is provisional, context-dependent and subject to revision. Postpositivist researchers value systematic observation and empirical evidence, while also acknowledging the interpretive nature of human experience and the influence of context on knowledge claims. This paradigm supports a reflexive and context-sensitive approach to research, which is important for addressing the complexity and situatedness of ethical issues in data science. For example, many of the data scientists I interviewed are based in South Africa, a context different from that in developed countries. In reading their stories, their ethical challenges might feel foreign to someone from a developed country, with well-developed infrastructure, higher levels of employment and lower levels of inequality.

By integrating elements from these paradigms, this study addresses both personal and the structural dimensions of ethical practice in data science. Critical theory provides the tools to analyse power and systemic injustice, interpretivism ensures sensitivity to participants' lived experiences, and postpositivism encourages reflexivity and recognition of the provisional nature of knowledge. This blended approach enables a nuanced and comprehensive analysis that is critically engaged and responsive to the complexities of the field.

Looking at prior research, several scholars have drawn on critical theory when researching data, big data, data science and artificial intelligence. For example, Neff et al. (2017) pose the question: What would data science look like if its key critics were engaged to help improve it, and how might critiques of data science improve with an approach that considers the day-to-day practices of data science? In their paper, they summarise four common critiques of data science, namely: data are inherently interpretive, data are inextricable from context, data are mediated through the socio-material arrangements that produce them, and data serve as a medium for the negotiation and communication of values (p. 85). Their findings emphasise that understanding and interpreting data is inherently a collective and context-dependent process, with data best seen as stories that are co-constructed and negotiated among multiple participants. This underscores the importance of incorporating collaborative, narrative and context-aware practices into both data science research and education, moving beyond a purely technical approach.

Moats and Seaver (2019) staged an intriguing research encounter. They asked practising data scientists to analyse a body of critical social science literature about their work, using data science tools. Their experiment was:

> …designed to probe the divide between data scientists who make algorithms and qualitative social scientists who study them by encouraging data scientists to reflect on the content of these criticisms and, in turn, the approaches they use to make sense of them (p. 9).

As things turned out, they found that matters were considerably more complex than they originally anticipated. For example, they did not anticipate data scientists being more self-critical than expected: 'often the critical literature on data scientists paints them in simplistic ways; they are far more critical of their own tools' (p. 9). They concluded that data scientists are often aware of the limitations and biases in their tools. This is a valuable insight because it challenges simplistic critiques of data science and suggests that there is room for reflexivity and self-critique within the

field. This suggests that efforts to improve data science ethics should build on existing critical awareness among practitioners, rather than assuming a lack of concern or understanding.

Tacheva (2022) argues that 'a resolutely transnational feminist approach can provide data theorists and practitioners with the hermeneutic tools necessary to identify and disrupt instances of injustice more inclusively and comprehensively' (p. 1). She identifies five ways transnational feminism can be leveraged as an intervention into the current data science canon. These include: the development of a framework for practicing data science based on feminist principles such as naming and resisting power in data science projects; identifying the interconnectedness between historically, geographically and socially distant struggles; acknowledging methods that exist beyond the 'Western Canon of the Enlightenment' (p. 2); diversifying knowledge; and demystifying the 'newness' of algorithms by acknowledging that algorithms can simply provide a different delivery system of oppression marked by invisibility and inscrutability (p. 3). She concludes that incorporating feminist principles, such as naming and resisting power, acknowledging interconnected struggles and diversifying knowledge, can transform data science practice and theory. This is noteworthy because it offers a concrete path for making data science more inclusive and just.

Desai et al. (2022) critically review the epistemological foundations of data science. They divide the epistemology of data science into five domains: (1) the constitution of data science; (2) the kind of enquiry that it identifies; (3) the kinds of knowledge that data science generates; (4) the nature and epistemological significance of 'black box' problems; and (5) the relationship between data science and the philosophy of science more generally (p. 468). Through a comprehensive literature review and analysis, they conclude that there are 'significant open problems and debates' in data science. Through their review – and the burgeoning evidence that data is always embedded in society and will always be interpreted by researchers – they restructure this division. Their reconstructed domains are: (1) descriptive and normative accounts of the composition of data science; (2) reflections upon the kind of enquiry that data science is; (3) the nature and genealogy of the knowledge that data science produces; (4) 'black box' problems; and (5) the nature and standing of a new frontier within the philosophy of science that is raised by data science (pp. 468-469). Their findings reinforce that data science is not just a technical field but is entwined with social, philosophical and ethical questions. This matters, because it demonstrates that data science's challenges are not just about better algorithms or more data, but about fundamentally rethinking how knowledge is produced, interpreted, and used in society.

These studies collectively show that data science is deeply social, interpretive and value laden. The most interesting and urgent conclusions are that context, collaboration and critical reflexivity are essential for ethical and meaningful data science. These findings underscore why drawing on the critical theory perspective is not just relevant but necessary: it provides the tools to question assumptions, address injustice and ensure that data science serves the broader public good rather than reinforcing existing power structures.

From paradigm to methodology and method(s)

As above, a paradigm is a way of thinking about research that influences methodology, which, in turn, influences the research method(s) used. It does this through the lens of a paradigm's key elements, including–but not necessarily restricted to – epistemology, ontology, ideology and axiology. These philosophical concepts are crucial in shaping how researchers approach and understand the nature of knowledge, reality, values, and belief systems.

Epistemology focuses on the nature of knowledge and how we believe that it is possible that we 'know' something. It explores questions related to the nature of truth, justification and the methods used to acquire knowledge. This raises further questions about the context of research, generalisability and transferability of 'results' and the role of the researcher. Critical theory challenges traditional positivist and objectivist epistemologies, emphasising that knowledge is socially constructed and influenced by power dynamics, ideologies, and historical contexts. Qualitative inquiry seeks 'to discover and to describe narratively what particular people do in their everyday lives and what their actions mean to them' (Denzin & Lincoln, 2018, p. 87). Qualitative research is an appropriate way to explore diverse viewpoints, narratives and lived experiences. My research is, therefore, qualitative, since the emphasis is placed on understanding how knowledge is produced, who controls it and how it can challenge and transform existing power structures.

Where epistemology is focused on *how* we can come to know, ontology is the study of what exists and *what* we can know. Researchers' ontological perspectives determine how they define and conceptualise the objects of their study and whether they view reality as objective, subjective or a combination of both. Critical theory recognises that reality is complex and socially constructed. This guided my research inquiry into the social forces that, through data science, can serve to sustain inequalities and injustices.

Finally, my research was grounded in a commitment to values such as equality, democracy, human rights and the reduction of power imbalances, which together form the axiological foundation of my inquiry. These values are commitments that shaped every aspect of my research, from the questions I asked to the methods I employed and how I interpreted findings. Specifically, my approach to axiology centres on advancing social justice and equity by critically examining how data science can either reinforce or challenge existing systems of oppression. I paid attention to power structures in technology, in data science and in capitalism with a view to revealing structures, legislation and practices that perpetuate injustice. By focusing on these values, my research contributes to academic debates and the broader project of creating more just and equitable social systems.

Methodology, methods and the collection and analysis of data and writing

Denzin and Lincoln (2018) describe research design as a 'flexible set of guidelines that connect theoretical paradigms, first, to strategies of inquiry and, second, to methods for collecting empirical material' (p. 58). In the research process, therefore, we move from paradigm to methodology, methods, the collection and analysis of data, and writing. Although we cannot assume a direct relationship between paradigms, methodologies, methods and writing, each element of the research process should accord with each other and the research question.

Moving from paradigm to the empirical world puts paradigms of interpretation into motion and connects the researcher to specific methods of collecting and analysing empirical materials. These methods can include, amongst others, observation, interviews, visual research, narrative inquiry, autoethnography, focus groups and collaborative inquiry. Because I wanted to find out more about the ethical issues that data scientists encounter in their field and how these might be addressed in ethical education, I chose interviews as my main data collection instrument. This is because I wanted to explore the perspectives, experiences and meanings that data scientists attribute to various aspects of their work as data scientists and as data science teachers. I selected the methods best suited to my research objectives: interviews are a valuable way to gather in-depth and nuanced information directly, to explore complex topics, understand interviewee perspectives and gain insights that might not be captured as effectively through other research methods.

Brinkmann states that 'the interview has become one of the most common ways of producing knowledge in the human and social sciences' (2018, p. 998). Although neither completely

structured nor completely unstructured interviews are possible (Brinkmann, 2018), interviews range in their relative degree of structure. In relatively structured interviews, standardised ways are used to ask questions, with the idea that answers can be compared across participants, and perhaps quantified. In relatively unstructured interviews, by contrast, the main role of the interviewer is that of a listener. Following an opening prompt, the only occasional questions are those for clarification. Semi-structured interviews find a middle ground between these two ends of the relative spectrum. The interviewer is more involved in the conversation's direction and becomes a 'knowledge-producing participant' (Brinkmann, 2018, p. 1002) rather than simply a listener or an issuer of survey-like questions.

According to Schostak (2006), too often interviewing is seen as simply a tool for data collection, while in reality it is a complex, subtle process that cannot be separated from the dynamic of the project or the multiple and changing contexts of everyday life. The interviewer's role in the interview and the subsequent analysis of the interview's contents raise important questions. Interviews in research are conducted to frame the interaction and raise several issues concerning power and control that are important to reflect upon for epistemic and ethical reasons. The interpretive practice of 'making sense of one's findings is artistic and political' (Denzin & Lincoln, 2018, p. 60). It is impossible to separate the researcher from the interpretation. The researcher becomes the interpreter, the 'writer-as-interpreter' (Denzin & Lincoln, 2018, p. 60). In Brinkmann's view:

> The interview is not a neutral technology to obtain people's descriptions but a kind of social practice that is historically constituted and with its own inbuilt presuppositions about human subjectivity that can be challenged (2018, p. 1011).

For these reasons, the personal biography of the researcher, who 'speaks from a particular class, gendered, racial, cultural, and ethnic community perspective' (Denzin and Lincoln, 2018, p. 52), is critical and demands a level of self-reflection and reflexivity by the researcher. There is an asymmetrical power relationship in the qualitative interview, since the interviewer initiates the interview, determines the interview topics and poses the questions. Post interview, too, the interviewer generally 'upholds a monopoly of interpretation over the interviewee's statements' (Brinkmann, 2018, p. 1017).

In reflecting on the aims of this study and my positionality as a researcher, I have noticed that I have moved from being embedded in an educational institution to a broader, more bird's-eye view as a researcher and observer of far broader systemic and global issues. As Chief Academic Officer

in my educational institution, I worked daily on hiring excellent teachers, building curricula with them, and discussing and building student assessments and rubrics. In data science, for example, my concerns included the following: How could we teach students detailed statistical methods and computer programming in a short period of time? How best could I persuade our professional data scientist teachers, who tend, in general, towards introversion and high attention to detail, to be vivacious and engaging in the classroom? What were the best exercises and assessments to determine how well and how quickly students were learning their statistics and programming? These are 'micro' issues, minutely concerned with the detail of teaching complex methodologies to young students.

The development of this research thesis shifted me from these 'micro' issues to a much broader 'macro' level. The research process forced me to confront my practices. It has led me to question, for example, what good it is to teach statistical concepts when they, excellently taught, could be used to reinforce structural injustices? My thoughts and practice have been irrevocably impacted by the process of researching and writing this dissertation. For the better, I trust.

The questions now concerning me could be called 'the macro made micro'. For example, I wanted to understand how best we can teach ethics to all students, how best we can convey to students how structural injustice works and explain to them how data science can be a tool to not only perpetuate social injustice but also entrench it, and the best ways that they can learn to avoid this. I want to understand how we can best assess how well we achieve this teaching in and for ethics and ethical conduct in data science – a far more complex endeavour than measuring understanding of statistical concepts and computer programming.

Reflexive thematic analysis

I used Braun and Clarke's six-phase process (Braun & Clarke, 2019) for reflexive thematic analysis to analyse the interview data. Reflexive thematic analysis is useful for addressing research questions related to 'people's experiences, understanding and representation, social processes, rules and norms, people's practices and behaviours and the construction of meaning' (Braun and Clarke, n.d.), and it was therefore suited to my research question, paradigm and methods. Reflexive thematic analysis is 'theoretically flexible' (Braun and Clarke, n.d.), which means it can be used within a range of theoretical frameworks.

Regarding data collection, Braun and Clarke (2022) highlight the broad spectrum of data collection approaches that have been used in published thematic analysis research, ranging from 'interviews and focus groups to story completion and visual methods' (p. 12). They emphasis a flexible approach to interviewing, which further reinforced my choice of using semi-structured interviews for data collection:

> A more flexible and fluid approach to interviewing that more closely resembles the 'messier' flow of real-world conversation: questions and topics are carefully considered but the interview centres the interaction and co-construction of meaning between researcher and participant; there is considerable scope for the researcher to be spontaneously responsive to the participant's unfolding account (Braun & Clarke, 2022, p. 13).

The semi-structured interviews I conducted were my main data collection instruments. I conducted interviews with nine data scientists. I began with a population of 20 potential participants, identified through my professional network as experienced data scientists who also taught data science. I made use of an online randomisation tool (wheelofnames.com) to determine the order in which I contacted potential participants. I then invited people sequentially in that randomised order until I had completed nine interviews.

The nine participants represented a diverse range of education and employment. All were practicing data scientists, but they worked across different sectors (including fintech, consulting, startups, health, research and non-profits) and in different capacities (for example, founders/CEOs, data scientists, researcher scientists and technical leads). All had experience teaching data science, either in universities, boot camps or corporate training. Their disciplinary backgrounds included economics, statistics, computer science, ecology, science and engineering. Geographically, participants originally came from America, Canada, Namibia, the Netherlands, South Africa and Zimbabwe, and they were working in their home country or working abroad.

These diverse educational and professional backgrounds shaped their reported understanding of ethical questions. For example, disciplinary training in statistics or computer science emphasises optimisation, efficiency, statistical performance and technical robustness as decision-making criteria. The type of organisation also determines the criteria for decision-making. For example, the decision-making logics of corporations, governments and large technical systems (e.g. platform infrastructures) can frame problems as commercial, operational or technical rather than ethical or political. In the analysis chapters, I examine how these diverse backgrounds contribute to the

implicit moral frameworks that participants drew upon, and how organisational logics enabled or constrained their expression of ethical concerns.

Researcher subjectivity and the researcher's role in knowledge production are at the heart of the reflexive thematic analysis approach. The researcher is asked to be cognisant of their assumptions, biases, and subjectivities throughout the research process. During the coding process, in particular, the researcher is required to question and query their assumptions when interpreting the data, since coding is the process whereby similar strands of data are ‘clustered together’ into themes by the researcher and are analytic outputs that are ‘actively created by the researcher at the intersection of data, analytic process and subjectivity’ (Braun & Clarke, 2019, p. 594).

My positionality in relation to the interviewees required explicit reflection. When I started the research, I was employed by the organisation that ran the data science teaching programme from which I recruited interview participants; I later left that organisation while my research was still in progress. This position created a potential power imbalance, particularly for participants who might have perceived me as having organisational power or being aligned with organisational interests. To mitigate this, I took several steps. I emphasised that each interviewee’s decision to participate and their answers would have no bearing on their relationship with the organisation. I used my university email for all research communication. I anonymised participants, organisations and specific projects in transcripts and analysis. During the analysis process, I repeatedly returned to the participants own language to ensure I was representing their views accurately.

I analysed the interviews thematically with coding emerging from and during the analysis. To remain faithful to and congruent with, my stated paradigmatic and methodological intent, I continuously cycled back to critical and liberal theory studies, and through my coding, to ensure that this dissertation does justice to and respects the voices of those I interviewed, and ultimately, those whom data science impacts – which is all of us, everywhere.

Reflexive thematic analysis provided a systematic yet flexible framework for engaging deeply with the interview data. I started with repeated readings of the interview transcripts to attain thorough familiarisation, followed by the generation of initial codes that captured content and references relevant to the research questions. Coding was an iterative process, and I refined and reorganised codes as new insights emerged. I then developed themes by clustering related codes, identifying patterns and constructing coherent narratives that reflected the participants’ experiences.

Throughout, I maintained reflexivity by keeping analytic memos, regularly questioning my assumptions and engaging in critical dialogue with the literature. This reflexive stance ensured that the analysis remained sensitive to context and that I continually scrutinised and refined my interpretations and my coding. In developing and refining themes, I also paid attention to the implicit moral theories evident in participants' accounts, noting where their explanations for their ethical choices aligned with, combined or resisted utilitarian, deontological and virtue-ethical perspectives.

The theoretical framework, drawing on both critical and liberal theory, played a central role in shaping the entire analytic process. Critical theory informed the design of interview questions by prompting a focus on issues of power, institutional responsibility and systemic inequalities within data science practice. Liberal theory, meanwhile, guided the exploration of individual agency, ethical judgement and the cultivation of professional responsibility. During coding and theme development, these theoretical perspectives provided concepts that helped identify and interpret data related to both the structural and the personal dimensions of ethical practice. For example, codes and themes were developed to capture not explicit references to organisational policies and power dynamics as well as participants' reflections on autonomy, fairness and moral reasoning. In interpreting the findings, integrating concepts from critical and liberal theory enabled a nuanced analysis that accounted for the broader institutional forces shaping ethical challenges as well as the individual capacities required for responsible action. This dual approach ensured that the analysis highlighted the interplay between systemic structures and personal agency.

In summary, Chapter 4 establishes the methodological foundation for this dissertation, clarifying how the research questions are addressed by drawing on critical theory and liberal theory, through qualitative empirical inquiry. The chapter begins by situating the study within a paradigm that draws on critical theory for power and structural analysis, while also engaging liberal theory for educational and ethical dimensions. This paradigm is significant for the project as it enables the research to move beyond surface-level description and instead interrogate how data science can both reinforce and challenge systems of inequality. Chapter 4 also outlines the empirical nature of the study, explaining the rationale for using semi-structured interviews with data scientists as the primary data collection method. This approach allows for the collection of rich, nuanced accounts of practitioners' lived experiences of encountering ethical challenges in their daily work.

Chapter 5: Interviews

As the preceding chapters have established, the ethical and political dimensions of data science are both urgent and complex, demanding more than technical solutions or abstract theorising. Having explored the theoretical foundations, literature and methodological approach underpinning this study, this chapter marks a pivotal transition: from conceptual groundwork to the lived realities of data scientists themselves. In this chapter, I present a detailed analysis of semi-structured interviews conducted with in-service data scientists. The aim is to illuminate how ethical challenges are experienced, interpreted and navigated in professional contexts, and to examine the adequacy of current ethics education and professional guidelines from the perspective of practitioners. By featuring the voices of those working at the coalface of data science, this analysis seeks to bridge the gap between theory and practice, revealing the nuanced, everyday dilemmas that arise in the field.

The interviews provide a rich, empirical foundation for understanding the types of ethical issues that data scientists encounter across the data science lifecycle, from data collection to model deployment and impact. They also shed light on the ways practitioners make sense of their responsibilities, negotiate organisational pressures and balance technical objectives with broader social values. The findings highlight perceived gaps in existing ethics education, and the strategies data scientists use to compensate for these shortcomings, as well as the tensions between individual agency, and institutional or regulatory frameworks. This chapter is structured thematically, with each theme illustrated with direct quotations and contextualised within the broader academic discourse. By focusing on the lived experiences of data scientists, this chapter grounds the dissertation's normative arguments in empirical reality, providing an important foundation for the recommendations and discussions in subsequent chapters.

5.1 Theme 1: The importance of ethics in data science

In this theme, I examine the importance of ethics in data science. A clear pattern emerged from all the data scientists I interviewed: ethical questions arise at every stage of the data science process, from data collection to final use. These concerns are significant, as each step directly affects people. As Data Scientist 1 put it:

> *Data Scientist 1: I think if you train accountants and finance students on the ethics of accounting and finance… it is arguable that data science can even have more of an impact in some areas of life. So, I think [ethical education for data scientists] is really important and should be introduced into the curriculum.*

Here, Data Scientist 1 emphasises that if ethics education is seen as essential for students in fields like accounting and finance, which already impact human well-being, then it is even more crucial for data science students, since the consequences of their work may be broader and more profound. This is a point underscored by Davis (2020), who argues that there is:

> Growing recognition of the importance of ethics education in data science programs. Recent news stories about data breaches and algorithmic biases indicate that big data projects raise ethical concerns with the potential to inflict harm on a wide societal or global scale (p. 2).

Davis is asserting that the potential for harm that data science holds is potentially societal or even global in scale.

The impact of data science on humans was a recurring concern among the data scientists I interviewed. As Data Scientist 4 noted, because data science is so focused on coding and numbers, it's easy to lose sight of the human beings affected, whether in data collection, use or model outcomes. This underscores the critical need for ethical education in the field:

> *Data Scientist 4: I would absolutely teach them [ethics]. For a lot of people, data and data science are just coding and numbers, and you forget about the person on the end of it, or the customer, or whoever it is that you are affecting, and there are some really good examples across the world where this has really gone pear-shaped. [Ethical training] is needed to give people an idea of the impact and the power that they have.*

This view aligns with Stoyanovich (2022), who highlights that even though data itself might seem inanimate, data science fundamentally impacts people's lives, because data scientists:

> Process data about people, some of which may be sensitive or proprietary, and help make decisions that are consequential to people's lives and livelihoods (p. 4).

This dissonance between the inanimate nature of data and the human-driven analysis of data in data science has also been referenced by Salz et al. (2018). They assert:

> While data science can bring objectivity to decision making, there is subjectivity within data science modelling in that decisions must be made about which algorithm to use, which data sources to use, whether one data point should be used as a proxy for a missing fact, and how to interpret results (p. 953).

Salz et al. are highlighting the subjective nature of data science practice. In their research paper, they investigate codes of ethical data science practice and review teaching curriculums globally.

Their analysis suggests that no existing code of ethics or ethics framework covers all the key topics that should be included within a data science curriculum. By matching the codes and the literature, and identifying gaps, the authors identified twelve key themes that they believe should be covered within a data science education. The authors cluster these twelve themes into three broad categories: general standards of professional conduct (such as duty to clients, colleagues, and the profession); data-related challenges (including privacy, misuse, and data validity); and model-related challenges (covering personal and group harm, subjective model design, and misuse). They also single out the 'newness of the field' as a distinct, standalone theme. The 'newness' of the data science field presents unique challenges for practitioners and students, as ethical frameworks and regulations often lag behind rapid technological developments, creating situations where existing codes offer little guidance and previously unconsidered ethical dilemmas can emerge (p. 956).

The recent appearance and rapid uptake of generative AI (gen AI) has made the results of data science available to any layperson with access to the internet, an interesting new development in the field of data science which has driven an increase in student awareness about ethics in data science, according to Data Scientist 3, below. Gen AI is a category of artificial intelligence that focuses on creating new data samples based on the patterns learned from existing datasets. Unlike discriminative models, which learn to classify or predict outcomes from input data, generative models aim to understand the underlying distribution of the data and generate new instances that could plausibly belong to that distribution. Working with gen AI can be much like having a conversation. Most laypeople are familiar with the gen AI applications that allow them to create recipes, ask conversational questions, or generate creative text through gen AI applications like ChatGPT or Perplexity. Art and design applications, like DALL-E and Midjourney, can create images based on text prompts, allowing artists and designers to generate visual content. It is now almost impossible to avoid gen AI content, or to avoid having gen AI interacting with your own work. For example, there are versions of Windows and Office – often institutionally selected – that use co-pilot constantly.

Data Scientist 3 notes that the rise of gen AI has sparked greater curiosity and more frequent ethical discussions among university students:

> *Data Scientist 3: Students are really interested about the impact of AI these days, because LLMs[6] are taking over. They care about how that is going to impact them.*

[6] LLM stands for large language model. In the context of gen AI, LLMs are advanced machine learning models trained on vast amounts of text data to understand, generate and manipulate human language. These models can perform various tasks, including text generation, translation, summarising, question-answering and more.

> *There has been increased demand for more discussion of the societal impacts. But I think there is also an awareness that data is more accessible – and data science is more accessible – than ever before. So, it is important that people do it responsibly.*

Data Scientist 3 highlights that the very accessibility of gen AI models, and their subsequent ubiquity, place importance on responsible use. This is a similar point made by Goldsmith and Burton (2017), who argue that students should be trained in ethical reasoning:

> So that they may make ethical design and implementation choices, ethical career decisions, and that their software will be programmed to consider the complexities of acting ethically in the world. (p. 4836)

Researchers from the University of Oxford's Institute for Ethics in AI argue that 'the idea that we start from an ethical blank slate in addressing the challenges and opportunities of this transformative technology is a fallacy' (Ober & Tasioulas, 2024, p. 2). They propose that Aristotle's approach provides a compelling, human-centred framework for AI ethics, as it represents:

> A truly 'human centred' approach to the ethics of AI, one that conceives both human flourishing and human morality as rooted in our nature as human beings whose fulfilment depends on the exercise of capacities for rationality, social engagement, and communication. (p. 2)

The authors place great importance on the development of a system of ethics for data science and AI. They argue that AI holds great promise for enhancing individual and collective flourishing, but in order for the promise to be realised and the pitfalls avoided, the world needs 'a compelling ethical framework to guide our choices about the development and deployment of AI systems' (p. 68). Although they acknowledge that no single philosophical tradition can fully address all the complex ethical challenges posed by AI, they maintain that we do not have to start from the beginning in constructing such a framework. Foundational insights for this task can already be found in Aristotle's ethical philosophy, which remains a landmark in both scientific and philosophical thought.

In summary, both practicing data scientists and the literature asset that ethics is essential in data science, because it impacts on the lives of humans. Ethical considerations must be addressed at every stage of the data science process; no step is exempt from scrutiny or the responsibility to engage in ethical reflection and discussion. Furthermore, unethical behaviour in data science will erode the public's trust in data science as a tool. Given the immense potential of data science to do good in the world, this would be an egregious outcome.

Further, the collection of data has inbuilt ethical considerations, as does the use of that data, the choices made in building any models, and the outcomes and impact of any models that have been built. Throughout, data scientists should make sure their choices are transparent, and that the workings of their model are replicable by anyone else. In summary, researchers and data scientist practitioners align on the importance of ethics in data science. How prepared – or not – do these data scientists feel to grapple with these important ethics? I examine this in Theme 2.

5.2 Theme 2: Ethical challenges encountered in practice

5.2.1 Feeling unprepared to grapple with ethical issues

A key aspect of the data from the interviews concerned the ethical challenges that data scientists encounter in practice, and how underprepared they felt to grapple with these ethical issues. Almost every respondent highlighted how little ethical training they'd had in their degree course.

> *Interviewer: If you do encounter ethical issues in your daily work as a data scientist, how prepared do you feel to grapple with these?*
>
> *Data Scientist 1: I guess the high-level takeaway is woefully unprepared.*

Data Scientist 1 went on to explain that the reason they felt woefully unprepared is for a number of reasons, including that data science is taught as a technical, mathematical and statistical subject and ethical education is not typically taught in subjects of that nature.

> *Data Scientist 1: Traditionally, how we teach data science is very technical. We learn a lot of statistics, a lot of math. You do not think about the societal ecosystem in which you are applying these models.*

Data Scientist 1 is explaining how the technical focus of data science education often neglects broader ethical and societal considerations.

Data Scientist 3 had a similar experience. They explain how, in their university training, ethical issues were only really taken seriously if they presented an interesting mathematical challenge:

> *Data Scientist 3: None of my graduate-level courses in optimisation or math really talked about ethics beyond, for example, if there is a fairness problem that can be modelled as a nice math problem. From an ethics perspective, I do not know how effective that is.*

Data Scientist 3's experience highlights that, during their graduate studies, ethical issues were usually addressed only when they posed an interesting mathematical or technical challenge. As a result, ethics was often treated as an abstract problem detached from real-world application, reinforcing the perception that data science is mainly about mathematics and technical problem-solving.

Stoyanovich (2022) asserts that on the contrary, responsible data science is 'not a purely technical discipline, rather, it is socio-legal-technical' (p. 4). She goes on to explain that when thinking about the responsible design of data science systems, students and practitioners should consider not only its technical components like the data and the model, but first and foremost, how it will be used. She suggests that data scientists ask themselves and their fellow researchers a series of questions such as: what goals is the project being set to achieve? Who are the stakeholders – individuals, groups or organisations – that are impacted (either directly or indirectly) by the model? What are the benefits when the model works well? Who benefits? What are the risks, and the actual or potential harms? These are questions that data science students should be taught to address, over and above learning the technical skills of data science, she argues. Borenstein and Howard (2022) echo this when they assert that 'the view that technology is 'value neutral' hides and obscures the reality that ethical issues are fundamentally embedded in the selection, design, deployment, and use of technology' (p. 63). Utts (2021) terms this a 'mysticism surrounding computers' that may lead to more credibility than is warranted for results of machine learning algorithms (p. 89). This echoes my own belief that recognising that technologies are not value-neutral is crucial for responsible development and use of AI and other advanced technologies, as it encourages continual ethical scrutiny. Challenging the assumption that technology is inherently positive or neutral is necessary to address the broader social impacts and unintended consequences that can arise from technological innovation.

In part, many data scientists feel underprepared for ethical challenges because their undergraduate background, in fields like engineering, economics, science, mathematics and computer science, rarely included substantial or seriously evaluated ethics education. Where ethics was included, it is often minimal and not integrated into the core curriculum. Data Scientist 3 explains why this was a problem for them:

> *Data Scientist 3: I was an engineering undergrad, so I took an engineering ethics course my first year. It was no credit, and it consisted of three lectures, essentially "do not do bad things" in nice language. It was definitely not extensive.*

Data Scientist 2 echoes the experience of not being taught ethics in their studies:

> *Data Scientist 2: I haven't been taught ethics. I did personally engage with that kind of stuff in my student days because I was quite interested in a broader set of issues, but I was never formally taught it.*

The experience of Data Scientist 5 was similar to that of Data Scientist 2, above, studying computer science at undergraduate level, where ethics were not taught. This meant that Data Scientist 5 had to learn on the job about how to deal with ethics and ethical issues:

> *Data Scientist 5: I did not actually study data science. I did computer science and applied maths. I got into data science by working in the field. But it was only when I was working with a certain company that I learned about data privacy and ethics.*

Tanweer et al. (2017) focus on 'learning on the job', highlighting 'the ways that ethics are implicated in the day-to-day work of data science, focusing on instances in which data scientists recognise, grapple with, and conscientiously respond to ethical challenges' (p. 1). They argue that ethical challenges are common in data science and suggest early ethical engagement and balancing priorities as key lessons.

