Ethical AI: Designing Responsible and Trustworthy Systems

"Shaping a Future with AI: Trustworthy, Responsible, and Fair by Design"

Dr.Suja Cherukullapurath Mana
Dr.G.Kalaiarasi
Dr.M.Selvi
Ms.R.Velvizhi
Ms.Manju C Nair

This Page Intentionally Left Blank

ETHICAL AI: DESIGNING RESPONSIBLE AND TRUSTWORTHY SYSTEMS

Dr.Suja Cherukullapurath Mana

Dr.G.Kalaiarasi

Dr.M.Selvi

Ms.R.Velvizhi

Ms.Manju C Nair

www.jpc.in.net

Ethical AI: Designing Responsible and Trustworthy Systems

Authors:

Dr.Suja Cherukullapurath Mana

Dr.G.Kalaiarasi

Dr.M.Selvi

Ms.R.Velvizhi

Ms.Manju C Nair

First Published: June 2023

ISBN 978-93-91303-66-2

9 789391 303662 >

ISBN: 978-93-91303-66-2

DOI: https://doi.org/10.47715/JPC.978-93-91303-66-2

Pages: 184 (Front pages 12 & Inner pages 172)

Price: 375/-

Publisher:
Jupiter Publications Consortium
22/102, Second Street, Virugambakkam
Chennai, Tamil Nadu, India.
Website: www.jpc.in.net
Email: director@jpc.in.net
Imprint:
Magestic Technology Solutions (P) Ltd.
Chennai, Tamil Nadu, India.

TITLE VERSO

Title of the Book:
Ethical AI: Designing Responsible and Trustworthy Systems

Author's Name:
Dr.Suja Cherukullapurath Mana
Dr.G.Kalaiarasi
Dr.M.Selvi
Ms.R.Velvizhi
Ms.Manju C Nair

Published By:
Jupiter Publications Consortium
Publisher's Address:
22/102, Second Street, Venkatesa Nagar, Virugambakkam
Chennai 600 092. Tamil Nadu, India.

Printer's Details:
Magestic Technology Solutions (P) Ltd.

Edition Details: First Edition

ISBN: 978-93-91303-66-2

COPYRIGHT

Jupiter Publications Consortium
22/102, Second Street, Virugambakkam
Chennai 600 092. Tamil Nadu. India

Printed on acid-free paper
International Standard Book Number (ISBN): 978-93-91303-66-2 (Paperback)
Digital Object Identifier (DOI): 10.47715/JPC.B.978-93-91303-66-2

This book provides information obtained from reliable and authoritative sources. The author and publisher have made reasonable attempts to publish accurate facts and information, but they cannot be held accountable for any content's accuracy or usage. The writers and publishers have endeavoured to track down the copyright holders of every content copied in this monograph and regret if permission to publish in this format was not acquired. Please notify us through email if any copyright-protected work has not been recognised so that we may make the necessary corrections in future reprints.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks and are used only for identification and explanation without intent to infringe.

Visit the Jupiter Publications Consortium Web site at
http://www.jpc.in.net

PREFACE

As we, the authors, stand on the precipice of the Fourth Industrial Revolution, the pervasive nature of Artificial Intelligence (AI) in our lives cannot be overstated. Our technological landscape is rapidly evolving, and AI is both the catalyst and the prodigious consequence of this transformation. It's an undeniable reality that with every leap of progress, new ethical conundrums arise. As AI proliferates, these ethical challenges have swelled in importance and complexity. Therefore, we found it imperative to put forth this monograph, "Ethical AI: Designing Responsible and Trustworthy Systems".

Through the collaborative efforts of Dr. Suja Cherukullapurath Mana, Dr. G.Kalaiarasi, Dr. M.Selvi, Ms. R.Velvizhi, and Ms. Manju C Nair, we hope to provide our readers, particularly B. Tech students, with a comprehensive overview of Ethical AI. Our objective is to inspire curiosity, prompt dialogue, and foster an ethical approach to AI development and utilization.

In Chapter 1, we lay the groundwork by exploring the definition, significance, opportunities, and challenges of Ethical AI, emphasizing the necessity for responsible and trustworthy AI systems.

This foundation paves the way for Chapter 2, where we delve into the understanding of AI Ethics, tracing its historical perspectives, ethical dilemmas, and applicable ethical frameworks.

Chapter 3 elaborates on the principles of Ethical AI—Transparency, Accountability, Privacy, Fairness, Beneficence, and Non-maleficence. We follow this with a detailed look at Ethical issues in AI design and implementation, exploring bias, privacy challenges, autonomy, consent, and misuse of AI in Chapter 4.

In Chapter 5, we turn to practical applications and real-world case studies, extracting invaluable lessons from the field. Chapter 6 transitions into a discussion on the regulation and governance of AI, focusing on current global regulatory landscapes, the role of policy and law, and the impact of regulation on innovation.

Chapters 7 and 8 focus on the design and evaluation of Ethical AI systems, outlining ethical considerations throughout the AI development lifecycle, the importance of explainable AI, algorithmic transparency, auditing tools, and techniques, and the role of third-party auditors.

We address the topic of trust in AI in Chapter 9, highlighting its importance for AI adoption and outlining strategies to enhance it. Chapter 10 peers into the future, identifying emerging trends in Ethical AI, its impact on future work, and potential ethical considerations for future AI technologies.

In the final chapter, Chapter 11, we sum up the key takeaways, issuing a call to action for Ethical AI practitioners, who hold the potential to shape an ethical, inclusive, and prosperous future enabled by AI.

We hope this monograph sparks thought-provoking discussions and guides you towards ethical practice in AI. The goal is to create AI technologies that benefit humanity while mitigating potential risks. The path to ethical AI may be complex, but it is certainly within our reach.

ABSTRACT

"Ethical AI: Designing Responsible and Trustworthy Systems" is a comprehensive and insightful exploration of the ethical considerations surrounding the design, implementation, and application of Artificial Intelligence (AI) technologies. The authors aim to inspire curiosity, promote dialogue, and encourage ethical practice in the AI landscape. This monograph elucidates the principles of ethical AI, probes into ethical dilemmas, outlines the development lifecycle, discusses transparency, highlights the role of third-party audits, underscores the importance of trust, and identifies emerging trends and future prospects. It also looks at the current global regulatory landscape and the impact of AI on future work. This treatise emphasizes the importance of responsible and trustworthy AI systems and outlines potential ethical considerations for future AI technologies, concluding with a call to action for Ethical AI practitioners. By contextualizing these themes within the rapidly evolving AI landscape, this monograph provides a substantial contribution to the ongoing dialogue surrounding Ethical AI.

