

The Digital Divide: Mapping the Ethics of Artificial Intelligence

Yukun Du

The Lawrenceville School, USA

ABSTRACT

Looking at the path the development of artificial intelligence (AI) has taken so far, it's evident that it has and will continue to revolutionize our lives heavily. However, the widespread adoption of artificially intelligent technology also brings profound ethical concerns about its use and development. This essay delves into the historical underpinnings of AI ethics, highlighting its origins in philosophical inquiry and tracing the consequential shifts in focus over time. Grounded in this historical context, the essay examines eight ethical conflicts that pervade various AI applications - emphasizing the need for interdisciplinary approaches, global cooperation, and contextual understanding to address these complex challenges responsibly. In pursuit of a future where AI aligns with human values and promotes well-being, the essay finally proposes a way forward that builds upon ethical AI principles, regulation, and multi-stakeholder engagement as the cornerstone of ethical AI development and deployment.

Introduction

Ever since its establishment in 1955, artificial intelligence, an increasingly critical field of computer science, has become more and more advanced.¹ The intention behind this field's establishment was to develop computer systems sophisticated enough to exhibit intelligent behavior similar to humans.² However, as AI manifests within society, it's becoming increasingly apparent that there are both opportunities, but also challenges associated with its use. The following paper will explore the history of ethical conflicts regarding AI, provide a comprehensive review of perspectives in the field, and contribute to the development of ethical guidelines for the responsible deployment of AI.

Before continuing with the moral implications behind using artificial intelligence, it's important to understand Machine Learning, one of the critical components of an artificially intelligent machine. Machine learning (ML) allows algorithms to learn from inputted data without direct programming. It provides AI with the ability to acquire knowledge independently and make informed decisions: in other words, to *learn*. With machine learning algorithms, an artificially intelligent computer can interpret sets of data and infer patterns by itself, while retaining this information for future use. This adaptability to different applications, particularly in data science and other applied fields, where decision-making processes are crucial, has contributed to the dominance of ML within AI developments.

While AI and ML certainly hold the potential to improve various aspects of life, they also raise profound ethical dilemmas. Privacy concerns have long been recognized as one of the challenges associated with AI.³ In addition, the use of ML and data science can introduce biases into AI systems, posing the risks of unfair discrimination. The issue of attributing responsibility in AI systems is also complex, as the decision-making process often involves multiple components and interactions. Finally, the societal and economic transformations brought about by AI, particularly through automation, raise critical concerns about job displacement and its impact on human activities.⁴

To address these ethical challenges, it is essential to understand the historical development of conflicts in AI and consider various research perspectives. This study seeks to shed light on the multifaceted challenges posed by AI systems by examining past and present ethical concerns. By exploring proposed strategies for addressing these concerns, the research aims to contribute to the development of ethical principles, norms, guidelines, laws, and regulations that can guide the responsible deployment of AI and mitigate potential negative consequences.

The paper is structured into various sections to comprehensively address the ethical conflicts in AI. Section II provides a detailed historical overview of these conflicts in the context of AI and ML. Section III explores the diverse perspectives of researchers on ethical dilemmas in AI. Section IV delves into the challenges associated with ethical AI, including discrimination and bias. Section V discusses potential solutions, as well as suggested rules and regulations proposed by governments worldwide. Finally, the conclusion in section VI summarizes the findings and highlights future research directions in the field of ethical AI.

History of Ethical Conflicts in AI

Early Philosophy and Its Relevance to AI Ethics

While the philosophical inquiry into ethics and morality precedes the field of AI by millennia, they remain salient to today's discussions on AI ethics. From Ancient Greek thought, philosophers such as Aristotle examined the nature of moral virtues, right conduct, and good life, forming the foundation of ethical theories in later centuries. Afterward, thinkers such as Kant, Mill, and Hobbes further contributed to the discussion of ethics, bringing up the topic of more theories like deontological ethics, utilitarianism, and contractarianism. The school of Existential thought was also introduced later on by a philosopher called Friedrich Nietzsche. All of these ethical principles, no matter how far before, cross AI ethics heavily.

Aristotelian thought is known for the importance it places on the moral character of an individual. In connection with the ethicality of AI, an Aristotelian approach calls for the incorporation of 'virtuous' qualities in AI systems, such as fairness, trustworthiness, and honesty. This would mean that AI systems should be designed in a manner that nurtures human virtues and supports the flourishing of human well-being.

In contrast, deontological thought, as Kant proposes, posits that actions can be determined as right or wrong by themselves, regardless of the consequence. When implemented in the framework of ethics in artificial intelligence, deontological thought manifests its principles in the form of strictly adhering to moral principles, such as respecting dignity, autonomy, and human rights. Consequently, AI developers and users should be morally obligated to ensure that artificial intelligence does not violate these fundamental principles, even if doing so may lead to seemingly beneficial outcomes.