Data Scientist 7 did receive a little training in ethics, but this only came at the end of their degree, and only addressed two ethical issues, those of fairness and bias, and carbon footprints:

> *Data Scientist 7: There were some courses in university that made me actually think about ethics, which was a good thing, but I thought, "Oh gosh, I have been doing this whole degree, and I am now nearly done, and we have to do this. They are talking about carbon footprints and bias, just let me get back to my final project". So yes, there was some [ethics teaching] in there – technically – but I do not think it was very strong.*

This kind of once-off ethics training has been likened to an ineffective 'vaccine' (National Academy of Engineering, 2009; Tractenberg, 2020; Davis, 2020). A preferable approach is to create a culture of ethical practice through repeated and consistent exposure, aligned with students' technical training, in order to produce data scientists better prepared to engage with the challenges to be faced in their future professional lives.

Despite this 'ineffective vaccine' nature of once-off training, Data Scientist 9 suggests that even a half-course in ethics would have been more beneficial than nothing:

> *Data Scientist 9: I have had no formal training [in ethics], and that does feel like a gap. I do not know how to think about things in an ethical framework because no one*

has ever exposed me to those frameworks. It feels like just a half-course in ethics in third year would have really been beneficial.

Data Scientist 9 goes on to explain that they see this lack of ethical education in other data scientists, too, in the course of their work and expresses amazement at the extent to which data scientists do not understand the ethical ramifications of their work.

> *Data Scientist 9: The lack of ethical understanding is shocking. The first thing is to know is that if you are building a machine learning model, it replicates the patterns it sees in the data. For example, there is that great example of Amazon discriminating against programmers because they are female.*[7] *I do not think people even think that far. They do not think about what the data-generating process looks like or whether it is ethical.*

Two other data scientists explained how, at the start of their careers, they felt unprepared to deal with ethical issues but have since discovered that experience is a great teacher. As they've matured and gained more work experience, they've felt better prepared to grapple with the ethical issues that they encounter in their daily work.

> *Data Scientist 1: I think it comes with experience. There might not be Ethics 101 in data science, but it is brought up in some of your training: "Remember, there is responsible AI and how to use it?" Then it comes from the workplace – the ethics of the company that you want to join and that you want to associate yourself with.*
>
> *Data Scientist 2: I would say that early on in my career, it was more difficult because I was not as confident in my judgement. So, at that stage, I was not well prepared to deal with ethical issues, but as one gets more mature, you have much more agency around that.*

In addition to often feeling unprepared, data scientists face ethical issues throughout the data science process, such as handling sensitive data, data labelling, model building, managing outcomes like bias and discrimination, responding to unethical client requests and dealing with lack of diversity in teams. This list of ethical issues, synthesised from the data scientists that I interviewed, echoes the literature, with possibly a more practical bent. The literature emphasises legal compliance, data quality, privacy, fairness, transparency and responsible data use. In what follows, I examine each of these areas in turn.

---

[7] Data Scientist 9 cited Amazon's recruitment AI, which was scrapped after it was found to disadvantage women. Trained on a decade of mostly male CVs, the tool learned to penalise terms like 'women's', reflecting gender bias in the tech industry. See: https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/

5.2.2 The source and nature of the data, and dealing with sensitive data

The source and nature of data raise three key ethical concerns. First, data collection can reflect existing social inequalities. For instance, urban health data tends to be more thorough than rural health data, and poor health data can lead to higher insurance premiums for disadvantaged groups. Historically, health data has also been biased, such as the tendency to collect medication data primarily from men, leaving gaps in knowledge about drug effects on women. Data Scientist 4 outlines a very common data collection bias:

> *Data Scientist 4: There is a big data collection bias that is also a very common one. It is just an everyday thing, where a man is just "Mr", but a woman often has to sign up as either "Miss", "Ms" or "Mrs". I decline every time because I can. What this means is that is we are collecting more information on women than on men. This is just normalised every day.*

Secondly, ethical data collection should involve informed consent. This includes giving consent to one's data being collected, but it is not as trivial as giving consent to the data collection itself. It extends to include informed consent about the potential uses for the modelling and analysis of the data. It's hard for the data scientist to know – unless they collected the data themselves – whether or not fully informed consent was given. As Data Scientist 1 outlines:

> *Data Scientist 1: Has someone given full consent? Do they know that you are doing it, or did they just click "yes" to the cookies, for example?*

Thirdly, the data itself can often be personal and/or sensitive in nature. Data Scientist 1 describes the issues with sensitive data:

> *Data Scientist 1: You are dealing with models that necessarily might use a geolocation of someone, like what area they live in. And maybe you are trying to influence the way in which they buy or the products that they are selling. [This is an ethical issue] for people in different race groups, if you are using that [as a feature of the model], such as race or areas in which people live in, especially in a country like South Africa, where we have a lot of inequality in the country.*

Here, Data Scientist 1 is foregrounding how data that might not necessarily be thought of as sensitive by the original project designers could actually be sensitive. Consider, for example, geolocation. This feels on the face of it, like fairly neutral information. How can an address be sensitive? But when one reflects more deeply, and considers that in South Africa during apartheid, many people of colour were forcibly removed from some locations and resettled in other locations, 'geolocation' can be laden with significance. For example, some geolocations have lower disposable incomes than others.

Data Scientist 5 gives an example of a breach in ethics on behalf of the company collecting data, but that was not picked up by data science students because they did not know it was a problem:

> *Data Scientist 5: I had an example of some companies giving students access to data. They just gave them the whole data set without anonymising it. The students did not see anything wrong with using the data because it was the company that gave them the data. But a student should be aware that, okay, now I should try at least to anonymise the data that I am given.*

In this example, Data Scientist 5 is explaining a common challenge that data scientists face, wherein organisations that collect reams of data on their customers and clients – like banks, for example, or health insurers – do not always consider the ramifications of greater sharing of that data. These organisations might have very strict internal controls over data privacy but unthinkingly hand over data sets to universities for students to use in their research. Students untrained in ethics may – similarly unthinkingly – assume that, since the organisation gave them the data, there is nothing wrong with using it as is.

Data Scientist 3 talks about the importance of privacy as an ethical consideration in the use of personal and/or sensitive data, in an interesting example about the collection and transfer of employees' diary information.

> *Data Scientist 3: Privacy considerations are getting a lot more traction these days. I was at Microsoft Research last summer, and we were working on building this meeting scheduling system. Essentially, you could talk to a chatbot (like an LLM), you say your preferences for a meeting, and it pops up a meeting time. One of the big issues that came up is privacy. A lot of the reasons people do not put information on their calendar is that they do not want their corporation knowing, for example, they need to get up an hour early to pick up their kid. Or, you are hungover, so you need a morning off. Things like that.*

Data Scientist 3 is highlighting the increasing attention to privacy in the context of developing a meeting scheduling assistant. The use of chatbots and AI in scheduling introduces new privacy challenges, as these systems often require access to sensitive personal information. Data Scientist 3 explains that users often avoid entering certain details into digital calendars, illustrating how privacy concerns can directly impact user behaviour and the completeness of data in such systems.

> *Data Scientist 3: One of the issues that we got pushback from our institutional review board was essentially that if you have a chat-based system, and somebody is communicating to schedule a meeting, oftentimes people inadvertently share that kind of information. And then the question is, where is that information being*

stored? Who has access to that information? Can it be used in ways that the user does not intend?

Data Scientist 3 is explaining the risks of unintentional data sharing when using chat-based scheduling tools. This raises critical questions about data storage, access and the potential for misuse, which are all issues central to ethical data management.

> *Data Scientist 3: For example, if you are a big e-commerce company, a customer might agree they can use the customer's data for one thing. Now the e-commerce company now has a database, and they use it to train some other machine learning models. This is a big ethical consideration: how do you handle users' data, especially when you want to use it for reasons that you might not have originally intended, or other than they originally gave consent for?*

Data Scientist 3 expands the conversation to domains like e-commerce, where user data is often reused for analytics or machine learning beyond its original purpose. This underscores the ethical issue of using data in ways not covered by users' initial, informed consent. For example, organisations such as Microsoft collect extensive personal information from emails and calendars; although users consent when purchasing software, it's unlikely they realise their calendar data could be used to train machine learning models. Similarly, Amazon collects extensive data on users, enabling it to predict purchases, estimate income from shipping addresses, and infer height, weight, health, ethnicity, political views and daily routines across its platforms. These examples show how broad data collection often happens without users' awareness, underscoring the crucial need for data privacy and ethical safeguards.

5.2.3 Data labelling and classification

Data science models run off variables, which means that all the data collected needs to be labelled or classified into variable categories. Inherent in this classification process are a number of ethical issues that data scientists have to grapple with in the course of their daily work. The first issue arises in the nature of the classification itself, as it can be tempting to reclassify data. As Data Scientist 6 explains:

> *Data Scientist 6: Some ethical issues relate to how a person identifies. I can give you an example: when we are talking about sex. For example, you see that someone is saying: I am identifying as female. But the data reflects something else.*

Data Scientist 6 notes that a person's identity may not be accurately captured in data collection or study design. Sometimes, if a chosen gender doesn't match other data, a data scientist might change

it during data cleaning. While this isn't necessarily wrong, all changes should be documented so others can replicate the analysis and results.

The second issue arises from the fact that classification has to be done by a human. Depending on the nature of the classification, the impact on the human doing it can be profound. As Data Scientist 7 explains:

> *Data Scientist 7: Many companies use people to provide feedback for tuning models. For example, for LLMs, you often start with a human-labelled solution. We had some interns – high school students who want to be involved in data science – who were assigned a data-labelling task. The data set that happened to need labelling was text descriptions to be classified by type of crime described. The interns were excited, thinking, "This is a noble cause! We can help the hotline prioritise important things to be flagged for immediate review".*
>
> *It seemed like a great idea. The labour is low-cost (they are actually free), and we thought we could have them just label a bunch of data. In this instance, many crime descriptions involved horrific sexual assault, or violence, or situations like, "This person said if I did not do this, I would not get the job, so I did this and now I am pregnant". If we hadn't checked, we could have had these poor innocent eighteen-year-olds who were so excited about data science exposed to extremely distressing material.*

Data Scientist 7 is foregrounding the ethical implications of the impact on the human doing data classification, where the categories of data are problematic, sometimes horrific. This is an ethical issue that has not appeared in the data science literature searches I have done, but has been addressed in literature in other academic fields, for example, psychology and social work, where the vicarious trauma experienced by social workers and child protection professionals has been well-documented, for example Molnar et al. (2020), Chouliara et al. (2009) and Bride (2007).

In this specific example offered by Data Science 7, there is an interesting 'switch' of ethical framing. Usually, when we consider ethical issues in data science, we consider the impact on humans of machine-driven analysis. In this case, however, there is an impact on the human from the need for a human to be involved in the loop. In a similar vein, Data Scientist 7 had two different examples where ethical challenges arise around data labelling. The first example involved the creation of safety filters:

> *Data Scientist 7: I worked with an image generation company. Obviously, you do not want to generate nasty stuff. One very admirable employee said, "I will work on the safety classifier" because the team felt existing filters missed some troubling images. He had to look at some pretty nasty stuff that risked slipping through the existing filters, but they ended up with a better safety filter.*

In the next example, also from Data Scientist 7, the issue with classification was not in the data labelling itself, but rather in what the data labelling would be used for:

> *Data Scientist 7: I left a project because of the downstream application was troubling. It was obviously going to be applied for military targeting, or discrimination in hiring.*

In the example above, Data Scientist 7 is considering the downstream harm of doing data labelling. A number of the data scientists I interviewed mentioned that it is possible to get so caught up in the technical aspects of a data science project that they forget to ask about the outcome and impact of the data science project. Here, the data scientist intentionally thought about the impact down-the-line and, as a result, left the project.

An ethical issue raised by Data Scientist 6 concerns biases in the classification process. As they explain:

> *Data Scientist 6: There is an example of a machine learning model to detect a person's race. But the data that was used was biased because it was trained on white people. Then the model sees a black person, maybe someone from Congo, it says, "Oh, this is a baboon". You know, any black person will feel offended. I would feel offended. I will give another example, for a wedding, say someone in a white dress and a guy in a suit. But if you take an Indian wedding, maybe it will not be recognised as a wedding. If you take an African wedding, where people are just going to sit and pay money, like sitting in a certain format or in a certain way, it will not be recognised as a wedding. But if someone had been taught to try and be inclusive, and know the implications of the bias, or which populations you are going to try and focus on, how they are affected by whatever outcome you have… that would be helpful.*

Data Scientist 6 is expressing concern about the hegemony of culture used in the training of data. Because most of the data that today's machine learning models were originally trained on came from developed nations, the models reflect the cultural norms of those countries. For example, models code images as 'weddings' where they reflect heterosexual norms and feature white people, with a man in a suit and a woman in a white dress. But the reality is that weddings look remarkably different around the world. A machine learning model trained on data from Europe, for example, would be hard-pressed to identify a traditional South African wedding where people (mostly men) sit and negotiate *lobola.*[8] This issue is similar in nature to an issue discussed later, of representation in data science teams – why it is an ethical issue, and why it matters.

[8] Lobola is a traditional practice in various African cultures, particularly among the amaZulu and amaXhosa, where a groom pays a negotiated bride price to the bride's family as a sign of respect and to formalise the marriage.

The next example of ethical issues in data classification was raised by Data Scientist 7. On a more macro scale, this issue concerns global inequality and data classification processes:

> *Data Scientist 7: Global companies pay for human feedback, usually outsourcing to places like Kenya, Nigeria and Southeast Asia where fluent English is available at low cost. Some try to pay fairly, but there is also sweatshop version. I prefer an open-source approach, with volunteers from all over the world with different backgrounds volunteering to label. That seems like a much healthier approach.*

Data Scientist 7 is expressing antipathy for the current global structural inequality, which can lead to low-cost labour in high English-speaking countries being exploited. On the other hand, this does provide jobs in developing economies, and it does allow for the issue raised by Data Scientist 6 above to be tackled. With different people from different countries all around the world involved in data labelling, cultural hegemonic norms around image classification may be slowly eroded, as people from different cultures train machines on image recognition.

5.2.4 The building of the model

Ethical challenges when building the model cluster into two broad areas. First, there are the questions about how the data is used to build the model and which variables are used. As Data Scientist 1 explains:

> *Data Scientist 1: What features are you putting into your model? Should you use that? Is it fair to use it? But there is no "responsible AI" course, which I think is really important. I think we are living in this ever-changing landscape, and it is very easy to cross that line and manipulate data points.*

Data Scientist 1 is talking about the ethical questions that arise when deciding which data points to use to build the model. These choices are not always clear-cut and can lead to manipulation by bad actors. For example, a bad actor might choose to use only that data in a model that reinforced outcomes they were looking for, for example, for commercial gain.

The second ethical challenge involved in the building of the model lies in the difficulty of communicating how the technical model works to non-technical people, who are often the ones who need to use the outputs of the model. I've dubbed this the ethical issue of the 'black box' of the model, following the language used by some of the interviewees. Data Scientist 2 describes the challenge:

> *Data Scientist 2: An interesting [ethical] case is due to knowledge differentials. People want to use data science, but no one really understands it properly. And no one really understands the modelling result, the modelling procedures, or the stats behind it. They don't understand the statistical language. So as a data scientist, you can basically tell people, "This thing's magic; it's going to solve this problem for you". And people will say, "All right, that sounds good to me. Go ahead".*
>
> *There's asymmetry in knowledge that's quite difficult to get around, and it's obviously bad actors who completely abuse that. You could completely pull the wool over people's eyes. Even if you're a good actor and you want to be perfectly aligned, you have to translate that language. And that's actually quite a tricky thing to do.*

This was an ongoing ethical issue that Data Scientist 2 faced in their work, centred around a knowledge differential and different levels of rigour:

> *Data Scientist 2: A key question is: to what extent are you bridging the gap between people are very eager to use the technology and want to believe that it can do anything, and the perspective of the technician, more aware of what it can and can't do, and where it might fail even though it looks like it's doing the right thing?*
>
> *There is a mismatch between people who don't think statistically but think in terms of stories. It can be tempting because the most compelling thing to communicate is to create a story around model outcome. You can always cherry-pick good and bad examples.*

The above explanation shows how Data Scientist 2 worries constantly about the asymmetry of information and the ethical dilemma that arises as a result. They feel responsible not just for building accurate models, but also for guiding others toward more informed and ethical use of those results.

This challenge was also raised by Data Scientist 9, who similarly worries about the black box of the model:

> *Data Scientist 9: There is an ethical question around truly understanding what your models are doing, especially with modern machine learning and AI, which can be black boxes. For example, you may not know why a neural network is encouraging someone to click an advert, or not, and you cannot explain to the client or the regulator how that is working. So, there is an ethical call to make: how much of a black box do you make your models? And how much explainability do you want to have, particularly if you are using these models to encourage different types of behaviours?*

Data Scientist 9 is expressing a similar concern to that raised by Data Scientist 2. Both data scientists know how their models work and what they can be used for, but they both find it difficult to explain to their non-technical clients and worry about misuse of the models as a result.

Data Scientist 3 identified an ethical issue arising in the use of a model in the hands of someone who might not necessarily understand all the statistics behind the model, and misuse it not for nefarious reasons, but rather out of ignorance:

> *Data Scientist 3: I work with practitioners who apply machine learning models, for example, doctors who use ML models to make clinical predictions. A frequent issue is misinterpreting causality models. For instance, consider a decision tree model predicting lung cancer. Often, the machine learning model has learned correlations. So, if you are a smoker, well, that relationship can be causal. But it can also pick up non-causal relationships, which practitioners may mistakenly treat as causal. This is where you get a lot of misuse, where the prescriptions that you get from this machine learning model are incorrect, like maybe "do not go for runs a lot because it uses your lungs more and it leads to lung cancer". The key challenge is teaching people to use these models responsibly, so they don't assume every correlation is causal.*

In this manner, worries Data Scientist 3, a model might be misused not by a bad actor but simply by someone who doesn't understand the statistics properly. An ethical issue therefore arises, believes Data Scientist 3, to responsibly educate the people using the models so that they do not misuse the models from sheer ignorance.

5.2.5 Outcomes, use and impact of the model

Almost every data scientist I interviewed expressed concern about the ethical implications that arise when data science models are used. They worry primarily about three things: the fact that the model they're building can impact human lives, unintended consequences of the model, and disparate impact.

Almost all of the data scientists I interviewed expressed a poignant concern that the work they're engaged with on a daily basis can have an impact on the lives of human beings, and that sometimes, this impact might be beyond what they can see in the moment. Data Scientist 2 and Data Scientist 1 expressed their concerns:

> *Data Scientist 2: I'm building a model that can influence outcomes for real people, and in some sense, that's an ethical issue. If you're in the finance space, you might build a model that profiles high-risk clients, and then you are actually doing something that changes your platform's behaviour, for example, maybe it puts forward an intervention for this client.*

> *Data Scientist 1: Is it fair to use something as a predictor of whether someone is going to buy my product, or of how I can influence the way you buy my product?*

Both Data Scientist 2 and Data Scientist 1 consistently refer to their worry that the models they are building are used to influence human beings in certain ways. They worry about the 'fairness' of this, the underlying issue of preventing people from being autonomous, and the impact on changes in behaviour for the human beings whose data they are using and for whom the model is predicting outcomes.

Data Scientist 5 highlighted the interesting point that doctors and lawyers, whose work similarly impacts on humans, have ethical guidelines, but data scientists do not:

> *Data Scientist 5: Doctors and lawyers have been operating in society for a long time, and they have ethical guidelines. Why can we not have ethical guidelines? Because we are dealing with problems that really affect people. I think we do need ethical guidelines.*

Data Scientist 5 suggests that the fact that other professionals whose work impacts on human beings are taught and held accountable to ethical guidelines suggests that data scientists, too, should be taught and held accountable to ethical guidelines.

Data Scientist 9 worries about the consequences of data science models on vulnerable people:

> *Data Scientist 9: Another ethical bit is understanding the data and what you are doing with it, and how you use it. You may be doing credit risk analysis, and maybe it is linked to gambling, but there are certainly some types of people who are more susceptible to nudges, and you can find that in the data. You can easily push people into circumstances that you might feel uncomfortable with.*

Data Scientist 9 is therefore also worried about impacts on human beings, but in this specific example, about the impacts on vulnerable human beings, for example those who might have addictive tendencies.

Data Scientist 2 expressed a pragmatism about the very nature of data science itself, being inherently imperfect but sometimes better than humans.

> *Data Scientist 2: Models dictate outcomes, real-world outcomes for people that may or may not be perfect, that may make errors. This model has to do something in the real world. No model is perfect. The questions to ask are: is this model systemically biased? If it's not biased, is it interpretable enough? To the extent that you're automating something, what's the trade-off you make? You ultimately are making peace with the non-perfect results. What's the harm that's done by that? How do you mitigate it?*

Data Scientist 2 is foregrounding their belief that while models aren't necessarily perfect, they're also not necessarily leading to a worse outcome than manual processes involving humans might lead to. This is a logical argument with pragmatic value, but I feel it ignores the fact that many people accord a reverence to data science that they might not accord to manual processes with human decision-making. In other words, there is a tendency to ascribe rationality to statistics and mathematics that I have demonstrated is not, in fact, the case.

Data scientists are united in their concern about the uses and impacts of their models, and the following example helps illuminate how they can differ as to the reasons. Data Scientist 4 shared an example of a company who were asking for machine learning to be applied to advertisements for cars and asking for the actors to be classified along demographics, including age, gender and race. Data Scientist 4 thought this was admirable, in that this could lead to marginalised groups feeling more comfortable with advertising in which they were represented. But their colleague refused to do the data classification, on the grounds that if the machine learning model showed that adverts with young, white, attractive female actors sold more cars, the model might be used in future to support hiring discrimination against actors of certain demographics.

> *Data Scientist 4: So, for me, I felt using race was an objective for inclusion so that people of colour in a certain area will receive ads that make them feel included. In my mind, I was thinking of the end outcome of the model being a more inclusive experience for people, whereas my colleague said, "I will not use race classification of variables, in any model, ever." I thought of the race variable as a positive to get people what they need, versus "I should never use that". It was a very interesting example of how people have different interpretations. And I guess therein lies the danger. I might have had good intentions, but in many examples, people may not.*

This example illuminates a critical ethical issue: Both data scientists believed they were 'being ethical'. Indeed, it is hard to argue that either of them is wrong, as both have admirable intentions.[9] However, using their own personal ethics as decision-making criteria leads them to make diametrically different choices.

Exploring this example gives rise to more questions. Given that these were two data scientists operating from similar backgrounds and in the same country, how are data scientists across the world, with different backgrounds and in different countries, making decisions? It must surely be the case that there is little alignment in the decisions and choices across countries and cultures. Consider, for example, data scientists raised in a traditional Western culture with a focus on

[9] Unless you're a consequentialist, where intentions are irrelevant to the evaluation of an action.

individual rights and freedoms, versus those raised in traditional Eastern culture with a focus on community cohesion and wellbeing. These data scientists would make different decisions at each step of the data science process.

Most of the data scientists whom I interviewed worried about unintended consequences. They expressed concern that they knew the stated aim of the models they were building, but it is inevitably hard to know the end use to which the models could be put. Data Scientist 7 expresses this concern:

> *Data Scientist 7: If you have more accurate facial recognition that allows you to catch criminals and stop them from performing more harm, I can see the arguments there. But whenever you start enumerating the potential use cases, you might say, "All of these use cases just sound like things that I am not too happy with existing."*

Data Scientist 7 is using the example of facial recognition to illustrate their point that you might build a model for one purpose. Still, it can end up having entirely different use cases and purposes. Facial recognition software is a tool that was originally intended to help with community safety by screening known criminals. However, it also has applications for societal control and other nefarious uses. So, Data Scientist 7 was outlining how a system purportedly for recognition of criminals could end up being used for social control.

Data Scientist 9 highlights how unintended consequences can stretch from the initial data collection to the final use of the model:

> *Data Scientist 9: The issue ranges from: did the person who gave the data give informed consent for the data collection, and informed consent for how it is being used? You may be weaponising the data, too. For example, figuring out that certain personality traits are more susceptible to certain types of things, and then serving adverts for gambling back to the person. There are all these things that we just do not think about, and which cannot be policed.*

Data Scientist 9 explains how the original model can promote products to people whose data variables show them vulnerable to certain types of advertising.

When it comes to disparate impact, Data Scientist 3 summarises the ethical issues that arise when different groups are impacted differently by data science models:

> *Data Scientist 3: A key ethical issue is how machine learning systems affect different demographic groups. One of the first projects I did at [Ivy League university] was with [a big commercial bank]. They were trying to build a loan prediction system to help a human loan officer decide whether or not to extend a loan.*

You have heterogeneous populations in your dataset. So, let's say you have white applicants, around 80% of the people who have applied for these loans, and black applicants, around 20%. And if you are using one model to fit both data sets, you will pick the one with more data points, because usually, how we train machine learning models is just as accurate as we can on the whole set. And the problem is that when you start deploying the model, each of these different groups has different error rates. So, for instance, they found that they had a higher false negative rate for black applicants. This means black applicants were unfairly denied loans more frequently than white applicants, which is disparate impact, which is something that banks try to avoid because extending loans has a lot of societal impact.

In this example, Data Scientist 3 shows how a model originally built to help a bank make loans discriminated against a particular demographic group – in this case, black applicants – because of patterns in the historical data used to train the model. This is what O'Neill refers to as 'math destruction' in her book – how data science can perpetuate vicious cycles owing to the historical data and the algorithms trained on that data.

5.2.6 Unethical requests from the client

Only two of the interviewees mentioned unethical requests from clients for which they were doing modelling. Data Scientist 7 explains:

Data Scientist 7: We were doing a research project on brain activity and imaging, called dynamic causal modelling. They didn't know much about coding because they were in the neuro-psych department. But the correlation they were looking for, between the different brain regions, was just not there. And [the client] said, "No, I am pretty sure there should be a thing there. Can you just re-run it a few times? Just keep re-running it until we get a statistically significant result, because I am sure it should be there". I was just, "Um, no".

In this example, Data Scientist 7 is relying on their code of ethics to reject an unethical request from their client to just 'keep re-running the model' until they get a statistically significant result.

Data Scientist 9 also had an example of ethical decisions needing to be made about the client:

Data Scientist 9: Some clients are in the gambling space, or are short-term predatory lenders, for example. They have got lots of data. There have certainly been several times when we had potential projects to work on those things. So, that is an ethical question: What projects do you work on? Where do you think you should be working? And you have to take some ethical stand on that.

In this example, Data Scientist 9 is not subject to a statistically unethical request but needs to make decisions or 'take some ethical stand' regarding the industry's nature and what the model would be

used for. The examples they give are gambling and short-term predatory lending – both industries with poor reputations in a country like South Africa with high levels of poverty and unemployment.

5.2.7 Lack of diversity and representation in data science teams

Several data scientists interviewed expressed concern that teams working on data science models often do not reflect the demographics of the people on whose data they're working. Data Scientist 4 outlines:

> *Data Scientist 4: The benefit of having diversity in teams means that many different perspectives about different issues come to the fore. Of course, we are middle-class white educated people. What do we know, really? So, I guess that is the trouble: It can be so subjective, and you never know the outcomes down the line that might result from it.*

Data Scientist 8 explains that lack of representation is true on a global scale, as well as within countries:

> *Data Scientist 8: Globally in data science, there is no interest in Africa; it is predominantly North America and Europe. So [the emerging world] is a population group that is underrepresented, not represented in anybody's thinking. Underrepresented is a euphemism, by the way.*

Data Scientist 9 offers a similar view:

> *Data Scientist 9: This is why this diversity in data science matters. For all white males building these models, we have no idea of the lived experience or the issues that people not from our background come from or deal with.*

In these examples, Data Scientists 4, 8, and 9 refer to the number of decisions and choices made at every stage of the data science pipeline and how these decisions and choices are intimately linked with the data scientist's background and life experience. This means that the choices can be subjective and negatively affect people from different backgrounds.

Diversity is not just about level of education, age, gender or race. It's also about how different people think and interact with the world. According to Data Scientist 9:

> *Data Scientist 9: I am banging the DEI drum, which is un-PC at the moment. Companies, based mostly in the Western half of the US, started by neuro-divergent men, mostly without these checks and balances, and their incentives are completely different for the people who are using the apps, or the games, or the data science models.*

Data Scientist 9 has extended representation beyond what is visible (age, gender, race) to what is invisible – ways of thinking, ways of learning and ways of responding to incentives. The point that they are making is the same. Without representation, people who are different from typical data scientists may not have their interests well represented in data science models and therefore may be negatively impacted by the outcomes of those models.

The data within this theme provides powerful statements about how data scientists encounter ethical challenges in their daily practice. In the interviews, many data scientists expressed how underprepared ('woefully underprepared') they felt to grapple with these issues and how non-existent or insufficient their ethical training during their data science programme was. Some described the steps they have personally taken to inform themselves about ethical issues in practice. All of the data scientists were able to describe ethical issues that they had encountered and detail how they had dealt with these issues. A few had broader suggestions for the data science field: about incorporating ethical training into data science programmes, about better representation in data science teams, and about trying to address global inequalities in data science. There was an intriguing difference between the level of agency each data scientist felt about improving their knowledge of ethical issues and acting on this. Some data scientists took this squarely upon themselves, and others seemed to feel it was a problem that organisations or legislations should solve. I turn to this theme next in Theme 3: The custodian of ethics.