Keywords: Artificial Intelligence, Ethical AI, Ethical Dilemmas, AI Development Lifecycle, Transparency, Third-Party Audits, Trust in AI, AI Regulation, Future of Work, Future AI Technologies.

This Page Intentionally Left Blank

INTRODUCTION

Artificial Intelligence (AI) has rapidly permeated every facet of our daily lives. From driving decision-making processes and business operations to redefining human-machine interactions, AI technologies have drastically altered the landscape of our contemporary world. These technologies range widely, from autonomous vehicles to personalized recommendation algorithms, each promising unprecedented convenience, efficiency, and a wealth of potential benefits. However, as with any monumental technological advancement, the rise of AI carries with it a series of intricate ethical dilemmas and societal implications.

The primary focus of "Ethical AI: Designing Responsible and Trustworthy Systems" is to address these concerns. This work embarks on an enlightening journey from the rise of AI, the opportunities it presents, the challenges it poses, into the nuanced world of AI ethics. It underlines the necessity for responsible and trustworthy AI systems, setting the stage for a deep-dive into the realm of ethical dilemmas and principles associated with AI.

The historical perspective of ethics in technology is discussed, broadening the perspective and shedding light on ethical quandaries specific to AI. It introduces the reader to the myriad ethical frameworks applicable to AI, effectively establishing a foundation for understanding the moral intricacies inherent in AI.

The monograph thoroughly explores the principles of ethical AI, diving into key concepts such as transparency, accountability, privacy, fairness, beneficence, and non-maleficence. These principles serve as the ethical compass guiding the implementation and design of AI systems.

We then delve into the ethical issues that arise in the design and implementation phase of AI. It elucidates the sources and consequences of bias in AI, the challenges of privacy and data protection, the implications of autonomy and consent, and the security concerns that stem from the misuse of AI.

The monograph also presents real-world case studies of both ethical and unethical AI usage, imparting vital lessons from the field. An in-depth discussion about the current global regulatory landscape follows, highlighting the vital role of policy and law in ensuring ethical AI and the impact of regulation on innovation.

The text also incorporates an insightful discussion about the future of AI. It elucidates emerging trends in ethical AI, how AI impacts the future of work, and points out ethical considerations that we must keep in mind for the technologies to come.

Finally, we consolidate our understanding, drawing key takeaways from the various facets of ethical AI discussed throughout the text. We conclude with a call to action for all AI practitioners, urging them to uphold these ethical standards as they continue to design, implement, and deploy AI systems.

Dr.Suja Cherukullapurath Mana
Dr.G.Kalaiarasi
Dr.M.Selvi
Ms.R.Velvizhi
Ms.Manju C Nair

TABLE OF CONTENTS

Chapter 1:

Introduction to Ethical AI

This page Intentionally Left Blank

Chapter 1: Introduction to Ethical AI

1.1 Definition and Significance of Ethical AI

Artificial Intelligence (AI) is a multidisciplinary field of science whose goal is to create intelligent machines, which can mimic, augment, or even surpass human intelligence **(Russell & Norvig, 2020)**. The capability of AI to learn, reason, perceive, plan, and process natural language offers transformative potential across diverse sectors including healthcare, finance, transport, and education **(Fountaine, McCarthy, & Saleh, 2019)**. However, this immense potential raises significant ethical challenges, which if left unchecked, could have adverse societal effects.

Ethical AI refers to the design, development, and deployment of AI systems in a manner that aligns with moral and ethical values and principles. Such values can encompass, but are not limited to, fairness, transparency, accountability, privacy, and beneficence. Ethical AI encourages designers and developers to consider the moral implications of their AI technology and to ensure these technologies respect human values and rights **(Floridi & Cowls, 2019)**.

Understanding the significance of Ethical AI begins with acknowledging the impact of AI systems on our society. According to PwC's "Sizing the Prize" report, AI could contribute up to $15.7 trillion to the global economy by 2030 (PWC, 2017). As AI proliferates across sectors, the ability to make autonomous decisions and predictions, sometimes with limited transparency, heightens ethical concerns.

AI decisions, if based on biased data or algorithms, can perpetuate social inequities. For instance, an AI recruiting tool may be biased against certain demographic groups if it was trained on historical hiring data that was itself biased **(Dastin, 2018).** Likewise, AI privacy infringements could lead to unauthorized personal data access, posing risks to individual liberties and societal trust.

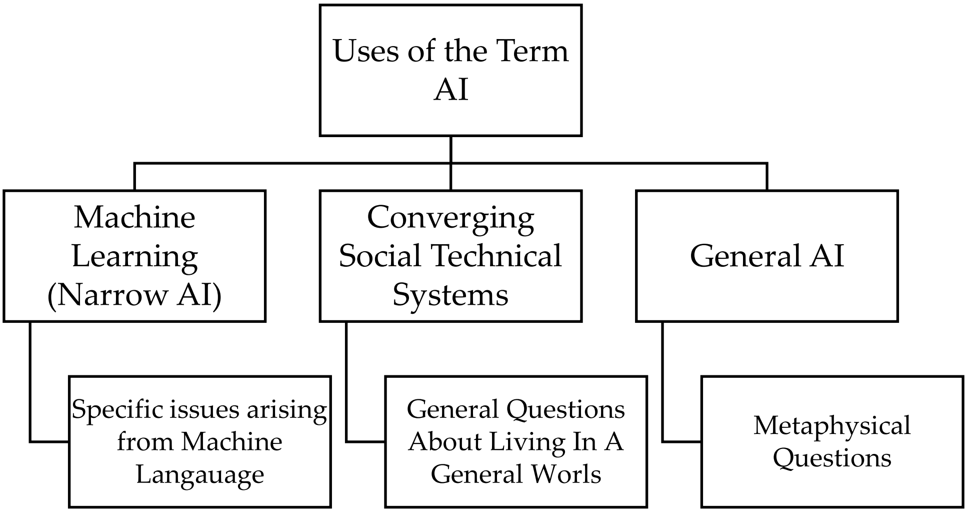

Uses of the Term AI
Machine Learning (Narrow AI)
Converging Social Technical Systems
General AI
Specific issues arising from Machine Langauage
General Questions About Living In A General Worls
Metaphysical Questions

Figure 1.1: A diagram illustrating potential ethical issues in AI

The essence of Ethical AI, therefore, lies in recognizing these risks and taking proactive steps to mitigate them, thereby ensuring AI serves as a tool for societal betterment rather than detriment. Only through prioritizing ethical considerations can we unlock the full potential of AI, fostering trust and encouraging its widespread adoption. These ethical concerns are not hypothetical; they are very real and have substantial implications.

