

# Examining the Advantages of Artificial Intelligence Alongside Its Potential Risks on Human Wellbeing, Data Privacy, and National Security

Olushola Agbaje\*

*Pursuing Masters at Harrisburg University of Science and Technology, 326 Mkt St, Harrisburg, PA 17101,  
United States*

*Email: agbajeshola@gmail.com, Email: oagbaje1@my.harrisburgu.edu*

## Abstract

This study seeks to comprehensively analyze the benefits and risks of artificial intelligence and discuss strategies and policies to balance them. The paper assesses AI's positive impact on four industries - healthcare, finance, transportation, and education – juxtaposed with its negative welfare, privacy, and security effects. The study utilizes a semi-systematic review methodology to explore diverse narratives surrounding AI's societal implications. Key findings suggest AI can improve decision-making, productivity, and quality of life but risks exacerbating bias, unemployment, and insecurity if not developed responsibly. The paper discusses practical strategies, policies, and regulatory interventions to help balance AI's pros and cons, including human-centered design, explainable AI, and governance frameworks. It also suggests actionable recommendations for individual, organizational, and national stakeholders. Suggestions for future research include developing robust AI resilient to attacks, increasing AI transparency and accountability, assessing long-term societal impacts, and addressing legal and ethical dilemmas. This timely study contributes a measured perspective to current debates on AI and provides a framework to help appropriate its advantages while mitigating its perils.

**Keywords:** artificial intelligence; benefits of AI; AI risks on human well-being; AI risks on data privacy; AI risks on national security; semi-systematic meta-narrative review.

## 1. Introduction

Artificial Intelligence (AI) has emerged as a transformative technology with the potential to revolutionize various sectors and improve human lives. However, as with any disruptive technology, AI brings both advantages and risks that need to be carefully examined. This research aims to explore the advantages of AI while also assessing its potential risks on human welfare, data protection, and territorial security.

---

*Received: 9/12/2023*

*Accepted: 10/17/2023*

*Published: 10/30/2023*

---

\* Corresponding author.

For this paper, a qualitative study was conducted to probe the nuances of these issues and provide valuable insights for policymakers, businesses, and individuals.

### ***1.1 Key definitions and essential terminologies***

To serve as a common lexical framework throughout the research paper and to ensure clarity and consistency in communication and analysis, it is important to establish the meanings of the four topical terms in the context of this study: artificial intelligence, human wellbeing, data privacy, and national security.

#### ***1.1.1 Artificial intelligence***

Though finding a generally accepted definition for Artificial Intelligence has remained elusive [1, 4], viewing it via the lenses of the four categorical approaches commonly adopted by its researchers – as if from the cardinal directions of the field – should suffice [5, p. 2]. From the perspective of "acting humanly" [5, pp. 2-3), AI is the practice of building devices that can carry out tasks that would typically require people to use their intelligence [6] and the study of programming computers to perform tasks that people are currently better at [7]. From the perspective of "thinking humanly" [5, p. 3), AI is the innovative new attempt to get computers thinking [8] and the automation of processes that are associated with thinking in humans [9]. From the perspective of "thinking rationally" [5, p. 4), AI is the use of computational models for the study of mental faculties [10] and the study of the calculations necessary for perception, reasoning, and action [11]. Finally, from the perspective of "acting rationally" [5, pp. 4-5), AI is concerned with how inanimate objects exhibit intelligent behavior [12], and Computational Intelligence is the study of intelligent agent design [13]. Alternatively, one could just settle for the simple and reader-friendly definition by [4]: Artificial intelligence is a branch of computer science in which researchers work to impart more human-like intelligence on computer systems.

#### ***1.1.2 Human wellbeing, data privacy, national security***

Just like the challenge with defining artificial intelligence, there is no consensus definition for “wellbeing” [14] (Simons & Baldwin, 2021). Nevertheless, Simons & Baldwin [14] goes ahead to give it an “operational definition” as a subjective and/or objective state of positive emotions and fulfillment of personal potential, measured by a salutogenic approach.

Likewise, [15] while conceding that the task is “notoriously complex”, defines data privacy as a subset of data security focused on the different aspects of data handling – consent (the ifs and hows of third-party data sharing), notice (the hows of legal data collection and storage), and regulatory requirements and restrictions, for example the GDPR or HIPAA.

National security is the capacity of a state to provide for the defense and protection of its populace [16].

### ***1.2 Background and context of the research topic***

Artificial Intelligence (AI) has rapidly evolved from being a topic limited to science fiction and technical jargon

for geeks in the mid-1900s to a widely discussed technology of the early 21st century [17]. AI is significant not only to the science and technology communities [18, 21], but also to consumers [22, 24], companies [25, 27], and countries [28]. The pace of research and development in this exciting field is matched only by the multitude of use cases in different industries, including healthcare, finance, transportation, and education [29].

However, like the bipolar Dr. Jekyll and Mr. Hyde character [30], AI also has its dark side. It is susceptible to algorithmic bias [31], particularly in law enforcement [32], HR recruitment [33], and credit assessment [34], resulting in discriminatory outcomes [35]. This "shadow" side of AI was the subject of a 2020 Netflix documentary, "Coded Bias," written, directed, and produced by Shalini Kantayya and featuring Joy Buolamwini, an MIT Media Lab researcher [36].

When misused or compromised, AI systems can result in data breaches and privacy violations [37]. A vivid example is China's Alipay Health Code app, launched in early 2020 to track the health status of Chinese residents during the Covid-19 pandemic. "China's coronavirus health code apps raise privacy concerns," claims a news article from April 1, 2020, written by Taipei-based Guardian UK correspondent Helen Davidson and posted on the British media outlet's website. The article claims that the app's collection of sensitive personal data, such as passport biodata, medical records, and travel history, allows civil liberties to be violated [38].

Most worrisome is the capacity for the "weaponization of AI" [39, 40] by malevolent state and non-state actors to attack critical infrastructure, wage high-tech warfare, or execute disinformation campaigns. In May 2019, Washington Post tech reporter Drew Harwell wrote about a doctored video of Nancy Pelosi that was widely circulated on social media [41]. The US House Speaker appeared inebriated in the video and slurred her words. Though the viral video was later revealed as a "deepfake" [42], Pelosi's reputation had already taken a significant hit. Deepfakes are AI-produced video manipulations or other digital representations that produce faked sounds and images that seem real [43].