5.3 Theme 3: The custodian of ethics

As I demonstrated in Theme 2, a theme that emerged in the interviews was who the custodian of ethics is. This theme arose as a direct result of how woefully under-prepared most of the data scientists that I interviewed felt to grapple with the ethical issues they faced daily in their work. In addition, the need for a custodian of data science ethics arises because although there is emerging legislation around ethics in data science, it is new and inadequate. It is entirely possible that all existing legislation is fully complied with, but the data science model and its uses are unethical. As Data Scientist 7 put it:

> *Data Scientist 7: "I feel like there is a larger decision boundary. You can fiddle around with questions like, "Is this technically against copyright?", or you can take one step back and ask, "Is this bad or not?" There is a difference between "I think this is technically fair use", but if you take a step back, you are actually ripping off an artist's style and making money from that. What's needed is taking a step back and thinking ethically. You need to turn on the part of your brain that is empathetic and that thinks of consequences, and ask, "Are we being the baddies?"*

Data Scientist 7 is explaining the gap between the letter and the spirit of legislation that governs data science. Although such legislation is in its infancy, there are still laws around copyright, data privacy and so forth that are globally applicable – but even so, such legislation can be complied with by the letter but still violate fundamental principles of ethics, such as copyright ownership of a certain style by an artist.

Data Scientist 3 also expressed frustration with the gap between legislation and operationalising the legislation. As they put it:

> *Data Scientist 3: I see a gap between the rules and principles that people want to be applied. It is really hard to operationalise a lot of these. For example, in the EU, there is a "right to explanation" when you are deploying models, but it is not really clear what the "right to explanation" means. Or "storing data sensitively", what does that really mean? Am I allowed to take it on my laptop and run a model on it? And if so, how do I ensure that no PII[10] is being leaked into these models? As a data scientist, I do not really know how to bridge that gap. I am not a lawyer. But when there is that gap, I am always going to err on the side that is easiest for me to do, rightly or wrongly. So, if there were guidelines or rules, I would be happy to follow them. But with ambiguity, I am usually going to resolve it with the path of least resistance.*

Data Scientist 3 finds the gap between legislation and the realities of a data scientist's work frustrating. Without legal training, they find it hard to interpret vague requirements, such as what 'storing data sensitively' really means, or how to apply these requirements in practice. They admit they don't use a personal ethical framework but instead take the 'path of least resistance'. While this honesty is striking, it is also troubling; if others share this approach, vague legislation combined with minimal effort could lead to weak ethical standards and poor practices.

5.3.1 The organisation as the custodian of ethics

Most of the data scientists felt that the organisation where they worked ought to be the custodian of ethics. In other words, the organisation should put in place rules, safeguards and decisions. In a way, these data scientists thereby 'outsourced' the custodian of ethics role and moved it away from their own accountability to that of the organisations they work for. As Data Scientist 2 describes it:

> *Data Scientist 2: You ultimately rely on organisational processes that are larger than yourself to provide guardrails for these kinds of things. To the extent that those guardrails aren't there, these [ethical] problems are probably not going to get satisfactorily solved because ultimately our capacity is limited. You've probably got*

[10] PII is an acronym widely used in data science and AI. It stands for Personally Identifiable Information.

> *multiple projects on the go, and what ends up happening is you give a best effort, given what you know, given your capacity, you'll do a best effort to do something. Speaking honestly, that best effort is limited.*

This echoes the point made by Data Scientist 3 about taking the 'path of least resistance'. Given that time, resources, energy and effort are limited, if data scientists the world over are making a limited best effort or taking the easiest path, the result is unlikely to be satisfactory from an ethical standards perspective.

Data Scientist 2 goes on to explain further what they mean:

> *Data Scientist 2: The sweet spot is probably for an organisation to have its own internal standards and processes that really help you to do your job well and help to give good evaluation criteria for these [ethical] things. If you try and go higher level than that, I think you get kind of clunky and into GDPR territory, which does a bunch of good things, but also, man, after the legal back and forth five years after the thing starts being contemplated, technology has moved a lot. There's a list of ills as well as a list of like pros from it. So, if your organisation can nail their own processes, I think that's the biggest thing. But it's a tricky thing to get correct.*

Data Scientist 2 highlights that legislation moves far more slowly than technology, and as a result, most legislation lags developments in technology and argues that organisational standards are vital for effective ethical practice, as legislation like GDPR is often slow-moving and can become outdated by the time it is implemented. This lag means new technologies, such as gen AI, lack governance when they first emerge. Rapid developments in gen AI, even prompting calls for a moratorium in 2023, demonstrate how legislation frequently trails behind technological advances, leaving gaps in oversight and regulation.[11]

It is intriguing that Data Scientists – typically intelligent people with graduate or postgraduate qualifications – feel a lack of agency around legislation and processes. One might expect individuals with advanced skills to feel like they have significant influence over decision-making, especially in areas directly related to their expertise. The contrast between their technical authority and their limited agency highlights the complex interplay between organisational power, compliance requirements and the sometimes-subjective nature of ethical and regulatory decisions in data science. This tension raises important questions about how expertise is valued and the extent to

[11] The six-month moratorium on AI development was proposed in March 2023 by over 1,000 signatories, who argued that it would allow time to establish necessary safety protocols and ethical guidelines to mitigate potential risks associated with advanced AI technologies. The moratorium called for a halt to the training of any AI systems more powerful than GPT-4. As of the end of 2024, this moratorium has not been enacted.

which organisational priorities can override even well-founded professional judgements. Data Scientist 2 goes on to express essentially how powerless they feel in the face of the more powerful organisation:

> *Data Scientist 2: Ultimately, the call is made at an organisational level. You can have your opinion on something, but it feels a little bit above your pay grade. I rely on organisational processes to help determine what's good enough and what is needed from a compliance perspective.*
>
> *An organisation can choose, to some extent, their internal political orientation and can say, look, this is what we've chosen, this is our sense of ethics, and to the extent that you're working here, we should be aligned on this.*

Data Scientist 2 appears to take comfort in the idea that the organisation, rather than they personally, should act as the custodian of ethics. They show little sense of personal agency in defining ethical boundaries. This contrasts with medical ethics, where professionals are expected to uphold ethical standards individually, not rely solely on their institutions.

Data Scientist 5, too, sees ethical boundaries as the organisation's responsibility, both at the client organisation when handing over data, and at their own organisation, which in the case of Data Scientist 5 is a university. Data Scientist 5 describes it like this:

> *Data Scientist 5: I think big companies should be training their people and have internal people who investigate this data and say, "Is it safe to give this data to a student?" Even though the student signs an agreement to say, "I am not going to share this data with anyone else", still… that data is still floating around. And if someone like me, who is an examiner, comes across this, maybe I run into a family member.*

Data Scientist 5, who works in a small country with relatively few organisations and few wealthy people, was concerned that they would come across personal information relating to a family member or friend, and that this potential occurrence was not mitigated sufficiently by the organisation handing information over to the university for analysis. At no stage did Data Scientist 5 express the view that they themselves had personal responsibility to raise this potential privacy violation, or to change internal organisational processes themselves.

Data Scientist 4 similarly believes that the organisation should be accountable for ethics, but they have never actually seen these ethical considerations built into workflows at their organisation:

> *Data Scientist 4: In all the places I have worked, I have never seen [ethics] built into a workflow. And so, there is a difference between education and implementation.*

Data Scientist 4 also believes organisations should be accountable for ethics but says they have never seen these principles built into workflows. While their organisation offered ethics training, none of its insights were incorporated into the data science project processes. This 'half-baked' approach delivers education without implementing the ethical principles in practice.

Data Scientist 3 works at a university and oversees academic data science research projects as well as data science projects for industry. When discussing the role of ethics in academic data science projects, Data Scientist 3 sees the custodial role being played by the university at which they work, but does not think that this custodial role is treated seriously beyond being a tick-box exercise:

> *Data Scientist 3: It is often just a box you check at the end. So, you write a paragraph, and as a reviewer, I always look through it, but I would say the general reviewer pool does not. There are not a lot of incentives to take that super seriously.*

Similar to the organisation where Data Scientist 4 works, ethical issues are raised, but the actual follow-through, adherence and building of ethical standards and principles into processes and incentives are absent.

5.3.2 The self as the custodian of ethics

In contrast, two of the data scientists thought of themselves as the custodians of ethics and did whatever they could to make up for gaps in their own ethical training. For example, Data Scientist 7 was exposed to excellent ethical decision-making in the organisations where they worked early on in their career, and from then on, took active measures to make sure that they were well-versed in ethical thinking and developments in data science and AI globally:

> *Data Scientist 7: From my first job onwards, a lot of my colleagues were exceedingly ethical, and perhaps atypically so. My first boss was amazing and doing incredible work in ecology and low-resource areas. I could have said, "Hey, I am a little bit uncomfortable about X," and she would say, "Okay, cool, burn the client". Part of it is privilege, being able to refuse things and not having big financial things hanging over me, like, "If our growth numbers do not look good, our plant is going to be downsized". I have never had to worry about that. Having that switch on in your brain to be self-aware, I think is the main trick.*

Data Scientist 7 shows strong personal agency in ethical decision-making, shaped by early exposure to highly ethical colleagues and a supportive boss willing to drop clients over concerns. They attribute part of their ability to act ethically to privilege, not having financial pressures like job security. Data Scientist 7 refers to an 'ethical switch', which is an awareness of self, of one's own ethical awareness and choices.

In a similar manner to Data Scientist 7, Data Scientist 9 also took accountability on themselves for ethical awareness, and was also fortunate to have work experience in that area that conscientised them to ethics in data science:

> *Data Scientist 9: For my sins, obviously, I come from an academic background. When I was at [university in South Africa], I sat on the Faculty Ethics Committee. So, I have run experiments in real life in academic environments, on which I have had to go to different ethics committees. So, obviously, I have thought about these things, and I thought about what is ethical or not. I think that was a really good preparation. In teaching, I had to teach a lot of impact evaluation, randomised control trials, that type of thing, so I did teach on some ethical aspects too.*

Data Scientist 9 explains how their previous experience taught them to think about ethical issues, and how the academic context gave them strong preparation for thinking about ethical challenges when they left the academic setting. Further, Data Scientist 9 went on to actually embed ethical teaching into their overall lecturing – something that many lecturers do not do, as Theme 1 and Theme 2 detailed.

Data Scientist 7 also saw themselves as a custodian of ethics:

> *Data Scientist 7: I like to think of myself as someone who has a really high moral compass, and I would say that I often tend to go on the side of caution, and that is just me. But not everyone is like me. I think there has to be some sort of accountability. People cannot say, "I did not learn that, so how am I supposed to know it is not right?"*

Data Scientist 7 is showing that they have reflected about themselves, and their attitudes and practices with regard to ethical challenges. The result of this reflection has been a sense of agency and a sense of themselves having 'a really high moral compass'. Furthermore, this reflection has led Data Scientist 7 to critique an attitude of passivity in the face of ethical challenges in daily working life by data scientists. Data Scientist is not letting other data scientists off the hook for making good ethical decisions merely because they were not specifically taught ethics as part of their technical training.

5.3.3 Blended custody of ethics

Although Data Scientist 7 came down quite strongly on personal accountability for ethical choices, they also saw the organisation as playing an important role in ethical practices in the workplace. In this way, they expressed a view that sees accountability as blended. As Data Scientist 7 explained:

> *Data Scientist 7: There are some companies that might push that limit and say, "Well, it is okay to do this," and there are other companies that say, "Let us use a moral compass. Where do we stand as a company?" I have luckily enough found myself in organisations that will always steer on the side of: rather not do it, in case that is not the right thing to do.*

In this quote, Data Scientist counts themselves as 'lucky' to have been in organisations that have made ethical choices; but in the same breath, they use the first person to describe their views on ethical choices. This implies that they believe both the organisation and the data scientist have roles to play in terms of ethical decision-making.

The data within this theme provides powerful statements about ethical accountability and differing views as to where this ethical accountability sits. Overall, the data scientists that I interviewed showed a continuum of beliefs about where the accountability for ethical decision-making lies. For those data scientists who had a strong sense of agency, they believed in educating themselves and in making ethical decisions that they were comfortable with. At the other end of the continuum were data scientists who did not seem to have a high sense of their own agency, and who believed that the custodian of ethics should be their organisation.

The distinction between the individual as custodian of ethics and the organisation as custodian of ethics highlights the interplay between concepts in liberal and critical theory. From a liberal theory perspective, ethical responsibility is rooted in the autonomy and moral reasoning of individuals, and emphasises personal integrity, agency and the capacity to make principled decisions. In contrast, critical theory draws attention to broader institutional and structural forces that shape ethical practices, highlighting how organisations establish the norms, policies and power dynamics that enable or constrain ethical action. By considering both perspectives, it becomes clear that ethical data science requires cultivating individual virtue and judgement as well as transforming institutional cultures and structures to support justice and accountability.

In the context of data science, ethical practices and outcomes are shaped by the actions of individual practitioners as well as by the broader institutional environments in which they operate. Institutions, whether they are universities, corporations or public agencies, play a pivotal role as custodians of ethics by establishing the norms, policies and cultures that guide decision-making. Organisational structures can either empower data scientists to act ethically or, conversely, create pressures and constraints that make it difficult to prioritise ethical considerations over technical or commercial objectives. As highlighted by participants in this study, the presence or absence of clear institutional

guidelines, leadership commitment to ethical values and mechanisms for accountability significantly influence how ethical dilemmas are navigated in practice. Recognising institutional responsibility means acknowledging that ethical conduct in data science is not solely a matter of personal virtue or compliance with professional codes but is also contingent on the collective commitments and systemic supports provided by organisations. Therefore, creating a culture of ethical awareness and responsibility at the institutional level is essential for ensuring that data science serves the broader interests of justice and social good.

For both groups of data scientists: those who took accountability for ethical choices upon themselves and those who 'outsourced' accountability for ethical choices to their organisation, ethical education during data science programmes was insufficient to non-existent. It is to the role of ethical education that I turn to next, in Theme 4: The role of ethical education, including suggestions for ways of teaching data science ethics.

5.4 Theme 4: The role of ethical education (including suggestions for ways of teaching data science ethics)

The data scientists that I interviewed had much to say about the role of ethical education for data scientists. They were aligned in their criticism of the current dearth of ethical education for data scientists. They all insisted that data scientists should receive ethical education, and that ethical education should not be a once-off session (or 'vaccine') but rather, it should be embedded into and throughout technical training. There was a wide range of suggestions about how best to teach ethics. These suggestions included: through case studies, learning by observation and/or experience, learning about practical ethical issues, and learning from the humanities about how ethical education is best done. I'll examine each of these sub-themes in turn.

5.4.1 Criticism of the current dearth of ethical education for data scientists

As Theme 1 detailed, almost all of the data scientists I interviewed felt unprepared to grapple with the ethical issues that they encountered in their daily work as data scientists. Many of them pointed to their own lack of formal training as a key reason for this unpreparedness. They also highlighted that this lack of ethical education persists in current technical data science programmes, as well as in all the disciplines that serve as pathways into data science, such as statistics, computer science, econometrics and engineering. As Data Scientist 3 describes:

> *Data Scientist 3: The people developing these courses are not ethicists, right? They are computer scientists who have maybe worked on a cool problem that has social impact. We do not have a data science school here at [Ivy League university]. We have an operations research department, a computer science department, a stats department. So, who takes responsibility for an ethics course?*

Data Scientist 3 explains that ethical education is often missing from data science programs because courses are usually designed by technical experts, such as computer scientists, rather than ethicists. As a result, ethics may not be prioritised or fully integrated. At [Ivy League university], where there is no dedicated data science school, responsibility for ethics education is unclear and falls between departments like operations research, computer science, and statistics.

Data Scientist 6 explained how their experience of ethical education was mostly based on legislation, and that this was conceptually inaccessible owing to the use of jargon and the fact that English was their fourth language. As they explain:

> *Data Scientist 6: There are a lot of policies. But oftentimes, you are just agreeing to things you do not even understand because it is written in a certain language you do not understand. Not like you do not understand English, but the jargon they use is for people who were legal experts at some point. So, for you to really fully understand it, it is hard.*

Data Scientist 6 is highlighting how training on legislation is insufficient for empowering data scientists to make ethical choices in their daily working lives as data scientists. For Data Scientist 6, for whom English is their fourth language, training on legislation can also be inaccessible – not just because they do not speak English as fluently as an English home language speaker – but also because legislation is written in language that makes sense to lawyers but can be intimidating to people who are not trained as lawyers.

Data Scientist 8 has a similar belief in the current level of ethical education being unsatisfactory:

> *Data Scientist 8: So now they have an ethics course, but it is totally separate from all the other courses. It is just one of these mandatory courses that you need to do, and hate doing, and you write a couple of essays, and that's it.*

Data Scientist 8 is expressing a common frustration: ethics courses are standalone courses and tend to cause students more resentment than anything else. Students comply because it is mandatory, but they tend to begrudge the time away from other studies, and as a result, they do not engage to the fullest extent possible with the material.

Data Scientist 7 also points to the fact that standalone, once-off content is insufficient and unsatisfactory for getting 'people thinking' about ethics. They believe that ethical education should be woven throughout any data science curriculum:

> *Data Scientist 7: It is all well and good to have a final chapter in the book called ethics, but that is often going to be skipped. Instead, you need examples that interweave through any content that you are preparing. You want to bring up examples to get people thinking.*

Similarly, Data Scientist 8 believes that ethical education needs to be meaningfully incorporated:

> *Data Scientist 8: The fact that we need [ethical education for data scientists] is a no-brainer. How you actually incorporate it in a meaningful way is important. You need to look at the STEM programmes, like for medical doctors, for instance, that incorporate an ethics component.*

Data Scientist 2 expresses a concern about the fact that ethics can differ considerably across countries and cultures and worries about the ethical education for data scientists potentially imposing ethical views from one cultural perspective onto other cultures. They explain it in this way:

> *Data Scientist 2: My political bias is libertarian, so I really don't like the idea of someone with a specific political programme trying to get me on that same page. Basically, ethics just seems to me such a fraught subject. People have very different ethical views.*

Here, Data Scientist 2 highlights how personal political beliefs can shape one's comfort with ethical instruction. They express discomfort with the idea of imposing a single ethical or political perspective, which is a real risk when designing ethics curricula. The diversity of ethical views among both instructors and students makes it difficult to create a universally accepted approach to teaching ethics in data science.

> *Data Scientist 2: Maybe this would be a no-brainer for some people who'd say, "Well, I know what's ethical and what's not" and just educate people along those lines. But in data science circles and in broad corporate circles, the dominant ethics are largely corporate and of an American political valence.*

Data Scientist 2 cautions that what is considered 'ethical' is often shaped by dominant corporate or national values. This can lead to a narrow or biased perspective being presented as universal, which may not be appropriate or fair in a global or multicultural context.

> *So, I really don't like the idea of making everyone conform to the same ethical programme. And I do think there are some tricky issues around. There are a lot of things that are contested, and it's good to do them justice. But it's not always simple.*

Data Scientist 2 argues against enforcing a single ethical standard, noting that many ethical issues in data science are contested and complex. Recognising and respecting this complexity is crucial for fair and effective ethics education:

> *What if I say, "Well, models should not display systemic bias in the outcomes that they predict"? That feels okay, but people can define that differently as well, "What is this thing, bias? What is good enough? Must everything be perfect parity?" Because some people would define it like that, and then that's probably just a no-go definition.*

Data Scientist 2 uses the example of bias in models to illustrate how even seemingly straightforward ethical principles can become complicated in practice. For example, most data scientists would agree that models should not be biased. But bias is not binary: there is significant debate about what constitutes bias, how much is acceptable and what standards should be applied, making it challenging to teach or enforce a single ethical framework.

Intriguingly, the difference between cultural values across the world and the problems this poses for potential global ethical education for data scientists seldom arose in the literature. This is possibly because most of the literature on ethical education and data science originates from researchers in developed economies such as North America and Europe. There is little to no literature from Africa, Latin America, Asia or other developing regions. This is what Data Scientist 2 is referring to when they say, 'the dominant ethics are largely corporate and of an American political valence'.

Data Scientist 7 emphasises that, for them, the most important outcome of ethical education for data scientists is the 'switching on' of ethical thinking. From there, everything else follows:

> *Data Scientist 7: Education is important in helping people learn how to follow these ethical reasoning steps, but to me, the most important, higher-order thing that it should do is build the habit of even considering that route.*

Here, Data Scientist 7 stresses that ethical education should not focus on content only but should cultivate a mindset where reflection and ethical reasoning become a habitual part of decision-making. The goal is to encourage students to consistently engage with ethical questions as a natural part of their work.

> *Thinking, for example, "Okay, I am about to work on this thing, there are certain questions I should ask, like, what data do I have? Have I checked what the class imbalance is?" These kinds of questions, so you can have your basic checklist, so that every single project you go through these steps.*

This highlights practical steps in ethical reasoning, such as critically examining the data itself. Questions about data quality and characteristics, like class imbalance, are essential to understanding potential ethical implications in data science projects. Data Scientist 7 suggests the use of a checklist as a tool to ensure that ethical considerations are systematically addressed in every project. This approach helps embed ethical thinking into routine practice rather than treating it as an afterthought:

> *On that checklist that you do every time, there should be: think through the ethical applications. Just having that switch turned on, I think, is 90% of the battle. Like we were saying earlier, it is like as soon as you step back, you consider, "Is this a good thing or a bad thing?" Hopefully, we are mostly moral humans who do this reasoning all the time in real life.*

This passage connects ethical reasoning in data science to everyday moral reasoning, suggesting that most people already engage in such thinking naturally. The challenge is to bring that same awareness into professional contexts:

> *I think the thing that I see missing most is not flawed reasoning or anything like that; it is just not even considering [ethical issues] at all. I think the important thing for the education side is to just turn on that pathway in the brain and have that become part of the process that you do every time.*

Data Scientist 7 identifies the biggest gap in ethical practice as the failure to consider ethical issues in the first place, rather than errors in reasoning once those issues are recognised. This underscores the importance of education that prompts ethical awareness from the outset. Data Scientist 7 is explaining the agency of the individual data scientist in the consideration of ethical issues daily. They point to the value of ethical education as turning the 'switch on', a metaphor that illustrates the potential of ethical education to ignite ethical thinking in the mind of the data scientist. For Data Scientist 7, the actual content of ethical education is important, but more important is the fact that ethical education, done properly, will 'switch' the data scientist on to ethical issues that they confront in their daily working lives – as opposed to going through technical data science tasks by rote without considering ethical issues at every stage.

5.4.2 Suggestions on how best to teach ethics to data scientists

There was a wide range of suggestions about how best to teach ethics. These suggestions included: through case studies, learning by observation and/or experience, learning about practical ethical issues, and learning from the humanities about how ethical education is best done.

Data Scientist 1 makes an argument for the use of case studies and precedents in the teaching of ethics, and draws parallels with how ethics is taught in law:

> *Data Scientist 1: I think it is almost like law, where you learn, "This precedent was ruled from this case". That is how new laws are developed, by using old law cases or old use cases. [For data science], we could use case studies to say, "What is our stance, how was it passed in the past? Are we manipulating the data? What was this use case, and what can the effects be?"*

Data Scientist 1 is comparing legal judgements in real-life cases as part of legal training to the potential of using ethical decisions in real-life data science case studies as part of ethical education for data scientists. In this way, even though students have not yet entered into the 'real' world of working as data scientists, they are exposed to actual examples of ethical decisions that were made–or not made–by other data scientists before. Consider, for example, the case (referenced earlier in this paper) of the algorithm that Amazon used to filter CVs of prospective software engineers based on the CVs of current software engineers in the company. The algorithm then started to filter out female candidates, as the current base of engineers was largely male. Data science students could examine this case, consider what they might do differently, and consider where and how data scientists had made ethical errors. The valuable use of case studies in ethics programmes has also been highlighted by the National Academy of Engineering (2009), stating that:

> Successful ethics programs generally require mandatory student participation, involve relevant faculty, use interactive formats and case materials, and are scheduled throughout the year (p. 34).

Data Scientist 3 also supports the use of case studies in ethical education for data scientists:

> *Data Scientist 3: Case studies are the most effective. When I teach data science classes, if I am introducing new concepts, we talk about examples and discuss some ethical considerations around them. That has been super helpful. It feels practical because it is much easier to map. For example, if you have a new ethical problem, you can map it to a previous case that you have seen, as opposed to more abstract principles.*

Data Scientist 3 both practices and teaches data science. As a result, they have witnessed firsthand how valuable the use of case studies can be in ethical education for data scientists, particularly by moving from more 'abstract' principles of ethics to real-world examples of how ethics were or were not applied in practice.

This move – from more 'abstract' principles of ethics to real-world examples of how ethics were or were not applied in practice – is also referenced by Data Scientist 8, who suggests that the use of case studies is an effective way to teach ethical education to data scientists:

> *Data Scientist 8: I would use database examples and case studies. There are plenty of examples out there where biases have crept into analyses. You know, like, the US Presidential election, I do not even want to think about that, but that is a great big case study of how biases can be intentionally introduced through a process to affect the outcome.*

Data Scientist 8 is making specific reference to a real-world case – in this instance, the 2016 US presidential election, where Donald Trump's campaign used Cambridge Analytica. Cambridge Analytica infamously employed psychographic profiling to target voters on platforms like Facebook, creating personalised messages that resonated with individual voters and influencing public sentiment.

For Data Scientist 4, the use of case studies and examples also features strongly in suggestions for teaching ethics to data scientists, and they highlighted, too, the importance of dialogue between data scientists tackling a challenge:

> *Data Scientist 4: I really like teaching through examples, and then also getting people just to challenge their own thinking and biases and have discussions around what does and does not work. I have had this in my past working with different people, for example, I had a variable in a model and my colleague said, "I am not doing that" and I said, "But why?" and he said, "Because it does not agree with my ethics", and I said, "But why?" I really had to confront things that I just took for granted that other people have strong opinions about. It is about opening that space for discussion of what does and does not fit with certain people.*

In this excerpt above, Data Scientist 4 is talking not only about classroom teaching, but also about teaching – and learning – on the job. They believe that through dialogue with colleagues, different ethical stances can be discussed and debated, and a stronger ethical decision reached as a result.

Data Scientist 7 uses case studies and examples in their work when coaching younger colleagues:

> *Data Scientist 7: For example, if we have a section on style transfer, I say, "Hey, but those are ethics. Consider where your source data is coming from. Obviously, do not take an artist's work and then claim it as your own. Ideally, stick to patterns from nature or open source like public domain stuff, and attribute if you are using a photograph, even if it is creative commons. If you are modifying it, check the licence".*

Data Scientist 7 also had some teaching experience as a teaching assistant (TA) on a data science course, which helped develop their own understanding of how case studies can help to illustrate ethics, for example, why biases might have crept in unseen into technology:

> *Data Scientist 7: I was a TA for Neuromatch Academy, which is close to a gold standard in terms of educational things.[12] It is a very intense course with a lot of technical content. Every day, they have a big section and a couple of questions for people to discuss, and they try to do a broad survey. So, it is not just "Consider what you are optimising for" but rather "If you are optimising for this, it might have these consequences". It includes things like, "History of the camera and why the defaults like ISO exposure, etc., were tuned on these test images, who were white females". Why does this matter? Look at facial recognition technology, look at the difference and accuracy [across races]. This is a legacy of Kodak in the '70s, choosing the default exposures. Every holistic technology has a little bit of history that might have dire consequences.*

Having had this experience, Data Scientist 7 has gone on to incorporate this kind of ethical teaching into their own teaching practice:

> *Data Scientist 7: I have been teaching some generative AI stuff, [and] I am putting those questions in the course material. To say to students: this has some potential applications that are amazing and some potential applications that are dodgy, and so you should think through the deployment phase if you are fine-tuning this on certain images. Consider where it is coming from [and] the downsides if they get reproduced.*

Interestingly, Data Scientist 7's experience of translating their own learning and teaching experience on Neuromatch Academy into their own teaching tackles a challenge highlighted earlier in this theme: that scientists, broadly speaking, are not trained as ethicists. Data Scientist 7's experience demonstrates that when data scientists themselves are exposed to ethical education, they can incorporate that ethical education into their own teaching.

Data Scientist 9 – who teaches data science as well as practices data science – outlines that their approach to teaching ethical issues to data scientists is to focus on practical ethical issues:

> *Data Scientist 9: I have taught on RCTs (randomised control trials) and ethical concerns, for example, why an RCT might be ethical when you are testing policy. It is clearly more ethical to randomise who gets access to a scarce resource than to give it to your friends.*
>
> *But these are tough questions, and one needs a framework to think about them. For example, how do you think about lifesaving medicine? Randomising to show it can*

[12] Neuromatch Academy is an online educational initiative that focuses on teaching computational neuroscience and deep learning, making use of interactive tutorials and collaborative projects for teaching. Students include undergraduate and graduate students, university faculty and industry professionals.

save lives is a great thing. But once you have learned that, when do you stop the experiment? Nobody teaches this in data science.

Data Scientist 9 explains how they use practical examples in their teaching, based on ethical challenges that arise in the daily work of a data scientist. This is similar to, but slightly different from, the use of case studies. In this example, when using randomised control trials to see what medicines work or don't work, a number of ethical challenges arise. Data Scientist 9 challenges students to think about the broad set of ethical issues that could arise, for example: when do you actually stop the trial? When you stop it, do you continue giving the medicine?

A number of the data scientists that I interviewed referred to the fact that they learned a great deal about ethical decision-making through observation and through on-the-job experience. For example, Data Scientist 9 highlights the kind of ethical decisions encountered daily in working as a data scientist. They explain:

Data Scientist 9: I asked our lawyer, "We are scraping websites – what are the regulations about scraping websites?" He said, "Do not get caught. Randomise your calls, send them from different IPs." Okay, great, but if we get caught, what happens? Is it illegal? Scraping a website for public data, to me, feels unethical. Then you start getting to issues like: some guy took a lot of effort to compile their pricing. If we scrape that data, is that ethical? Or let's say we do scrape, and then a website goes down, and they're just a small Airbnb – that is not great. But this happens all the time. Some guy says, "I am going to write my scraper, and I am going to test it", but without thinking about what this means.