For instance, the COMPAS system, a risk assessment tool used by US courts to inform parole decisions, was found to be biased against African-American defendants **(Angwin, Larson, Mattu, & Kirchner, 2016).** Similarly, an investigation revealed that Amazon had to abandon an AI recruitment tool because it was biased against women **(Dastin, 2018)**. These cases highlight the pressing need to infuse ethics into AI, for failing to do so can reinforce societal biases and lead to unfair outcomes.

As AI systems continue to penetrate our daily lives, the significance of Ethical AI becomes even more pronounced. We are at a critical juncture where our actions (or inactions) regarding Ethical AI can shape the future of our society. If we navigate this path well, we stand to gain from AI's benefits while minimizing potential harms. However, if we ignore the ethical dimensions of AI, we risk creating systems that undermine the very values we cherish, such as fairness, justice, and autonomy.

Ethical AI is not an option—it is a necessity. As AI technologies evolve, so should our efforts to understand and address the ethical challenges they pose. By integrating ethical considerations into AI design, development, and deployment, we can ensure that AI technologies serve the broader interests of society, thus earning public trust and acceptance. Table 1.1 provides a brief summary of the key terms and their definitions related to Ethical AI:

Table 1.1: Key Terms and Definitions Related to Ethical AI

Term	**Definition**
Artificial Intelligence (AI)	A multidisciplinary field of science that focuses on creating intelligent machines capable of performing tasks that require human intelligence.
Ethical AI	The design, development, and deployment of AI systems in a manner that aligns with moral and ethical values and principles.
Bias in AI	Biased outcomes from AI systems, often resulting from biased data or algorithms.
Ethical AI Principles	Fundamental values that guide the ethical development and use of AI, such as fairness, transparency, accountability, and privacy.

The importance of Ethical AI cannot be understated. It is a cornerstone of responsible AI development and a prerequisite for building trustworthy AI systems. As we delve deeper into this monograph, we will explore these concepts in greater depth, providing a robust understanding of Ethical AI's principles, practices, and implications.

1.2 The Rise of AI: Opportunities and Challenges

The rise of artificial intelligence (AI) marks one of the most significant technological advancements of our era. A report by **McKinsey Global Institute (2017)** estimates that AI techniques have the potential to create between $3.5 trillion and $5.8 trillion in value annually across nine business functions in 19 industries. This tremendous growth stems from the ability of AI to transform numerous sectors, create novel opportunities, and redefine the way we live and work.

The opportunities that AI presents are vast and varied. In healthcare, AI can improve diagnosis, predict patient outcomes, and personalize treatment plans **(Jiang et al., 2017).** In education, AI-powered personalized learning platforms can adapt to students' unique learning styles, thereby improving the learning process **(Luckin et al., 2016).** In transportation, self-driving vehicles hold the potential to significantly reduce accidents, improve traffic flow, and revolutionize mobility **(Fagnant & Kockelman, 2015).**

AI has also opened up new frontiers in scientific research. For example, Google's DeepMind developed AlphaFold, an AI system that can predict protein structures with remarkable accuracy, a breakthrough that could accelerate research in various fields, from drug discovery to environmental sustainability **(Senior et al., 2020).**

Yet, with these opportunities come significant challenges, the most pressing of which revolve around ethical issues. As AI systems become more autonomous, questions arise about accountability and transparency. For instance, if an autonomous vehicle is involved in an accident, who is responsible – the manufacturer, the software developer, or the owner of the vehicle? **(Nyholm & Smids, 2016).**

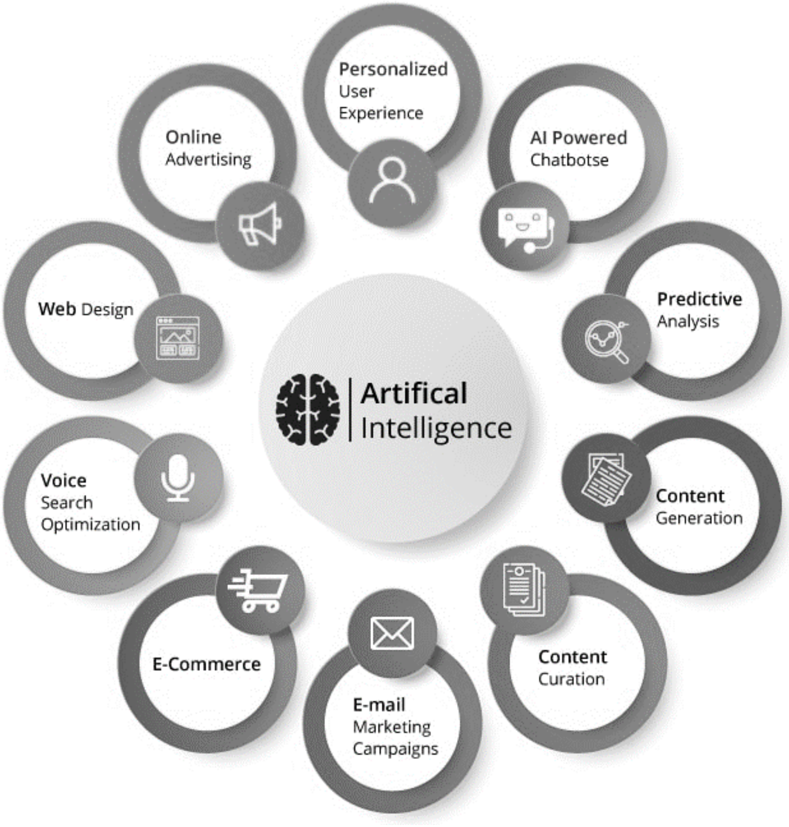


Figure 1.2: Diverse Applications of AI across Various Sectors

Further, AI systems are often perceived as "black boxes," with decision-making processes that are complex and opaque. This lack of transparency can impede trust and acceptance among users and the broader public, especially in high-stakes domains such as healthcare and criminal justice **(Castelvecchi, 2016).**