### ***1.3 Purpose statement and research problem***

AI's advancement is happening quickly, and its influence on society is increasing exponentially. Therefore, it is imperative to conduct extensive assessments and comparative analyses of its benefits and risks that provide users, practitioners, and policymakers with a balanced yet informed perspective [44]. This exercise is important because people, businesses, institutions, and governments need to urgently develop behaviors, strategies, policies, and regulations that would help control AI's untoward effects without stifling the technology's progress [45]. Furthermore, this paper contributes to the ongoing public discourse on AI's societal implications and the need to promote responsible and ethical development [46].

The central problem necessitating this study arose because of the lip service paid by stakeholders towards the sustainable development and deployment of powerful AI technologies. In 2017, a conference was held to tackle these problems, leading to the creation of the AI Asilomar Principles by the Future of Life Institute [47]. However, some worry that the actions being taken now are too little too late, particularly when compared to the biotech industry stakeholders' response to a similar gathering at the exact location in 1975, as noted by [48].

#### **1.4 Research questions and objectives**

The research questions to be investigated in this study include:

- What are the reported advantages of AI in each of the selected industries (healthcare, finance, transportation, and education)?
- What are the known risks of AI on human wellbeing, data privacy, and national security?
- How are these risks currently being mitigated through existing government policies and corporate strategies?
- What future policies and strategies could be implemented to further mitigate these risks?
- How can a balance be achieved in the development, deployment, and adoption of AI that maximizes its benefits while minimizing its risks?

Based on the research questions, the aims of this research are:

1. To identify and analyze the key advantages of artificial intelligence across different domains.
2. To examine the risks associated with the widespread adoption of AI on human well-being.
3. To investigate the implications of AI on data privacy and the challenges it poses for individuals and organizations.
4. To assess the risks AI poses to national security and critical infrastructures.
5. To conduct a comparative analysis of the prospects and dangers of AI thus encouraging balance in future reporting of the technology.
6. To compile and evaluate some of the current interventions that have been implemented for responsible, ethical, and sustainable development and deployment of AI technologies.

#### **1.5 Scope of the study**

The delimitations of this study are as follows:

1. **Timeframe:** The study focuses on an eight-decade period from 1940 to 2020. While this timeframe allows for a comprehensive – albeit not exhaustive – examination of historical and contemporary moments in the struggle between technology (AI) and humanity (security), it does not capture seminal and future developments before 1940 and beyond 2020 respectively.
2. **Scope:** The study primarily focuses on the advantages of AI within the healthcare, finance, transportation, and education sectors as case studies as well as potential risks of AI on human well-being, data privacy, and national security specifically. Other aspects related to AI, such as its impact – positive or negative – on politics or macroeconomics, are not extensively explored within the limits of

this study.

3. **Geographical Focus:** The study does not specifically focus on a particular region or country. It aims to provide a global perspective on the advantages and potential risks of AI. However, regional variations in the adoption and regulation of AI may not be extensively analyzed within the boundaries of this paper.
4. **Specific AI Technologies:** The study does not delve into specific AI technologies or applications in detail. It aims to provide a broader understanding of AI's advantages and potential risks rather than focusing on specific AI systems or algorithms.

It is important to acknowledge these delimitations as they provide a clear focus to the study. Researchers and stakeholders should consider these limitations when interpreting the findings and applying them to specific contexts or future research endeavors.

### ***1.6 Overview of the structure of the research***

This research paper comprises six main chapters and an appended reference list:

Chapter 1 (Introduction) serves as the orientation point for the topic discussed throughout the entire research paper. It communicates what the study is about and why it is important, gives a bird's-eye view of the whole study, lays the groundwork for what is to come, gives a teaser of what is discussed within, and provides the underlying motivation for such a demanding project.

Chapter 2 (Theoretical Frameworks) serves as the structural platform of the study via broad overviews of two seminal publications and two pivotal events that underpin the two topical variables of advancement in technology and impact on humanity.

Chapter 3 (Methodology) describes the research process selected for the study, including the search strategy and methods of data extraction. The adoption of a qualitative, narrative research approach will be justified, and the broader research design will be discussed, including the limitations thereof.

Chapter 4 (Findings) identifies, evaluates, and synthesizes prior studies and landmark publications that address the research paper topic. It also presents the main findings from the reviewed literature, as synthesized from the results of the various studies under review. It also discusses common themes, trends, or conclusions drawn from the literature, and identifies any gaps or areas for future research.

Chapter 5 (Discussion) juxtaposes the findings in Chapter 5, identifying trade-offs and possible mitigations. It also explores policy implications, regulatory considerations, and strategic recommendations for stakeholders.

Chapter 6 (Conclusion) serves as the capstone by summarizing the overall findings, recommendations,

limitations, gaps in the current literature, and suggestions for future research while recapping the critical talking points of the entire research project.

Finally, the References section includes a list of all the sources consulted (in alphabetical order) while writing this paper, following the APA style guide and convention.

## 2. Theoretical Frameworks

This section features meta-narrative reviews on key historical and contemporary moments in the struggle for supremacy between technology (AI) and humanity (security) within a historic eight-decade period (1940-2020).

### 2.1 *Man vs. machine: Maslow vs. Turing*

Abraham Maslow (1908-1970) was an American psychologist renowned for his theories on human motivation, particularly his well-known "Hierarchy of Needs." Maslow was one of the pioneers of humanistic psychology [49, p. 39], focusing on the individual's potential for personal growth. Maslow's hierarchy of needs theory proposes that human needs are arranged in a pyramid, with basic "physiological" needs at the bottom, ascending towards higher-level "safety," "love and belonging," "esteem," and "growth" needs [49, pp. 40–41]. In 1943, Abraham Maslow published "A Theory of Human Motivation," a landmark work that contextualizes the ongoing discourse about the connection between human beings and machines. According to Maslow's theory, human needs are arranged hierarchically, starting from the most basic needs for survival and culminating in self-actualization [50]. Alan Turing (1912-1954) was a renowned British mathematician and logician. He is widely recognized as the pioneer in modern computing and artificial intelligence. Turing was arguably the first person to contemplate the possibility of computer intelligence [10]. He started his pioneering 1950 paper "Computing Machinery and Intelligence" by asking, "Can machines think?" a rather provocative question at the time, and then sought to answer it with a thought experiment that he called The Imitation Game [51, pp. 433–434]. In the remaining pages of his Mind magazine article, Turing proposed programming a computer to exhibit intelligent behavior. He also countered contemporary and hypothetical arguments against machine intelligence while predicting that by the turn of the millennium, human minds and machines would be indistinguishable [51]. In 1950, Alan Turing published "Computing Machinery and Intelligence," which proposed the Turing Test, also known as the Imitation Game [52]. The Turing Test is a method for determining whether a computer can achieve "indistinguishability" from a human [53].