Data Scientist 9 emphasises that they believe ethical education about practical issues should be embedded in the technical training that data scientists receive:

Data Scientist 9: I do think that if you are doing a master's in data science, you should have a module on "practical ethical issues you might come across". That could range from questions like: how do I randomise, who should I randomise, when is the right time to switch, what do I do about consent? Is it ethical to be running an experiment on my website? Is it ethical to track people on the website? Is it ethical to put a cookie to see what people are doing on other websites?

Data Scientist 9 goes on to explain how this intersects with legislation, particularly in a burgeoning area like gen AI, which is currently largely unregulated:

Data Scientist 9: I would approach it as the ethics and regulation of artificial intelligence, data science and ML. This will be a huge area. No one actually has an idea what GDPR actually says, what you can do, and what you cannot do. I think framing it as ethics and regulation gives a practical tool.

Data Scientist 9 highlights how navigating ethics and regulation, especially with emerging technologies like gen AI, is difficult because legislation like GDPR can be unclear. They point out that compliance with laws doesn't always guarantee ethical behaviour and emphasise the need to teach students about both the ethical and legal aspects of data science within their technical training.

Data Scientist 2 highlights that ethical education for data scientists needs to bridge the technical-non-technical divide. In other words, they explain how important the ability to communicate how the model works is for ethical reasons:

> *Data Scientist 2: Education on communication is a basic thing that, at least in my experience, was not taught. It is like, "Hey, you're going to come out of this degree as a very technical person, and then you're actually going to be interfacing with very non-technical people, people that are very business-minded and practical, and you need to bridge that gap effectively". This is not a neutral thing.*
>
> *There are a variety of misaligned outcomes that can stem from not being able to bridge that gap well enough. In my experience, many technical people would benefit from being able to better bridge that gap, both from their own perspective and from an alignment perspective.*

Data Scientist 2 is highlighting how ethical issues can arise when data scientists do not or cannot explain their models to the very people who will be using them in decision-making. Data Scientist 2 believes that data scientists should, as part of the ethical education they receive, also receive training on how to bridge this divide, on how best to communicate the inner workings of a model to lay people, so that these lay people are, in turn, informed, educated and empowered to make sound ethical choices about the use and impact of the model.

Bridging this divide is a topic covered by Stoyanovich (2022) in their course on 'Responsible Data Science'. In Module 4: Transparency and Interpretability, they discuss various stakeholders of data science models, and methods for bringing them 'into the loop' of automated decision-making (p. 6). Technical topics that Tractenburg believes should be the subject of education for various stakeholders include feature-based explanations of black-box models, discrimination in online ad delivery, and 'nutritional labels' for public disclosures, as well as education on current international and local regulatory frameworks.

Data Scientist 9 suggests that the humanities have much to offer about conducting ethical education. They reiterate – as did Data Scientist 5, who came to data science from computer science – that computer science does not teach ethics at all:

> *Data Scientist 9: I will not hire someone from a computer science background for a data science role. I think they are not a fit and they do not understand. People are much better fits if they come through a humanities or a science background where they have had to grapple with some of these issues.*

Indeed, the literature highlighted that most of the (few) researchers and academics teaching ethical education to future data scientist students drew significantly on ethical and philosophical teaching from the liberal arts (humanities). Baumer et al. (2022) highlight the potential of possible inter-divisional synergies in the teaching of data science ethics in a liberal arts context. I investigate this further in Chapter 6: Analysis and discussion.

Data Scientist 9 goes on to explain what they see as an ideal approach to ethical education:

> *Data Scientist 9: This is why I like the American approach to university, where in your first year, you do a wide range of subjects. It feels like, particularly in South Africa, what we do not do is study things outside of our specialisation. Adding a half course in ethics would be great and would benefit everybody, regardless of whether you become a data scientist, an engineer or an accountant. I do not believe in "Let us teach ethics to data scientists". I believe that we should teach ethics to everybody, regardless of their background.*

Data Scientist 9 contrasts South Africa's specialised, technical-first university model with the U.S. liberal education approach, where students study a broad range of subjects before specialising. In the U.S., for example, future doctors, engineers and lawyers typically complete a general programme before professional school, exposing them to ethics and critical thinking. In South Africa, however, technical students begin their specialisation immediately, often missing this broader foundation. Data Scientist 9 argues that a liberal approach of teaching ethics to all students, not just data scientists will better prepare graduates across disciplines.

Data Scientist 7 outlines the potential that ethical writings – in the vein of liberal arts research and literature – hold for ethical education for data scientists. They describe, for example, the value they found for themselves in exploring the writing of an author called Jonathan Haidt[13] and his exploration of religion and ethics:

> *Data Scientist 7: There are a variety of ethical frameworks and approaches. I get really interested in religion and ethics and how people build those systems. For example, there is an author called Jonathan Haidt. One of his papers is called "The*

[13] Jonathan Haidt is a prominent American social psychologist known for his work on the psychology of morality and moral emotions. His research examines moral foundations theory, which explores the evolutionary origins of human moral reasoning, emphasising that moral judgements often stem from intuitive feelings rather than logical reasoning.

> *Emotional Dog and its Rational Tail", as well as a book called "The Righteous Mind" that expands his thinking further. It is fantastic – his cause is the social relativism of morality.*

This comparison highlights the idea that ethical standards are not universal but are deeply influenced by cultural context*:*

> *So, all I have to say is yes, there are some grand things in ethics that are somewhat universal, but I feel like an avoidance of harm and suffering, that I think that is quite universal.*

Data Scientist 7 acknowledges that while many ethical norms are culturally specific, some principles, such as the desire to avoid harm and suffering, may be almost universal. This suggests that although differences do exist, there are foundational ethical values that can guide data scientists and others in navigating complex moral landscapes.

In summary, the data in this theme has illuminated the views of data scientists about the role that ethical education can, and should, in their view, play in the training of data scientists. They discuss the critical need for ethical education in data science, emphasising that current training is inadequate and often treated as a one-time requirement rather than an integral part of the curriculum. Suggestions for effective ethical education include using case studies, experiential learning and interdisciplinary collaboration with ethicists to create a robust curriculum that echoes real-life dilemmas. This theme also touches on challenges, like possible differences in cultural perspectives on ethics, as well as the need for a shift from rote learning to cultivating reflection and critical ethical thinking among data scientists. Finally, this theme considered the rich potential that liberal education holds for ethical education for data scientists, with both data scientists and the literature providing suggestions for how this seam could be mined for ethical education for data scientists to create ethically educated, reflective and critically thinking data scientists.

5.5 Theme 5: Implications of gen AI and emerging technologies

I did not originally intend to ask the data scientists that I interviewed about gen AI, but in almost every interview, data scientists raised gen AI as a new, burgeoning field that is unregulated and unpredictable, and as such, a field where a whole new set of ethical issues arises. I've provided definitions of data science, AI and gen AI in Chapter 3 of this paper, but for the purposes of this theme, I will remind the reader that gen AI is a subset of AI that focuses on creating ('generating',

hence the name) new content and ideas, such as text, images, videos, and music. Gen AI uses advanced machine learning models to learn patterns from existing data and generate original outputs, based on user prompts. Gen AI models that lay people use daily include ChatGPT, Perplexity and Dall-E, for example.

The data scientists that I interviewed pointed to the new and changing nature of gen AI, and the ethical issues and challenges the field raises. As Data Scientist 1 explains:

> *Data Scientist 1: In generative AI, it is just changing as we go, and the rules need to change. We are living in a time where innovation is almost instant. ChatGPT became available, and people started using it straight away. It was just like, "Oh, I can use it, it is ChatGPT". There are no guidelines.*

Data Scientist 1 goes on to explain that besides being available for use immediately, by anyone, a technology like ChatGPT opens up a minefield of potential ethical issues:

> *Data Scientist 1: Gen AI is essentially using the data that it is trained on to create new content. You see videos, for example, that look like someone. They look so real but are actually fake. It is quite intense that there can be a video of someone who looks exactly like me, that could be spewing hate speech. There are real concerns about issues like impersonation or stealing someone's character. These are big concerns.*

In both these excepts, Data Scientist 1 is highlighting firstly, the concern that change in this field is so quick ('almost instant') that no guidelines are available for use yet. Certainly, there is no legislation yet in this space. A second concern for Data Scientist 1 is that gen AI can be fake. Gen AI can 'do' immoral things, but it is not a moral agent like a human being. So how is it held accountable?

Data Scientist 1 also expresses the concerns not only about the result of gen AI, but also about the source of the data that feeds the gen AI model. As Data Scientist 1 explains:

> *Data Scientist 1: There are also issues with what data it is being trained on. For example, creating an image of a dog between a husky and a golden retriever by sourcing the internet to get this whole bunch of data. Where are they getting all these images from? Source is a big thing. What about my images that I load to Instagram? Are those getting used in algorithms? In chatbots, for example, there have been instances where people upload their code, and it gets stolen. So, gen AI is really a whole different world.*

Data Scientist 3 also highlights the source, or input, of data into gen AI models as an area of ethical concern:

Data Scientist 3: The big thing is the level of abstraction with respect to input data. Gen AI is trained on the corpus of the internet. So, it is really hard to get accountability for the data that is going into it and how it is being managed. When I teach Intro to Data Science, we talk about where your data is from, who gave consent for it, all of that good stuff. But these gen AI systems are just run at a scale where it is really hard to do that. It is also the most accessible of all tools. For example, if you want to train a deep learning network, you need to go to school and take a whole semester-long course to learn how to do it. With gen AI, anybody can just pop open ChatGPT and start using it. I think there is a scary level of accessibility and also a scary level of lack of accountability in terms of the training data and the guardrails around it. Gen AI can spit out hate speech, it hallucinates. It is hard to use responsibly, and everyone can use it. So, it is bonus dangerous over data science, in my opinion.

Data Scientist 3 is expressing concern over the high level of abstraction in gen AI, which is trained on vast internet data, making it difficult to ensure accountability regarding the data used and its management. They highlight the easy accessibility of these systems, explaining that while traditional data science requires formal education, gen AI tools like ChatGPT can be used by anyone, which raises issues of responsible use and the potential for harmful outputs such as personally identifiable information and hallucinations.[14]

Another ethical issue that arises with gen AI is that students are using it. As Data Scientist 3 highlights:

Data Scientist 3: It is changing grade distribution curves on everything. For example, I taught a course in Spring last year, which was all coding homework. ChatGPT took off around February/March. And now ChatGPT can solve all of it.

The way of combating this is not to forbid the use of ChatGPT, but rather to assume students are using it and to make assignments harder as a result, explains Data Scientist 3:

Data Scientist 3: My philosophy is that it is like a calculator. It is a tool, right? People are going to use it. So, the onus is on us to design more realistic assignments that use that. For instance, I was talking to the PhD student who took over my course this year. Instead of doing homework coding tests every two weeks, they did two or three big projects during the course that were more realistic. And because people were expected to use ChatGPT, the deliverables were more ambitious. But I do not think anyone has really figured it out, to be honest.

Data Scientist 3 suggests that instead of banning the use of ChatGPT, educators should design challenging and realistic assignments that acknowledge its existence and usage, in the same way

[14] In the context of gen AI, 'hallucinations' refer to instances where the AI generates information that is false, misleading or nonsensical.

that calculators are used as tools in education. They advocate for assignments that create a more ambitious learning environment that incorporates the use of AI tools like ChatGPT.

Data Scientist 3 believes that increasing the scope of courses and assignments is an adequate solution for university, but that it gets harder the lower down in schooling you go:

> *Data Scientist 3: Even at graduate level or undergrad level education, it is tough. But in elementary school, it is even tougher. What is essential learning? If you never write an essay and you just use ChatGPT, are you missing out on fundamental knowledge? I do not know what the answer is.*

Data Scientist 3 is voicing an uncertainty that worries parents and educators alike. What is the impact of using tools like ChatGPT? Will students miss developing essential skills, such as essay writing, if they rely solely on AI for their work, similar to how calculators impact learning basic arithmetic? How is it possible to know what is necessary in basic education these days? Do students still need to learn their times tables, for example, and if so, why, in a world with gen AI?

Data Scientist 6 points to a more existential issue that impacts many people when faced with gen AI:

> *Data Scientist 6: There is an existential crisis that people face. I do not know if that is the right word. I think that people feel like AI is going to replace them at some point. But how do you tackle that? Do you include that within the curriculum?*

Data Scientist 6's concern is deeply empathetic in nature and acknowledges the vulnerability that many people feel in the face of the immense computing power of gen AI tools. Data Scientist 6 suggests that addressing this anxiety within educational curricula is important.

Data Scientist 7 points out that a fundamental difference between pure data science and gen AI is the nature of the task that the tools are built for. Data science models are usually pinpoint-specific in their final application, whereas gen AI models can be used for almost anything. As Data Scientist 7 explains:

> *Data Scientist 7: With data science, you often are building some set of features that are very customised to a company and a problem, and you are building one model that predicts one thing. It is very focused. For example, you can build a predictor based on seasonal rainfall, satellite imagery and soil change capacity. That means you build one thing, with specific applications, and we get to predict the yield so that we can ensure the farm is better. It seems like having better predictions would be better for everyone. So, it is a very single special-purpose technology.*

Here, Data Scientist 7 is highlighting how traditional data science projects are typically narrow in scope. The models are tailored to solve a specific problem for a particular organisation, for example, predicting crop yields for a farm or predicting patterns of consumer spending at different times of the year. This clear focus allows for a clear understanding of the model's purpose and its potential impact, making it easier to anticipate and address ethical concerns.

> *Gen AI, on the other hand, is more, "I am building a vaguely intelligent box that can take in any text and follow any instructions". This is very general, and so it is like saying, "What are the potential ethical implications of a pen or a photocopier?"*

In contrast, explains Data Scientist 7, gen AI is a general-purpose tool, capable of handling a wide range of tasks. The analogy to a pen or photocopier emphasises the versatility and corresponding unpredictability of gen AI. Because it can be used for almost anything, it is much harder to foresee all the possible ethical implications:

> *This can be used for writing college essays and composing fake apologies to your spouse, ranking CVs, filtering through legal documents, and helping people who are not First-Language English speakers. When you are dealing with something this general, it is kind of impossible to think through all the downstream applications.*
>
> *If you are building a single-use data science object, I feel like it is very easy to reason about ethics. With gen AI, on the other hand, you have many of the same ethical questions, but at the research and development level, they are much more complex.*

Data Scientist 7 notes that while ethical issues in narrow, single-purpose data science projects are usually clear, gen AI's versatility as a general-purpose tool makes anticipating all ethical implications far more difficult. Obvious problems, like automated CV grading, can be addressed, but the broader research and development context is highly complex and uncertain. This uncertainty suggests that deploying gen AI requires even more thorough and proactive ethical consideration than traditional data science.

Data Scientist 9 makes a similar point:

> *Data Scientist 9: Gen AI is a completely different thing to data science. In gen AI, the big ethical question is the data it is trained on, ownership, IP, bias and all of that stuff. Gen AI works because of the training data it has been trained on. So, there is that ethical aspect of when you use it to draw a picture, is it based on some other artist's work? There is also obviously the bigger black box aspect. We do not know what is going on and why it is suggesting things.*

Here, Data Scientist 9 argues that gen AI differs significantly from traditional data science, raising ethical concerns related to the ownership of training data, intellectual property, potential biases, and the opacity of AI decision-making processes.

Data Scientist 9 explains how the absence of a human in the loop can be perceived as neutral or as accurate, but AI often hallucinates or provides false information:

> *Data Scientist 9: I think what is going to happen is that it will be so commonplace that there is not going to be a human in the loop. Your AI will speak to my AI. Unless we have something monitoring it, there may be bad outcomes. If you ask ChatGPT to write an essay, it just makes things up that are clearly wrong. It is not too far-fetched that it makes something up that is wrong, and a human perceives it as right and acts on that. It is early days, so we do not really know how all these things are going to fit together and what is built on what. It is the Wild West, and there is likely to be some sort of crisis if we build like this.*

Data Scientist 9 warns that as gen AI becomes ubiquitous, interactions may occur entirely between machines, without any human oversight, creating a false perception of neutrality or accuracy. Without monitoring, these systems could produce fabricated or incorrect information that users accept and act upon, leading to harmful outcomes. They caution that the current unregulated, 'Wild West' development of AI is likely to trigger a major crisis if human oversight and safeguards are not built in.

Data Scientist 9 also worries about the impact on mental health:

> *Data Scientist 9: And the other thing (sorry to just rant about this) is that we see trends that seem now well-established, like increasing male suicide, as they lose their connection to the community. We have hectic incidents at the school, like teen suicide, self-harm, mental breakdowns, and it is driven by social media being always on. Contributing to that are the algorithms that feed you stuff. Even though I know this stuff, I can see how it is operating on me, and there are many people who cannot see it. You are being served up this content of good-looking, seemingly happy people, and now you compare yourself to these people. It feels like human coping mechanisms have not kept up. We want those dopamine hits and end up just going after them.*

Data Scientist 9 is worried about the rising levels of mental health issues among youth and believes that this trend is a result of the pervasive influence of social media. They argue that the algorithms underpinning social media drive feelings of inadequacy and unhappiness by continuously showing idealised images of other people. This challenges human coping mechanisms that have not adapted to this new reality, which is driven by technology and capitalism.

Data Scientist 7 highlights a concerning trend in the gen AI landscape, where power is concentrated in a few global organisations. This raises issues regarding the ethical implications of gen AI development, as it reflects a narrow perspective on morality and ethics that may not take diverse cultural viewpoints and the complexities of global ethical standards into account:

> *Data Scientist 7: How do you make sure you have broad standards of ethics, such as consideration of harms, that are somewhat universal, but you also have a way for people to interact with things that are representative of where they are, their culture, their upbringing, their context? I think you do not get that if you have a concentration of power in a couple of organisations. This is one of the reasons I support open-source data and models.*

Data Scientist 7 discusses the challenge of establishing broad ethical standards in gen AI while ensuring that these standards are representative of diverse cultural contexts and individual backgrounds. They argue that the concentration of power within a few organisations limits this representation, advocating for open-source models and data to promote inclusivity and a broad range of ethical considerations in AI development.

Data Scientist 3 points out that because gen AI is accessible to everyone, this means that everyone should be getting ethical education:

> *Data Scientist 3: There is a question about data ethics for everyone, because AI is becoming so accessible. My mom can use ChatGPT now. How do you give her an understanding of the ethics associated with it? Or, for another example, I have undergrads writing papers with ChatGPT. How do you talk about plagiarism in that context? It almost feels like it should be gen ed at this point.*

In poignant words, and in a fitting summary for this theme, Data Scientist 9 reminds us that AI is just a machine, which has learned how to mimic human language, but does not, in the end, have any of the empathy and compassion that a human might have:

> *Data Scientist 9: There is no ethics in gen AI. If you had a human, there may be empathy and compassion. Whereas in gen AI it is just a minimising loss function, and that loss function does not have ethics in it.*

The data within this theme provides a powerful view of the worries and concerns expressed by the data scientists that I interviewed about the untrammelled and unregulated continuous development of gen AI models in the world. This rapid development and lack of regulation give rise to ethical concerns, including worries about impersonation, data sourcing, and accountability. There are few papers in the literature about ethics and gen AI so far, likely owing to the very newness of the gen AI, even within a ‘new’ field like data science. Of those papers that do exist, some authors advocate

for incorporating Aristotelian virtue ethics into education to prepare future generations for the evolving ethical landscape of AI, rather than relying solely on principlist approaches that may become outdated. Finally, data scientists express anxiety about the easy accessibility of gen AI tools like ChatGPT, which can produce misleading or harmful outputs. They emphasise the need for educational reforms that acknowledge the role of AI in learning and, increasingly, in humans' lives, but pay attention to deeply human concerns, including mental health impacts and existential fears about job displacement due to automation.

Conclusion

Each of the themes has identified key concepts that will provide the framework for my discussion in this analysis chapter. In 'Theme 1: The importance of ethics in data science', the data scientists that I interviewed emphasise the critical role of ethics in data science, highlighting its impact on most facets of human life. In this theme, some data scientists discuss how data is not objective but influenced by human biases, which can be perpetuated by algorithms and models. The theme stresses the need for ethical education in data science curricula, comparing it to established ethical frameworks in fields like medicine. A few of the data scientists that I interviewed suggest that unethical behaviour in data science could erode public trust, potentially limiting the field's potential for positive impact.

In 'Theme 2: Ethical challenges encountered in practice', data scientists describe the ethical challenges that they face in their daily work. Notably, this is a lengthy section, as the ethical issues are numerous and detailed. Data scientists express almost to a person how unprepared they feel to handle ethical issues, often because of their lack of formal training in this topic. The data scientists that I interviewed have ethical concerns about every stage of the data science process, including data collection, labelling, model building, and the use and impact of models. Some of the specific issues that they outlined include dealing with sensitive data, potential biases in data and models, the 'black box' nature of some models, and the unintended consequences of data science applications. The theme also touches on the challenges of communicating complex technical concepts to non-technical stakeholders and the ethical implications of lack of diversity in data science teams.

'Theme 3: The organisation as custodian of ethics' discusses the role of organisations in managing ethical considerations in data science. It points to the gap between emerging legislation around data science ethics and the practical implementation of these principles. A number of the data scientists that I interviewed felt that their organisations should be the primary custodians of ethics,

responsible for establishing rules, safeguards and ethical decision-making processes. Others felt that they were the custodians, and still others felt that both the individual data scientist and the organisation for which they work should be the custodian of ethics. This theme also investigates the challenges of interpreting vague legislation and the tendency of some data scientists to follow the 'path of least resistance' when faced with ethical ambiguities. It also touches on the need for clear organisational guidelines and the importance of fostering a culture of ethical awareness and responsibility within data science teams.

'Theme 4: The role of ethical education' explores the acute need for ethical education in data science training. In their interviews, data scientists emphasised that current ethical education is inadequate and often treated as a once-off, rather than an integral and integrated part of the curriculum. Their suggestions for effective ethical education included using case studies and real-world examples, learning through observation and hands-on experience and drawing on concepts and teaching from the humanities and liberal arts. Theme 4 also highlights the importance of interdisciplinary collaboration between data scientists and ethicists to develop robust ethical curricula. It also discusses the challenge of bridging the technical-nontechnical divide, emphasising the need for data scientists to effectively communicate complex ethical issues to non-technical stakeholders.

Finally, in 'Theme 5: Implications of gen AI and emerging technologies', the data scientists I interviewed described the ethical challenges presented by the rapid development and widespread adoption of gen AI. They expressed deep concerns about the unregulated nature of gen AI and its potential for misuse, including the 'instant' nature of gen AI innovation and adoption, which is rapidly outpacing the development of ethical guidelines, concerns about data sourcing and accountability in gen AI models, and ethical issues related to impersonation and the creation of fake content. The theme highlights the difficulty in anticipating the universe of ethical implications of rapidly developing technologies and highlights the problem of concentration of power in a few large tech companies and the ethical implications of their dominance in AI development. The interviewees spoke about the need for broader ethical education, not just for AI developers but for all users of AI technologies and stressed the importance of maintaining human oversight and decision-making in AI systems.

The empirical insights presented in this chapter expose the high level and degree of ethical challenges that data scientists encounter in data science practice. These findings not only highlight

the limitations of current ethics education but also raise new and important questions about institutional responsibility, individual agency and the broader social impact of data-driven technologies. The next chapter builds on these insights by situating them within the wider academic literature, offering a comparative analysis that deepens understanding and outlines theoretical findings and practical recommendations.

Chapter 6: Analysis and discussion

This chapter serves as a bridge between the findings and insights from my interviews with data scientists and the literature. In this section, I compare my findings with existing research in order to contextualise my results within ongoing debates about ethics, education and justice in data science. This analytical lens enables a deeper exploration of the structural, institutional and personal dimensions of ethical practice, identifying critical gaps and opportunities for reform. The chapter is structured around the five themes that I generated through the thematic analysis from the interviews. These five themes are 1) the importance of ethics in data science; 2) ethical challenges encountered in practice; 3) the custodian of ethics; 4) the role of ethical education; and 5) the implications of gen AI.

6.1 The importance of ethics in data science

A consistent finding across all the data scientists that I interviewed was that every step that is involved in data science holds ethical issues – from the collection of data all the way through to how the data is used. And these ethical issues matter, because every single step impacts humans.

There is a common misconception that data science yields objective and scientifically neutral results because it uses statistical and mathematical modelling. This notion is flawed for several reasons. Firstly, the notion of 'raw data' is a misnomer. Data is not an objective entity that exists independently of human influence. Rather, it is generated, collected and interpreted by humans, and thereby inherently incorporates potential mistakes, prejudices and biases. As Crawford (2013) outlines:

> Data and data sets are not objective; they are creations of human design. We give numbers their voice, draw inferences from them, and define their meaning through our interpretations.

Crawford explains that data is inherently subjective as it is shaped by human decisions at every stage of collection, interpretation and use. O'Neill (2016) succinctly encapsulates the subjective nature of mathematical models with the observation, 'Models are opinions embedded in mathematics'.

Secondly, the algorithms fuelling mathematical and statistical models are chosen by humans, with all the accompanying challenges that an imperfect human brings to any decisions and choices.

Thirdly, the ‘model’ is a so-called ‘black box’ for most people who are not data scientists, and indeed, often for other data scientists, too. Speaking about science broadly, but in terms that apply to data science, too, Nead (2021) outlines that:

> On one level, science is a collection of facts about the world, and adding to that collection does require discoveries. But science is also something larger. It’s a mindset, a process, a way of reasoning about the world that allows us to expose wishful thinking and biases and replace them with deeper, more reliable truths.

Nead highlights that science is not simply the accumulation of facts but rather, a systematic process and mindset aimed at uncovering reliable truths by challenging biases and wishful thinking. Given the impossibility of verifying every experiment individually, trust in the integrity and credibility of others' work is essential to the scientific endeavour. Part of building this trust is the ability to explain the ‘black box’ of the model in a way that is transparent and understandable for other data scientists and lay people alike. Ethics is necessary precisely because of the ‘black box’ nature of many data science models, which potentially makes data science outcomes open to manipulation by bad actors.

Finally, recommendations arising from data science models are applied to humans and impact every facet of our lives. As Buijsman, Klenk and van den Hoven (2025) explain,

> AI is used to support fraud detection, credit risk assessments, education, healthcare diagnostics, recruitment, autonomous driving, and much more. Actions and decisions in these areas have a high impact on individuals, and therefore AI becomes more and more impactful every day (p. 59).

As pervasive as the use of data science and AI is in our daily lives, its recommendations can only consider historical data. This means that they inherently reproduce structures that have existed historically – and as such, can reproduce unfairness, inequality and structural imbalances. O’Neill (2016) explains:

> Big Data processes codify the past. They do not invent the future. Doing that requires moral imagination, and that's something only humans can provide.

While data science and algorithms have immense power to shape human lives, they often reinforce existing biases and inequalities rather than creating a fairer or more innovative future. As such, ethics in data science should be taken very seriously indeed.

The significance of ethics in data science extends into areas that may not initially be apparent. For example, the potential for bias in data labelling is often acknowledged because human annotators

bring their subjective perspectives to the labelling task. However, a different but no less important ethical dimension that can be unheeded is the impact of the labelling work on the people doing the labelling. Consider, for example, a dataset containing descriptions of child abuse that must be labelled for police records and analysis. While the purpose of this labelling is important and serves a greater societal good, the process itself can have an intense psychological impact on the people reading and categorising the content. This example raises ethical questions regarding the well-being of data labellers, and the potential for trauma highlights the need to prioritise not only fairness and impartiality in data labelling but also the mental health of those involved in the labelling. Ethical frameworks in data science need to address the societal implications of biased and flawed datasets, as well as the individual experiences of those contributing to their creation.

Ethics in data science is a relatively new topic, largely because data science itself is a relatively new field. Utts (2021) argues that the burgeoning nature and the exponential growth in data science in the past decade have led to more opportunities, but 'with opportunity comes responsibility' (p. 2). As a profession, argues Utts, data scientists need to put more emphasis on ethical guidelines and procedures. Utts argues:

> We need to train our students and practitioners to ask 'why' before asking 'how'. As statisticians and data scientists, we need to question the ethics of our work. We need to ask who benefits and who might be hurt (p. 1).

In sharp contrast, the field of ethics is an extremely old field of study. As Colando and Hardin (2024) state:

> A challenge for anyone teaching data science ethics is to be accomplished in the ever-changing and growing field of data science while simultaneously being knowledgeable and well-versed in hundreds of years of ethical theory (p. 361).

Ethics have long been applied in other professional fields. In the field of medicine, for example, the importance of ethics has long been acknowledged. The Oath of Hippocrates, a brief exposition of principles for physicians' conduct, dates from the fifth century BCE. Its statements protect the rights of the patient and oblige the physician voluntarily to behave in an altruistic manner towards patients (Riddick, 2003). Since then, medical practice and research today have benefited from a system of ethical frameworks, in which philosophers, practitioners and lawmakers have all played a role in the development. For example, Thomasma (2004) outlines how philosophers analysed traditional ethical theories, such as utilitarianism, deontology, and virtue ethics, and applied these concepts to medical dilemmas. This philosophical groundwork helped articulate principles that now govern

patient care and informed consent, emphasising the importance of moral reasoning in clinical settings.

A practical example of how traditional ethical theories have been applied to medical dilemmas can be seen in the case of end-of-life decision-making for terminally ill patients. In a very rough account to illustrate application, from a utilitarian perspective, the decision to continue life-sustaining treatment weighs overall societal benefit: if treatment consumes significant resources without improving quality of life or recovery prospects, discontinuation may be favoured to benefit patients with better prognoses. A deontological approach centres on respecting patient autonomy, underpinning informed-consent practices: even if a doctor deems treatment futile, they must fully inform patients of their options and honour their choices. Virtue ethics focuses on the moral character of the healthcare provider, guiding them to act with compassion, honesty, and wisdom, such as holding difficult prognosis discussions with empathy and clarity.