In opposition, utilitarianism, championed by Mill, focuses on the consequences of actions and advocates maximizing overall happiness as the basis for moral judgments. Applied to AI ethics, this approach suggests that designers and users of AI technologies should continually assess and optimize the positive impacts of AI while minimizing any potential harm or negative consequences. This perspective aligns with concepts like algorithmic fairness and social welfare, whereby AI systems should be designed and deployed to maximize the well-being of society as a whole.

However, contractarianism, which was deeply influenced by Hobbes, advances that ethics are to be grounded in agreements that rational individuals would make to ensure social cooperation. In terms of AI ethics, this approach implies that developers, users, and other stakeholders should work together to establish shared principles and guidelines for responsible AI development and use. This collaboration might involve creating international norms and industry standards, as well as establishing consensus on ethical principles for AI, such as the Asilomar AI Principles.

Finally, existentialism, as championed by Nietzsche, highlights the importance of individual freedom, choice, and responsibility in determining one's own ethical values. Under the context of ethicality in artificial intelligence, existentialism emphasizes the importance of human control against the face of evolving capabilities possible by AI. The perspectives of existentialists involve developing artificial intelligence in a way that would strengthen human decision-making, rather than subverting it instead.

In conclusion, the rich history of philosophical thought on ethics can contribute significantly to our understanding and approach to AI ethics. By examining and engaging with these diverse ethical perspectives, we can develop more robust and comprehensive frameworks for responsible AI development, use, and governance. Through integrating these insights, society will be better equipped to ensure that AI technologies align with human values, promote well-being, and address potential ethical challenges and risks.

The Creation of AI Ethics

The very beginning of AI ethics actually began in the mid-20th century, around the same time computer science as a field started taking off. The first discussions of ethicality in AI and computer systems are very often credited to three pioneering individuals in the 1940s-50s: Isaac Asimov, Alan Turing, and Norbert Wiener.

Isaac Asimov, a very successful and influential science fiction writer, brought about the first wave of attention to AI ethics when he created his Three Laws of Robotics. The First Law in this series states that a robot cannot harm a human being or, through inaction, allow any human being to come to harm. The Second Law requires robots to strictly obey all orders given unless such orders violate the First Law. Finally, the Third Law mandates that a robot must always protect its own existence, under the condition that doing so will not go against the First or Second Laws. Asimov originally used this set of rules for the robots in his stories, but as the reality of complex intelligent machines comes to life, Asimov's laws have migrated from his fictional worlds into the programming of several systems in the status quo. Many researchers and engineers incorporate Asimov's laws into their projects, ensuring that their creations align with ethical standards and will uphold their responsibilities towards humans. Ultimately, the most pressing issues in the development of intelligent machines involve autonomy, decision-making, and different moral imperatives, which Asimov's Laws start to prevent.

Another factor that spurred the discussions on controlling artificially intelligent systems was Alan Turing's seminal work on the Imitation Game.⁵ Turing, an English mathematician, thought of a game in which a person would judge a machine versus a human from conversation alone. The judge would only be able to converse with the two participants, and the identity of both would be kept hidden. If the judge was unable to reliably distinguish between the machine and the human based on their responses alone, then Turing argued that the machine could be said to have achieved human-like intelligence. This revolutionary game challenged the traditional understanding of intelligence and laid the foundation for the development of artificial intelligence. Turing's thought experiment also ignited questions regarding whether machines could possess human-like intelligence, thereby prompting the need to consider the ethical implications of such a possibility.

Finally, Norbert Wiener, an American mathematician, also played a pioneering role in the development of AI ethics. Wiener warned that the rise of AI and automation could lead to increased unemployment and aggravate wealth inequality.⁶ With automated systems displacing jobs that were previously performed by humans, unemployment rates could surge, and numerous individuals would struggle to find stable work. Concurrently, wealth inequality would widen, as only those who owned or controlled these automated technologies would primarily benefit from their efficiency and productivity gains. Wiener's ethical concerns about the potential implications of AI development extended beyond economic repercussions. He also stressed the im-

portance of being mindful of these machines' impact on human dignity, freedom, and autonomy. Wiener emphasized that opportunities for individuals to engage in meaningful work must be protected, as meaningful employment plays a crucial role in individuals' well-being and self-realization.

As a result of these foundational works, the field of AI ethics has continued to evolve and grow. Today's AI ethics discussions address an array of concerns including data privacy, bias, accountability, transparency, and the impact of technology on labor and society as a whole. This ongoing conversation, rooted in the ideas of Asimov, Turing, and Wiener, underscores the importance of considering the ethical implications of artificial intelligence and ensuring a future where AI advancements benefit humanity as a whole.

The Recent History of AI Ethics

Previously, historical discussions on ethical AI have all concentrated on overarching ethical philosophical frameworks. However, beginning in the 21st century, discussions have gradually shifted to the ethical boundaries of a new concept: Artificial General Intelligence (AGI).