It is interesting to note that [54] "rediscovered" the concept of "transcendence" - in which humans attain '*selflessness*' (emphasis mine) - originated by [55] within a year's interval of [56] predicting the imminence of the "singularity" - in which machines attain '*selfishness*' (emphasis mine) - a concept introduced into the public consciousness by [57]. Perhaps even more interesting is that the 2014 science fiction thriller film written by Jack Paglen, directed by Wally Pfister, and headlined by Johnny Depp, "Transcendence" was really about the Singularity by [58]. Most of the arguments and analysis featured throughout this study would lean on the constructs framed by these two seminal publications, authored by two pioneers whose respective fields are worlds apart, to encourage the pursuit of technical advancements in ways that are not only beneficial to the

human condition but also sustainable for the society.

## ***2.2 Machine vs. man: McCarthy and his colleagues vs Tegmark and his colleagues***

The 1956 Dartmouth College AI Conference and the 2017 Asilomar Conference on Beneficial AI are two pivotal events in the evolutionary history of artificial intelligence. The Dartmouth Conference proposed the creation of intelligent machines capable of mimicking human intelligence. The Asilomar Conference emphasized the need for ethical and responsible AI development. The former is considered the foundational event that birthed the field of artificial intelligence. Conversely, the latter was positioned as the epochal event that would "save" AI or, more accurately, humans from machines. The seminal ideas that gave impetus to the 2017 Conference were contained in a document known as the 1955 Dartmouth Proposal, or officially "The Proposal for the Dartmouth Summer Research Project on Artificial Intelligence." The late emeritus Stanford professor John McCarthy authored the proposal in collaboration with fellow cognitive scientist Marvin Minsky, fellow computer scientist Nathaniel Rochester, and fellow mathematician Claude Shannon. The landmark study was predicated on the assumption that it should be possible, in theory, to precisely describe the features of human intelligence - including learning - in ways that a machine can imitate. The pioneer researchers were also interested in investigating "how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves" [59, pp. 12-13]. Ultimately, the agenda was to develop machines that could think, learn, and solve problems like humans, a radical albeit controversial concept at the time. J. McCarthy and his colleagues M. Minsky, N. Rochester, and C. Shannon organized the Dartmouth Conference in the 1950s. This conference was significant as it brought together experts from various fields, such as mathematics, psychology, and computer science, to explore the possibility of creating machines that could simulate human intelligence. It also marked the first time the term "artificial intelligence" was used to describe the field and established AI as a distinct area of research. The conference laid the foundation for future collaborations and advancements in the field [4, 5]. In 2017, the Asilomar Conference on Beneficial AI was organized by the Future of Life Institute at the Asilomar Conference Grounds in Pacific Grove, California. Over 800 participants from academia, industry, government, and civil society groups attended the conference. It aimed to discuss the ethical and societal implications of AI and develop a set of guidelines that would promote the long-term well-being of humanity [60]. The Asilomar AI Principles, consisting of 23 guidelines for developing beneficial AI, were released at the end of the conference [61]. Many well-known individuals in artificial intelligence gathered at the Asilomar Conference. Demis Hassabis from DeepMind, Elon Musk from Neuralink, and Stuart Russell from UC Berkeley were in attendance. The conference was particularly noteworthy for its emphasis on exploring the impact of AI on society and how it could shape humanity's future. It helped raise awareness about AI's challenges and opportunities and sparked ongoing discussions and debates about the responsible development and use of the technology [62]. These two pivotal gatherings by AI stakeholders two generations apart reflect the now-familiar dilemma: the tension between the desire for technological progress and the need for ethical and responsible development. While the Dartmouth Conference focused primarily on the technical aspects of AI development, the Asilomar Conference highlighted the importance of considering the broader societal implications of AI. Later, specifically in Chapter 5, this paper will discuss strategies for alleviating these difficult trade-offs toward a win-win situation.

### **3. Methodology**

#### ***1. Research approach***

A semi-systematic review approach was adopted to select the literature used for this study. This approach was chosen as it enables the thorough investigation of the different accounts and viewpoints connected to AI and its societal effects, with a loosely structured method and iterative process for gathering, analyzing, and synthesizing data to uncover and probe the various narratives and perspectives related to a specific topic. This approach acknowledges the complexity and diversity of the extant literature and seeks to capture the different perspectives and debates within the field [63].

#### ***2. Data collection***

A wide range of data sources was studied to thoroughly evaluate the literature on the influence of AI on society. The sources used were academic papers, industry reports, case studies, government documents, policy papers, and regulations, with preferences for publications considered landmark, seminal, or pivotal by experts in the field and industry. This wide range of materials enabled a thorough knowledge of the various perspectives and storylines around AI and its impact on society. A systematic search of academic databases such as Google Scholar, JSTOR, and Project Muse and a targeted search of relevant industry and government websites were done to obtain these materials. Keywords used in the search included "artificial intelligence," "AI," "wellbeing", "welfare", "privacy", "security", "national security", "ethics," "society," "policy," and "regulation". Data sources were selected based on their relevance to the research questions and the quality of research, as indicated by the number of citations. A stringent screening process was used to ensure the pertinence of the literature sources. An initial screening of titles and abstracts was followed by a full-text screening to ensure relevance to the study questions.

#### ***3. Data processing***

The data analysis process utilized qualitative techniques, such as content analysis, thematic analysis, discourse analysis, and grounded theory. The analyses involved a systematic review of the literature to identify the key themes, narratives, and perspectives surrounding AI and its impact on society. The qualitative analysis involved structured data extraction, coding, and synthesis. The data extraction process involved identifying relevant information from the selected data sources, such as critical arguments, evidence, and examples. The coding process involved categorizing the extracted data into themes and sub-themes. The synthesis process involved integrating the coded data into an overarching narrative that captured the perspectives and debates surrounding AI and its impact on society.