Cohen-Amalgor (2017) traces theoretical foundations of medical ethics that stem from the philosophies of Aristotle, Immanuel Kant, John Stuart Mill and John Rawls. In the article, Cohen-Amalgor discusses the concept of autonomy according to Kant and Mill, Kant's concepts of dignity, benevolence and beneficence, Mill's Harm Principle (nonmaleficence), the concept of justice according to Aristotle, Mill and Rawls and Aristotle's concept of responsibility. Hofmann (2021) analyses the role of philosophy and ethics at 'the edge of medicine', and demonstrates that although centuries of work have gone into defining the current frameworks in medical ethics, there is still a great deal of work to be done, and the task is never-ending:

> This does not only underscore and invigorate existing roles of philosophy and ethics but also that completely new tasks are needed at "the edge of medicine." There is a lot of work to be done – for the improvement of health and wellbeing of living beings – now and in the future (p. 10).

Other researchers, too, have picked up on the similarities between the fields of medicine and AI, and the corresponding need for a similar ethical code for data scientists. For example, Borenstein and Howard (2020) state:

> Taking the example of medicine, physicians may promise to uphold the Hippocratic Oath. While a professional oath is not a panacea, it can serve as a statement of and a commitment to a social contract between a profession and the public (p. 63).

Borenstein and Howard proceed to suggest that the Hippocratic Oath is a reminder to physicians of their ethical obligation to improve the health of the public, and that, because AI provides similar

benefits and potential harms, we should expect similar ethical responsibilities of those who develop the technology.

It is striking that even in a field like medical ethics, where ethical issues have been debated for centuries and ethical frameworks and legislation have been developed and refined over time, the work is never truly complete. This is as it should be, given the immense stakes involved. The contrast with data science is clear: as a field still in its infancy, data science has a long way to go before it can approach the level of comprehensiveness, structure, and governance achieved by medical ethics.

Today, the world is at a critical point where a comparable ethical intervention is needed to shape ethics in data science and AI. New ethical issues arise daily, ranging from facial recognition and voter profiling to brain-machine interfaces and weaponised drones, and ongoing debate about gen AI's global impact on employment emphasises the need to establish robust ethical frameworks. So far, however, the approach to ethics in data science has been largely driven by regulation, which has mostly been outpaced by developments. Current regulations include the EU's General Data Protection Regulation (GDPR), China's Personal Information Protection Law, California's Consumer Privacy Act, Canada's Consumer Privacy Protection Act, Australia's Privacy Act, India's Personal Data Protection Bill and South Africa's Protection of Personal Information Act.

Three insights emerge. Firstly, most regulations in data science are recent, with little to no regulation dating back further than a decade. Secondly, there is no global regulation in the field. Most countries have established some form of personal information protection laws, but the effectiveness and comprehensiveness of these regulations can differ widely based on local legal frameworks and enforcement capabilities. Thirdly, most regulations concern data privacy, with almost no regulations governing the decisions or outputs of data science models.[15]

This means that regulation alone cannot ensure ethical decision-making in data science, in part because data science technology is advancing faster than regulation can adapt. There are few global standards, and those that do exist legislate for technology that has already changed. There are

[15] The two exceptions here are Canada's Directive on Automated Decision-Making, which sets rules for how government agencies should use automated systems to make decisions, ensuring that these systems are clear, fair, and respect people's rights, while requiring assessments of their impact on individuals; and S-11-7, which gives guidance for financial institutions on managing model risk, emphasising the need for robust governance, independent validation and ongoing monitoring of models used in decision-making processes.

inconsistent rules and practices across countries and regions, which gives rise to opportunities for bad actors to practice exploitation in regions with little or no regulation. Most global legislation focuses on data privacy, which ignores issues like bias, fairness and transparency. Ironically, a compliance-driven 'checkbox' mindset can actually deter deeper ethical reflection in the erroneous belief that all the boxes are checked.

A few technology companies follow a self-regulation path. Consider, for example, Steve Jobs' stance towards privacy when he worked at Apple (quoted in Swisher, 2024):

> Silicon Valley is not monolithic. We (Apple) have always had a very different view of privacy than some of our colleagues in the Valley. We take privacy very seriously (p. 123).

But this example represents an exception rather than the rule, as Swisher (2024) goes on to explain:

> I have spent an increasing amount of time talking to government officials and legislators in recent years, since no significant US laws have been passed to rein in tech… ever (p. 248) … Most regulators and politicians are utterly missing in action.

Swisher notes that her discussions with government officials and lawmakers have increased significantly over the years, particularly because no major legislation in the U.S. has successfully restricted the power of tech companies. She points out that most regulators and politicians appear inactive in this domain.

Buijsman et al. (2025) highlight two specific features of AI that means it differs from other technologies: first, AI systems can have a greater degree of agency than other technologies, in that AI systems can, in principle, make decisions on their own and act in dynamic fashion, responding to the environment they find themselves, and second, AI systems have a higher degree of epistemic opacity than other technical systems (p. 61). Both of these specific features of AI were referenced by the data scientist that I interviewed. Indeed, Data Scientist 9 sees a key feature of ethical education for data scientists being the need to teach data scientists how to explain their models to a lay person. Taken together, the authors explain:

> These two features of AI systems make it difficult to develop, deploy, and use them responsibly. They have more agency than other technologies, which exacerbates the challenge – though we should be clear that AI systems do not have moral agency … and thus should not be anthropomorphized and cannot bear responsibility for results of their outputs. In addition, even its developers struggle to anticipate (due to the opacity) what the AI system will output and why.

As a result, argue Buijsman et al., familiar ethical problems that arise out of irresponsible or misaligned action are repeated and exacerbated by the speed, scale, and opacity that come with AI systems. This is precisely the effect that O'Neill references in her book 'Weapons of Math Destruction', wherein the three characteristics of opacity, scale and damage enable data science algorithms to affect large populations rapidly. Feedback loops can trap victims in systems they cannot escape, while bias amplification entrenches existing inequalities. Scalable AI spreads these harms rapidly to larger populations, and self-fulfilling prophecies allow algorithms to shape realities without correction. At scale, such efficiency can damage livelihoods or penalise the poor with little transparency. Together, these dynamics create a vicious cycle where inequality becomes embedded in future algorithmic iterations, further deepening injustice.

Ethical failings in data science have already caught the attention of the media. Examples include the Cambridge Analytica scandal mentioned early in this paper; the Dutch childcare benefits scandal[16]; the Grindr privacy breach[17]; Project Nightingale, the partnership between Google and Ascension[18]; and others. Data is not just information; it is power. It can shape behaviours, influence decisions and even define the course of our lives. It is in this context that the importance of ethics is hard to overstate.

6.2 Ethical challenges encountered in practice

This section examines the ethical dilemmas data scientists face in daily practice, as reported by study participants and supported by academic literature. Many interviewees felt unprepared to address these challenges, noting minimal ethical training in their degree courses, a finding echoed in recent studies (Utts, 2021; Davis, 2020; Tractenberg, 2020; Salz, 2019; Tanweer, 2017). As a relatively new field, data science has given limited attention to ethical issues and even less to formal ethics education (Salz et al., 2018). This gap leaves graduates ill-equipped for complex moral

[16] Between 2005 and 2019, approximately 26,000 parents were falsely accused of welfare fraud by the Dutch Tax and Customs Administration. Affected families were forced to repay childcare benefits in full. Tens of thousands of families, often with lower incomes or belonging to ethnic minorities, were pushed into poverty because of debts to the tax agency. Some victims committed suicide. More than a thousand children were taken into foster care.

[17] The dating app Grindr shared sensitive user data, including HIV status and GPS location, with third-party advertisers using AI analytics, violating user privacy and trust.

[18] Project Nightingale granted Google access to the complete health records of over 50 million patients without their explicit consent. The project, which aimed to improve patient care through advanced data analytics, raised significant privacy concerns and led to investigations by the U.S. Department of Health and Human Services due to the sensitive nature of the data involved and the lack of patient notification.

dilemmas, especially as technological advances outpace ethical guidance. To ensure responsible use of data and AI, integrating robust ethics training into data science curricula is essential.

Utts (2021) investigates the 'newness' of ethical issues arising in data science, involving data quality and privacy, and the analysis, interpretation and dissemination of data-driven decisions, as sources of data become more plentiful and massive datasets are easier to acquire (p. 1). This echoes Tractenberg (2020), who examines and refutes the assumption that 'everyone who enters a profession has been sufficiently prepared to take on the responsibilities for the standards that define that profession or discipline' (p. 3), highlighting two reasons why this is not the case. Firstly, many who enter into the domain of data science do so after training in a variety of other fields, and secondly, there are many new jobs emerging every day with the specific title/duties of 'data science' and many more jobs where data science tools, techniques, and methods are only a part of daily practice. Tractenberg (2020) claims that the data science domain is too new, and practitioners operate – and join the practice – from too many other domains for there to be a single community to drive the creation and adoption of ethical education. There is a lack of both academic and professional communities of expertise.

Oliver and McNeil (2021) also point to the 'newness' of data science as a possible reason for the dearth of ethical teaching in data science programmes:

> The lack of attention to ethical considerations in undergraduate data science programmes could be a consequence of the recent rise of data science (p. 7).

However, it is unlikely that the relative 'newness' of the field is the sole reason. Wolpe (2006) highlights that scientists, broadly, are often wary of ethical scrutiny, and generally reluctant to engage the public in moral conversation about their work and points out several reasons that scientists avoid thinking about ethics. The first reason, he asserts, is that scientists believe that they are 'not trained in ethics':

> Ethics as an academic field has an established body of knowledge, a set of disciplinary concepts, a canon, and many other trappings of an intellectual discipline. Most scientists are not formally trained in ethics (p. 1023).

Wolpe is intimating that scientists are wary of treading into an academic field not 'their own', and that because they themselves did not receive ethical training, they tend to avoid thinking about ethics. Baumer et al. (2022) go a step further. Instead of foregrounding the lack of training in ethics on the part of scientists themselves, they discuss 'the default position' of 'indifference to ethics in

data science' and counter this with a mission to encourage data science students to grapple with the often not-so-obvious ramifications of their data science work and to develop their own compasses for navigating these waters. This echoes my own thinking, as I believe that the reluctance of scientists to engage with ethics in data science stems from a combination of academic siloing and lack of formal training in ethical considerations. This matters because the reluctance on the part of scientists to engage with ethical issues has led to a deficiency in ethical teaching in the field. This is concerning, given the significant societal impact of data science work. The data scientists that I interviewed echoed their experience of this reluctance, explaining how their training focused heavily on the technical aspects of data science (such as the statistics, maths, models and programming) and largely ignored the broader societal issues of how data science is used. Garzcarek and Steuer (2019) suggest that the morality of the data science community is evolving and that it is a 'shared task to develop it' (p. 14). This is a theme I will return to later, when I look at the possibilities for education in and for ethics and ethical conduct in data science.

The absence of ethical training in their data science teaching experienced by Data Scientists 2, 5, 7 and 9, described in my findings section, mirror the research by Oliver and McNeil (2021), who examine 25 data science or comparable programmes across universities. They conclude that that current data science undergraduate programs provide solid grounding in computational and statistical approaches yet may not deliver sufficient context in terms of ethical considerations necessary for appropriate data science applications. This matters, because best practices in ethics are critical for the reputation of data science, for minimising social harms caused by bad data science and for 'maintaining the quality of data science applications' (Saltz, Dewar & Heckman, 2018). Oliver and McNeil argue that the omission of ethical study from undergraduate data science programs potentially 'creates a Promethean workforce prepared to use a variety of computational and statistical tools in socially inappropriate ways' (p. 11). In this way, Oliver and McNeil (2021) argue that ignoring ethics in data science teaching at universities will not only create students who are ignorant about the potential ethical implications of their work, but more importantly, it could create highly technically-skilled students with great power to wreak devastation – much in the way that Prometheus did with fire. They harness the analogy of the story of Prometheus, best known for stealing fire – an immensely powerful tool of both creativity and destruction – from the gods and giving it to humans, to illustrate their point. Like fire, data science can be used as a tool for innovation and enlightenment. It can also be used as a tool of great destruction.

Following their sense of being unprepared, data scientists report facing ethical issues at every stage of the process: from the source and sensitivity of data, labelling and classification, and model

building, to the outcomes, uses, and impacts of models – where cultural bias and discrimination are common. Other challenges include unethical client requests and a lack of diversity in teams. These practitioner insights, distilled from interviews, align with the literature but offer a more practical perspective.

In her paper about teaching responsible data science, Stoyanovich (2022) recounted her experience in developing, teaching, and refining a technical course called 'Responsible Data Science', in which she covered issues of ethics in AI, legal compliance, data quality, algorithmic fairness and diversity, transparency of data and algorithms, privacy, and data protection. Davis (2020) highlights how ethics permeates every phase of the life cycle, from acquisition, cleaning, using/reusing, and publishing, to preserving/destroying data (p. 2). Tractenberg (2020) researches the 'many ethical considerations that arise as workers engage in data science', which include:

> Deciding what data to collect, obtaining permissions to use data, crediting the sources of data properly, validating the data's accuracy, taking steps to minimize bias, safeguarding the privacy of individuals referenced in the data, and using the data correctly and without alteration (p. 3).

The above topics highlighted in the literature and the issues discussed by the data scientists that I interviewed overlap to a great extent. The literature identified a wide range of ethical challenges, which include legal compliance, data quality, privacy and data protection, algorithm fairness, transparency, labelling and reusing data, minimising bias, and the use, preservation or destruction of data. These concerns closely aligned with the issues raised by the data scientists I interviewed, who also identified challenges such as the source and nature of data, issues with handling sensitive data, data labelling and classification, model development, and the outcomes and impacts of models – including the risks of cultural bias and discrimination. Further, they mentioned facing unethical client requests and noted a lack of diversity and representation within data science teams.

Swisher (2024) addresses the significant issue of lack of diversity and representation in the tech industry, noting that the predominance of white males in leadership positions leads to a lack of awareness regarding the problems this creates. She argues that the backgrounds and identities of those who develop technology play a crucial role in shaping the products themselves. When the creators of these products lack diverse perspectives, it can result in harmful outcomes:

> More and more, the white male homogeneity in tech was creating problems that the people at the top could not perceive or understand. Who makes products and what characteristics they have matters a great deal as to how products turn out—especially when those products become damaging.

> A truism began to form in my brain about the lack of women and people in the leadership ranks of tech: The innovators and executives ignored issues of safety not because they were necessarily awful, but because they had never felt unsafe a day in their lives.

She suggests that those at the helm often overlook safety concerns – not out of malice, but because they have never personally experienced feelings of vulnerability. Their own backgrounds heavily influence the way they design and promote technology, leading to platforms that may not adequately consider user safety or potential risks.

The data scientists that I interviewed highlighted that the source and nature of the data give rise to three main areas of ethical concern. Firstly, the data collection process can often mirror pre-existing structural inequalities and imbalances. Secondly, ethical data collection should involve informed consent. For example, Utts (2021) makes the point that today, in most clinical trials and other designed experiments, standards such as informed consent and data privacy are well established, but this is not necessarily the case in observational studies, especially when data sources include web scraping, purchasing data or harvesting data collected by one's employer. Utts urges data scientists to ask questions such as:

- Is informed consent possible? If so, is it used?
- Is anonymity guaranteed? Or is there a risk that identity could be revealed?[19]
- Are there structural biases built into the data, such as when zip codes or post codes are used as a proxy for ethnicity?
- Are certain subgroups disadvantaged, for instance, because of past discriminatory behaviours? (p. 7).

Thirdly, the data itself can often be personal and/or sensitive in nature.

Regarding data labelling and classification, the data scientists that I interviewed explained that data science models run off variables, which means that all the data that is collected needs to be labelled, or classified, into the variable categories. Inherent in this classification process are a number of ethical issues that data scientists have to grapple with in the course of their daily work. There can be vicarious trauma, too, on the part of the classifier, if the data that they have to classify concerns

[19] Consider, for example, the case where Netflix released a dataset as part of the Netflix Prize contest, which aimed to improve its recommendation system. This dataset included viewing habits of nearly 500,000 customers, revealing sensitive information such as movie titles, genres, rental dates, and ratings without proper anonymisation or consent. A closeted lesbian mother from Ohio filed a lawsuit against Netflix in 2009, claiming that the company violated her privacy rights.

horrific crimes. The classification process often requires discrete decisions on the part of the data scientist, which data scientists grapple with, particularly when it comes to sensitive information like gender, or identity, or when it comes to trade-offs in the labelling. This issue about ethical trade-offs is outlined by Buijsman et al. (2025):

> For the actual implementation of values there are a number of additional challenges to consider. Most prominently is the fact that conflicts can occur between different design requirements, which is more often referred to as value conflicts or trade-offs. These [can be] conflicts between accuracy and fairness or between privacy and fairness. If we want to use statistical fairness measures to promote equal treatment of, for example, men and women, then they need datasets labelled with gender, thus reducing privacy (p. 74).

The authors describe how pursuing an overarching 'value' such as equal treatment means that data has to be identified by gender. This labelling by gender reduces privacy, which is a different but no less worthy value. Making this trade-off can be a tricky ethical choice for a data scientist.

Regarding the building of the model, the data scientists' concerns clustered into two broad areas. First, there are questions about how the data is used to build the model and which variables are used. The second ethical challenge involved in the building of the model lies in the difficulty of communicating how the technical model works to non-technical people, who are often the ones who need to use the outputs of the model. The first broad area of concern, also noted by Utts (2021), is that many decisions made during the planning of a data science project, though not always obviously ethical, should be treated as such if they affect the validity of the model and results. She highlights several examples: ensuring ecological validity so that study conditions reflect real-world settings; avoiding interventions without consent, as in the 2012 Facebook experiment involving nearly 700,000 unwitting users whose news feeds were manipulated, potentially causing psychological harm; and conducting a power analysis beforehand, since underpowered studies waste resources and risk drawing false conclusions. Utts argues that raising these issues is an ethical responsibility for data scientists.

Utts (2021) tackles the second area of concern with the suggestion that the population at large should be educated about ethical issues in data science, and states that, 'now more than ever statistical education at all levels should include a dialogue about ethics along with discussions of statistical ideas and methods' (p. 12). Humans have a poor intuitive understanding of statistics, and one of the most common mistakes made in the media is to attribute cause-and-effect relationships

when they are not warranted, either deliberately or through ignorance. Utts has developed a list of 10 topics that she believes are important for an educated populace:

1. Observational studies, confounding and causation
2. The problem of multiple testing
3. Sample size and statistical significance:
4. Poor intuition about probability and risk
5. Why many studies fail to replicate
6. Does decreasing risk actually increase risk?
7. Personalised risk versus average risk
8. Using expected values to make decisions
9. Surveys and polls – good and not so good
10. Confirmation bias and selective consumption of news.

The 'black box' nature of data science models makes it difficult to explain how bias arises, even when algorithms appear to use objective data. As Utts (2021) notes, the absence of obvious bias in input data can lead to the mistaken belief that results are unbiased. The Austrian Public Employment Service (AMS) algorithm illustrates this problem: using factors like gender, age, citizenship, and childcare responsibilities, it predicts job placement chances but consistently disadvantages women, older workers, non-EU citizens and caregivers. Despite its seemingly neutral inputs, the model reflects and potentially amplifies societal biases, and its complex inner workings obscure how these outcomes are weighted or produced.

Tanweer et al. (2017) describe how, sometimes, when a data science team responds to one set of ethical dilemmas, those responses gave rise to still more dilemmas. In their research, they investigate specifically a case study of ethical dilemmas that arose in a 'data science for social good' project focused on improving navigation for people with limited mobility. They focus on instances in which data scientists recognise, grapple with and conscientiously respond to ethical challenges, foregrounding the ways that ethics are implicated in the 'quotidian work of data science' (p. 1). The key ethical dilemmas they highlight that face data science practitioners include:

> The risk of exacerbating disparities; the thorniness of algorithmic accountability; the evolving opportunities for mischief presented by new technologies; the subjective and value laden interpretations at the heart of any data-intensive project; the potential for data to amplify or mute particular voices; the possibility of privacy violations; and the folly of technological solutionism (p. 2).

The researchers in the case study analysed by Tanweer et al. were trying to build routing software that would meet the specific needs of individuals with limited mobility. One of the first ethical challenges they faced was the availability of data. The team quickly realised that OpenStreetMap data (crowd-sourced data) overrepresented certain areas of the city and underrepresented others. Relying only on this data for the project, then, could serve to further entrench or exacerbate existing inequities. This particular ethical dilemma they faced is a common one in data science projects, in which the reliance on conveniently available data can serve to create or compound disparities. In their conclusion, the authors encourage approaching ethical thinking as a thoughtful and intentional balancing of priorities rather than a binary differentiation between right and wrong.

Almost every data scientist that I interviewed expressed concern about the ethical implications that arise when data science models are put to use. They worry primarily about three things: the fact that the model they're building can have an impact on the lives of human beings, unintended consequences of the model and disparate impact. The fact that some of the data scientists that I interviewed had differing ethical views on the same issue (in one particular example, the use of race as a variable to assess the effectiveness of a car commercial) highlights the fact that data scientists are often making ethical choices in their daily work based on their own ethical frameworks. Goldsmith and Burton (2017) highlight how practitioners' ethical frameworks shape their decisions, using the example of whether to sign an open letter urging the UN to ban weaponised AI. Their case study shows that most AI practitioners use utilitarian reasoning, and they compare decisions made under utilitarianism with those informed by deontology and virtue ethics.

6.3 The custodian of ethics

The data scientists that I interviewed held differing views on where the accountability, or 'custody' of ethics lies, and whether this was with the individual data scientist themselves, or the institution that they worked for, or a blend of the two. Data Scientist 3 explained that they 'take the path of least resistance', Data Scientist 2 says 'ethics feels above their pay grade' and 'you ultimately rely on organisational processes that are larger than yourself to provide guardrails for these kinds of things', and Data Scientist 7 talks about their personal 'moral compass'. These are all expressions of an individualistic orientation to ethics.

The Data Scientists did not explicitly reference any moral or ethical frameworks, but implicit moral theories were evident in their accounts. For example, Data Scientist 4's reasoning about the car commercial race variable contains a consequentialist, utilitarian bent, as they're weighing outcomes

(inclusion for people of colour) against potential harms. Their colleague's refusal ('I will not use race classification of variables, in any model, ever') is a deontological stance, as it is a categorical prohibition. Data Scientist 7's language about a 'moral compass', being 'self-aware' and having an ethical 'switch' in the brain aligns with virtue ethics, with the emphasis on character, habituation and moral exemplars (their first boss). Yet none of these data scientists explicitly recognised that they were operating from specific philosophical traditions.

Those data scientists who felt 'woefully under-prepared' to grapple with ethical issues in data science were the most likely to feel that their institution should put ethical guard rails in place, whereas those data scientists who had made a point of educating themselves on ethical issues felt a stronger individual accountability for ethics. Data Scientist 3 says that they take the 'path of least resistance'. Data Scientist 2 explains that ethics feels 'above your pay grade', exhibiting 'ethical fading' where the moral dimension drops out of view when reframed as a technical or organisational problem.

Two data scientists intimated that accountability should be a blend of institutional and individual accountability. In an example from the literature of blended 'ethical custody', Utts ascribes the accountability for ethics to the full team working on a data science project. This includes organisational representatives, data scientists and statisticians. She provides guidance on important ethical considerations, especially when working as part of a multidisciplinary team, which she suggests should be discussed with teammates to make sure everyone is focused on the ethical implications of the team's work. She says that statisticians and data scientists can and should play a leadership role in these discussions (2021, p. 89). Utts proposes a number of ethical guidelines for data scientists. She emphasises the importance of team discussions about ethics, urging teams to consider 'why' before 'how' and to evaluate the societal implications of their work, for example, the fairness of algorithms that suggest driving or walking routes for people. Utts asserts that transparency is an important principle: data scientists should explain algorithms and acknowledge statistical uncertainty. Utts also encourages human oversight in algorithmic decisions, as human judgement can outperform algorithms in individual cases. Finally, she highlights the potential for data insights to drive positive social change, encouraging data scientists to investigate underlying causes of disparities rather than just accepting correlations.

Possible sources of the implicit moral frameworks expressed by the data scientists include their formal ethics education (or lack thereof), their disciplinary training, the culture of the organisation

in which they work, and cultural and societal norms. When formal ethics education is absent, data scientists necessarily fall back on intuition, personal values and organisational norms. Similar to Haidt's social intuitionist model, my participants' moral reasoning seems to be largely post-hoc justification of intuitive judgements.

Data scientists come to data science from a variety of disciplinary backgrounds, and this likely influences their implicit moral frameworks. For example, disciplinary training in computer science teaches students to think in terms of optimisation, efficiency and loss minimisation, which possibly nudge a data scientist towards a consequentialist/utilitarian stance. Data Scientist 7's ethical sensibility was shaped by early career exposure to a boss willing to 'burn the client', and Data Scientist 9's came from sitting on an academic ethics committee. There are forms of moral habituation and Aristotelian in nature, although neither of them explained their choices in this manner.

Finally, all of my data scientists work within capitalist economies, and the associated cultural and societal norms. This can give rise to 'ethical fading', where a focus on dimensions of decision-making other than ethics means that the ethical dimensions of the decision fade from view. For example, a focus on optimising technical choices (like loss minimisation within machine learning) can give rise to a fading of the human implications of model decisions. A focus on profit maximisation can do the same. Data Scientists 2 and 3 implicitly illustrate this dynamic when they say they take 'path of least resistance' or that ethics feels 'above your pay grade'.

The interviews provide empirical evidence that practising data scientists make decisions in ways that align with recognisable ethical traditions, but often without knowing it. The direct implication for ethical education and teaching data science ethics is that ethics education need not start from scratch. Rather, it needs to make the implicit explicit and give data scientists a vocabulary for what they are already doing (but doing unreflectively). Further, the interviews show that no single ethical theory is sufficient. For example, the car commercial shows that utilitarian and deontological reasoning can lead to different conclusions, each defensible. This supports an argument for virtue ethics, but more broadly suggests that what is needed is the capacity for navigating between ethical frameworks, implying phronesis, or practical wisdom.

There is also evidence of a gap between individual moral reasoning and team-based decision making. Data science is done in teams in organisations, but my interviewees' moral reasoning was

almost entirely individualistic. This means that ethical education for data science also needs to consider collective decision making within power dynamics, not just individual data scientists taking autonomous choices. Finally, the phenomenon of 'ethical fading' suggests that ethical education for data scientists should challenge the assumption that data scientists will know an ethical dilemma when they see it. A more fundamental problem in data science is not simply choosing the right action but recognising the ethical dimension in the first place.

6.4 The role of ethical education and teaching data science ethics

The data scientists that I interviewed had much to say about the role of ethical education for data scientists. They were aligned in their criticism of the current dearth of ethical education for data scientists. They all insisted that data scientists should receive ethical education, and that ethical education should not be a once-off session (or 'vaccine') but rather, it should be embedded into and throughout technical training. There was a wide range of suggestions about how best to teach ethics. These suggestions included: through case studies, learning by observation and/or experience, learning about practical ethical issues, and learning from the humanities about how ethical education is best done.

One of my interviews detailed how, even at an Ivy League university, there is no ethical education for data scientists, partly owing to the fact that most of the scientists developing courses in data science are not themselves ethicists, and partly because there are different departments (computer science, operations, statistics) who present technical data science teaching, and so there is no overall accountability for data science as a discipline, per se. For the data scientists that I interviewed, the actual content of ethical education is, of course, important, but more important is the fact that ethical education, done properly, will 'switch on' ethical thinking in data scientists, so they consider ethical issues at every stage. This is a view echoed strongly by Tractenberg (2020), who emphasises that data scientists should be taught to utilise judgement, rather than memorisation, as they learn about what constitutes ethical data science.

The data scientists that I interviewed were united about the importance of integrating ethics education into data science and AI curricula. Three of them insisted that stand-alone ethics courses are insufficient, and instead, ethical teaching should be woven through the curriculum. The literature highlights the need for interdisciplinary collaboration, particularly with philosophers and ethicists, to develop teaching about ethical frameworks for data scientists. Both the data scientists that I spoke with and the literature suggest various approaches to teaching ethics, including case

studies, practical applications and learning from humanities disciplines, including the potential of Aristotelian virtue ethics as a framework for AI ethics. The data scientists highlighted the importance of developing critical thinking and ethical reasoning skills in data science students to prepare them for the complex ethical challenges they may face in their professional lives. I examine ethical theories, and the data scientists' suggestions, in more detail in the next chapter.

6.5: Implications of gen AI and emerging technologies

When conceiving of this research project, I did not intend to ask the data scientists that I interviewed about gen AI specifically, but in almost every interview, data scientists raised gen AI as a new, burgeoning field that is unregulated and unpredictable, and as such, a field where a whole new set of ethical issues arises. I've provided definitions of data science, AI and gen AI earlier in this paper, but for the purposes of this theme, I will remind the reader that gen AI is a subset of AI that focuses on creating ('generating', hence the name) new content and ideas, such as text, images, videos, and music. Gen AI uses advanced machine learning models to learn patterns from existing data and generate original outputs, based on user prompts. Gen AI models that lay people use daily include ChatGPT, Perplexity and Dall-E, for example.

Smith and Vickers (2024) argue that because artificial intelligence technologies have become a ubiquitous part of human life, we should ask, 'how should we live well with artificial intelligence?' (p. 19). In particular, they argue, the world needs:

> Effective and flexible ethical training that can prepare future generations for living in a world in which novel ethical situations crop up with novel technologies and their applications and for designing AI systems that are likely to benefit rather than to harm (p. 20).