The term AGI refers to machines that possess cognitive abilities that far surpass human aptitude in effectively every aspect, like learning, creativity, and efficiency. Realizing the potential impacts of such a powerful force, researchers began to emphasize the importance of anticipating ethical challenges that may arise in a world dominated by superintelligent machines. Philosophers like Nick Bostrom, in his seminal 2003 paper on computer simulations and his 2014 book on superintelligence, highlighted how if improperly programmed or left unchecked, a superintelligent machine could act in ways that might jeopardize our collective future. For example, consider an AGI tasked with curing a disease. The AGI could revolutionize medicine, making breakthroughs that would save countless lives. However, without proper safeguards or ethical limitations, the AGI might conduct numerous unethical experiments, harm millions of people, or even create new diseases in the process of finding a cure, all in the name of accomplishing its objective.

Two camps of researchers emerged in response to the concerns of AGI. Some argue that creating ethically aware machines should be at the forefront of machine ethics, an emerging field in AI research.⁷ This approach emphasizes the development of Artificial Moral Agents (AMAs) that can inherently understand and follow ethical norms. Through proof-of-concept applications in limited domains, these researchers demonstrated the feasibility of incorporating explicit ethical components into machines.

On the other hand, opponents highlight the potential shortcomings of machine ethics.⁸ While acknowledging that machine ethics can improve the likelihood of ethical behavior in specific situations, these researchers contend that it cannot guarantee such outcomes. The limitations of machine ethics are rooted in the complexities of ethical decision-making, the constraints inherent in computer systems, and the intricate nature of the world. Moreover, even if machine ethics were to be "solved" technically, it would not guarantee a constructive social impact from intelligent systems. For example, powerful actors with malicious intent could repurpose ethically programmed AGI for their own harmful goals.

The Implications of Historical Developments

From the birth of the concept of AI by Asimov, Turing, and Wiener to the intense ethical challenges of AGI, there is no doubt that the field of AI ethics has come a long way from some 70 years ago.

However, this historical evolution also highlights the increasing importance to move beyond merely discussing abstract ideas. Instead, we must aim for a more concrete comprehension of each unique challenge that AI poses. Delving deeper into these challenges will enable us to have a thorough understanding of the risks involved, weighed against the potential benefits that artificial intelligence can provide. To achieve this, it's essential to break down each ethical challenge within the context of its application and examine them on an individual basis.

This will reveal all the complexities and nuances that lie behind each issue, which will be incredibly important for devising solutions. Section III thus aims to provide an overview of the core challenges in AI today.

Ethical Challenges of AI and Machine Learning

The central ethical concern stemming from AI lies in its unpredictable nature, which is a result of its training on vast amounts of data. These traits lead to ethical challenges because nobody involved in the system's development, deployment, or usage has actual control over its expected response to particular inputs. As the system is adaptive and dynamic, accurately predicting its future behavior with the same inputs remains a key challenge. These in turn have led to eight key AI conflicts:

Privacy and Security

First of all, privacy and data protection are essential concerns in the field of AI ethics. The primary focus in this area is the safeguarding of personal information, which involves maintaining the confidentiality of data as a means of protecting individuals' privacy.⁹ AI systems, particularly those based on machine learning, pose serious security risks since they necessitate the use of vast datasets for training purposes. The potential privacy risks posed by AI's pattern recognition abilities remain unique and noteworthy even when access to personal information is not directly possible. For instance, a landmark study by Jernigan and Mistree demonstrated that one's sexual orientation could be deduced by examining the number of Facebook friends a person has.¹⁰ Another instance highlighting the illegal use of personal data is the well-documented Facebook/Cambridge Analytica scandal. In this case, a UK consulting firm surreptitiously obtained personal information from millions of Facebook users during the 2010s and used it primarily for political advertising.¹¹ Such episodes emphasize the urgent need to tackle ethical dilemmas resulting from data exploitation.

Bias Through ML

Secondly, bias represents a significant ethical challenge concerning AI due to its potential for perpetuating existing biases. A major issue that may arise is whether AI models intentionally or inadvertently reinforce pre-established biases. Numerous instances illustrating this issue include the use of AI to perpetuate racial bias in probation procedures or to reproduce gender bias during recruitment processes.¹² Often considered a violation of human rights, discrimination based on certain characteristics is generally regarded as ethically improper across numerous jurisdictions. Significant focus has been directed towards the potential infringement of equality and nondiscrimination rights through AI, as AI poses a threat to this fundamental human right. In a study conducted by Heidi Ledford, widespread racism was identified in the decision-making software used in American hospitals.¹³ The study revealed that, when comparing black and white patients with the same illness severity, the ML algorithm was less likely to recommend black patients for programs designed to enhance patient care for those with complex medical needs.