#### ***4. Methodology constraints***

The research design choices made for this study and the opportunity costs that arise as a result inadvertently imbue this research with unavoidable tactical limitations. The author of this study acknowledges the implicit limitations associated with using a semi-systematic literature review methodology. These constraints include:



### ***1. Limitations due to not using quantitative review***

Below are the implicit constraints that accompanied the author's choice of qualitative research in lieu of quantitative research:

#### **1. Subjectivity and Interpretation:**

- ***Qualitative Research:*** Involves gathering and interpreting non-numerical data.
- ***Limitation:*** Results are more subjective and influenced by the researcher's perspective. Personal bias can impact interpretation.

#### **2. Time-Consuming:**

- ***Qualitative Research:*** Requires time for data collection and analysis.
- ***Limitation:*** Can be more time-consuming compared to quantitative methods.

#### **3. Depth vs. Breadth:**

- ***Qualitative Research:*** Provides in-depth insights into concepts or experiences.
- ***Limitation:*** May not cover a wide range of topics or phenomena as comprehensively as quantitative research.

### ***3.4.2 Limitations due to not using systematic review***

Below are the implicit constraints that accompanied the author's choice of semi-systematic literature review in lieu of systematic literature review:

#### **1. Methodological Rigor:**

- ***Systematic Review:*** Follows a rigorous and standardized process, including explicit criteria for study selection, data extraction, and quality assessment.
- ***Semi-Systematic Review:*** May lack the same level of rigor. Selection criteria and data extraction might be less systematic, leading to potential bias.

#### **2. Comprehensiveness:**

- ***Systematic Review:*** Aims to include all relevant studies on a topic, minimizing bias.
- ***Semi-Systematic Review:*** May miss important studies due to less comprehensive search methods.

#### **3. Transparency and Reproducibility:**

- ***Systematic Review:*** Detailed methods are documented, allowing others to replicate the process.
- ***Semi-Systematic Review:*** May lack transparency in search strategies or inclusion/exclusion criteria.

#### 4. Risk of Bias:

- **Systematic Review:** Efforts are made to minimize bias (e.g., publication bias, selective reporting).
- **Semi-Systematic Review:** Greater risk of bias due to less stringent procedures.

#### 5. Time and Resources:

- **Systematic Review:** Time-consuming but thorough.
- **Semi-Systematic Review:** Faster but sacrifices depth and rigor.

#### 4. Findings

This section assesses the existing body of research around the conversations and debates surrounding the benefits of AI technologies and the drawbacks of AI adoption. It also presents the results of the content analysis, thematic analysis, discourse analysis, and grounded theory performed on literature sources.

##### 4.1. Assessment of Benefits

Research has documented several advantages of AI technologies over the last decade. It could be said that artificial intelligence has aided humans in ascending the Maslow "pyramid" by fulfilling survival, safety, social, success, and self-actualization needs.

##### 4.1.1 Evaluation of AI's Positive Impact on Society

The following are some AI use cases across four specific domains selected for this study:

**1. Healthcare:** AI is transforming the health sector by enabling faster and more accurate medical diagnosis [64] and facilitating personalized medicine [65]. It also helps drug discovery by identifying potential drug candidates [66, 67]. Additionally, AI is utilized in developing wearable medical devices that monitor health conditions and provide real-time feedback to patients and healthcare providers [68].

**2. Finance:** AI is used in the financial services industry to process large amounts of data, identify patterns and anomalies within them, and thus improve fraud detection and risk assessment [69]. Furthermore, AI is used in investment management to identify investment opportunities [70] and in personal finance to provide customized financial advice to customers [71, 72].

**3. Transportation:** AI is used in the transportation industry to develop autonomous vehicles and traffic management solutions. It helps develop autonomous vehicles by analyzing data from sensors and cameras to inform driving decisions, helping to reduce accidents usually caused by human error [73, 74]. AI helps to optimize traffic flow by analyzing data from traffic cameras and sensors, thus helping to reduce traffic congestion and improve mobility [75].

**4. Education:** AI is used to develop personalized learning programs. AI analyzes student performance data to identify strengths and weaknesses, thus helping develop personalized learning plans [76].

#### **4.1.2 Analysis of Economic Benefits and Growth**

AI has the potential to drive growth in various economic sectors. The following are some of AI's economic use cases:

**1. Job creation and automation of mundane tasks:** AI automates repetitive and iterative tasks, thus improving productivity and freeing time for knowledge workers to focus on higher-value tasks like creativity and teamwork [77]. It also reduces costs by automating labor-intensive functions [78]. AI creates new jobs in industries that require AI expertise, such as data science and machine learning, thus combating unemployment [79]. Additionally, it helps reduce the skills gap in an economy by providing training and upskilling opportunities [80].

**2. Economic productivity and innovation:** AI drives efficiency gains due to the optimization of existing processes [81]. It also improves supply chain management and logistics processes, increasing efficiency and reducing costs [82, 83]. Furthermore, AI helps accelerate the speed and accuracy of business decision-making using big data analytics [84].

**3. Market competitiveness and business opportunities:** AI helps identify new market opportunities and innovative solutions to existing problems, leading to the creation of new products and services [85]. It also helps identify customer needs and preferences, informing product development and marketing strategies. AI also helps analyze market trends and customer behavior data for business development strategy [86].

#### **4.1.3. Geopolitical advantages**

AI confers comparative benefits to countries that invest in its research and development. The following are some use cases:

**1. Military applications:** AI enhances national security and defense capabilities. AI helps develop autonomous weapons systems and improve situational awareness on the battlefield [87]. The results are a reduction in the death toll of soldiers and improved military operations' effectiveness.

**2. International relations:** AI helps facilitate communication and understanding between countries. It is used in natural language processing and translation, which helps to break down language barriers [88]. The results are improved peaceful relations between countries and reduced likelihood of misunderstandings and conflicts.

**3. Scientific research:** AI accelerates research and development in various fields, such as medicine, physics, and astronomy. The results are advancing scientific knowledge and developing beneficial new technologies [89].

It is important to conclude this section by communicating the observation that the benefits of AI are unevenly distributed. Some concerns about AI adoption may exacerbate societal inequalities, particularly regarding access to jobs, education, and healthcare. For example, automation may lead to job displacement for specific sectors of the population while only creating new jobs for those with the necessary skills and expertise [90]. Similarly, using AI in medicine may lead to better health outcomes for those who can afford it while leaving others behind [91].