Moreover, as AI systems typically learn from existing data, they can unwittingly perpetuate and amplify existing biases present in the data. For example, an AI system trained on hiring data from a company that historically favored male candidates may continue to favor male candidates, even if unintentionally **(Dastin, 2018).**

Table 1.2 summarizes the opportunities and challenges brought about by the rise of AI:

Table 1.2: Opportunities and Challenges in AI

Opportunities	**Challenges**
Improved healthcare diagnostics and treatment plans	Accountability in autonomous systems
Personalized education	Transparency and trust in AI decision-making
Safer and efficient transportation	Bias and fairness in AI outcomes
Accelerated scientific research	Privacy and data protection

In light of these challenges, there is a growing consensus among researchers, practitioners, and policymakers about the need for Ethical AI – AI systems that not only maximize benefits but also respect and uphold ethical principles and societal values. The succeeding chapters of this monograph will delve into these ethical considerations and explore strategies for designing responsible and trustworthy AI systems.

Lastly, there is the issue of privacy and data protection, which has become increasingly significant in the digital age. AI systems typically rely on large amounts of data to learn and make decisions. In many cases, this data includes sensitive personal information, raising concerns about data misuse, unauthorized access, and potential violations of privacy **(Burt, 2020).**

Moreover, as AI becomes more pervasive, new ethical dilemmas and social challenges may emerge that we cannot yet fully foresee. There are concerns about the impact of AI on jobs and the workforce, as automation could potentially lead to job displacement **(Arntz, Gregory, & Zierahn, 2016).** Similarly, there are concerns about the use of AI in warfare, with the development of autonomous weapons systems sparking debates about their potential risks and ethical implications **(Horowitz & Scharre, 2015).**

Therefore, it is crucial that as we harness the benefits of AI, we also actively address these challenges. Ethical AI is about finding a balance – maximizing the benefits of AI while minimizing its potential harms. This requires a multidisciplinary approach, combining insights from computer science, social sciences, law, and philosophy, among others.

This dynamic interplay of opportunities and challenges brought about by AI forms the impetus for this monograph. In the subsequent chapters, we will delve deeper into the concept of Ethical AI, examining the principles and practices that can guide the development of responsible and trustworthy AI systems.

1.3 Necessity for Responsible and Trustworthy AI Systems

As AI continues its pervasive integration into our lives, ensuring responsible and trustworthy AI systems becomes more than just a technical requirement—it is a social, moral, and legal necessity. The implications of AI decisions are far-reaching and can significantly affect individuals and society at large. Therefore, it is essential that these systems are designed and implemented in a responsible and trustworthy manner.

Firstly, trustworthiness in AI systems is crucial for widespread acceptance and adoption of AI technologies. As we saw in previous sections, AI's decision-making processes are often complex and not easily interpretable, leading to perceptions of AI as "black boxes" **(Castelvecchi, 2016**). This lack of transparency can lead to distrust and scepticism among users and the general public. Thus, to build trust in AI, systems should be designed to be explainable and understandable to their users, adhering to the principle of transparency.

Responsible AI also demands accountability. As AI systems increasingly make decisions traditionally made by humans, we must ensure that there are mechanisms in place to hold these systems and their operators accountable for their actions **(Nyholm & Smids, 2016).** This involves establishing clear guidelines and regulations on AI behaviour and usage, and implementing methods for tracing back AI decisions when things go wrong.

Bias and fairness are additional critical concerns in AI systems. As AI systems learn from data, they can often reflect and amplify existing biases in the data, leading to unfair outcomes **(Dastin, 2018).**

Thus, ensuring fairness in AI systems involves not just correcting biased algorithms, but also scrutinizing the data used to train these algorithms.

Furthermore, responsible AI must respect privacy and protect data. AI technologies often require substantial amounts of personal data to function effectively. However, without proper safeguards, this can lead to privacy infringements and data misuse **(Burt, 2020).**

Lastly, while AI brings many benefits, it also poses risks, such as job displacement due to automation (Arntz, Gregory, & Zierahn, 2016) and the development of autonomous weapons systems **(Horowitz & Scharre, 2015).** Therefore, responsible AI also involves assessing and mitigating these potential risks.

Table 1.3: Key Aspects of Responsible and Trustworthy AI

Aspect	**Description**
Transparency	AI systems should be designed to be explainable and understandable.
Accountability	Mechanisms should be in place to hold AI systems and their operators accountable for their actions.
Fairness	AI systems should be free from bias and ensure fair outcomes.
Privacy and Data Protection	AI systems must respect privacy and protect personal data.
Risk Mitigation	Potential risks posed by AI, such as job displacement and misuse, should be assessed and mitigated.

The responsible and trustworthy AI is not a luxury—it is a necessity. As AI becomes increasingly integrated into our lives, we must ensure that these systems uphold our values and principles, and serve the interests of all stakeholders. The following chapters of this monograph will delve deeper into these aspects, offering a comprehensive exploration of Ethical AI.

In the upcoming chapters, we will be examining in detail how to design and implement AI systems that are ethical, responsible, and trustworthy. We will explore the principles of ethical AI, delve into strategies to address bias and ensure fairness, and discuss frameworks for accountability and transparency in AI. Furthermore, we will look at how privacy and data protection can be ensured in AI systems, and consider how to assess and mitigate the risks posed by AI.

The goal is not only to understand the necessity for responsible and trustworthy AI systems but to provide you, the reader, with the knowledge and tools to contribute to the development of such systems. We aim to foster an understanding that the design of ethical AI systems is a shared responsibility, requiring the input and collaboration of AI developers, users, regulators, and broader society.

In a sense, the design and use of ethical AI can be likened to the construction and driving of a car. Just as we want cars to be safe, reliable, and follow traffic rules, we want AI systems to be trustworthy, responsible, and follow ethical guidelines. And just as driving safely involves not only the driver but also traffic laws, road infrastructure, and car maintenance, ensuring ethical AI involves not only the AI developer but also ethical guidelines, regulatory frameworks, and ongoing monitoring and maintenance of AI systems.