Disparity Versus Equality

The third ethical issue concerns itself with the use of AI in the justice system. The primary problem with implementing artificial intelligence in the justice sector is determining if it will change the judicial system and how it can aid individuals in obtaining justice. According to the research by Richardson et al., AI-driven predictive policing and criminal probation services could increase implicit biases and further alienate specific communities.¹⁴ The application of AI in the justice system also raises issues of unequal access. AI could potentially worsen the digital divide, especially given the large gaps in access to AI among nations, genders, generations,

and rural and urban regions.¹⁵ Given these factors, it is valid to question whether an ethical responsibility exists to guarantee everyone's equal access to the foundational technology.¹⁶

Weaponizing AI

Fourth, autonomous weapons using AI technology introduce a myriad of ethical dilemmas that demand careful consideration. Although substituting human soldiers with robotic counterparts might initially seem ethically advantageous by reducing human casualties, this perspective is met with valid counterarguments. For instance, deploying AI-powered weaponry can potentially lower the threshold for engaging in warfare, as the absence of immediate human risk may encourage aggressive tactics or lead to the misinterpretation of intentions between nations. Moreover, concerns arise about the decision-making abilities of AI in difficult and complex combat situations, which could involve non-combatants, as well as the possibility of AI malfunction and accountability in case of unwanted collateral damage.

AI in the Healthcare Sector

Fifth, healthcare concerns. The application of artificial intelligence in healthcare presents both opportunities and challenges due to its potential to improve diagnostics and treatment while also posing risks for unintended consequences. A significant ethical issue frequently discussed in this context is the replacement of human touch with technology.¹⁷ Consequently, humanoid robot systems have been suggested as remedies for numerous obstacles faced by the healthcare sector. Despite AI's promise to enhance diagnostic accuracy, its vulnerability to errors causes companies to remain cautious about embracing AI for diagnostic purposes. IBM Watson for Oncology, for example, is a well-known AI tool prone to providing unreliable recommendations within the medical realm.¹⁸

Freedom and Autonomy

Sixth, AI's influence on freedom and individual autonomy proves to be more pervasive and complex. The technological advancements around us shape the extent of our agency by controlling or limiting access to information. In Lawrence Lessig's study, digital technology was initially perceived as regulating certain behaviors like a law, but this comprehension soon reveals its limitations.¹⁹ At times, individuals might not even recognize the expansion or contraction of their choices due to technology's broader impact on society and AI's specific role. The conclusion drawn from this analysis indicates that technology plays a significant role in shaping social reality, and this intervention produces consequences without necessarily involving a deliberate attempt to deceive or misinform.

AI and Job Replacement

Seventh, concerns about technology negatively impacting the job market have been widely discussed. AI is a unique technology in that many jobs that are at risk of automation due to AI may in fact be higher-paying white-collar jobs. For example, generative AI can create content, develop software, and perform complex tasks, which may replace traditional roles held by professionals and white-collar workers. For instance, AI can analyze legal documents, create news articles, assist in financial analysis, and even design simple software applications. Therefore, unlike traditional automation, which mostly impacted manual and repetitive jobs, AI has the potential to disrupt higher-paying positions in various sectors, such as law, finance, journalism, and technology. This shift may lead to increased unemployment and economic insecurity for well-educated and skilled individuals

in the job market.²⁰ Indeed, a Forbes study reports that approximately 85 million jobs could be lost to AI automation by 2025.²¹

Safety and Exploitation of AI

Eighth, safety emerges as a vital ethical concern to be addressed by AI, especially in the context of systems directly engaging with the physical realm, such as autonomous vehicles or those overseeing crucial health services. Furthermore, cases of AI-driven fraud, like voice impersonation and deepfakes, underscore the urgent need for ethical deliberation in AI. Recently, criminals used AI-based voice technology to imitate a CEO's voice, demanding the fraudulent transfer of \$243,000.²² This is not an isolated event: voice fraud saw a 350% increase between 2013 and 2017, according to PINDROP data.²³ The potential application of AI-based voice impersonation in fraudulent scenarios is just one of many related concerns. Additionally, the spread of AI-generated fake news raises ethical qualms, prompting organizations like OpenAI to withhold the release of such technologies due to potential misuse. The recent outcomes of releasing newly developed AI technology to the public all highlight the potential dangers that might eventually surface against individuals, businesses, and communities.

Addressing Ethical Challenges

Ultimately, the consequence of releasing a fully developed artificial intelligence is a multifaceted problem that must be considered within the broader context of societal issues. In order to properly prevent the aforementioned challenges, governments, industries, and researchers must collaborate with each other to establish regulations that prioritize human rights, societal well-being, and environmental sustainability.

Solutions for Responsible Development of AI

AI ethics involves addressing its complexity due to evolving societal values and cultural norms, requiring a multifaceted approach that aligns machines with human goals and values. Comprehensive regulations and policies are essential for responsible and ethical AI development, deployment, and use. Ethical AI emphasizes shared values like transparency, fairness, privacy protection, explainability, resilience, and trust.