#### **4.2 Assessment of Risks**

The rapid adoption of AI technologies also brings risks that have been discussed in recent literature. However, this section of the paper will pay particular attention to the extent to which the acquisition of "intelligence" by "computing machinery" [51] affects the human pursuit of "safety needs" [50].

#### ***4.2.1 Ethical considerations***

**1. Algorithmic bias and fairness concerns:** As brought up in Chapter 1, one of the risks of AI, especially the one in the eye of a media storm, is its susceptibility to algorithmic bias, in which AI systems perpetuate and amplify existing biases and discrimination in society [92, 93]. The lack of diversity in the AI development teams and the biased data used to train AI models contribute to algorithmic bias [94, 95]. Algorithm bias results in unfair outcomes in healthcare [91]; hiring [96]; lending [97]; ad targeting [98, 99]; social media targeting [100]; judicial decision-making [101, 95]; and criminal sentencing [102].

**2. Accountability and Transparency of AI Systems:** The opacity and lack of interpretability of complex AI systems can make it difficult to trust the decisions made by these systems [103, 104]. This lack of transparency poses a significant challenge when applying AI tech to life-or-death decision-making situations, such as critical surgeries. As the stakes are high and the repercussions of errors might be severe, the potential for problems in such scenarios is substantial [105].

**3. Legal and Ethical Implications of AI Decisions:** The use of AI in decision-making raises concerns about accountability and liability, which are closely connected to points 1 and 2. For instance, if an autonomous vehicle gets into an accident, it is unclear who would be responsible: the manufacturer, the programmer, or the user. This lack of clarity regarding liability could lead to severe legal and financial repercussions, particularly as AI systems become more widespread in society [106, 107].

#### ***4.2.2 Privacy breaches and data protection challenges***

It has been noted that AI systems rely heavily on data to make decisions [108]. However, this data often contains sensitive personal information vulnerable to unauthorized access and misuse. Such actions can result in severe consequences like privacy violations, identity theft, and financial fraud [109]. For instance, hackers often target healthcare records to steal personal information. A data breach at a credit reporting organization might also reveal millions of people's sensitive financial information. These occurrences highlight the importance of strong security measures to protect personal data from potential attackers [110].

#### ***4.2.3 National security implications***

**1. Artificial Intelligence in National Security:** Allen and Chan [87] discusses the implications of AI in the context of national security. It explores the potential benefits and risks associated with AI technologies, including the use of autonomous weapons, surveillance capabilities, and cyber threats. It also examines the challenges of adapting national security strategies to leverage AI capabilities while addressing ethical, legal, and policy concerns.

**2. Cybersecurity and Global Threats:** Carlin and Graff [111] focuses on the global cyber threat landscape and the ongoing battle between nations, particularly the United States, Russia, and China. It delves into what the

authors term “the code war” and discusses the increasing importance of cybersecurity in the context of national security. It also emphasizes the need for robust defense mechanisms to protect critical infrastructure, private data, and national interests.

**3. Autonomous Weapons and Future Warfare:** Scharre [112] explores the development and implications of autonomous weapons systems. The book discusses the ethical and legal considerations surrounding the use of AI-enabled weapons and the potential for autonomous decision-making in warfare. It also examines the challenges of ensuring human control, accountability, and adherence to international humanitarian law in the context of autonomous weapons.

#### ***4.3 Previous studies***

This section assesses the landmark, seminal, and pivotal literature sources under the umbrella of what the author terms “social construction of artificial intelligence”, an embryonic field of AI research so obscure that an initial google search of the term in quotes returned just “4 results”, only to reduce to “2 results” the second time around. It is worth noting that though this is a comprehensive list, inclusive of key publications that have significantly contributed to the understanding of the research topic, it is by no means an exhaustive list. Through the combined application of content analysis, thematic analysis, discourse analysis, and grounded theory, the following key themes and coding categories have emerged from the selected literature sources:

#### **The Second Machine Age and Progress**

Brynjolfsson and McAfee [77] discusses the concept of the second machine age and its impact on work, progress, and prosperity. They analyze the potential benefits and challenges of AI technologies in transforming industries and the economy.

#### **The Singularity and Superintelligence**

Several sources, including [113, 6, 56, 57], discuss the concept of the technological singularity, where AI surpasses human intelligence. They explore the potential implications associated with the development of superintelligent AI systems.

#### **Socioeconomic Implications**

[77, 114, 115] examine the economic and social consequences of AI. They explore the potential for job displacement, the role of AI in reshaping industries, and the geopolitical implications of AI superpowers. [116, 117, 79, 90, 118] discuss the socioeconomic implications of AI, automation, and algorithmic decision-making. They explore the potential for job displacement, inequality, and the concentration of power, emphasizing the importance of addressing these challenges to ensure equitable outcomes. Zuboff [119] focuses on the emergence of surveillance capitalism and its impact on society, democracy, and power dynamics. The book examines how data collection and analysis by tech companies have transformed into a new form of economic and social control. It also highlights the need to address the ethical and social implications of surveillance capitalism and protect individual privacy and autonomy.

### **AI Applications in Governance and Economics**

[120, 121] explore the use of AI in governance and economic systems. They discuss the potential for AI-driven tax policies and the concept of "hypernudge" as a mode of regulation. These sources examine how AI can be leveraged to improve equality, productivity, and decision-making in governance and economics.

### **Explainability and Interpretability of AI**

[103] focuses on the opacity and interpretability challenges of machine learning algorithms. The source highlights the importance of understanding how AI systems make decisions and the potential biases that can emerge because of that. [122] delves into the challenge of explaining AI decisions and ensuring interpretability. They discuss the importance of developing methods and techniques to provide meaningful explanations for AI outputs, particularly in high-stakes domains. These sources emphasize the need for transparency and understandability in AI systems.

### **Algorithmic Bias and Discrimination**

Several sources, including [100, 101, 92, 94] highlight the issue of algorithmic bias and discrimination in machine learning systems. They discuss cases where algorithms have exhibited biased outcomes, particularly against certain racial, gender, or socioeconomic groups, and emphasize the need for fairness and accountability in algorithmic decision-making. Other literature, including [91, 97, 102, 123, 124], focus on the impact of algorithmic decision-making in the law enforcement and criminal justice system. These sources examine the potential biases in recidivism prediction, pretrial risk assessment, and law enforcement practices, highlighting the need for transparency and fairness.