As we embark on this journey towards understanding and implementing ethical AI, we hope to inspire and equip the future leaders and innovators of AI with the knowledge and tools to ensure that AI serves the best interests of society.

In the end, the objective is to ensure that the rise of AI brings about a future that reflects our shared values and aspirations, a future where AI is used not only to improve efficiency and productivity but to enhance the wellbeing of all people and to foster a fair, inclusive, and just society.

Chapter 2:

Understanding AI Ethics

This page Intentionally Left Blank

Chapter 2: Understanding AI Ethics

2.1 Historical Perspective of Ethics in Technology

Understanding the ethical considerations of artificial intelligence (AI) requires a foundational knowledge of ethics in technology as a whole. Over the course of human history, technology and ethics have been inextricably intertwined, with the development and use of technology often raising important ethical questions and challenges.

The ethical implications of technology date back to the inception of human civilization itself. Consider, for instance, the invention of fire and the wheel. While these innovations undoubtedly brought numerous benefits, such as improved survival and transportation, they also introduced new ethical dilemmas. Fire could be used for warmth and cooking but also for destruction and warfare. Similarly, the wheel enabled faster travel but also led to new types of accidents and injuries **(Tennant, 2018).**

Fast forward to the industrial revolution of the 18th and 19th centuries, and we witness an unprecedented surge in technological development that dramatically reshaped society. Innovations such as the steam engine, the telegraph, and later the telephone and electricity, transformed the way people lived, worked, and communicated. However, these advancements also led to significant social and ethical challenges. The rise of factories, for instance, resulted in poor working conditions and child labor. Meanwhile, the advent of electricity sparked debates about resource use and environmental impact **(Misa, 1995).**

In the 20th century, the introduction of nuclear technology epitomized the ethical dilemmas that can arise from technological advancement. While nuclear technology has significant potential for energy generation, its use in warfare and the associated threats of nuclear fallout and radiation sickness demonstrated the devastating potential of the technology **(Rhodes, 1986).**

The advent of computer technology and the Internet brought forth a new era of ethical challenges. Concerns about data privacy, cyber security, and digital divide became prominent. The revelation of widespread surveillance by governments, as brought to light by **Edward Snowden in 2013**, underscored the privacy implications of digital technology (Greenwald, MacAskill, & Poitras, 2013). Meanwhile, the digital divide—the gap between those who have access to information and communications technology and those who do not—raised questions about equity and social justice in the digital age **(Norris, 2001).**

The development of AI represents the latest frontier in the intersection of technology and ethics. As we have seen in the previous chapter, AI has significant potential for societal benefit but also presents new ethical challenges. These include issues of transparency, accountability, fairness, privacy, and the potential risks associated with AI.

From this historical perspective, it is clear that ethical considerations have always been an integral part of technology. The rise of AI and the associated ethical challenges is thus not a new phenomenon, but rather the latest chapter in the long-standing interplay between technology and ethics.

This historical perspective offers us a framework for understanding and addressing the ethical implications of AI. Like the technologies before it, AI brings great promise, but also great responsibility. The lessons learned from previous technological developments can guide us in creating a future where AI is used ethically and responsibly, enhancing societal wellbeing while mitigating potential harms.

In the following sections of this chapter, we will delve into the ethical principles that guide the design and use of AI. Drawing on lessons from the past and present, we will explore how these principles can help us navigate the ethical landscape of AI and shape a future where technology serves the best interests of humanity.

2.2 Ethical Dilemmas in AI

The development and deployment of AI technology have surfaced numerous ethical dilemmas. In this section, we will delve into some of the key ethical challenges that arise from AI, which underscore the necessity for a strong ethical framework for its development and use.

1. **Bias and Fairness:** AI systems, particularly those based on machine learning, are often trained on large datasets that may reflect existing biases in society. This can result in AI systems that perpetuate or even amplify these biases, leading to unfair outcomes. For instance, an AI system used for hiring might be biased against certain demographic groups if the training data reflected discriminatory hiring practices **(Barocas & Selbst, 2016).** Similarly, facial recognition systems have been found to have higher error rates for people of color and women, raising significant fairness concerns **(Buolamwini & Gebru, 2018).**

2. **Transparency and Explainability:** AI systems, particularly those using deep learning techniques, are often described as "black boxes," as their internal workings can be difficult to understand. This lack of transparency can be problematic in many contexts. For example, if an AI system denies a person's loan application or makes a medical diagnosis, it is important for the person to understand why that decision was made **(Castelvecchi, 2016).**

3. **Accountability:** Given the autonomous nature of AI systems, determining responsibility when things go wrong can be challenging. If an AI system makes a mistake or causes harm,

who is accountable - the AI developer, the user, the owner, or someone else? This dilemma is especially prominent in the case of autonomous vehicles, where determining liability in the event of an accident is a complex issue (Goodall, 2014).

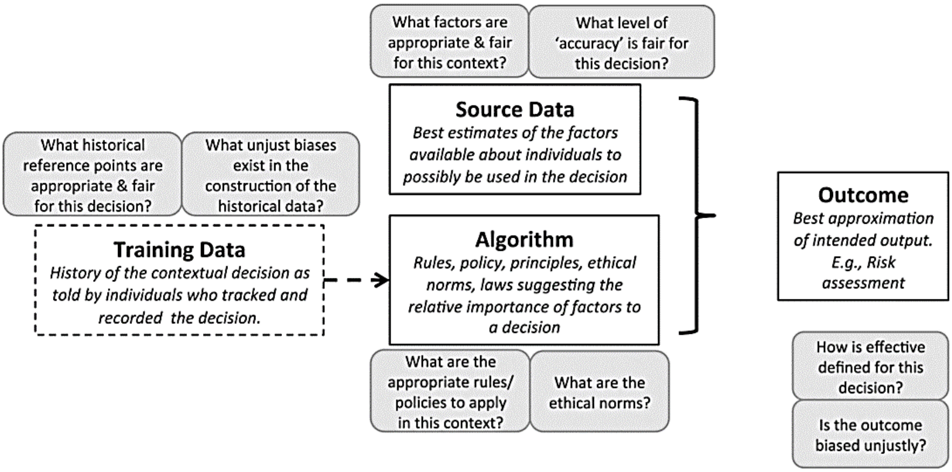
4. **Privacy and Data Protection:** AI systems often rely on large amounts of data, which can include sensitive personal information. This raises concerns about privacy and data protection. How is data collected, stored, and used? How is privacy protected when AI is used for surveillance or data analysis? These are critical ethical issues in the era of AI **(Taylor, Floridi, & Van der Sloot, 2017).**