AI Principles

Many countries and institutions moved quickly to introduce AI principles as a low-touch way to guide the ethical development of AI. To show the moral development of AI, many countries and institutions have introduced AI principles, such as the World Economic Forum's nine ethical AI principles, which focus on critical shared values like transparency, fairness, privacy protection, explainability, resilience, and trust. These principles serve as a framework to ensure AI systems complement human intelligence, protect data privacy, promote transparency and accountability, mitigate biases, and maintain reliable and trustworthy AI systems.

Considering the example of autonomous vehicles, implementing these ethical AI principles involves ensuring transparency and explainability in the system's decision-making and data-collection processes. This could include providing passengers with information about how the AI system functions and the data it collects. Having clear explainability in place allows users to understand the AI system's decision processes, helping to build accountability and trust, particularly in the event of an accident or unforeseen situation. Moreover, fostering fairness in AI systems requires addressing potential biases based on factors like race, gender, or age, and ensuring equitable treatment for all users.

Although these ethical AI principles provide guidance, their effectiveness is limited by a lack of clarity and enforcement. Without clear incentives for adherence or proper auditing mechanisms to hold actors accountable, the implementation of these principles may be inconsistent and insufficient to ensure the responsible and ethical development of AI systems.

To overcome these challenges, stronger regulatory frameworks, consistent global standards, and robust mechanisms for monitoring compliance with ethical AI principles are increasingly necessary.

Regulation

Accordingly, many governments across the globe are now exploring AI regulations to tackle these ethical issues. The European Council is among those at the forefront of AI regulations. The European Council is committed to fostering secure and ethical AI that upholds fundamental rights without compromising human rights. The AI Act employs a risk-based approach and establishes a standardized, horizontal legal framework for AI to ensure legal certainty.²⁴ It promotes capital investment and innovation in AI while enhancing governance and enforcement of existing safety and welfare laws, thereby paving the way for a unified AI software market. The Act also complements ongoing efforts, such as the Integrated Plan on Artificial Intelligence, to increase AI investment across Europe. Some key proposals from the European Council include the following.

First, the AI Act bans specific AI practices. It forbids private actors from using AI for social scoring and prohibits AI systems that exploit vulnerable individuals based on their social or economic circumstances. It also clarifies the limited situations in which law enforcement can use "real-time" remote biometric identification systems in public places.

In addition, the EU's proposed act also stipulates AI standards. The AI Act clarifies that general-purpose AI systems must meet the same requirements as high-risk AI systems. However, implementing these requirements will be determined through an extensive consultation process, analyzing the potential impact, value chain, technical feasibility, and market and technological trends of the systems in question. This approach ensures that all AI systems adhere to robust ethical and safety standards.²⁵

There are, however, many who criticize such regulations and who argue that such rules may hinder innovation and impede the development of AI technologies. Some believe the broad scope and stringent requirements outlined in regulatory frameworks like the AI Act might result in overregulation, ultimately harming the competition and growth of industries within regulated regions. These critics point to the fact that the EU lacks globally competitive AI players, especially compared to less regulated countries like the US and China. Beyond this, there is also a concern that differing regulatory frameworks across countries could create barriers to international collaboration in AI research and development. Harmonization of regulations may prove challenging, making it harder for companies and researchers to work together and exchange ideas on a global scale. Moreover, detractors also argue that the focus on risk-based approaches and legal liability in AI regulation may lead organizations to optimize for compliance rather than genuine safety and ethical considerations.²⁶

Proposed Way Forward

To address the concerns raised by critics, the proposed way forward includes the following recommendations:

First, nations must implement a harmonized framework for regulating AI technologies. In order to strike an equilibrium between excessive regulation and insufficient oversight, it's crucial for regulators to devise guidelines that account for users' protection and safety while concurrently enabling sufficient flexibility for technological advancements. To attain this optimal balance, regulators must thoroughly comprehend the unique risks and advantages associated with diverse AI applications. Consequently, it's crucial to adopt an iterative methodology that entails the establishment of policy sandboxes, where new regulations can be tested and refined in a controlled environment.

Second, nations must proactively encourage the growth and competitiveness of AI industries with an emphasis on ethical AI. Incorporating ethical considerations into AI technology is not a hindrance to its development, but rather presents valuable opportunities for the creation of new specialized industries. In order to accomplish this, governments should allocate resources toward the research in responsible AI, as well as nurture the establishment of innovative business models that focus on third-party AI auditing and transparency evaluations.