### **Methodological Approaches**

[33, 99, 123, 125] focus on the methodological approaches for studying algorithmic bias and discrimination. They discuss techniques for detecting and auditing discrimination in algorithms, as well as the challenges and limitations of current methodologies.

### **Big Data Applications and Challenges**

[126, 127] discuss the applications and challenges associated with big data. They highlight the potential of big data in various domains and the need for advanced techniques and technologies to handle the volume, velocity, and variety of data. These sources provide an overview of the concepts, methods, and analytics used in big data research. [128] focuses on the potential privacy harms associated with predictive analytics and big data. They emphasize the need for a framework to address the ethical, legal, and social implications of using big data for decision-making. These sources highlight the importance of due process and the protection of individuals' rights in the context of data-driven decision-making.

### **Privacy and Data Protection**

[98, 104, 129, 132] examine the privacy and data protection implications associated with algorithmic decision-making. These sources discuss the need for transparency, informed consent, and the responsible use of personal data to mitigate the risks of privacy violations and surveillance.

### **Mitigation and Prevention of Malicious Use**

[133] discusses the malicious use of AI and the need for forecasting, prevention, and mitigation strategies. They explore the potential risks and challenges associated with AI in the context of cybersecurity and national security.

### **AI Policy and Regulation**

[106,107, 134] discuss the need for policy and regulation to address the challenges and risks associated with artificial intelligence (AI). These sources explore the regulatory strategies, competencies, and methods required to govern AI systems effectively. They emphasize the importance of considering ethical, legal, and social implications in shaping AI policies.

### **Ethical and Moral Considerations**

[32, 35, 117, 135, 136] explore the ethical dimensions and fairness considerations surrounding algorithmic decision-making. They discuss the limitations and opportunities for promoting fairness and accountability in the design, implementation, and evaluation of algorithms. [137, 138, 121] also focus on the ethical dimensions of AI. They propose frameworks and principles for a responsible and beneficial AI society, addressing issues such as transparency, accountability, fairness, and privacy. These sources highlight the need to develop guidelines and standards that prioritize human values and mitigate potential harm. [139, 140] on their part delve into the ethical and moral dimensions of artificial agents and the impact of AI on human reality. These sources address the need for ethical frameworks and guidelines to ensure the responsible design and use of AI technologies. [141] focuses on ethical considerations in healthcare by examining, case-study style, the role of Google DeepMind and the use of algorithms in healthcare. It discusses the ethical considerations surrounding data privacy, consent, and transparency in the context of healthcare algorithms. It also highlights the importance of ensuring that the use of big data and algorithms in healthcare aligns with ethical principles and safeguards patient privacy.

### **Philosophical Foundations of AI**

[4] explores the philosophical underpinnings of AI and the quest for creating intelligent machines. The source provides a historical overview of the field and discusses the fundamental challenges and advancements in AI research.

### **Research Priorities for Beneficial AI**

[45] outlines research priorities to ensure the robustness and beneficial use of AI. They highlight the importance

of safety, value alignment, and long-term societal impact in AI development. These sources emphasize interdisciplinary collaboration and responsible AI research agendas. These themes can further inform research and discussions surrounding the subject matter of artificial intelligence.

## 5. Discussion

### 5.1. Comparative Analysis

**Table 1:** Here’s a comparative analysis chart highlighting the advantages of artificial intelligence (AI) and alongside its implicated risks:

Aspect	Advantages	Risks
<b>Efficiency and Automation</b>	<ul style="list-style-type: none"> <li>- Automates repetitive tasks, reducing human error.</li> </ul>	<ul style="list-style-type: none"> <li>- Job displacement due to automation.</li> </ul>
	<ul style="list-style-type: none"> <li>- Streamlines processes, saving time and resources</li> </ul>	<ul style="list-style-type: none"> <li>- Dependence on AI systems can lead to vulnerabilities.</li> </ul>
<b>Data Analysis</b>	<ul style="list-style-type: none"> <li>- Enables informed decisions through data analysis.</li> <li>- Assists in medical diagnosis and treatment planning.</li> </ul>	<ul style="list-style-type: none"> <li>- Biased algorithms can perpetuate existing inequalities.</li> <li>- Privacy concerns due to extensive data collection.</li> </ul>
<b>Autonomous Vehicles</b>	<ul style="list-style-type: none"> <li>- Advances self-driving cars and transportation systems.</li> </ul>	<ul style="list-style-type: none"> <li>- Safety risks if AI systems fail or are hacked.</li> </ul>
<b>National Security</b>	<ul style="list-style-type: none"> <li>- Enhances defense strategies through predictive analytics.</li> <li>- Improves surveillance and threat detection capabilities.</li> </ul>	<ul style="list-style-type: none"> <li>- Potential misuse by malicious actors (e.g., cyberterrorism).</li> <li>- Lack of transparency in AI decision-making processes.</li> </ul>
<b>Human Wellbeing</b>	<ul style="list-style-type: none"> <li>- Supports healthcare with personalized treatment plans.</li> <li>- Assists people with disabilities (e.g., speech recognition for communication).</li> </ul>	<ul style="list-style-type: none"> <li>- Ethical dilemmas (e.g., AI deciding life-and-death situations).</li> <li>- Emotional disconnect (AI lacks empathy and creativity).</li> </ul>
<b>Data Privacy</b>	<ul style="list-style-type: none"> <li>- Facilitates secure data storage and access control.</li> </ul>	<ul style="list-style-type: none"> <li>- Risk of data breaches or unauthorized access (e.g., Cambridge Analytica scandal).</li> </ul>
	<ul style="list-style-type: none"> <li>- Streamlines processes, saving time and resources</li> </ul>	<ul style="list-style-type: none"> <li>- Dependence on AI systems can lead to vulnerabilities.</li> </ul>



### ***5.1.1. Balancing Benefits and Risks in AI Adoption***

When adopting AI, it is important to balance its advantages and potential risks. To fully reap the benefits of AI while avoiding any negative impacts, it is important to concentrate on implementing it in areas where it can be most useful. This requires collaboration between policymakers, industry experts, and the general public to guarantee that AI is developed and utilized responsibly and ethically. A human-centric approach to creating and implementing AI is crucial. This approach prioritizes people and ensures that AI systems align with human values and ethics. Additionally, ethical frameworks for AI can guide the responsible development and deployment of AI.