5. **Risk and Safety:** As AI systems become more powerful and autonomous, they present potential risks. An AI system could cause harm if it malfunctions or is used maliciously. More broadly, there are concerns about the long-term future of AI and the possibility of artificial general intelligence (AGI) that surpasses human intelligence, which presents profound safety and existential risks **(Bostrom, 2014).**

These ethical dilemmas highlight the complexity and importance of ethics in AI. Addressing these issues requires a multifaceted approach that includes technical solutions, ethical guidelines, regulatory frameworks, and public engagement.

Each of these dilemmas offers unique challenges and requires dedicated attention to navigate. As we become increasingly dependent on AI systems, these ethical considerations will continue to be of paramount importance. In the following sections, we will delve into these ethical challenges in more detail and explore strategies for addressing them.

Figure 2.1: Ethical dilemmas in AI – Bias and Fairness, Transparency and Explainability, Accountability, Privacy and Data Protection, and Risk and Safety – and their implications (Martin, 2021)


What factors are appropriate & fair for this context?
What level of 'accuracy' is fair for this decision?
Source Data
Best estimates of the factors available about individuals to possibly be used in the decision
What historical reference points are appropriate & fair for this decision?
What unjust biases exist in the construction of the historical data?
Training Data
History of the contextual decision as told by individuals who tracked and recorded the decision.
Algorithm
Rules, policy, principles, ethical norms, laws suggesting the relative importance of factors to a decision
What are the appropriate rules/ policies to apply in this context?
What are the ethical norms?
Outcome
Best approximation of intended output. E.g., Risk assessment
How is effective defined for this decision?
Is the outcome biased unjustly?

We will look at principles for ethical AI, methods for ensuring fairness, transparency, and accountability in AI systems, ways to protect privacy and data, and approaches to assess and mitigate risks posed by AI. By understanding and addressing these ethical dilemmas, we can help ensure that AI technology is developed and used in a manner that aligns with our societal values and serves the best interests of humanity.

2.3 Ethical Frameworks Applicable to AI

Responding effectively to the ethical dilemmas presented by AI requires a robust ethical framework. Several ethical theories and principles that have been developed over centuries of philosophical thought are directly applicable to AI. This section introduces some of these frameworks that guide ethical considerations in AI.

1. **Consequentialism:** Consequentialist theories, such as utilitarianism, judge the rightness or wrongness of an action based on its outcomes. In the context of AI, a consequentialist might argue that an AI system is ethical if its deployment results in a net positive outcome, such as increased productivity or improved quality of life, even if there are potential negative consequences **(Bryson, 2018).**

2. **Deontological Ethics:** In contrast to consequentialism, deontological ethics focuses on the inherent rightness or wrongness of actions, regardless of their outcomes. A deontologist might argue that certain uses of AI are inherently wrong, such as uses that infringe upon privacy or autonomy, even if they lead to beneficial outcomes **(Moor, 2006).**

3. **Virtue Ethics:** Virtue ethics emphasizes the importance of character and virtues in ethical decision-making. In the AI context, a virtue ethicist might focus on the virtues or values that should guide the design and use of AI systems, such as fairness, transparency, and respect for human dignity **(Vallor, 2016).**

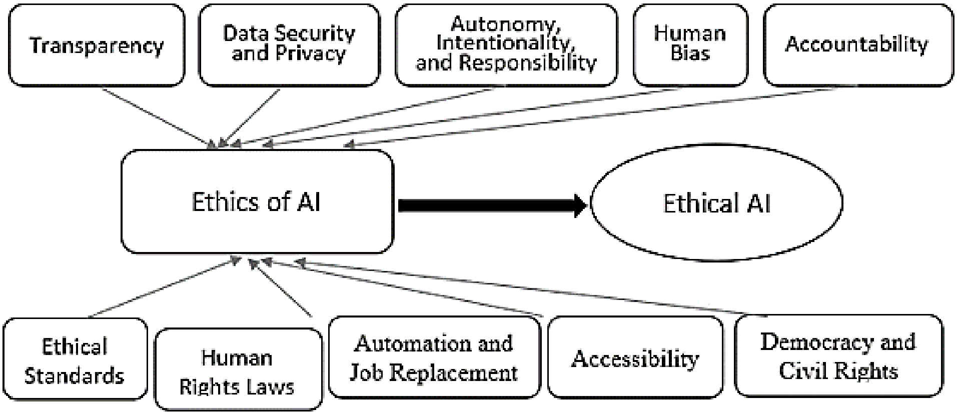
4. **Rights-based Ethics:** Rights-based ethics focuses on respecting and protecting individual rights. In terms of AI,

this might involve ensuring that AI systems respect human rights, such as the right to privacy, freedom of expression, and non-discrimination (Taddeo & Floridi, 2018).

5. **Contractualism:** Contractualism emphasizes mutual agreement as the basis for ethical norms. In the AI realm, this could mean involving stakeholders in decisions about how AI systems are designed and used to ensure that the systems respect everyone's interests **(Etzioni & Etzioni, 2017).**

6. **Ethics of Care:** The ethics of care emphasizes relationships and care for others. In the AI context, this might involve designing AI systems in a way that promotes human relationships and care, such as AI in healthcare or social robotics (Van Wynsberghe, 2013).

Figure 2.2: **AI Ethics: Framework of building ethical AI**


Transparency
Data Security and Privacy
Autonomy, Intentionality, and Responsibility
Human Bias
Accountability
Ethics of AI
Ethical AI
Ethical Standards
Human Rights Laws
Automation and Job Replacement
Accessibility
Democracy and Civil Rights

These ethical frameworks provide different perspectives on what constitutes ethical AI. They can be used individually or in combination to guide the design and use of AI systems. However, applying these ethical frameworks to AI also raises new challenges and requires thoughtful interpretation and adaptation.