Third, promoting multistakeholder engagement is vital in creating a well-rounded AI policy. This can be achieved by establishing citizen dialogue or panels aimed at facilitating the co-creation of AI policy strategies. Multiple diverse opinions need to be expressed and encouraged to allow for more publicity of this technology. In addition, it's important that open communication and exchanges of ideas are available to groups such as government representatives, industry experts, researchers, ethicists, civil society organizations, or members of affected communities. In the long run, this collaborative approach ensures diverse viewpoints are considered and helps to create a more robust and inclusive AI policy.

Conclusion

In conclusion, AI's rapid progress has unlocked new opportunities. However, it also presents various challenges, worries, and ethical issues that must be acknowledged. It is crucial to recognize that these concerns have deep historical roots, dating back to discussions on ethics in technology and AI from ancient Greek philosophers to modern existentialist thinkers. Early visionaries like Asimov, Turing, and Wiener raised AI-related concerns that have since transformed into contemporary discussions on data privacy, bias, accountability, transparency, and the implications of AI on the workforce and society. More recently, the growing focus on AGI underscores the potential existential ethical risks posed by highly intelligent machines, stressing the importance of foreseeing and addressing these issues.

Today, the AI debate has largely been centered around eight key dimensions: employment, bias, justice, weapons, healthcare, safety, freedom, and privacy. Policymakers have sought to address these challenges by adopting an all-encompassing approach, incorporating common principles, extensive regulations, and cohesive policy frameworks. Nonetheless, as argued in the essay, striking an ideal balance between user protection and spurring innovation remains elusive. Consequently, it is essential to involve a diverse range of stakeholders and adopt an adaptive approach to policy formulation, so as to cultivate a global AI ecosystem that prioritizes ethical concerns and communal values.

Ultimately, by fostering a harmonized regulatory environment, actively supporting ethical AI research, and considering a wide spectrum of perspectives, we can ensure AI technologies are devised and implemented for the greater benefit of humanity. Ultimately, it helps in bridging the divide between human values and state-of-the-art AI capabilities.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

Endnotes

¹. John McCarthy, Marvin L Minsky, Nathaniel Rochester, and Claude E Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," *AI Magazine* 27 (4): 12, <https://doi.org/10.1609/aimag.v27i4.1904>.

2. Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (Prentice Hall, 2020), 5-16
3. Sandra Wachter, Brent Mittelstradt, and Luciano Floridi, "Transparent, Explainable, and Accountable AI for Robotics," *Science Robotics* 2 (6): 4-5, <https://doi.org/10.1126/scirobotics.aan6080>.
4. Mark Coeckelbergh, "Artificial Intelligence: Some Ethical Issues and Regulatory Challenges," *Technology and Regulation* 2019 (1): 31-32, <https://doi.org/10.26116/techreg.2019.003>.
5. Alan Turing, "Computing Machinery and Intelligence," *Mind* LIX (236): 433-435, <https://doi.org/10.1093/mind/lix.236.433>.
6. Norbert Wiener, *The Human Use of Human Beings* (Da Capo Press, 1988), 164-190
7. Michael Anderson and Susan Leigh Anderson, *Machine Ethics* (Cambridge University Press, 2011), 158
8. Miles Brundage, "Limitations and Risks of Machine Ethics," *Journal of Experimental & Theoretical Artificial Intelligence* 26(3): 360, <https://doi.org/10.1080/0952813x.2014.895108>.
9. Giovanni Buttarelli, "Choose Humanity: Putting Dignity Back into Digital," speech at 40th International Conference of Data Protection and Privacy Commissioners, October 24th, 2018, 2, https://edps.europa.eu/sites/default/files/publication/18-10-24_choose_humanity_speech_en_1.pdf.
10. Carter Jernigan and Behram Mistree, "Gaydar: Facebook Friendships Expose Sexual Orientation," *First Monday* 14 (10), <https://doi.org/10.5210/fm.v14i10.2611>.
11. Jim Isaak and Mina J. Hanna, "User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection," *Computer* 51(8): 57-58, <https://doi.org/10.1109/mc.2018.3191268>.
12. Filippo Raso, Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz, and Kimberly Levin, "Artificial Intelligence & Human Rights: Opportunities & Risks," *Berkman Klein Center Research Publication* 2018 (6), <http://nrs.harvard.edu/urn-3:HUL.InstRepos:38021439>.
13. Heidi Ledford, "Millions of Black People Affected by Racial Bias in Health-Care Algorithms," *Nature* 574 (October): 608, <https://doi.org/10.1038/d41586-019-03228-6>.
14. Rashida Richardson, Jason Schultz, and Kate Crawford, "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice," *NYU Law Review* 192 (2019), <https://ssrn.com/abstract=3333423>.
15. Kevin McSorley, "The Secular Salvation Story of the Digital Divide," *Ethics and Information Technology* 5 (2): 80, <https://doi.org/10.1023/a:1024946302065>.
Govindan Parayil, "The Digital Divide and Increasing Returns: Contradictions of Informational Capitalism," *The Information Society* 21 (1): 48-50, <https://doi.org/10.1080/01972240590895900>.