The authors explain how, increasingly, universities are requiring that students who might one day build new AI are required to take courses in ethics. However, explain the authors, most teaching rests on the principlist approach. This approach, which is also known as principle-based ethics, is a widely used framework in applied ethics, especially in fields like healthcare, biomedical research, and increasingly, technology and AI. Rather than relying on a single overarching ethical theory, the principlist approach uses a set of core moral principles to guide ethical decision-making in complex situations.

The principlist approach is built around four main principles. The first principle is respect for autonomy, which means that individuals should have the freedom to make their own choices and

control their own lives. In practice, this means respecting people's decisions and ensuring informed consent. The second principle is beneficence, wherein there is a duty to promote good, act in the best interests of others. and contribute to their welfare. The third principle is non-maleficence, which means that one must avoid causing harm. This principle is often summarised as 'do no harm'. The fourth and final principle is justice, wherein ethical decisions should be fair, ensuring equal treatment and a fair distribution of benefits and burdens. These principles are considered prima facie, meaning they are binding unless they conflict with one another. When principles do conflict, they must be weighed and balanced to determine the most ethical course of action.

The principlist approach to ethics suggests that to live well with AI, we need to establish and follow certain rules and principles that guide our interactions with it, in the belief that if we can identify the right rules and make sure everyone understands them, we will be able to coexist well with AI. The principlist approach has challenges, though. Rules can quickly become outdated, and rules often react to problems instead of anticipating them. Instead of relying solely on rules, argue Smith and Vickers (2024), a better way to prepare future generations for living and working with AI is through Aristotelian virtue ethics. This approach focuses on developing good character traits and making wise choices, which can help people navigate the challenges posed by AI in a more thoughtful and reflexive way.

Borenstein and Howard (2020), believe that the ubiquitous nature of AI means that we have to confront its impacts:

> Artificial Intelligence (AI) is becoming pervasive. The technology is reaching into so many facets of our lives that we have no choice but to confront its impacts. The creation and deployment of AI is changing our lives and communities in countless ways (p. 61).

The authors go on to explore exactly the point raised above and hinted at by Data Scientist 1: AI (and gen AI, as a subset of AI) is capable of producing fake information, of producing horrible and immoral things, but it is not, ultimately a moral agent. The 'root of the problem', argue Borenstein and Howard, is people. As they explain:

> Addressing these and other ethical concerns requires starting with the root of the problem (i.e., people). Tackling the problem head-on requires educating ourselves at the beginning stages of our interaction with AI–irrespective of whether we are developers, first learning about AI, or users, just starting to interact with AI.

This approach highlights the need for widespread understanding of AI technologies, not just among developers but also among users. The authors further explain:

> The opportunity to learn about how data are used to train AI, about the applications that the AI can enable, etc., should be available to any person that interacts at any stage with AI.

Borenstein and Howard particularly stress the importance of ethical considerations in the development process:

> If we focus just on those designing AI technology, there is tremendous potential to shape what developers are learning and encourage them to embrace the crucial message that ethics is intertwined with the entire design process (before, during, and after) (p. 62).

Borenstein and Howard suggest that tackling the problem starts with educating ourselves – the data scientists who produce the models, and the people who use the models – about issues including how data are used to train AI, how the algorithms work and how fairness should be assessed in the outputs of models.

The concern about the fact that gen AI is used by everyone, everywhere is gaining prominence in the literature, too. As Horvitz and Mitchell (2024) express it:

> As AI systems become more integrated into daily life, ensuring their reliability and safety is paramount, especially when the methods are applied in high-stakes areas like medicine, criminal justice, education, and industrial process control (p. 168).

These authors express similar trepidation about the fact that, given the rise of the use of AI in daily life, more controls and safeguards should be put in place. In a similar vein, Borenstein and Howard (2020) also believe that any person interacting with AI should understand the ethical issues implicit in AI. They argue for a comprehensive approach to AI education:

> Tackling the problem head-on requires educating ourselves at the beginning stages of our interaction with AI–irrespective of whether we are developers, first learning about AI, or users, just starting to interact with AI. The opportunity to learn about how data are used to train AI, about the applications that the AI can enable, etc., should be available to any person that interacts at any stage with AI.

This inclusive approach to AI education emphasises the importance of ethical considerations throughout the development process. Borenstein and Howard (2020) further explain:

> If we focus just on those designing AI technology, there is tremendous potential to shape what developers are learning and encourage them to embrace the crucial message that ethics is intertwined with the entire design process (before, during, and after). Moreover, ethics should not be a slapped-on component after-the-fact, a standalone lesson, or a second thought. It is integral at every stage when learning about AI.

The authors provide a concrete example of how ethical considerations can be integrated into technical education:

> When we teach the mathematical derivations of a linear regression function for supervised learning in AI, we can also mention the use of disparate impact as a metric to evaluate fairness of the output in the hopes that we move closer to a result that is "correct" and "fairer" (p. 62).

Borenstein and Howard are expressing the view that to effectively address the ethical challenges associated with AI, it is important to educate all individuals interacting with AI, including developers and users, about data usage, applications and ethical considerations. Ethical education should be integrated throughout technical training rather than treated as an isolated component, because ethics is essential at every stage of AI development and application.

Smith and Vickers (2024) suggest that virtue ethics can guide us how to 'live well with AI' and provide stronger guidelines for living well with AI than the principlist or consequentialist approaches. They argue that virtue ethics offers a flexible framework for addressing the ethical landscape of AI, which is valuable because, as AI becomes increasingly integrated into daily life, rigid rules may become outdated. Virtue ethics encourages individuals to develop practical wisdom (phronesis) that helps them assess situations thoughtfully and act in ways that promote well-being. By creating virtuous agents with both ethical understanding and technical expertise, society can better ensure that AI is developed and used in ways that contribute positively to our lives.

Aristotelian virtue ethics is also examined by Ober and Tasioulas (2024) as a framework for ethics in data science and AI. They contend that virtue ethics emphasises a human-centred approach to AI ethics, with a focus on human flourishing and morality, rather than strictly maximising preferences or abiding by human rights laws. They suggest that rather than being seen as a replacement for human endeavours, AI should be seen as 'intelligent tools' that enhance human capabilities and promote democratic engagement. Lastly, they support Aristotelian-based regulation that can successfully combine market forces, state action and rights protections while stimulating international collaboration in AI governance.

However, argue Smith and Vickers (2024), despite the importance of a virtue ethics approach to living well with AI, actually doing this in practice is challenging. Virtue ethics, rooted in the philosophy of Aristotle, emphasises the importance of developing good character traits, or 'virtues', as the foundation for ethical behaviour. Unlike rule-based approaches that focus on specific actions

or principles, virtue ethics is agent-based, meaning it centres on the moral character of individuals. Virtues such as courage, justice, compassion, integrity and generosity are seen as stable traits that guide people (agents) in making ethical decisions across different contexts. These virtues are nurtured and inculcated through a process called ‘habituation’, where people learn from moral exemplars and practice virtuous actions until they become second nature. This approach encourages a deep understanding of ethical contexts and the ability to respond appropriately to novel situations. To give a practical example, a data scientist who cultivates the virtue of justice might seek to correct an algorithm that unfairly discriminates against a specific gender in hiring.

Smith and Vickers conclude their article on the application of virtue ethics to the field of AI with each author providing their own answer to the question of whether and how such a system might be implemented. Nicholas Smith argues that there is, in theory, little hope for implementing this virtue ethics approach. Based on her experience in teaching virtue ethics, Darby Vickers argues that there is some hope for educating students to live well with AI, even if it is extremely challenging.

Conclusion

In this chapter, I have drawn connections between the findings from my interviews with data scientists and the existing body of literature. Through my interviews with data scientists about the ethical challenges that they face in their daily practice, I uncovered four more themes, which include: the importance of ethics in data science, who the custodian of ethics is, the role of ethical education and the data scientists’ ideas for how this could be introduced and taught to data scientists, and the implications of gen AI. In the discussion of generative AI ethics, the analysis reveals how traditional ethical frameworks both apply to and fall short of addressing novel challenges posed by these emerging technologies.

This chapter has established new connections between practitioner experiences and theoretical frameworks, contributing to our understanding of ethics in data science. The comparative analysis presented in this chapter reveals how practitioners' experiences with ethical challenges throughout the data science lifecycle substantiate theoretical concerns raised in the literature while introducing nuanced perspectives often absent from academic discussions. Further, this chapter offers a novel examination of the contested ‘custodianship’ of ethics in data science, presenting an empirical investigation that maps practitioners' varied perspectives on whether ethical responsibility lies primarily with individuals, institutions or in a balanced approach, deepening understanding of the complex interplay between personal ethical agency and institutional frameworks.

The comparative analysis of practitioners' ethical education experiences with curriculum research reveals critical gaps between what data scientists need and what current education provides. By identifying specific pedagogical approaches – such as case studies, experiential learning, and humanities integration – that align with both practitioner needs and research recommendations, this chapter starts to provide actionable insights for transforming data science education, which I will further develop in the next chapter.

By analysing the intersection between practitioner experiences and theoretical frameworks, this chapter has established a comprehensive understanding of the ethical landscape in data science that neither perspective could provide alone. The analysis reveals critical gaps in current approaches while identifying practical pathways toward more ethical data science practice through education, institutional frameworks and theoretical grounding. This analysis directly addresses the research questions by demonstrating that: 1) ethics education is indeed essential for in-service data scientists, as evidenced by practitioners' consistent experiences with ethical unpreparedness; 2) liberal education concepts offer substantial resources for addressing ethical challenges, particularly through virtue ethics frameworks and interdisciplinary approaches; and 3) these concepts could be effectively introduced through integrated, case-based learning rather than stand-alone ethics courses.

Through this analysis, it seems clear that addressing ethical challenges in data science requires more than technical solutions or the virtue of individual data scientists. The discussion highlights the need for comprehensive, justice-oriented frameworks that integrate concepts from both critical and liberal perspectives. In the next chapter, I build upon these analytical insights and develop specific recommendations for professional practice and education, investigate implications for my own professional development and identify promising directions for future research that could further advance ethical data science education and practice. The framework established here provides a solid foundation for translating theoretical insights and connections into practical actions so that ethics training can be reconceptualised to address broader structural concerns, informing a more politically engaged vision for the future of ethics education in data science.

Chapter 7: Education in and for ethics and ethical conduct in data science

Informed by the empirical findings as well as the theoretical analysis, this chapter offers practical recommendations for integrating ethics and ethical conduct into data science education and practice. Emphasising the importance of interdisciplinary approaches, political consciousness and institutional responsibility, the chapter advocates for a shift from compliance-based models to a more holistic, justice-driven vision for the field.

In this dissertation, I sought to draw on the experiences of data scientists to examine whether education in and for ethics and ethical conduct in data science should be introduced into in-service training for data scientists. Both the insights from my interviews with data scientists and the findings from the literature support an argument that education in and for ethics and ethical conduct in data science should be introduced into in-service training for data scientists, addressing my first research question. This chapter presents recommendations for incorporating ethics and ethical conduct into the training of future data scientists, based on literature findings and interview results from Chapter 5 and Chapter 6, addressing my second and third research questions. I then take the argument further, suggesting that education for data scientists should also help them move beyond considering only ethical principles, encouraging them to recognise and engage with the broader political and social implications of their work as active participants in shaping society.

Drawing from a politics of justice perspective, these recommendations address the technical aspects of ethics in data science as well as the broader political and social justice implications, suggesting that data scientists move towards seeing themselves not just as technical experts but also as political actors whose work shapes society in normative ways. This moves beyond ethical frameworks to embrace a deeper understanding of the social impact of the work that data scientists do. This approach matters because it goes beyond simply teaching ethical principles to cultivate a reflective practice that encourages data scientists to continually examine their own decisions and the broader impact of their work. Without this reflective process, data scientists may overlook complex ethical dilemmas or fail to recognise the societal consequences of their actions. Developing these reflective skills enables them to navigate ambiguity, make more informed decisions and contribute responsibly to society. In this way, ethical education can serve as a form of critical activism that empowers data scientists to challenge systemic injustices and cultivates a sense of resistance against exploitation, discrimination and exclusion.

7.1 Recommendations for education in and for ethics and ethical conduct in data science

Two of the recommendations arising from the literature and from the interviews with data scientists about education in and for ethics and ethical conduct in data science include ensuring course content is embedded throughout the programme (rather than having standalone courses) and promoting an interdisciplinary approach towards both content and the people teaching the content. In particular, regarding this last point, there is strong support for drawing from the liberal arts, broadly understood. There are also recommendations about the ways in which ethics teaching is done – for example, making extensive use of case studies. Case studies are valuable for teaching data science students about ethics because they present real-world scenarios that make ethical concepts tangible and relevant, allowing students to see the real impact of ethical decisions in data science. By analysing these cases, students develop critical thinking and decision-making skills, learning to identify, evaluate and address the kinds of ethical challenges they may face in their careers.

Regarding embedding education in and for ethics and ethical conduct in data science throughout training programmes, Utts (2021) recommends that ethical principles be integrated across all statistics and data science courses, rather than confined to standalone ethics classes. Embedding ethics throughout the curriculum encourages students to continuously reflect on the ethical implications of their work and the reasons behind their actions before focusing on technical solutions:

> With that kind of training, students may remember to ask and answer questions about 'why' before answering questions of 'how' (p. 6).

The data scientists I interviewed also believed that ethical content should be woven throughout curricula. They viewed a standalone course as insufficient and containing the possibility of being treated by students as unnecessary and extraneous, whereas ethical education woven throughout the curriculum helps emphasise that ethical challenges arise at *every* stage of the data science process.

All the data scientists that I interviewed expressed a belief that ethical education needs to be meaningfully incorporated so that it is not just once-off, pass/fail training. This sentiment is echoed by participants at a workshop on ethics education and scientific and engineering research (National Academy of Engineering, 2009) who asserted that:

> Stand-alone, online programs that students, post-docs, and faculty take on a "pass/fail" basis do not provide an adequate introduction or enough practical experience to prepare them for ethical problems that arise in academic and professional life (p. 36).

The data scientists' responses and the statements in the literature repeatedly use the words 'stand-alone', 'pass/fail', 'once-off', and 'totally separate' as descriptors of courses or content on ethics that are inadequate or insufficient. In contrast, words like 'interwoven' and 'embedded' were used by both data scientists and in the literature to illustrate the preferable nature of ethical education.

Borenstein and Howard (2022) also argue that ethics should not be an afterthought or a separate component in education, but rather integrated throughout all stages of learning, particularly in AI. For instance, when teaching the mathematical foundations of algorithms, instructors can simultaneously address ethical issues such as fairness by discussing metrics like disparate impact to help ensure outcomes are both correct and fair:

> Moreover, ethics should not be a slapped-on component after-the-fact, a standalone lesson, or a second thought. It is integral at every stage when learning about AI (p. 62).

These views were echoed by the data scientists whom I interviewed. Several interviewees expressed that standalone ethics courses are insufficient, often perceived as extraneous or irrelevant to the technical core of their studies. For example, Data Scientist 1 described feeling 'woefully unprepared' to handle ethical issues because their education focused almost exclusively on technical skills, with little attention to the societal impacts of their work. Similarly, Data Scientist 3 noted that ethical discussions were only addressed when they could be framed as technical challenges, resulting in a disconnect from real-world consequences. The lived experiences of practising data scientists thus underscore the necessity of integrating ethical reasoning and reflection into every stage of data science education, as both the literature and the interviews make clear that ethical challenges arise continuously throughout the data science process.

The data scientists that I interviewed might be somewhat reassured by the view of Baumer et al. (2022), who emphasise that teaching ethics to data science students is 'not about indoctrinating students about *what* to think, but rather to force students to grapple with the often not-so-obvious ramifications of their data science work' (p. 4). Gille & Nardo (2020) propose a similar approach to teaching ethics (in this case, to medical students). Drawing on educational theory, they propose that 'negative experiences' marked by 'confusion and a sense of not-knowing' can be 'fertile ground for

learning' (p. 2). This approach suggests that structured dilemmas can develop ethical understanding. This is similar to the view of Data Scientist 7, who expressed in their interview that they believe that the role of ethical education is, first and foremost, to 'switch on' ethical thinking. Data Scientist 7 does not propose a specific ethical framework or ethical teachings but instead proposes that data scientists be conscientised by ethical education so that they are sensitive to and aware of ethical issues in their daily work.

A key challenge highlighted in the literature and by the data scientists I interviewed is that data science faculty members are not typically comfortable with the content of ethics courses. This is a common challenge at universities. But as Baumer et al. (2022) point out, it is not a challenge that is insurmountable:

> One of the primary challenges is that while educators are typically well-trained in the ethics of human subjects' research, few have specific training in, say, algorithmic bias, or even general ethical philosophy. But why should a lack of training prevent us from teaching our students? (p. 10).

The authors make a strong case that ethical issues are a general issue with the impact of technology on society, and data scientists might make up for a lack of ethics-specific training by partnering with philosophers and ethicists to develop a robust ethical curriculum.

Harvard is an example of a university that uses interdisciplinary faculty teaching, using philosophy PhD students to deliver guest lectures in various computer science classes to help students think normatively about their responsibilities, as part of Harvard's Embedded EthiCS initiative. One computer science student (2021) described their experience:

> Weighing competing notions of preference-utilitarianism and freedom of choice was unfamiliar territory for many of those who majored in Computer Science, Applied Math, or Statistics. In class discussions, students often were visibly uncomfortable about being asked to reason with definitions and concepts that were perceived to be "vague" or "arbitrary", especially in contrast with the mathematical ones we were used to encountering. "Who gets to decide what counts as ...?", "Where should we draw the line?" were questions that were often uttered. Given the (relative) lack of easily operationalizable terminology, as time went on, our conversations would frequently not converge on definite agreeable answers but rather surface even more questions (p. 4).

It is clear that some of the students majoring in computer science, applied maths, or statistics found it challenging to weigh different philosophical ideas like preference-utilitarianism and freedom of choice. They felt uneasy when asked to think through definitions and concepts that seemed unclear or subjective. This matters because if students cannot grapple with complex and subjective ethical

concepts, they may struggle to identify and address real-world ethical challenges in data science, leading to decisions that overlook critical societal impacts. However, when students develop these skills, they are better equipped to navigate ambiguity and make responsible, thoughtful choices in their professional practice. In this way, education for data scientists in and for ethics can confront structural power, not just individual moral decisions, moving beyond narrow ethical codes to address collective harms.

Wolpe (2006) also examines this issue and asserts that while ethics as an academic field has an established body of knowledge, a set of disciplinary concepts and many other trappings of an intellectual discipline, 'most scientists are not formally trained in ethics' (p. 1023). However, he points out that scholars trained in ethics sometimes work with scientists and scientific societies, helping to set guidelines and assess the impact of new technologies. Wolpe and Baumer et al. point to the possibility of interdisciplinary collaboration between ethicists and philosophers, for example, and scientists for developing and implementing ethical education programmes for science students. Similarly, Ober and Tasioulas (2024) suggest that Aristotelian virtue ethics offer a compelling framework for ethics in AI. A number of the data scientists I interviewed suggested that a liberal education has much to offer data scientists by way of ethical thinking, ethical frameworks, and ethical education.

Regarding how to teach ethics to data scientists, the data scientists I interviewed had a wide range of suggestions, including learning from the humanities about how ethical education is best done. The potential of learning from the humanities about how to teach ethics to data scientists is highlighted in the literature, too. For example, Baumer et al. (2022) specifically explored potential points of intersection between the data science school and the philosophy department at the university where they worked. Further, all entering students are required to take first-year seminars as part of the liberal arts curriculum, and these are often interdisciplinary and help students who might not otherwise have a fully developed interest in data science connect data science to larger issues in society. They are currently discussing a standalone course on data science ethics co-taught by the data science school and the philosophy school.

Borenstein and Howard (2022) highlight the importance of having interdisciplinary teams who create ethics content and teach it, as the challenges emerging in relation to AI cross over disciplinary lines and are too complex for any single type of expertise to handle. They propose drawing heavily on liberal arts content and practitioners:

> Insights from lawyers, sociologists, policy scholars, philosophers, and others along with scientists and engineers can be especially valuable when determining how to educate students about AI ethics (p. 64).

They highlight that a central goal of AI ethics education is developing students' critical thinking and ethical reasoning skills that apply across professions.

Crawford (2013), too, highlights the potential of including methodologies from the social sciences in data science:

> We know that data insights can be found at multiple levels of granularity, and by combining methods such as ethnography with analytics, or conducting semi-structured interviews paired with information retrieval techniques, we can add depth to the data we collect.

This integration of methods allows for a more comprehensive understanding of the data, moving beyond simple quantitative analysis. Crawford argues that qualitative approaches can provide crucial context:

> We get a much richer sense of the world when we ask people the why and the how not just the "how many". This means complementing data sources with rigorous qualitative research.

By incorporating these methodologies, researchers can address limitations in big data analysis and gain deeper insights. Crawford concludes:

> Social science methodologies may make the challenge of understanding big data more complex, but they also bring context-awareness to our research to address serious signal problems. Then we can move from the focus on merely "big" data towards something more three-dimensional: data with depth (p. 1).

Crawford highlights that the inclusion of methodologies from liberal education will provide more nuance, more granularity and less bias than the use of data alone. By combining quantitative analytics with qualitative methods such as ethnography and semi-structured interviews, Crawford argues, researchers can not only deepen their understanding of complex social phenomena but also encourage students to engage with diverse perspectives – a central principle of liberal education.

Another illustration of the integration of liberal education principles into ethics education for data scientists is the four-module 'Responsible Data Science' semester-long programme that Stoyanovich (2022) developed:

- Module 1: Algorithmic fairness (4 weeks)

- Module 2: The data science lifecycle (2 weeks)
- Module 3: Data protection (3 weeks)
- Module 4: Transparency and interpretability (4 weeks)

In Module 1 on Algorithmic Fairness, she covered classification and risk assessment, as well as fairness in set selection and ranking, intersectional discrimination, and connections between algorithmic fairness and diversity, with an important component of this module being the introduction of equality of opportunity doctrines from political philosophy. She drew on the political philosophy of John Rawls (1971), 'A Theory of Justice'; Ronald Dworkin (1981), 'What is Equality? Part 1: Equality of Welfare'; John Roemer (2002), 'Equality of opportunity: A progress report'; and Joseph Fishkin (2014), 'Bottlenecks: A New Theory of Equal Opportunity'.

In this way, Stoyanovich's course integrates principles of liberal education by emphasising critical thinking and ethical reasoning within its curriculum, particularly in the first module on Algorithmic Fairness, which explores complex concepts such as intersectional discrimination and equality of opportunity through philosophical frameworks. This approach develops a deeper understanding of ethical implications in data science and encourages students to engage with diverse perspectives.

Buijsman et al. (2025) make a compelling case that although there is a pressing need to find concrete and justifiable answers to the problems posed by AI, we need not reinvent the wheel:

> When one wants to tackle these ethical challenges, the first place to look is the vast philosophical literature centred around the main ethical theories. We have millennia of thinking about right and wrong action (p. 68).

They highlight the three main ethical theories from the history of ethics in philosophy (virtue ethics, consequentialism and deontology) and also examine how these theoretical perspectives can be trade-offs: each championing one particular type of value at the expense of other types. For example, some take agent relative perspectives into account, but others disregard the individual's perspective and consider the agent's place in a social network or champion a universalistic perspective.

Buijsman et al. (2025) ask us to consider how education might foster essential virtues like recognising the broader social implications of technology, prioritising the public good and being attentive to the needs of others. Their perspective also encourages us to examine whether the process leading to the development of an AI system was guided by ethical considerations. For instance, did diverse stakeholders have genuine opportunities to influence important design

decisions, reflecting principles found in value sensitive design and participatory design methodologies?

Using the lens of consequentialism, Buijsman et al. (2025) argue that it is important to consider the consequences of developing an AI system, just as it is important for those involved in the operation of the system to consider the consequences of the individual decisions made once the AI is up and running. Deontology is complex because it often lacks clear guidance for resolving competing rights, such as privacy versus protection from harm. The authors argue that all three ethical theories are too 'coarse-grained' to be of use at an abstract, general level, and rather need to be applied to specific use cases and then built up:

> We need an idea of how we go from the philosophical, conceptual, analysis to the design of a specific AI system. For that, the (relatively recent) design approaches to (AI) ethics are crucial. They require input from all the different parts of philosophy mentioned in this section but add to that a methodology to make these ethical reflections actionable in the design and use of AI (p. 72).

This aligns with the views of most of the data scientists that I interviewed. They did not all consider broad philosophical ethical theories useful, but they all made strong arguments for the use of examples, case studies, learning by observation and/or experience, learning about practical ethical issues, and practical application to real-world examples in a contained teaching environment as the ideal way to teach and learn ethics.

In the literature, several authors put forward an Aristotelian virtue ethics approach to ethics for data science and AI – see, for example, Smith and Vickers (2024) and Ober and Tasioulas (2024). Focusing on Aristotelian virtue ethics for helpful guidance has a precedent in medical ethics, as Cohen-Almagor (2017) describes in his paper, 'On the philosophical foundations of medical ethics: Aristotle, Kant, JS Mill and Rawls'. Cohen-Almagor views Aristotle's contribution to medical ethics as pivotal, particularly through his emphasis on justice and responsibility. Aristotle's philosophy suggests that ethical behaviour in medicine stems from the cultivation of virtues that guide practitioners in their duties towards patients. This focus on rationality and moral responsibility has influenced medical ethics, focusing on the need for healthcare providers to act in the best interests of their patients while maintaining respect for their autonomy and dignity.

Smith and Vickers (2024) believe that Aristotelian virtue ethics provides concrete and actionable, flexible guidance, which makes it well-placed to deal with the forward-looking and rapidly changing landscape of life with AI. However, virtue ethics is agent-based rather than action-based,

which means that using virtue ethics requires ‘ensuring that at least some virtuous agents also possess the relevant scientific and technical expertise’ (p. 19). Since virtue ethics does not prescribe a set of rules, it requires role models who can serve as exemplars for those learning to be virtuous. The authors admit this is a challenge, since ‘no system of training can guarantee the production of virtuous agents’. Interestingly, the article concludes with differing views from each author about the practicality of such an approach. Smith argues there is little hope for practical application due to the difficulty in cultivating virtuous agents who can navigate complex ethical landscapes; Vickers is more optimistic, based on her teaching experience, believing that while educating students to live well with AI is highly challenging, there remains some potential for success in instilling virtue ethics through effective moral education.

Ober and Tasioulas (2024) argue that Aristotle's ethical theory offers the most compelling framework for addressing AI challenges and opportunities. They propose human-centred AI ethics grounded in human nature, with a broader conception of ethics than preference satisfaction or wealth maximisation. Their strategy emphasises the connection between ethics and politics, viewing AI systems as tools of human flourishing. It aims to be more encompassing in regulatory imagination than existing state-, market- or rights-based solutions, engaging global AI regulation challenges in a way that is respectful of state sovereignty. They also recommend a new human right in the AI era: the qualified right to human decision-making in particular domains. Aristotelian virtue ethics is aligned with the objectives of liberal education through its emphasis on character development, practical wisdom, rational discourse, civic engagement and interdisciplinary learning. Both liberal education and Aristotelian virtue ethics aim at the formation of thoughtful, ethical human individuals who can contribute valuably to society.

In practice, Aristotelian ethics in data science is about cultivating virtues, such as honesty, humility, rigour and compassion, in the day-to-day work of data scientists. As opposed to being solely motivated by strict regulations or checklist compliance, this approach encourages data scientists to apply practical wisdom (phronesis) and take broader implications of their actions, especially in cases that are not absolute and do not have a clear-cut right or wrong answer. For example, data scientists ought to pause and think about whether the models reflect their professional and personal beliefs, value the limitations of their data, and prioritise human flourishing and well-being when developing and implementing AI systems. This virtue-based approach accommodates a human-centred approach, where data and AI are developed to enhance human potential and to benefit the common good, rather than to optimise efficiency or profit. Nicolson (2008) suggests ethical education should aim at ‘inspiring an interest in ethics; illuminating the general and professional

ethical tools available to resolve issues of professional ethics; illustrating these tools and issues through exposure to situations involving moral dilemmas; and inculcating the habit of identifying, evaluating and caring about ethical issues so that this becomes a more or less spontaneous response in practice' (p. 162).

There is, however, a tension here that I think is worth naming rather than trying to resolve. The Aristotelian tradition holds that ethical formation requires sustained practice, community and habituation over time, the kinds of conditions a virtuous person grows up within and continues to inhabit throughout their working life. As Nicolson (2008) puts it, 'virtue, practical wisdom and a person's overall moral character are gradually developed through actual engagement with moral issues' (p. 157). A curriculum-based training programme, whether at undergraduate level or in-service, simply cannot replicate those conditions. This does not mean that ethics education for data scientists is futile, but it does mean that realistic expectations of what any curriculum can achieve must be matched by attention to the workplace cultures, professional communities and mentorship structures within which data scientists actually develop their ethical sensibilities. Both the literature I have drawn on and the experiences of the data scientists I interviewed point in this direction.

Nicolson (2008) makes a similar argument in the context of legal education, suggesting that university law clinics, in which students work on real cases with mentorship and supervision in an apprenticeship-type model, offer the kind of sustained, situated engagement with moral issues that classroom-based ethics teaching cannot replicate. Data science currently has no equivalent.

7.2 Transitioning from ethics to political consciousness in data science education

While valuable for reflection, relying on ethical codes and principles as training for data scientists is not sufficient for addressing broader social justice concerns. In Chapter 3, I contended that frameworks based in deontology, utilitarianism and virtue ethics shared a broadly liberal-individualist assumption, and that addressing the structural harms that data science can produce requires drawing on critical theory as well. Drawing on critical theory reveals that data science is not a value-neutral or purely technical field, but one deeply entangled with questions of power, justice and social impact. The findings from this research highlight that ethical challenges in data science often stem from structural inequalities and institutional dynamics, rather than isolated individual decisions. For example, interview participants described how organisational cultures,

lack of clear guidelines and the myth of technical neutrality can obscure the broader societal consequences of data science. These insights align with the literature, which argues that current ethics education, which is primarily focused on personal virtue or compliance with professional codes, is insufficient to address the systemic harms that data science can produce. Gray and Witt (2021) argue that when AI ethics content 'fails to equitably and explicitly assign responsibility to actors in the machine learning economy, there is a risk of implicitly reinforcing the status quo of gender power relations and other substantive inequalities' (p. 1). Instead, there is a need for ethics education that foregrounds political consciousness: the ability to recognise, interrogate and challenge the power structures and social injustices embedded in data science practice.