In addition to these traditional ethical frameworks, several organizations and initiatives have proposed principles for ethical AI, which offer practical guidance for developing and using AI in an ethical manner. Some of these principles include:

1. **Fairness:** AI systems should be designed and used in a way that treats all individuals and groups fairly. This includes avoiding and mitigating bias and ensuring that the benefits and burdens of AI are shared equitably **(High-Level Expert Group on Artificial Intelligence, 2019).**

2. **Transparency:** AI systems should be transparent. This means that individuals should be able to understand how an AI system works, how decisions are made, and what data is used. Transparency is critical for trust, accountability, and the ability to challenge and correct decisions made by AI **(The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019).**

3. **Accountability:** Those who develop and use AI systems should be held accountable for their actions. This includes taking responsibility for the outcomes of AI systems and having mechanisms in place to address any harm caused **(Partnership on AI, 2019).**

4. **Privacy and Data Protection:** AI systems should respect privacy and protect personal data. This involves collecting, storing, and using data in a way that respects individual

privacy and complies with data protection laws and standards (The Public Voice, 2018).

5. **Safety and Beneficence:** AI systems should be safe and beneficial. This includes ensuring that AI systems do not cause harm, that risks are properly managed, and that AI is used for the benefit of all **(Asilomar AI Principles, 2017).**

By combining these principles with the ethical frameworks discussed earlier, we can form a comprehensive and robust approach to ethical AI. This approach recognizes the complexity and multifaceted nature of AI ethics and offers a balanced and nuanced way of navigating ethical dilemmas in AI.

https://www.edelman.com/research/2021-trust-barometer-technology

Etzioni, A., & Etzioni, O. (2017). Incorporating ethics into artificial intelligence. The Journal of Ethics, 21(4), 403-418.

European Commission. (2021). Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence

European Commission. (2021). Proposal for a Regulation on a European approach for Artificial Intelligence. Retrieved from www.ec.europa.eu

European Parliament and Council. (2016). General Data Protection Regulation. Retrieved from https://gdpr-info.eu/

Faden, R. R., & Beauchamp, T. L. (1986). A history and theory of informed consent. Oxford University Press.

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transportation Research Part A: Policy and Practice, 77, 167-181. https://doi.org/10.1016/j.tra.2015.04.003

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication.

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review. https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., & Taddeo, M. (2016). What is data ethics? Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2083), 20160360.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689-707.

Fosch-Villaronga, E., & Heldeweg, M. (2021). Designing for Trust: Embedding Trust by Design in Artificial Intelligence Systems. IEEE Transactions on Technology and Society, 2(1), 25-33.

Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-Powered Organization. Harvard Business Review, 97(4), 62–73.

Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. (2016). On the (im)possibility of fairness. arXiv preprint arXiv:1609.07236.

Goodall, N. J. (2014). Machine ethics and automated vehicles. Road Vehicle Automation, 1, 93-102.

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". AI magazine, 38(3), 50-57.

Google. (2020). AI for social good. Google AI. https://ai.google/social-good/

Green, F. (2020). The Pillars of Just Transition: A Framework for Policy. ILO Brief.

Greenleaf, G. (2017). Global Data Privacy Laws 2017: 120 National Data Privacy Laws, Including Indonesia and Turkey. Privacy Laws & Business International Report, 7-28.

Greenwald, G., MacAskill, E., & Poitras, L. (2013). Edward Snowden: the whistleblower behind the NSA surveillance revelations. The Guardian. https://www.theguardian.com/world/2013/jun/09/edward-snowden-nsa-whistleblower-surveillance

Grgić-Hlača, N., Redmiles, E. M., Gummadi, K. P., & Weller, A. (2018). Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. In Proceedings of the 2018 World Wide Web Conference (pp. 903-912).

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM computing surveys (CSUR), 51(5), 1-42.

Gupta, A., Verma, S., & Sood, S. (2021). Artificial Intelligence: A Review of its Applications and Legal and Ethical Considerations. International Journal of Computer Applications, 176(18), 7-15.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 30(1), 99-120.

High-Level Expert Group on Artificial Intelligence. (2019). Ethics Guidelines for Trustworthy AI. European Commission. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Hill, K. (2020, June 24). Wrongfully accused by an algorithm. The New York Times. https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html

Horowitz, M. C., & Scharre, P. (2015). An introduction to autonomy in weapon systems. Center for a New American Security, 13.

Horowitz, M. C., & Scharre, P. (2015). Meaningful human control in weapon systems: A primer. Center for a New American Security. https://s3.amazonaws.com/files.cnas.org/documents/Ethical_Autonomy_Working-Paper_031315.pdf

https://data-flair.training/blogs/ai-in-healthcare-sector/

https://hcil.umd.edu/tutorial-human-centered-ai/

https://www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/innovation/Intelligence-Unleashed-Publication.pdf

IBM. (2019). AI adoption advances, but foundational barriers remain. Retrieved from https://newsroom.ibm.com/2019-02-11-IBM-Study-AI-Adoption-Advances-But-Foundational-Barriers-Remain

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399.

Kearns, M., & Roth, A. (2020). The Ethical Algorithm: The Science of Socially Aware Algorithm Design. Oxford University Press.

Kifer, D., & Machanavajjhala, A. (2011). No free lunch in data privacy. Proceedings of the 2011 ACM SIGMOD International Conference on Management of data, 193-204.

Kroll, J. A. (2018). The fallacy of inscrutability. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133), 20180084.

Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016). How We Analyzed the COMPAS Recidivism Algorithm. ProPublica.

Latonero, M. (2018). Governing Artificial Intelligence: Upholding Human Rights & Dignity. Data & Society. https://datasociety.net/library/governing-artificial-intelligence/

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50-80.

Lipton, Z. C. (2018). The mythos of model interpretability. Queue, 16(3), 31-57.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence Unle & Wang, Y. (2017).

Lundberg, S. M., & Lee, S. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774.

Luxton, D. D. (2016). Recommendations for the ethical use and design of artificial intelligent care providers. Artificial Intelligence in Medicine, 64(1), 19-25.

Martinez-Martin, N., Greely, H. T., & Cho, M. K. (2021). Ethical development of digital phenotyping tools for mental health applications: Delphi study. JMIR mHealth and uHealth, 9(7), e27343.

Microsoft. (2020). AI for Accessibility. Microsoft AI. https://www.microsoft.com/en-us/ai/ai-for-accessibility

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1-38.