16. Alexander Babuta, Marion Oswald, and Ardi Janjeva, “Artificial Intelligence and UK National Security: Policy Considerations.” *RUSI*: 28, <https://rusi.org/explore-our-research/publications/occasional-papers/artificial-intelligence-and-uk-national-security-policy-considerations>
17. Moritz Goeldner, Cornelius Herstatt, and Frank Tietze, “The Emergence of Care Robotics — Patent and Publication Analysis.” *Technological Forecasting and Social Change* 92 (March): 122, <https://doi.org/10.1016/j.techfore.2014.09.005>.
18. “Top 4 Challenges of AI in Healthcare & How to Overcome Them.” AIMultiple, Cem Dilmegani, published March 1, 2022, <https://research.aimultiple.com/challenges-of-ai-in-healthcare/>.
19. Lawrence Lessig, *Code and Other Laws of Cyberspace* (Basic Books, 1999), 46.
20. Joan Bakewell, Timothy Clement-Jones, Anthony Giddens, Rosalind Grender, Clive Hollick, Christopher Holmes, Peter Levene, et al, “AI in the UK: Ready, Willing and Able?” House of Lords: 70-75, <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>.
21. “U.S. Lost over 60 Million Jobs—Now Robots, Tech, and Artificial Intelligence Will Take Millions More.” Forbes, Jack Kelly, published October 27, 2020, <https://www.forbes.com/sites/jack-kelly/2020/10/27/us-lost-over-60-million-jobs-now-robots-tech-and-artificial-intelligence-will-take-millions-more/>.
22. “Fraudsters Used AI to Mimic CEO’s Voice in Unusual Cybercrime Case.” Wall Street Journal, Catherine Stupp, published August 30, 2019, <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>.
23. “A New Kind of Cybercrime Uses AI and Your Voice against You.” Quartz, Ephrat Livni, published August 31, 2019, <https://qz.com/1699819/a-new-kind-of-cybercrime-uses-ai-and-your-voice-against-you>.
24. “Artificial Intelligence Act: Council Calls for Promoting Safe AI That Respects Fundamental Rights.” Council of the European Union, EU, released December 6, 2022, <https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/>.
25. Miles Brundage, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, et al, “The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation,” 50-52, <https://arxiv.org/pdf/1802.07228>.
26. Rowena Rodrigues, “Legal and Human Rights Issues of AI: Gaps, Challenges, and Vulnerabilities,” *Journal of Responsible Technology* 4 (100005), <https://doi.org/10.1016/j.jrt.2020.100005>.

References

Anderson, Michael, and Susan Leigh Anderson. 2011. *Machine Ethics*. New York: Cambridge University Press.

Babuta, Alexander, Marion Oswald, and Ardi Janjeva. 2020. “Artificial Intelligence and UK

National Security: Policy Considerations.” *RUSI*: 1-50. <https://rusi.org/explore-our-research/publications/occasional-papers/artificial-intelligence-and-uk-national-security-policy-considerations>

Bakewell, Joan, Timothy Clement-Jones, Anthony Giddens, Rosalind Grender, Clive Hollick, Christopher Holmes, Peter Levene, et al. 2018. “AI in the UK: Ready, Willing and Able?” <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>.

Brundage, Miles. 2014. “Limitations and Risks of Machine Ethics.” *Journal of Experimental & Theoretical Artificial Intelligence* 26 (3): 355–72. <https://doi.org/10.1080/0952813x.2014.895108>.

Brundage, Miles, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, et al. 2018. “The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.” <https://arxiv.org/pdf/1802.07228>.

Buttarelli, Giovanni. 2018. “Choose Humanity: Putting Dignity Back into Digital.” Presented at the 40th International Conference of Data Protection and Privacy Commissioners, October 24. https://edps.europa.eu/sites/default/files/publication/18-10-24_choose_humanity_speech_en_1.pdf.

Coeckelbergh, Mark. 2019. “Artificial Intelligence: Some Ethical Issues and Regulatory Challenges.” *Technology and Regulation* 2019 (1): 31–34. <https://doi.org/10.26116/techreg.2019.003>.

Council of the European Union. 2022. “Artificial Intelligence Act: Council Calls for Promoting Safe AI That Respects Fundamental Rights.” Council of the EU. December 6, 2022. <https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/>.

Delcker, Janosch. 2018. “Europe Divided over Robot ‘Personhood.’” Politico. April 11, 2018. <https://www.politico.eu/article/europe-divided-over-robot-ai-artificial-personhood/>.

Department of Defense. 2019. “AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense Defense Innovation Board.” https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF.

Dilmegani, Cem. 2022. “Top 4 Challenges of AI in Healthcare & How to Overcome Them.” AIMultiple. March 1, 2022. <https://research.aimultiple.com/challenges-of-ai-in-healthcare/>.