The implications for data science education are twofold. First, a shift from individualistic approaches to ethics toward a politics of justice is called for, where data scientists are equipped to see themselves as civic actors with the capacity and responsibility to shape more equitable outcomes. This echoes Chapter 3's claim that liberal-individualist frameworks, while necessary, are not sufficient on their own. This means moving beyond technical fixes and compliance checklists to cultivate critical reflection, civic responsibility and ethical reasoning habits. This is an approach supported by Gille and Nardo (2020) writing about the teaching of medical ethics, emphasising that students need to do more than simply understand ethical concepts. Instead, students must be able to critically examine their own beliefs and values and learn to apply them thoughtfully in a variety of real-world situations.

Second, concepts in critical theory suggest that education in data science should explicitly address how algorithms and data infrastructures can reinforce or disrupt existing social hierarchies. This includes teaching students to conduct power analyses of their projects and consider who benefits and who may be harmed. The literature further supports this approach, arguing that attempts to remain apolitical or neutral can preserve the status quo and perpetuate systemic injustices. As Green (2021) asserts:

> Ethics can help data scientists reflect on certain normative aspects of their work, [but] such efforts are ill-equipped to generate a data science that avoids social harms and promotes social justice (p. 249).

In other words, Green says that ethics training can encourage data scientists to think about the values and principles behind their work, but these efforts alone are not enough to ensure that data science prevents social harm and supports social justice.

Green (2021) further argues that data science education should incorporate not only a professional ethics approach but also teach a more politically conscious framework, helping data scientists to recognise themselves as political actors engaged in 'normative constructions of society and evaluate their work according to its downstream impacts on people's lives' (p. 249). In practical terms, educational programmes should incorporate critical analyses of how data science shapes power dynamics, resource distribution and social opportunities. Rather than framing ethical decisions as merely personal or professional choices, educators should help students understand the structural implications of their work.

Grounding data science in a politics of justice addresses three arguments 'that data scientists commonly invoke when challenged to take political positions regarding their work' (Green, 2021, p. 249). The first is the tendency of data scientists to shield themselves behind technical neutrality. As Green argues, the common defence of 'I'm just an engineer' represents a troubling abdication of responsibility. Data science education can actively challenge this stance by developing course content and teaching that exposes the political nature of seemingly technical decisions, from feature selection to algorithm deployment. Training programs should explicitly discuss how attempting to remain 'apolitical' is, in fact, a political stance – and often a conservative one that preserves existing power structures.

The second argument that data scientists invoke, when challenged to take political positions regarding their work, is that it is not their job to take political stances. But, argues Green, their desire for neutrality suffers from two important failings:

> First, neutrality is an unachievable goal, as it is impossible to engage in science or politics without being influenced by one's background, values, and interests. Second, striving to be neutral is not itself a politically neutral position. Instead, it is a fundamentally conservative one (p. 252).

In other words, conducting science in a truly neutral way is impossible. As Green explains:

> Neutrality may appear apolitical, but that is only because the status quo is considered a neutral default. Anything that challenges the status quo—which efforts to promote social justice must do by definition—will therefore be seen as political. But efforts for reform are no more political than efforts to resist reform or even the choice simply to not to act, both of which preserve existing systems (p. 253).

Green acknowledges that normative ideals can be complicated and sometimes clash with each other and that personal beliefs might change over time. He argues that data scientists do not need to have

clear answers to every political issue but should make decisions based on their stated values and be willing to deal with the uncertainty that comes with these ideals.

An anonymous computer science student articulates this well, describing their experience of an interdisciplinary course on ethics, delivered by teachers from the Philosophy department:

> As we were debating which of these criteria (if any) is just, the realms of probably theory and moral philosophy blurred in a way that felt quite disorienting to me, given how compartmentalised I experienced these two academic disciplines most of the time. Our conversations in case certainly dispelled, in my eyes, the notion that technology is, or even can be, neutral, and therefore challenged the idea that practitioners could evade responsibility by somehow not putting our "thumb on the scales" (Anonymous, 2021, p. 1).[20]

The student realised that technology is not and cannot be truly neutral, which made them question the idea that professionals can avoid responsibility simply by claiming they are not influencing outcomes – or by reverting to a 'neutral' stance, which, as Green argues, is inherently conservative and preserves the status quo.

The third argument that data scientists invoke, when challenged to take political positions regarding their work, is that 'we should not let the perfect be the enemy of the good', in which data scientists fall back on a seemingly pragmatic position: because data science tools can improve society in incremental but important ways, society should support the development of such tools regardless, rather than argue about what a perfect solution might be. Green counters that several underdeveloped ideas limit this viewpoint. To begin with, data science does not have strong, well-defined theories about what counts as 'perfect' or even 'good'. Because of this, the field often relies on shallow reforms, making broad and sometimes self-evident statements about which social outcomes are desirable. Additionally, this perspective does not clearly explain how to assess or balance the pursuit of perfection with achieving what is merely good. As a result, attempts to advance social good often assume that gradual, technology-driven changes are the right path for societal improvement. However, when viewed through the lens of genuine equality and anti-oppression, many data science initiatives that claim to promote social good are not achieving those aims in a meaningful or consistent way.

[20] This article was accepted as part of a *Journal of Social Computing* special issue on 'Technology ethics in action: Critical and interdisciplinary perspectives'. However, it had to be withdrawn because the journal could not publish the article with an anonymous author.

In summary, Green makes a strong case for grounding data science in a politics of justice. Ethics can help data scientists think about the moral aspects of their work, but relying on ethics alone is not enough to prevent social harms or advance social justice. Data scientists play a significant role in shaping society and must recognise that their work has political implications. To truly address social justice, data science must move beyond narrow ethical codes and actively engage with the broader social and political impacts of its practices.

Benthall and Goldenfein (2020) also highlight that data science is deeply political and that addressing social justice requires moving beyond narrow ethical codes to engage with broader social and political impacts. The authors argue that the dominant frameworks for data science ethics, rooted in liberal law and philosophy, are fundamentally limited. This is because these frameworks focus on protecting individual autonomy (through privacy and consent), rationality (by preventing manipulation) and property (by treating data as a market commodity), but fail to address structural power, cannot keep pace with technological change and overlook collective and systemic harms.

The frameworks fail to address structural power because liberal ethics centres on the individual, but data science operates at scales and through mechanisms like platforms and algorithms that far exceed individual agency, producing new forms of power and control that traditional ethics cannot contain. Further, dominant frameworks for data science ethics cannot keep pace with technological change, not just because law and ethics are 'catching up' to technology, but they are working with fundamentally different assumptions and materials from those of contemporary data science. Finally, these frameworks fail to consider collective and systems harms, because by focusing on individual rights and harms, liberal ethics misses the broader, collective and often opaque social harms produced by data-driven platforms, including information asymmetries, market distortions and the privatisation of public goods.

Benthall and Goldenfein (2020) argue that data science should be understood as a 'techno-political and techno-economic phenomenon' (p.1) rather than merely a technical or ethical field. They explain that, over the twentieth century, developments in computer science and cognitive psychology, combined with neoliberal legal frameworks, have led to the supremacy of private corporations over individuals who would know and defend their own individual interests. In this context, platforms have 'inverted the relationship between individuals and the market, making the former public and the latter private' (p.1).

The authors emphasise that platforms act as political actors since data science is typically practised within private companies to reshape the traditional roles of individuals and markets, with implications for power and governance. They contend that data science has made liberal frameworks outdated by enabling new collective rationality and control forms that bypass individual autonomy. This shift has profound political implications, as it redistributes power and control in society. The authors further argue that data science has rendered liberalism (and its ethical frameworks) obsolete as a mode of governance. Finally, the paper calls for new ethical and legal theories that emerge from within data science, rather than being imposed from outside via outdated liberal concepts (Benthall & Goldenfein, 2020).

In summary, the authors argue that data science is inherently political, and therefore, its impacts extend beyond individual moral choices. They conclude that data science ethics 'cannot be addressed from within liberalism; it requires new theory that builds on data science itself' (p. 1). This means that achieving social justice requires moving beyond narrow ethical codes to confront the structural, organisational and political dimensions of data science practice.

The data scientists I interviewed also highlighted the structural, organisational and political dimensions of data science practice. Chapter 5 shows how the data scientists I interviewed mentioned several points related to the role of ethics, power and social impact in data science education. For example, Data Scientist 4 highlighted how practitioners often 'forget about the person on the end of it...or whoever it is that you are affecting', focusing instead on 'just coding and numbers'. Similarly, Data Scientist 1 noted they 'do not really think about the societal ecosystem in which you are applying these models'. This disconnect between technical work and its social impact highlights the need for a critical lens.

Further, the data scientists' testimonies reveal a professional environment where ethical indifference is structurally embedded. Most received minimal ethics training. Data Scientist 3 described their ethics course as 'no credit, like three lectures, essentially "do not do bad things" in nice language'. This aligns with practices that preserve the status quo rather than challenging power structures. The 'woeful unpreparedness' expressed by Data Scientist 1 reflects the systemic neglect of ethical education, which can be emblematic of neoliberal education approaches. Overall, the data scientists' recognition that their field requires integration of technical expertise with ethical reasoning supports

a call for an educational approach that transcends disciplinary boundaries to address systemic issues.

As highlighted in Chapter 3, the rapid pace of technological development, the opacity and scale of algorithms and the profound societal impacts of data-driven decisions mean that ethical issues in data science are both pervasive and urgent. Data scientists routinely face dilemmas related to bias, fairness, privacy and the well-being of data subjects and practitioners. Yet many data scientists report feeling underprepared to navigate these challenges.

Ethics training could involve prep work that focuses on the relational, political and power-laden quality to better prepare data scientists to navigate the difficult ethical landscape that they will be facing. Taking a cue from medical ethics models, these could involve simulation immersion, interdisciplinary exposure to adjacent disciplines (e.g., environmental justice), and role-playing exercises in adversarial settings. As Gille and Nardo (2020) describe, these approaches work less as instruments of testing and more as prompts for discussion, encouraging students to challenge their own values against the opinions of others. This context puts ethics education less in the realm of compliance training and more as praxis, a dialectic of thinking and doing required to counteract the structural damages enabled by illiberal data science practice.

Ethical training for data scientists should not be treated as a once-off 'vaccine' – a single course or workshop to be completed and then forgotten. Stand-alone ethics courses risk commodifying ethics as a technical skill, an apolitical practice that preserves the status quo. Instead, ethical training for data scientists should foster critical consciousness, not compliance. Data science education should prioritise critical activism over procedural ethics. To operationalise these insights, practical recommendations for embedding political consciousness and justice into data science education could involve interdisciplinary teaching with the humanities and social sciences, case-based learning that foregrounds real-world power dynamics and reflective practice exercises that challenge students to identify and question the political assumptions underlying technical decisions. By integrating these elements, data science education can move toward a more justice-oriented and socially engaged practice, preparing practitioners not only to identify and mitigate bias or privacy risks but also to interrogate and transform the broader structures that shape their work.

Summary

While many data scientists may receive training in ethical principles, they often remain unprepared to address the full spectrum of ethical challenges that arise in real-world practice. In truth, few people are. However, the recommendations presented in this chapter point toward a reorientation of data science education, moving from a technical focus enhanced with ethical teaching towards a politically conscious practice grounded in social justice commitments. This reorientation addresses all three research questions by affirming that ethics and ethical conduct should be introduced into in-service training for data scientists, identifying liberal education concepts and practices that can inform ethical data science, but further arguing that this training must go beyond traditional professional ethics to include political consciousness.

My findings suggest that institutions should take proactive steps to embed ethical considerations throughout all levels of data science training and practice. This includes developing comprehensive ethics policies, providing ongoing training and support and creating spaces for open dialogue about ethical challenges. Institutions should also ensure that ethical responsibilities are articulated and integrated into organisational processes, from project design to implementation and evaluation. By embracing their role as custodians of ethics, institutions can help to create environments where data scientists are empowered to act with integrity and are supported in addressing the complex ethical and social implications of their work.

I have also given practical recommendations for course content and teaching methods, such as interdisciplinary teaching and teachers, learning by doing, and case studies. Some recommendations for how training programmes for data scientists could stimulate a shift towards political consciousness include: incorporating specific modules on ‘Data Science as Political Action’ that analyse case studies of data systems and their societal impacts, including reflective practice exercises where students identify the political assumptions underlying technical decisions, inviting speakers from communities affected by data science decisions to provide firsthand accounts of impacts and requiring students to conduct power analyses of proposed data science projects, identifying who benefits and who might be harmed by different approaches.

In summary, the recommendations presented here aim to cultivate a data science profession that is not only technically proficient but also socially responsible and attuned to the demands of justice in a digital world. These proposals lay the groundwork for the concluding chapter, synthesising the

key findings, reflecting on their implications for professional practice, and outlining directions for future research and reform in data science ethics education.

Chapter 8: Conclusions

My research study has focused on the ethical and political dimensions of data science practice and education, highlighting the lived experiences of data scientists and the adequacy of current ethics education. Specifically, in my research, I wanted to investigate the overarching question: How do data scientists experience and respond to ethical challenges in their work, and how can these experiences inform the development of ethics education that addresses structural and political dimensions of data science? I examined whether education in and for ethics and ethical conduct in data science should be introduced into in-service training for data scientists. In my research, I aimed to answer the following broad questions:

1. How do in-service data scientists experience and respond to ethical challenges in their work?
    a. What role, if any, has formal ethics education played in shaping their responses?
    b. What gaps do data scientists identify in their existing training regarding ethics, and how do they articulate the need for structured ethical and political education within the field?
2. How might concepts from liberal theory – such as critical thinking, civic responsibility and ethical reasoning – inform the design and delivery of ethics education in data science?
3. What are the perceived needs and potential benefits of incorporating ethics and ethical conduct into in-service training for data scientists?
4. What are the perceived challenges and possibilities for incorporating political and social justice considerations into in-service ethics training for data scientists?

8.1 Summary of key findings

In addressing the overarching research question, 'How do in-service data scientists experience and respond to ethical challenges in their work?', the interviews provide a detailed understanding of the ethical challenges faced by in-service data scientists and the implications for ethics education in the field. The insights from the interviews reveal that data scientists frequently encounter complex ethical dilemmas in their daily work, ranging from issues of data privacy and informed consent to bias, discrimination, and concerns about the unintended consequences of algorithmic decision-making. These findings are discussed in detail in Chapter 5, particularly in *Theme 2: Ethical challenges encountered in practice*, and are further analysed in Chapter 6, particularly *6.2 Ethical challenges encountered in practice*.

The data scientists I interviewed reported feeling 'woefully unprepared' to address ethical dilemmas, attributing this unpreparedness to the lack of formal ethics teaching in their education. Where ethics was included, it was often limited to standalone modules perceived as superficial or irrelevant. Ethical reasoning was usually developed informally, through workplace experience or personal initiative, rather than structured education. This is explored in Chapter 5, *Theme 2: Ethical challenges encountered in practice*, particularly *5.2.1 Feeling unprepared to grapple with ethical issues* and further discussed in Chapter 6, particularly *6.2 Ethical challenges encountered in practice.*

Because the data scientists I interviewed often relied on informal, on-the-job learning or personal initiative to develop their ethical reasoning, they articulated the need for structured, ongoing, politically informed ethics education that moves beyond technical compliance to address real-world complexities. Regarding research question 1b, 'What gaps do data scientists identify in their existing training regarding ethics?', my research exposes significant gaps in current training, with data scientists consistently identifying the need for more comprehensive, practical and context-sensitive ethics education. This aligns with the literature, which warns that without meaningful ethics education, data science risks perpetuating social harm and eroding public trust. These points are discussed in Chapter 5, particularly *Section 6.4 The role of ethical education (including suggestions for ways of teaching data science ethics),* as well as in Chapter 6, particularly *Section 6.4 The role of ethical education and teaching data science ethics.*

Regarding research question 2, 'How might concepts from liberal education – such as critical thinking, civic responsibility and ethical reasoning – inform the design and delivery of ethics education in data science?', my findings highlight the value of critical reflection and reflexive practice. Both the literature and data scientists emphasise the importance of teaching data scientists to navigate ambiguity and complexity, see themselves as civic actors with societal and political responsibilities, and develop robust ethical reasoning skills through exposure to diverse frameworks and real-world case studies. This is discussed in Chapter 5, particularly *Section 5.4.2 Suggestions on how best to teach ethics to data scientists,* in Chapter 6, *Section 6.4 The role of ethical education and teaching data science ethics*, and with further recommendations in Chapter 7, *Section 7.1 Recommendations for education in and for ethics and ethical conduct in data science.*

My research also demonstrated that an effective approach to integrating liberal education principles into ethics training for data scientists could involve several strategies. First, ethics should be

embedded throughout training content, rather than being relegated to standalone modules, ensuring that ethical reflection becomes part of everyday practice (Chapter 7, Section 7.1). Second, the use of case studies and real-world examples helps make ethical issues tangible and relevant, allowing data scientists to grapple with the complexities they are likely to encounter in their professional lives (Chapter 5, Section 5.4, and Chapter 7, Section 7.1). Third, promoting interdisciplinary teaching, particularly by involving the humanities, enlarges perspectives and deepens ethical reasoning, as these disciplines offer valuable frameworks for ethical analysis (Chapter 6, Section 6.4, and Chapter 7, Section 7.1). Finally, encouraging critical thinking and reflexive practice as ongoing habits can help data scientists to develop the capacity to navigate ambiguity and complexity by cultivating continual ethical awareness and self-examination (Chapter 5, Section 5.4 and Chapter 7, Section 7.1).

Regarding research question 3, 'What are the perceived needs and potential benefits of incorporating ethics and ethical conduct into in-service training for data scientists?', there is a clear and urgent need for comprehensive, embedded and interdisciplinary ethics education that not only addresses technical compliance but also foregrounds the political and social justice dimensions of data science practice. The potential benefits of such an approach include better preparation for real-world ethical challenges, increased public trust, and the cultivation of data scientists who are both technically proficient and socially responsible. These points are discussed in Chapter 6 and Chapter 7.

Regarding research question 4, 'What are the perceived challenges and possibilities for incorporating political and social justice considerations into in-service ethics training for data scientists?', my research demonstrated that the persistent belief that data science is value-neutral prevents critical engagement with the political and social dimensions of the field. Differences in ethical perspectives across regions and cultures complicate the development of universal frameworks, and organisational inertia and lack of clear guidelines can hinder the integration of social justice into ethics training. Moreover, the pace of innovation, especially with gen AI, currently outstrips the development of ethical standards and regulatory frameworks. These challenges and possibilities are explored in Chapter 5, *Section 5.5 Theme 5: Implications of gen AI and emerging technologies,* Chapter 6, *Section 6.5: Implications of gen AI and emerging technologies*, and Chapter 7, *Section 7.2 Transitioning from ethics to political consciousness in data science education.* My research advocates for a shift from individual virtue and compliance to a politics of justice, where data scientists are equipped to interrogate power structures and challenge

systemic harms. Continuous learning and reflective practice, by data scientists themselves, are essential for keeping pace with evolving ethical challenges.

In summary, my research demonstrates that data scientists face pervasive and complex ethical challenges for which current education is mostly inadequate. I argue that the current state of ethics education for data scientists is not just insufficient – it is fundamentally misaligned with the realities and responsibilities of the profession. The evidence from my research is unequivocal: data scientists are routinely confronted with ethical dilemmas that are deeply entangled with questions of power, justice and social impact, yet their training remains overwhelmingly technical, fragmented and detached from these broader concerns. This is not a minor gap. It is a structural failing that leaves practitioners unprepared to navigate the profound consequences of their work.

What is needed is a radical reimagining of ethics education in data science – one that is comprehensive, embedded and staunchly interdisciplinary. Ethics cannot be relegated to a single module or a box-ticking exercise at the margins of a technical curriculum. Instead, ethical reflection and critical engagement should be woven throughout every stage of training, from the very first encounter with data all the way through to the deployment of models in the real world. This means moving beyond compliance and technical fixes, and instead cultivating the habits of critical thinking, civic responsibility and ethical reasoning, which are the hallmarks of a liberal education.

I go further and contend that ethics education should explicitly foreground the political and social justice dimensions of data science practice. Data scientists must be equipped not only to identify and mitigate bias or privacy risks, but also to interrogate the power structures their work sustains, to challenge the myth of neutrality and to see themselves as civic actors with the capacity and the obligation to shape more just and equitable outcomes. This requires drawing on the insights of philosophy and social sciences and demands case-based pedagogies rooted in real-world complexity. Anything less risks perpetuating the very harms and injustices that data science has the potential to address.

8.2 Contributions to my professional practice

This research has generated key insights relevant to my role in creating and curating data science curricula. Conducting interviews and engaging with concepts from critical and liberal theory has profoundly deepened my awareness of the ethical and political implications of data science,

prompting ongoing self-reflection on my teaching and leadership decisions. Drawing on the insights from the interviews and the literature, I have integrated ethics throughout the technical modules of our programmes, including case studies, practical ethical dilemmas and interdisciplinary perspectives across not only the data science curriculum, but our other curricula too.

Recognising the discomfort many technical educators feel with ethics content, I have suggested workshop sessions with faculty from the humanities to co-develop and co-teach ethics modules for data scientists. This research has also motivated me to advocate for clearer institutional guidelines and support structures for ethical decision-making, both in building curricula and in our own organisational practices. Further, I am trying to develop my own daily critical thinking and reflexive practice habit, so that I am not only 'telling' data scientist teachers what to do, encouraging them to see themselves not only as technical experts but also as political actors whose work shapes society, but also living my own suggestions.

8.3 Broader implications for practice

My findings hold implications for the broader landscape of data science education and professional practice. One of the most pressing insights is the need to integrate ethics education throughout the entire data science curriculum. Rather than relegating ethical considerations to isolated modules, ethics should be interwoven with technical content, using real-world case studies, experiential learning opportunities and content and teaching from interdisciplinary teaching teams. This approach stimulates critical reflection and ethical reasoning to help data scientists navigate the complex moral terrain they encounter in practice.

Institutional accountability also emerges as an important consideration. Organisations must develop explicit, actionable guidelines and robust support structures to help practitioners address ethical dilemmas. At the same time, it is essential to empower individuals within these organisations to exercise moral judgement, ensuring that ethical responsibility is not simply outsourced to policy documents but is actively embodied in daily practice.

Moreover, the research underscores the importance of explicitly addressing the political and social justice dimensions of data science, challenging the persistent myth of neutrality and encouraging practitioners to interrogate power structures and recognise the potential for data-driven systems to perpetuate or exacerbate systemic harms. By foregrounding these issues, data science education can

cultivate a generation of professionals who are not only technically proficient but also critically engaged with the broader societal impacts of their work.

As technological advances continue to reshape the field, ongoing professional development and reflective practice become vital. Ethical challenges will inevitably evolve alongside new tools and methodologies, making it necessary for practitioners to engage in continuous learning to maintain high ethical standards. In addition, increasing diversity and representation within data science teams is critical, as a wider range of perspectives can drive more inclusive ethical decision-making, where the impacts of data-driven systems are more equitably distributed.

Finally, with the rapid rise of AI and other emerging technologies, the need for ethical education extends beyond data scientists to encompass all AI users. Equipping society with the knowledge and tools to navigate new risks and responsibilities is essential for cultivating a more just, trustworthy and socially responsible digital future.

The findings from this study suggest practice implications in relation to data science education. For example, the finding that data scientists' implicit moral reasoning draws on recognisable ethical traditions without an awareness of doing so suggests that ethics education need not start from scratch. Curricula could examine realistic scenarios, and ask students what they would do (or, in executive education, what they may already have done) and then introduce ethical theory as a lens for examining those decisions. The ethics professor could make explicit the implicit frameworks that the students used and ask them to consider and reflect upon the merits or demerits of each.

The finding in my research that data scientists develop ethical sensibility through workplace socialisation (for example, Data Scientist 7's previous boss and Data Scientists 9's ethics committee experience) suggests that ethical education for data scientists should include structured mentorship and exposure to ethical role models. This could include internship placements with organisations with recognised ethical cultures, pairing students with practitioners who can model ethical reasoning.

The tension my findings revealed between the fact that data science is usually done in teams in organisational settings but that the interviewees' responses reflected individualistic reasoning suggests that ethics education for data scientists should explicitly teach students how to navigate ethical disagreement in teams when power relations are unbalanced. This means practicing

structured ethical deliberation, learning how to see when colleagues are operating with different ethical frameworks and developing protocols for resolving (or learning to live with) ethical disagreement.

Data Scientists who framed ethical decisions as their organisation's remit exhibit 'ethical fading' when they reframed ethical decisions as matters of workload and process ('path of least resistance' and 'above your pay grade'). This finding suggests that ethical education curricula should directly address the issue of 'ethical fading', including teaching how to recognise when and how ethical issues are reframed as technical or operational ones. This could include exercises similar to those suggested by Bazerman and Tenbrunsel (2011) that are specifically designed to illustrate ethical fading and practice identifying the moment when the moral dimension fades out of a decision.

Data Scientist 4 highlighted how they have 'never seen [ethics] built into a workflow', which suggests that ethical education should include practical training on how to embed ethical checkpoints across data science workflows, at data collection, feature selection, model building, testing and implementation stages.

Consider a concrete example, taking the car commercial example as a case study. An ethical education curriculum could ask students to break into groups to consider how they might approach using race as a variable. In role plays, students could be assigned managerial and operational titles with differing organisational goals (for example, profit maximisation, most effective ads, best technical model and so forth). The teams could then develop decision protocols, identifying where ethical fading might enter, and highlighting ethical check points sit in each of the workflows. In the report back plenary, an experience ethics professor could highlight the ethical frameworks implicit in their choices and task them with creating a reflection journal wherein they consider the decision points from different ethical framework perspectives. The professor could also ask students to reflect on who benefits and who is harmed by each decision protocol, examining how their choices might reinforce or disrupt existing power relations. In this way, a case study can address the dynamics that my research findings highlight: implicit moral frameworks, ethical fading and the gap between individual reasoning and team-based decision making, at the same time moving the students towards a political consciousness of how their technical choices distribute power and shape social outcomes.

8.4 Limitations and recommendations for future research

While my study was primarily qualitative and focused on a specific group of in-service data scientists, which may limit the generalisability of the findings, this limitation also highlights a rich landscape for future research, which could be ambitious in both scope and method. For example, the sample could be broadened to include data scientists from a broader range of regions, industries, and cultural contexts, especially those underrepresented in current literature, such as practitioners in the Global South, the public sector and grassroots technology initiatives. Comparative studies across regulatory environments and organisational cultures could illuminate how context shapes ethical reasoning and practice. Quantitative studies could complement these insights by measuring the impact of different ethics education interventions on practitioner behaviour and outcomes. For example, large-scale surveys and experimental studies could systematically measure the effects of various ethics education interventions, such as embedded curricula, interdisciplinary teaching or case-based learning, on practitioner behaviour, ethical decision-making and organisational outcomes. Longitudinal studies could track how ethical awareness and agency develop over time, and whether interventions lead to sustained changes in practice.

A further limitation concerns what any curriculum can realistically achieve. The moral education literature I have briefly drawn on suggests that ethical formation requires sustained community, shared practice and habituation, the kinds of conditions that modular, assessment-driven higher education does not easily provide. This does not mean that ethics education for data scientists is futile, but it does mean that the recommendations I make in Chapter 7 need to be read alongside the recognition that workplace learning, professional communities and organisational culture may matter as much as, or more than, formal teaching. The data scientists I interviewed bear this out: it was workplace exposure, mentorship and concrete cases encountered at work, much more strongly than their university curricula, that shaped their thinking on ethical issues.

There is also a political dimension to the recommendations themselves, in that decisions about what an ethics curriculum teaches and what it leaves open, are not just pedagogical or epistemological questions but reflect and reproduce power relations. Who decides that a particular ethical claim is 'settled'? On what authority, and in whose interests? These are questions that the critical theory framing of this dissertation invites, and that any curriculum reform of the kind I have argued for would need to take seriously.

Building on the findings of this dissertation, I advocate for research that moves beyond description and suggestion to practical intervention and evaluation. Action research and participatory approaches could directly involve data scientists, educators and affected communities in co-designing and testing new models of ethics education and organisational governance. For example, a justice-oriented, interdisciplinary ethics curriculum could be built as a pilot and tested in both university and in-service settings, with a focus on embedding critical reflection and political consciousness throughout technical training.

Given the rapid evolution of generative AI and other emerging technologies, future research must proactively address new ethical frontiers. This includes investigating the unique challenges of teaching and regulating ethics in the context of gen AI, where downstream applications and risks are difficult to anticipate. Further future research could look at evaluating the societal impacts of AI systems in high-stakes domains (for example, healthcare, criminal justice and education), with particular attention to issues of bias, accountability and the erosion of human agency. A particular focus for future research must be on investigating the global governance of AI, including the development of transnational ethical standards, participatory regulatory frameworks and mechanisms for including marginalised voices in decision-making. Only through such ambitious inquiry and intervention can the field of data science move beyond technical compliance to become a force for social responsibility and transformative change.

8.5 Final reflections

The stakes for getting ethics in data science wrong are high. Data science shapes the decisions that govern our lives: who gets a job, who receives medical care, who is under surveillance, who is heard and who is silenced. When ethical considerations are sidelined, the consequences are not abstract but devastatingly real. We have already seen algorithms amplify discrimination, entrench inequality and erode public trust. The unchecked deployment of data-driven systems has led to wrongful arrests, denied opportunities and the deepening of social divides. Suppose we continue to treat ethical education for data scientists as an afterthought or implement it as a ‘vaccine’. In that case, we risk building a world where injustice is automated and harm is scaled at the speed of computation.