### ***5.1.2. Strategies for Addressing Ethical, Privacy, and Security Challenges***

There are several ways to tackle ethical, privacy, and security issues related to AI. First, Explainable AI (also known as Interpretable AI or Explainable Machine Learning) can enhance transparency and accountability in AI systems, which can help identify and address ethical concerns more efficiently. Explainable AI means that AI systems can explain their decision-making process in a way humans can comprehend. Second, ethical frameworks for AI, such as the Asilomar Principles [60], can be developed to guide the responsible creation and implementation of AI. These frameworks can ensure that AI aligns with human values and ethics and is developed and deployed responsibly and ethically. In AI systems, ethical frameworks are important in addressing algorithmic bias and ensuring fairness and impartiality. Furthermore, privacy-enhancing technologies such as differential privacy can be employed to protect sensitive personal data. Differential privacy is a technique that adds noise to data to anonymize it and protect individuals' privacy while still enabling data analysis. Other privacy-enhancing technologies include homomorphic encryption and federated learning. Finally, secure coding practices and robust cybersecurity measures can help protect AI systems from cyberattacks and other security threats. Ensuring that AI systems are developed with security in mind is important. Regular audits and updates should be conducted to address any potential vulnerabilities. Implementing cybersecurity measures is crucial in preventing data breaches and safeguarding confidential data.

### ***5.1.3. Mitigating Risks of Job Displacement***

Many people are worried about AI technology's impact on job opportunities. With the ability to automate tasks once performed exclusively by humans, some industries may experience significant job losses. Fortunately, some strategies can be implemented to help reduce the risks of job displacement. For example, reskilling and upskilling programs can help workers whom AI displaces to develop new in-demand skills suitable for the emerging digital economy. Policies such as universal basic income can also provide a safety net for workers whom AI displaces.

### ***5.1.4. Addressing Bias in AI***

Algorithmic bias is a significant concern in AI development and deployment. Developing and training AI systems on diverse and representative data is crucial to avoid unfair outcomes for certain groups of people, perpetuating existing biases and discrimination in society and undermining the credibility of AI systems.

Regular auditing of AI systems is also necessary to identify and address potential biases.

### ***5.1.5. Ensuring Transparency and Accountability***

Ensuring responsible and ethical development and deployment of AI systems requires transparency and accountability. It is crucial that the decision-making process of AI systems is transparent and can be audited, explained, and understood by humans. This helps foster trust in AI systems and ensures their fair and ethical use. In addition, accountability measures should be in place to hold both AI developers and users responsible for any adverse consequences resulting from the use of AI.

## ***5.2. Frameworks, Policies, and Recommendations***

### ***5.2.1 Ethical Frameworks for AI Evaluation***

Ethical frameworks offer a collection of principles and values that can guide the creation and utilization of AI. By following these frameworks, AI can be developed and utilized ethically and in line with human values. Exemplar: the IEEE Global Initiative for Ethical Considerations in AI and Autonomous Systems.

### ***5.2.2 Privacy and Data Protection Frameworks***

Privacy and data protection frameworks aim to safeguard personal information when AI is utilized. These frameworks ensure that AI is developed and used in a way that respects people's privacy rights. One notable example is the General Data Protection Regulation of the European Union.

### ***5.2.3 National security risk assessment policies***

There are many ways to evaluate the potential risks of using AI in national security, particularly in military and intelligence contexts. The US Department of Defense has created ethical guidelines to guarantee that AI is used appropriately in military environments.

### ***5.2.4 Recommendations for Stakeholders***

Every stakeholder needs to play a part in ensuring that AI is developed and used responsibly and ethically.

## **1. Government:**

Governments must be at the forefront of creating policies and regulations that govern the development and use of artificial intelligence. Governments must prioritize developing ethical guidelines, responsible AI frameworks, and data protection and privacy laws. It would be beneficial for governments to offer financial research support to explore AI's potential benefits and risks while ensuring that it is developed and utilized in a manner that aligns with human values. In addition, establishing regulatory bodies to oversee the development and implementation of AI systems would help ensure adherence to ethical and legal standards.

## **2. Industry:**

Industry leaders must prioritize responsible and ethical development and the use of AI. This can be done by creating best practices and self-regulatory mechanisms for AI development and implementation. In addition, the industry should prioritize the development of explainable, transparent, accountable AI. The industry needs to recognize the potential negative impact of AI on employment and work towards equitable distribution of its advantages.

## **3. Individuals:**

Individuals need to be informed and empowered to make decisions regarding the use of AI. This means they should know digital literacy, privacy awareness, and the advantages and disadvantages of AI. Individuals need access to the necessary tools and information to understand how AI is used. They should also have control over their personal information. Additionally, people should be included in discussions regarding the development and implementation of AI, and their perspectives should be considered when making decisions.

## **6. Conclusion**

### ***6.1 Summary of Key Findings and Contributions***

The quick progress of Artificial Intelligence (AI) is causing major transformations in economies, societies, and political systems around the globe. Nonetheless, the creation and implementation of AI also pose substantial hazards and obstacles. This paper examines the opportunities and risks associated with AI, compares their benefits and drawbacks, identifies potential solutions to mitigate risks, discusses frameworks for risk assessment, and explores policies and regulations for development and use. Lastly, recommendations are provided for different stakeholders, such as governments, industries, and individuals.

Our key findings suggest that AI has proven beneficial to humankind – facilitating the smooth advancement through the ranks of the Maslow Hierarchy of Needs. The impact of AI technology on the world has been mostly positive, with significant contributions toward achieving the UN Sustainable Development Goals. However, policymakers, industry leaders, researchers, and the general public must work together to ensure the responsible and ethical development, deployment, and usage of AI.

### ***6.2 Gaps in Current Literature***

The research paper has identified gaps in the current literature on the effects of AI on society. It highlights the need for further research on how AI could impact geopolitics and governance. Moreover, there is a need for more research on how AI could affect marginalized communities, especially those residing in low-income areas. Another area for improvement is the narrow technical focus of most research overlooking social and ethical considerations. Another gap is the need to understand better the long-term economic, social, and political impacts of widespread AI usage. A final area for improvement in the existing literature (given the time constraint) was the limited research in the emerging field of Critical AI Studies. The findings of this study can

inform future research and policy decisions.