Ministry of Electronics and Information Technology. (2023). AI Strategy. Retrieved from www.meity.gov.in

Misa, T. J. (1995). A Nation of Steel: The Making of Modern America 1865-1925. Johns Hopkins University Press.

Mittelstadt, B. (2019). AI Ethics. Stanford Encyclopedia of Philosophy.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501-507.

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 205395171667967.

Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. Proceedings of the Conference on Fairness, Accountability, and Transparency, 279-288.

Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. IEEE Intelligent Systems, 21(4), 18-21.

National AI Initiative Act. (2021). National AI Initiative. Retrieved from www.congress.gov

Norris, P. (2001). Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide. Cambridge University Press. https://doi.org/10.1017/CBO9781139164887

Nyholm, S., & Smids, J. (2016). The Ethics of Accident-Algorithms for Self-Driving Cars: an Applied Trolley Problem? Ethical Theory and Moral Practice, 19(5), 1275-1289. https://doi.org/10.1007/s10677-016-9745-2

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.

OpenAI. (2020). OpenAI and GPT-3. OpenAI Blog. https://www.openai.com/blog/openai-and-gpt-3/

Partnership on AI. (2019). Partnership on AI Tenets. Partnership on AI.

Pew Research Center. (2019). Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information. Retrieved from https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/

PricewaterhouseCoopers. (n.d.). Accelerating innovation through responsible AI. PwC. https://www.pwc.co.uk/services/risk/insights/accelerating-innovation-through-responsible-ai.html

PWC. (2017). Sizing the prize: What's the real value of AI for your business and how can you capitalise? Retrieved from https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. Ethics and Information Technology, 20(1), 5-14.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI Accountability Gap: Def ining an End-to-End Framework for Internal Algorithmic Auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 33-44).

Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reshaping Business With Artificial Intelligence. MIT Sloan Management Review and Boston Consulting Group.

Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability. AINow Institute.

Renda, A. (2019). AI and Digital Policy: The role of Regulatory Sandboxes in Experimentation and Compliance. Centre for European Policy Studies, 5, 1-17.

Rhodes, R. (1986). The Making of the Atomic Bomb. Simon & Schuster.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144).

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

Rocher, L., Hendrickx, J. M., & de Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications, 10(1), 1-9.

Rossi, E. (2020). AI in healthcare: A risk management approach. IT Professional, 22(2), 29-35.

Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. Penguin.

Russell, S., & Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Pearson.

Ryan, M. D. (2020). Privacy and AI. In Ethics of Artificial Intelligence and Robotics (Stanford Encyclopedia of Philosophy, Fall 2020 Edition), Edward N. Zalta (ed.). Retrieved from https://plato.stanford.edu/archives/fall2020/entries/ethics-ai/#PrivAI

S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. Stroke and Vascular Neurology, 2(4), 230-243. http://dx.doi.org/10.1136/svn-2017-000101

Scherer, M. U. (2015). Regulating artificial intelligence systems: risks, challenges, competencies, and strategies. Harvard Journal of Law & Technology, 29, 353.

Schwartz, A. (2018). Artificial Intelligence—With Very Real Biases. Wall Street Journal.

Selbst, A D., & Powles, J. (2021). Meaningful Information and the Right to Explanation. International Data Privacy Law, 7(4), 233-242.

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., ... & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. Nature, 577(7792), 706-710. https://doi.org/10.1038/s41586-019-1923-7

Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. JAMA, 320(21), 2199-2200.

Stahl, B. C., & Wright, D. (2018). Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation. IEEE Security & Privacy, 16(3), 26-33.

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. Science, 361(6404), 751-752.

Taylor, L., Floridi, L., & Van der Sloot, B. (Eds.). (2017). Group privacy: new challenges of data technologies. Springer.

Tennant, C. (2018). The Wheel: A Great Innovation. A Great Leap Forward. Medium. https://medium.com/@christennant/the-wheel-a-great-innovation-a-great-leap-forward-9a2adb736038

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. IEEE.

The Public Voice. (2018). Universal Guidelines for Artificial Intelligence. The Public Voice.

Turilli, M., & Floridi, L. (2009). The ethics of information transparency. Ethics and Information Technology, 11(2), 105-112.

Turkle, S. (2011). Alone together: Why we expect more from technology and less from each other. Basic books.

Vallor, S. (2016). Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future of character. Philosophy & Technology, 31(1), 107-124.

Van Wynsberghe, A. (2013). Designing robots for care: Care centered value-sensitive design. Science and engineering ethics, 19(2), 407-433.

Veale, M., & Brass, I. (2019). Administration by Algorithm? Public Management Review, 21(6), 857-874. https://doi.org/10.1080/14719037.2018.1529345

Voigt, P., & Von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR): A Practical Guide. Springer International Publishing.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and accountable AI for robotics. Science Robotics, 2(6), eaan6080.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76–99. https://doi.org/10.1093/idpl/ipx005

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. Harvard Journal of Law & Technology, 31(2), 841-887.

Wakefield, J. (2018). Uber halts self-driving car tests after death. BBC News. Retrieved from https://www.bbc.com/news/business-43459156

Webber, W. (2021). AI Strategies: National and International Policy Considerations. Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, 149-155. https://doi.org/10.1145/3461702.3462624

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. Nuffield Foundation.

World Economic Forum. (2020). The Future of Jobs Report 2020. World Economic Forum.

Wyden, R., & Clarke, Y. (2019). Algorithmic Accountability Act of 2019. US Senate, Washington, DC, USA.

Yu, H. (2020). The Role of Law in Governing AI Systems. The European Journal of Law and Technology, 11(1), 1-14. http://ejlt.org/article/view/730

Yu, K. (2020). A review on the EU General Data Protection Regulation (GDPR) and its impact on AI. Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI).

Yu, K. H., Beam, A. L., & Kohane, I. S. (2020). Artificial intelligence in healthcare. Nature Biomedical Engineering, 4(10), 973-981.

Zhao, B., Wang, Y., Amini, A., Aberer, K., & Wang, F. (2021). Fairness in deep learning: A computational perspective. IEEE Signal Processing Magazine, 38(4), 126-138.

Zliobaite, I., & Custers, B. (2016). Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models. Artificial Intelligence and Law, 24(2), 183-201.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.