This research demonstrates that technical proficiency alone is not enough. To realise the field’s potential for public good, data science must encompass not only technical excellence and ethics, but also collective political consciousness: an awareness that every technical choice is embedded and

contestable in relations of power, governance and justice. Data scientists must be equipped with the critical, ethical and political awareness necessary to navigate the profound societal impacts of their work. The evidence is clear: current approaches to ethics education fail to prepare practitioners for the complexity and gravity of the decisions they face. The findings of my dissertation highlight a need for a considerable shift in how ethics is taught and practised in data science. By embedding ethics education throughout the curriculum, supporting interdisciplinary collaboration and embracing a politics of justice, the field can move toward a more responsible, inclusive and socially engaged practice. Only through such engagement can data science challenge, rather than reinforce, the inequities of our datafied world.

The importance of this work extends beyond the classroom or the lab. The future of data science is inextricably linked to the future of democracy, equity and human dignity. If we get this wrong, we risk ceding control of our most vital institutions to opaque systems that reflect and reinforce the biases and power imbalances of the past. But suppose we get it right by nurturing a data science profession grounded in ethical reflection, civic responsibility and social justice? In that case, we can harness the power of data science to create a more just, inclusive and humane world for future generations.

List of references

Alexander, L., & Moore, M. (2021). Deontological ethics. In E. N. Zalta (Ed.), *Stanford encyclopaedia of philosophy* (Winter 2021 ed.). Stanford University. https://plato.stanford.edu/archives/win2021/entries/ethics-deontological/

Andersen, M. L., & Klamm, B. K. (2018). Accounting students' ethical decisions: The influence of intuition and social interaction. *Journal of Accounting Education, 44*, 35–46.

Angwin, J., & Larson, J. (2016, December 30). Bias in criminal risk scores is mathematically inevitable, researchers say. *ProPublica.* https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say

Appel, F. (2005). Ethics across the computer science curriculum: Privacy modules in an introductory database course. *Science and Engineering Ethics, 11*(4), 635-644. https://doi.org/10.1007/s11948-005-0031-2

Bartneck, C., Lutge, C., Wagner, A., & Welsh, S. (2021). *An introduction to ethics in robotics and AI.* Springer.

Baumer, B. S., Garcia, R. L., Kim, A. Y., Kinnaird, K. M., & Ott, M. Q. (2022). Integrating data science ethics into an undergraduate major: A case study. *Journal of Statistics and Data Science Education, 30*(1), 15-28. https://doi.org/10.1080/26939169.2022.2038041

Bazerman, M. H., & Tenbrunsel, A. E. (2011). *Blind spots: Why we fail to do what's right and what to do about it*. Princeton University Press.

Benthall, S. & Goldenfein, J. (2020). Data science and the decline of liberal law and ethics. *SSRN.* http://dx.doi.org/10.2139/ssrn.3632577

Bietti, E. (2021). From ethics washing to ethics bashing: A moral philosophy view on tech ethics. *Journal of Social Computing, 2*(3), 266-283. https://doi.org/10.23919/JSC.2021.0031

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: Handbook I: Cognitive domain.* David McKay.

Boddington, P. (2023). *AI ethics: A textbook*. Springer. https://doi.org/10.1007/978-981-19-9382-4

Bohman, J. (2021). Critical theory. In E. N. Zalta (Ed.), *Stanford encyclopaedia of philosophy* (Winter 2021 ed.). Stanford University. https://plato.stanford.edu/archives/spr2021/entries/critical-theory/

Borenstein, J. & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI Ethics, 1,* 61–65. https://doi.org/10.1007/s43681-020-00002-7

Borg, J. S., Sinnott-Armstrong, W., & Conitzer, V. (2024). *Moral AI: And how we get there.* Pelican.

Braun, V. & Clarke, V. (2019, April). Thematic Analysis: A reflexive practice. https://www.psych.auckland.ac.nz/en/about/thematic-analysis.html

Bride, B. E. (2007). Prevalence of secondary traumatic stress among social workers. *Social Work (New York), 52*(1), 63-70. https://doi.org/10.1093/sw/52.1.63

Bridges, D. (1992). Enterprise and liberal education. *Journal of Philosophy of Education*, *26*(1), 91-98. https://doi.org/10.1111/j.1467-9752.1992.tb00267.x

Bridges, D. (1997). *Education, autonomy, and democratic citizenship: Philosophy in a changing world.* Routledge.

Brinkmann, S. (2018). The interview. In Denzin, N. K., & Lincoln, Y. S (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 997-1038). SAGE Publications, Inc.

Brock, G. (2022, February 22). Liberalism. *The Stanford encyclopaedia of philosophy.* https://plato.stanford.edu/entries/liberalism/

Buijsman, S., Klenk, M., & van den Hoven, J. (2025). Ethics of AI: Toward a "Design for Values" Approach. In N. A. Smuha (Ed.), *The Cambridge handbook of the law, ethics and policy of artificial intelligence* (pp. 59–78). Cambridge University Press.

Burton, E., Goldsmith, J., & Mattei, N. (2018). How to teach computer ethics through science fiction. *Communications of the ACM, 61*(8), 54-64. https://doi.org/10.1145/3154485

Carr, D. (2007). Character in teaching. *British Journal of Educational Studies, 55*(4), 369–389. https://doi.org/10.1111/j.1467-8527.2007.00386.x

Chouliara, Z., Hutchison, C., & Karatzias, T. (2009). Vicarious traumatisation in practitioners who work with adult survivors of sexual violence and child sexual abuse: Literature review and directions for future research. *Counselling and Psychotherapy Research, 9*(1), 47-56. https://doi.org/10.1080/14733140802656479

Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology*, *12*(3), 297-298. https://doi.org/10.1080/17439760.2016.1262613

Cohen-Almagor, R. (2017). On the philosophical foundations of medical ethics: Aristotle, Kant, JS Mill and Rawls. *Ethics, Medicine, and Public Health, 3*(4), 436-444. https://doi.org/10.1016/j.jemep.2017.09.009

Colby, A., Ehrlich, T., Sullivan, W. M., Dolle, J. R., Shulman, L. S. (2011). *Rethinking undergraduate business education: Liberal learning for the profession* (1st ed.). Jossey-Bass.

Crawford, K. (2013). The hidden biases in Big Data. *Harvard Business Review.* https://hbr.org/2013/04/the-hidden-biases-in-big-data

Curren, R., & Metzger, E. (2017). *Living well now and in the future: Why sustainability matters*. MIT Press.

Dastin, J. (2018, October 11). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Davenport, T. H., & Patil, D. J. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review, 90*(10), 70-128.

Davis, K.C. (2020, June 22-26). *Ethics in data science education* [Conference presentation]. American Society for Engineering Education's virtual conference. https://doi.org/10.18260/1-2—34589

Dencik, L., Hintz, A., Redden, J., & Treré, E. (2019). Exploring data justice: Conceptions, applications and directions. *Information, Communication & Society, 22*(7), 873–881. https://doi.org/10.1080/1369118X.2019.1606268

Denzin, N. K., & Lincoln, Y. S. (2018). *The SAGE handbook of qualitative research* (5th ed.). SAGE Publications, Inc.

Desai, J., Watson, D., Wang, V., Taddeo, M., & Floridi, L. (2022). The epistemological foundations of data science: A critical review. *Synthese (Dordrecht), 200*(6), 469. https://doi.org/10.1007/s11229-022-03933-2

Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way.* Springer International Publishing. https://doi.org/10.1007/978-3-030-30371-6

Emmerich, N. (2013). *Medical ethics education: An interdisciplinary and social theoretical perspective.* Springer.

Fest, I., Wieringa, M., & Wagner, B. (2022). Paper vs. practice: How legal and ethical frameworks influence public sector data professionals in the Netherlands. *Patterns (New York, N.Y.), 3*(10), 100604. https://doi.org/10.1016/j.patter.2022.100604

Fiesler, C., Garrett, N., & Beard, N. (2020). What do we teach when we teach tech ethics? A syllabi analysis. Paper presented at the 51st ACM Technical Symposium on computer science education. https://doi.org/10.1145/3328778.3366825

Floridi, L. & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical, and Engineering Sciences, 374*(2083), 20160360. https://doi.org/10.1098/rsta.2016.0360

Fotopoulou, A. (2019). Understanding citizen data practices from a feminist perspective: Embodiment and the ethics of care. In H. Stephansen & E. Treré (Eds.), *Citizen media and practice* (pp. 227–242). Routledge.

Gille, F., & Nardo, A. (2020). A case for transformative learning in medical ethics education. *Journal of Medical Education and Curricular Development, 7,* 1–2. https://doi.org/10.1177/2382120520931059

Goldsmith, J. & Burton, E. (2017). *Why teaching ethics to AI practitioners is important.* Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17). https://doi.org/10.1609/aaai.v31i1.11139

Grammarly. (2025). *Grammarly* (Aug 2025 version) [AI writing assistant]. https://app.grammarly.com/

Gray, J., & Witt, A. (2021). A feminist data ethics of care framework for machine learning: The what, why, who and how. *First Monday, 26*(12). https://doi.org/10.5210/fm.v26i12.11833

Green, B. (2021). Data science as political action: Grounding data science in a politics of justice. *Journal of Social Computing, 2*(3), 249-265. https://doi.org/10.23919/JSC.2021.0029

Greene, T. (2020, April 2). What would an ethics of data science look like? *Towards Data Science.* https://towardsdatascience.com/what-would-an-ethics-of-data-science-look-like-e9d4e9ddc2b3

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834.

Haidt, J. (2013). *The righteous mind: Why good people are divided by politics and religion*. Vintage.

Haidt, J., & Bjorklund, F. (2008). Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Volume 2: The cognitive science of morality* (pp. 181–217). MIT Press.

Hale, S. A. (Director). (2017). *What is data science?* [Video/DVD]. SAGE Publications Ltd. https://methods-sagepub-com.ezproxy.lib.gla.ac.uk/video/what-is-data-science

Hammersley, M. (2005). Should social science be critical? *Philosophy of the Social Sciences, 35*(2), 175-195. https://doi.org/10.1177/0048393105275279

Havill, J. (2019, February 27–March 2*). Embracing the liberal arts in an interdisciplinary data analytics program.* [Conference presentation]. SIGCSE 2019, Minneapolis, MN, United States. https://doi.org/10.1145/3287324.3287436

Hofmann, B. (2021). The role of philosophy and ethics at the edges of medicine. *Philosophy, Ethics, and Humanities in Medicine: PEHM, 16*(1), 14. https://doi.org/10.1186/s13010-021-00114-w

Horvitz, E. & Mitchell, T.M. (2024). Scientific progress in artificial intelligence: History, Status, and Futures. In K.H. Jamieson, A.M Mazza, & W. Kearney (Eds.) *Realizing the promise and minimizing the perils of AI for science and the scientific community.* University of Pennsylvania Press.

Hosseini, M., Wieczorek, M., & Gordijn, B. (2022). Ethical issues in social science research employing big data. *Science and Engineering Ethics, 28*(3), 29. https://doi.org/10.1007/s11948-022-00380-7

Hursthouse, R. & Pettigrove, G. (2022). Virtue ethics. In E.N. Zalta & U. Nodelman (Eds.), *Stanford encyclopaedia of philosophy* (Winter 2022 ed.), Stanford University. https://plato.stanford.edu/entries/ethics-virtue/

Husu, J., & Tirri, K. (2003). A case study approach to teachers' ethical dilemmas. *Teaching and Teacher Education, 19*(3), 345–357.

Kincheloe, J.L., McLaren, P., Steinberg, S.R., & Monzó, L.D. (2018). Critical pedagogy and qualitative research: Advancing the bricolage. In Denzin, N. K., & Lincoln, Y. S (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 418-465). SAGE Publications, Inc.

Kohlberg, L. (1958). Development of moral character and moral ideology. *Review of Child Development Research, 1*, 383–431.

Kohlberg, L. (1963). The development of children's orientations toward a moral order: I. Sequence in the development of moral thought. *Vita Humana, 6*(11–12), 11–33.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Rand McNally.

Lazari-Radek, K. d., & Singer, P. (2017). *Utilitarianism: A very short introduction.* Oxford University Press.

Lundie, D. (2024). The ethics of research and teaching in an age of big data. In H. G. Schuetze, W. de Vries, & G. Álvarez Mendiola (Eds.), *Digitalization of higher education* (pp. 86-94). *Journal of Comparative & International Higher Education*, *16*(2). https://ojed.org/jcihe

MacCoun, R. J. (2022). P-hacking: A strategic analysis. In L. Jussim, J. A. Krosnick, & S. T. Stevens (Eds.), *Research integrity: Best practices for the social and behavioural sciences.* Oxford Academic. https://doi.org/10.1093/oso/9780190938550.003.0011

Markham, A. N., & Pereira, G. (2019). Experimenting with algorithms and memory-making: Lived experience and future-oriented ethics in critical data science. *Frontiers in Big Data, 2*, 35-35. https://doi.org/10.3389/fdata.2019.00035

Martens, D. (2022). *Data science ethics: Concepts, techniques, and cautionary tales.* Oxford University Press. https://doi.org/10.1093/oso/9780192847263.001.0001

Merriam-Webster. (n.d.). Algorithm. In *Merriam-Webster.com dictionary.* Retrieved April 30, 2023, from https://www.merriam-webster.com/dictionary/algorithm

Merriam-Webster. (n.d.). Artificial intelligence. In *Merriam-Webster.com dictionary.* Retrieved April 30, 2023, from https://www.merriam-webster.com/dictionary/artificial%20intelligence

Merriam-Webster. (n.d.). Ethics. In *Merriam-Webster.com dictionary.* Retrieved April 30, 2023, from https://www.merriam-webster.com/dictionary/ethics

Mill, J. S. (1879). *Utilitarianism.* Floating Press.

Moats, D., & Seaver, N. (2019). "You social scientists love mind games": Experimenting in the "divide" between data science and critical algorithm studies. *Big Data & Society, 6*(1), 205395171983340. https://doi.org/10.1177/2053951719833404

Molnar, B. E., Meeker, S. A., Manners, K., Tieszen, L., Kalergis, K., Fine, J. E., Hallinan, S., Wolfe, J. D., & Wells, M. K. (2020). Vicarious traumatization among child welfare and child protection professionals: A systematic review. *Child Abuse & Neglect, 110*(Pt 3), 104679. https://doi.org/10.1016/j.chiabu.2020.104679

National Academy of Engineering. (2009). *Ethics education and scientific and engineering research: What's been learned? What should be done?* The National Academies Press. https://doi.org/10.17226/12695

Neff, G., Tanweer, A., Fiore-Gartland, B., & Osburn, L. (2017). Critique and contribute: A practice-based framework for improving critical data studies and data science. *Big Data, 5*(2), 85. https://doi.org/10.1089/big.2016.0050

Nicolson, D. (2008). 'Education, education, education': Legal, moral and clinical. *The Law Teacher, 42*(2), 145–172. https://doi.org/10.1080/03069400.2008.9959773

Nussbaum. M. C. (1997). *Cultivating humanity: A classical defence of reform in liberal education.* Harvard University Press.

Nussbaum, M. C. (2010). *Not for profit: Why democracy needs the humanities*. Princeton University Press.

Ober, J., & Tasioulas, J. (2024). *AI ethics with Aristotle* [White paper]. University of Oxford Institute for Ethics in AI. https://www.oxford-aiethics.ox.ac.uk/lyceum-project-ai-ethics-aristotle-white-paper

Oliver, J.C. & McNeil, T. (2021). Undergraduate data science degrees emphasize computer science and statistics but fall short in ethics training and domain-specific context. *PeerJ Computer Science 7*(441). https://doi.org/10.7717/peerj-cs.441

O'Neil, C. (2017, November 14). The ivory tower can't keep ignoring tech. *The New York Times.* https://www.nytimes.com/2017/11/14/opinion/academia-tech-algorithms.html

O'Neil, C. (2016). *Weapons of math destruction: How Big Data increases inequality and threatens democracy.* Crown.

Paul, K. (2019, October 25). Healthcare algorithm used across America has dramatic racial biases. *The Guardian.* https://www.theguardian.com/society/2019/oct/25/healthcare-algorithm-racial-biases-optum

Perplexity AI. (2025, August 30). Perplexity (Aug 2025 version) [Large language model]. https://www.perplexity.ai/

Poirier, L. (2021). *Reading datasets: Strategies for interpreting the politics of data signification. Big Data & Society,* 8(2), 205395172110293. https://doi.org/10.1177/20539517211029322

Power, F. C., Higgins, A., & Kohlberg, L. (1989). *Lawrence Kohlberg's approach to moral education*. Columbia University Press.

Prinsloo, P., & Slade, S. (2017). Big data, higher education and learning analytics: Beyond justice, towards an ethics of care. In B. Kei Daniel (Ed.), *Big data and learning analytics in higher education: Current theory and practice* (pp. 109–124). Springer International Publishing. https://doi.org/10.1007/978-3-319-06520-5_8

Ratti, E., & Graves, M. (2021). Cultivating moral attention: A virtue-oriented approach to responsible data science in healthcare. *Philosophy & Technology, 34*(4), 1819-1846. https://doi.org/10.1007/s13347-021-00490-3

Riddick F. A., Jr (2003). The code of medical ethics of the American medical association. *Ochsner journal*, *5*(2), 6–10.

Saltz, J. S., & Dewar, N. (2019). Data science ethical considerations: A systematic literature review and proposed project framework. *Ethics and Information Technology, 21*(3), 197-208. https://doi.org/10.1007/s10676-019-09502-5

Saltz, J., Dewar, N., & Heckman, R. (2018, February 21-24). *Key concepts for a data science ethics curriculum.* Paper presented at the 49th ACM Technical Symposium on Computer Science Education, Baltimore, MD, USA. https://doi.org/10.1145/3159450.3159483

Sandel, M. J. (2010). *Justice: What's the right thing to do?* Farrar, Straus and Giroux.

Savulescu, J., Persson, I., Wilkinson, D. (2020). Utilitarianism and the pandemic. *Bioethics, 34*(6), 620-632. https://doi.org/10.1111/bioe.12771

Schostak, J. F. (2006). *Interviewing and representation in qualitative research.* Open University Press.

Shane, S., & Wakabayashi, D. (2018, April 4). 'The Business of War': Google employees protest work for the Pentagon. *The New York Times.* https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html

Shapiro, B., Meng, A., O'Donnell, C., Lou, C., Zhao, E., Dankwa, B., & Hostetler, A. (2020, April 25-20). *Re-shape: A method to teach data ethics for data science education.* [Conference presentation]. Paper presented at the 2020 CHI Conference on Human Factors in Computing

Systems. https://doi.org/10.1145/3313831.3376251

Smith, N., & Vickers, D. (2024). Living well with AI: Virtue, education, and artificial intelligence. *Theory and Research in Education, 22*(1), 19-44. https://doi.org/10.1177/14778785241231561

Smuha, N. A. (2025). An introduction to the law, ethics, and policy of artificial intelligence. In N. A. Smuha (Ed.), *The Cambridge handbook of the law, ethics and policy of artificial intelligence* (pp. 1–14). Cambridge University Press.

Stanley, M. L., Yin, S., & Sinnott-Armstrong, W. (2023). Moral psychology in the classroom: Implications for ethics education. *Journal of Moral Education, 52*(1), 1–16.

Stark, L., & Hoffmann, A. (2019). Data is the new what? Popular metaphors and professional ethics in emerging data culture. *Journal of Cultural Analytics 4*(1), 1-22. https://doi.org/10.22148/16.036

Stoyanovich, J. (2022). Teaching responsible data science. *1st International Workshop on Data Systems Education*.

Swisher, K. (2024). *Burn book.* Little, Brown.

Tacheva, Z. (2022). Taking a critical look at the critical turn in data science: From 'data feminism' to transnational feminist data science. *Big Data & Society, 9*(2), 205395172211129. https://doi.org/10.1177/20539517221112901

Tanweer, A., Bolten, N., Drouhard, M., Hamilton, J., Caspi, A., Fiore-Gartland, B., & Tan, K. (2017). Mapping for accessibility: A case study of ethics in data science for social good. *ArXiv, abs/1710.06882*.

Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society, 4*(2), 1–14. https://doi.org/10.1177/2053951717736335

Touma, R. (2022, July 19). TikTok has been accused of 'aggressive' data harvesting. Is your information at risk? *The Guardian.* https://www.theguardian.com/technology/2022/jul/19/tiktok-has-been-accused-of-

aggressive-data-harvesting-is-your-information-at-risk

Thomasma, D.C. (2004). Theories of medical ethics: The philosophical structure. In Thomasma, D.C. & Graber, G.C. (Eds.), *Theory and practice in medical ethics.* (pp. 23-38). Continuum International Publishing.

Tractenberg, R.E. (2020). *Ten simple rules for integrating ethical reasoning into quantitative instruction.* [Conference presentation]. 2020 Joint Statistical Meetings, Alexandria, VA, United States.

Tripp, D. (1992). Critical theory and educational research. *Issues In Educational Research, 2*(1), 1992, 13-23.

Utts, J. (2021). Enhancing data science ethics through statistical education and practice. *International Statistical Review, 89*(1), 1-17. https://doi.org/10.1111/insr.12446

Van Noorden, R. (2020). The ethical questions that haunt facial recognition research. *Nature. 587*(7834), 354-358. https://doi.org/10.1038/d41586-020-03187-3

Weaver, K., & Olson, J. K. (2006). Understanding paradigms used for nursing research. *Journal of Advanced Nursing, 53*(4), 459-469.

Wolpe, P. R. (2006). Reasons scientists avoid thinking about ethics. *Cell, 125*(6), 1023-1025. https://doi.org/10.1016/j.cell.2006.06.001

Wylie, C. D. (2020). Who should do data ethics? *Patterns (New York, N.Y.), 1*(1), 100015. https://doi.org/10.1016/j.patter.2020.100015

Zinser, E. (2004). Making the case for liberal education. *Liberal Education, 90*(1), 38.

List of accompanying material

Appendix A: Participant Information Sheet

Participant Information Sheet

Study title: An exploration of the ethical issues that data scientists encounter in their field, and how these might be addressed in ethical education.
Researcher name: Nicola Weaver

You are being invited to take part in a research study. Before you decide to take part, it is important for you to understand why the research is being done and what it will involve. Please read the following information carefully and discuss it with others if you wish. Ask the researcher if there is anything that is not clear or if you would like more information. Take some time to decide whether or not you wish to take part.

Thank you for reading this.

This study focuses on a group of data science teachers and their experiences of ethical issues in teaching and practising data science.

As a data science teacher, you are also a data science professional and are, therefore, accustomed to dealing with ethical issues in your daily work. However, in your courses, we focus only on teaching the technical skills involved in data science. This almost exclusively technical focus is not unusual, and mirrors that of most data science degrees, diplomas, and training programmes around the world.

My research will examine the experience that you have of ethical issues in your field, how prepared you feel to grapple with these issues, and – if you feel unprepared – how might this be addressed for future data scientists in ethical education.

In my research, I will examine whether education in and for ethics and ethical conduct in data science should be introduced into in-service training for data scientists. My study will aim to answer the following broad questions:

1. Should education in and for ethics and ethical conduct in data science be introduced into in-service training for data scientists?
2. In what ways – if any – can the concepts in liberal education be bought to bear on ethical issues in data science?
3. In what ways – if any – can these concepts be introduced into in-service training for data scientists?

Participation is voluntary. The estimated time commitment required from you and all participants is between 40-60 minutes for the interview. All interviews will be conducted by me and will be recorded via Zoom recordings. The interviews will be planned to take around 40-60 minutes. If the participant wishes to continue past the 40 minutes, the interview will continue with your permission for no longer than 60 minutes.

When I transcribe the interviews, I will remove all references to participants' names, and instead assign each interview a pseudonym, which I will cross-reference separately to the interviewee's consent form and identification. In this way, I will address the ethical issues of confidentiality and security of data and ensure that the participants will not be able to be identified from their interviews by any reader not deeply familiar with the context.

You have a right to withdraw at any time without prejudice.

The purposes of my research mean that I need not collect much personal data and I will collect as little personal data as possible. I will destroy any sensitive personal data, or directly identifiable personal data, once de-identification has been completed, by shredding any paper documents and deleting electronic files, using secure removal software. These actions will be taken at the end of my research project, by December 2024. De-identified data will be deposited in the University of Glasgow's Enlighten: Research Data server, and may be shared/archived or re-used. This data will be retained and disposed of in line with University protocols after 10 years.

Confidentiality will be respected unless there are compelling and legitimate reasons for this to be breached. If this was the case, I will inform you of any decisions that might limit your confidentiality.

This project has been considered and approved by the College Research Ethics Committee.

To pursue any complaint about the conduct of the research: please contact the College of Social Sciences Lead for Ethical Review, email socsci-ethics-lead@glasgow.ac.uk

_____End of Participant Information Sheet_____

Appendix B: Privacy Notice

PRIVACY NOTICE

Privacy Notice for Participation in Research Project: An exploration of the ethical issues that data scientists encounter in their field, and how these might be addressed in ethical education.

Your Personal Data

The University of Glasgow will be what's known as the 'Data Controller' of your personal data processed in relation to your participation in the research project "An exploration of the ethical issues that data scientists encounter in their field, and how these might be addressed in ethical education". This privacy notice will explain how The University of Glasgow will process your personal data.

Why I need it

I am collecting basic personal data such as your name and contact details in order to conduct my research. I need your name and contact details to arrange interviews.

I will only collect data that I need for the research project and I will de-identify your personal data from the research data (your answers given during the interview, for example) through pseudonymisation.

Please see accompanying **Participant Information Sheet**,

Legal basis for processing your data

I must have a legal basis for processing all personal data. As this processing is for Academic Research, I will be relying upon **Task in the Public Interest** in order to process the basic personal data that you provide. For any special categories data collected, I will be processing this on the basis that it is **necessary for archiving purposes, scientific or historical research purposes or statistical purposes.**

Alongside this, in order to fulfil our ethical obligations, I will ask for your **Consent** to take part in the study. Please see accompanying **Consent Form**.

What I do with it and who I share it with

All the personal data you submit is processed by: me, Nicola Weaver, the researcher on this project.

In addition, security measures are in place to ensure that your personal data remains safe: pseudonymisation, secure storage, and encryption of files and devices. Further, I will destroy all personal data once de-identification has been completed, by shredding any

paper documents and deleting electronic files, using secure removal software. These actions will be taken at the end of my research project, by December 2024.

Please consult the **Consent form** and **Participant Information Sheet** which accompanies this notice.

I will provide you with a copy of the study findings and details of any subsequent publications or outputs on request.

What are your rights?*

GDPR provides that individuals have certain rights including: to request access to, copies of and rectification or erasure of personal data and to object to processing. In addition, data subjects may also have the right to restrict the processing of the personal data and to data portability. You can request access to the information I process about you at any time.

If at any point you believe that the information I process relating to you is incorrect, you can request to see this information and may in some instances request to have it restricted, corrected, or erased. You may also have the right to object to the processing of data and the right to data portability.

Please note that as I am processing your personal data for research purposes, the ability to exercise these rights may vary as there are potentially applicable research exemptions under the GDPR and the Data Protection Act 2018. For more information on these exemptions, please see UofG Research with personal and special categories of data.

If you wish to exercise any of these rights, please submit your request via the webform or contact dp@gla.ac.uk

Complaints
If you wish to raise a complaint on how I have handled your personal data, you can contact the University Data Protection Officer who will investigate the matter.
The Data Protection Officer can be contacted at dataprotectionofficer@glasgow.ac.uk

If you are not satisfied with the response or believe I am not processing your personal data in accordance with the law, you can complain to the Information Commissioner's Office (ICO) https://ico.org.uk/

Who has ethically reviewed the project?

This project has been ethically approved via the College of Social Sciences Research Ethics Committee or relevant School Ethics Forum in the College.

How long do I keep it for?

Your **personal** data will be retained by the University only for as long as is necessary for processing and no longer than the period of ethical approval, until December 2024. After this time, personal data will be securely deleted.

Your **research** data will be retained for a period of ten years in line with the University of Glasgow Guidelines. Specific details in relation to research data storage are provided on the Participant Information Sheet and Consent Form which accompany this notice.

End of Privacy Notice _____

Appendix C: Consent Form

University of Glasgow
College of Social Sciences

Consent Form

Title of Project: An exploration of the ethical issues that data scientists encounter in their field, and how these might be addressed in ethical education.

Name of Researcher: Nicola Weaver

Please tick as appropriate

Yes ☐ No ☐ I confirm that I have read and understood the Participant Information Sheet for the above study and have had the opportunity to ask questions.

Yes ☐ No ☐ I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

Yes ☐ No ☐ I consent to interviews being audio-recorded.

Yes ☐ No ☐ I acknowledge that participants will be referred to by pseudonym.

I agree that:

Yes ☐ No ☐ All names and other material likely to identify individuals will be pseudonymised.

Yes ☐ No ☐ The material will be treated as confidential and kept in secure storage at all times.

Yes ☐ No ☐ De-identified material will be retained in secure storage for use in future academic research and may be shared/archived or re-used in accordance with Data Sharing Guidance provided on Participant Information Sheet. This data will be retained and disposed of in line with University protocols after 10 years.

Yes ☐ No ☐ The material may be used in future publications, both print and online.

Yes ☐ No ☐ I waive my copyright to any data collected as part of this project.

Yes ☐ No ☐ Other authenticated researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.

Yes ☐ No ☐ Other authenticated researchers may use my words in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form.

Yes ☐ No ☐ I acknowledge the provision of a Privacy Notice in relation to this research project.

I agree to take part in this research study ☐

I do not agree to take part in this research study ☐

Name of Participant …… Signature ……

Date ……

Name of Researcher ……Signature ……

Date ……