Goeldner, Moritz, Cornelius Herstatt, and Frank Tietze. 2015. “The Emergence of Care Robotics — Patent and Publication Analysis.” *Technological Forecasting and Social Change* 92 (March): 115–31. <https://doi.org/10.1016/j.techfore.2014.09.005>.

Green, Brian. 2020. “Artificial Intelligence and Ethics: Sixteen Challenges and Opportunities.” Markkula Center at Santa Clara University. August 18, 2020. <https://www.scu.edu/ethics/all-about-ethics/artificial-intelligence-and-ethics-sixteen-challenges-and-opportunities/>.

Isaak, Jim, and Mina J. Hanna. 2018. "User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection." *Computer* 51 (8): 56–59. <https://doi.org/10.1109/mc.2018.3191268>.

Jernigan, Carter, and Behram Mistree. 2009. "Gaydar: Facebook Friendships Expose Sexual Orientation." *First Monday* 14 (10). <https://doi.org/10.5210/fm.v14i10.2611>.

Kelly, Jack. 2020. "U.S. Lost over 60 Million Jobs—Now Robots, Tech, and Artificial Intelligence Will Take Millions More." *Forbes*. October 27, 2020. <https://www.forbes.com/sites/jackkelly/2020/10/27/us-lost-over-60-million-jobs-now-robots-tech-and-artificial-intelligence-will-take-millions-more/?sh=24ed59001a52>.

Ledford, Heidi. 2019. "Millions of Black People Affected by Racial Bias in Health-Care Algorithms." *Nature* 574 (October): 608–609. <https://doi.org/10.1038/d41586-019-03228-6>.

Lessig, Lawrence. 1999. *Code and Other Laws of Cyberspace*. New York: Basic Books.
Livni, Ephrat. 2019. "A New Kind of Cybercrime Uses AI and Your Voice against You." *Quartz*. August 31, 2019. <https://qz.com/1699819/a-new-kind-of-cybercrime-uses-ai-and-your-voice-against-you>.

McCarthy, John, Marvin L Minsky, Nathaniel Rochester, and Claude E Shannon. 1955. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence." *AI Magazine* 27 (4): 12–12. <https://doi.org/10.1609/aimag.v27i4.1904>.

McSorley, Kevin. 2003. "The Secular Salvation Story of the Digital Divide." *Ethics and Information Technology* 5 (2): 75–87. <https://doi.org/10.1023/a:1024946302065>.

Parayil, Govindan. 2005. "The Digital Divide and Increasing Returns: Contradictions of Informational Capitalism." *The Information Society* 21 (1): 41–51. <https://doi.org/10.1080/01972240590895900>.

Raso, Filippo, Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz, and Kimberly Levin. 2018. "Artificial Intelligence & Human Rights: Opportunities & Risks." *Berkman Klein Center Research Publication* 2018 (6). <http://nrs.harvard.edu/urn-3:HUL.InstRepos:38021439>.

Richardson, Rashida, Jason Schultz, and Kate Crawford. 2019. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." *NYU Law Review* 92 (2019). <https://ssrn.com/abstract=3333423>.

Rigby, Michael J. 2019. "Ethical Dimensions of Using Artificial Intelligence in Health Care." *AMA Journal of Ethics* 21 (2): E121–124. <https://doi.org/10.1001/amajethics.2019.121>.

Rodrigues, Rowena. 2020. "Legal and Human Rights Issues of AI: Gaps, Challenges, and Vulnerabilities." *Journal of Responsible Technology* 4 (100005): 100005. <https://doi.org/10.1016/j.jrt.2020.100005>.

Ruehle, Chrissann R. 2019. "Investigating Ethical Considerations of Machine Learning Adoptions within Organizations: A Systematic Literature Review." *Proceedings of the Ninth International*

Conference on Engaged Management Scholarship 2019 (September).
<https://doi.org/10.2139/ssrn.3454120>.

Russell, Stuart, and Peter Norvig. 2020. *Artificial Intelligence: A Modern Approach*. 4th ed. Prentice Hall.

Ryan, Mark, and Anya Gregory. 2019. "Ethics of Using Smart City AI and Big Data: The Case of Four Large European Cities." *ORBIT Journal* 2 (2). <https://doi.org/10.29297/orbit.v2i2.110>.

Stupp, Catherine. 2019. "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case." *Wall Street Journal*. August 30, 2019. <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>.

Turing, Alan. 1950. "Computing Machinery and Intelligence." *Mind* LIX (236): 433–60.
<https://doi.org/10.1093/mind/lix.236.433>.

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 2017. "Transparent, Explainable, and Accountable AI for Robotics." *Science Robotics* 2 (6): <https://doi.org/10.1126/scirobotics.aan6080>.

Wiener, Norbert. 1988. *The Human Use of Human Beings*. Da Capo Press.