### **6.3 Suggestions for Future Research**

Future research should focus on the following areas:

Developing ethical frameworks for AI aligned with human values and ethics. It is important to consider various factors to ensure the responsible and ethical development and deployment of AI systems. One such factor is the impact of AI on employment and the labor market. Strategies should be developed to address adverse impacts and ensure equitable sharing of AI benefits. It is important to assess the long-term effects of AI on society, including its impact on social structures, culture, and politics, to anticipate and mitigate unintended consequences. To improve our ability to do so, new methods for assessing AI risks and benefits, such as predictive analytics and machine learning, should be developed. Investigating the use of AI in emerging technologies like autonomous vehicles, robotics, and augmented reality will help identify potential risks and benefits and develop strategies to address them.

Additionally, addressing challenges in AI governance and regulation, such as international cooperation and coordination, is crucial to ensure responsible and ethical development and deployment of AI. Additionally, AI systems should be developed to be transparent, explainable, and interpretable to build trust and accountability and ensure they are auditable and accountable.

In AI research, creating systems that can withstand adversarial attacks is important. Such attacks are deliberate attempts to manipulate data inputs to deceive an AI system and cause it to make incorrect decisions. For instance, an image can be manipulated to appear as something else to an AI system, leading to misclassification. This is particularly dangerous in applications like medical diagnosis and autonomous vehicles. Researchers are exploring ways to develop robust AI systems, including adversarial training and model-based verification and validation. However, more research is needed to address the issue of adversarial attacks effectively.

Another important area of focus in AI research is the development of explainable and interpretable AI systems. Understanding how decisions are made and why is crucial, especially in fields like healthcare, where AI systems can make decisions with life-or-death consequences. Various methods for creating explainable and interpretable AI systems are being studied by researchers, such as model-based explanations and post hoc analysis. However, further research is necessary to interpret and explain the decisions made by AI systems effectively.

To sum up, there is a need for further research to investigate the ethical, legal, and social implications of artificial intelligence. For example, should AI systems have legal personhood? How should we ensure proper regulation when it comes to AI in warfare? Additionally, what impacts will AI have on privacy and data protection? To answer these questions, we need interdisciplinary research considering the perspectives of ethics, law, social sciences, and humanities.

## References

- [1] J. F. Allen, "AI Growing Up: The Changes and Opportunities," *Ai Magazine*, vol. 19, no. 4, pp. 13–23, Dec. 1998, doi: 10.1609/aimag.v19i4.1422.
- [2] S. Bhatnagar *et al.*, "Mapping Intelligence: requirements and possibilities," in *Studies in applied philosophy, epistemology and rational ethics*, 2018, pp. 117–135. doi: 10.1007/978-3-319-96448-5\_13.
- [3] R. J. Brachman, "AA)AI More than the Sum of Its Parts," *Ai Magazine*, vol. 27, no. 4, pp. 19–34, Dec. 2006, doi: 10.1609/aimag.v27i4.1907.
- [4] N. J. Nilsson, *The quest for artificial intelligence*. 2009. doi: 10.1017/cbo9780511819346.
- [5] S. J. Russell, P. Norvig, and E. Davis, *Artificial intelligence: A Modern Approach*. Prentice Hall, 2010.
- [6] R. Kurzweil, *The age of intelligent machines*. Cambridge, Mass. : MIT Press, 1992.
- [7] E. Rich and K. Knight, *Artificial intelligence*. 1991.
- [8] J. Haugeland, *Artificial intelligence: The Very Idea*. MIT Press, 1989.
- [9] R. Bellman, *An Introduction to Artificial intelligence: Can Computers think?* 1978. [Online]. Available: <https://ci.nii.ac.jp/ncid/BA03493655>
- [10] E. Charniak, D. McDermott, and D. V. McDermott, *Introduction to artificial Intelligence*. Addison Wesley Publishing Company, 1985.
- [11] P. H. Winston, *Artificial intelligence*. Addison-Wesley, 1992.
- [12] N. J. Nilsson, *Artificial intelligence: A New Synthesis*. Morgan Kaufmann, 1998.
- [13] D. I. Poole *et al.*, *Computational intelligence: A Logical Approach*. Oxford University Press on Demand, 1998.
- [14] G. Simons and D. S. Baldwin, "A critical review of the definition of 'wellbeing' for doctors and their patients in a post Covid-19 era," *International Journal of Social Psychiatry*, vol. 67, no. 8, pp. 984–991, Jul. 2021, doi: 10.1177/00207640211032259.
- [15] M. Buckbee, "Data Privacy Guide: Definitions, Explanations and Legislation," *Varonis blog.*, Jun. 02, 2023. <https://www.varonis.com/blog/data-privacy> (accessed Oct. 20, 2023).
- [16] S. Osisanya, "National Security versus Global Security | United Nations," *United Nations*. <https://www.un.org/en/chronicle/article/national-security-versus-global-security> (accessed Oct. 20,

2023).

- [17] O. Shani, "From science fiction to reality: the evolution of Artificial intelligence," *WIRED*, Aug. 07, 2015. Accessed: Oct. 20, 2023. [Online]. Available: <https://www.wired.com/insights/2015/01/the-evolution-of-artificial-intelligence/>
- [18] S. Bianchini, M. Müller, and P. Pelletier, "Artificial intelligence in science: An emerging general method of invention," *Research Policy*, vol. 51, no. 10, p. 104604, Dec. 2022, doi: 10.1016/j.respol.2022.104604.
- [19] A. Clyde, "AI for science and global citizens," *Patterns*, vol. 3, no. 2, p. 100446, Feb. 2022, doi: 10.1016/j.patter.2022.100446.
- [20] M. Krenn *et al.*, "On scientific understanding with artificial intelligence," *Nature Reviews Physics*, vol. 4, no. 12, pp. 761–769, Oct. 2022, doi: 10.1038/s42254-022-00518-3.
- [21] Y. Xu *et al.*, "Artificial intelligence: A powerful paradigm for scientific research," *The Innovation*, vol. 2, no. 4, p. 100179, Nov. 2021, doi: 10.1016/j.xinn.2021.100179.
- [22] C. Chan, J. Moore, C. Chan, and J. Moore, "Generative AI: the next consumer platform," *Andreessen Horowitz*, Sep. 2023, [Online]. Available: <https://a16z.com/2023/02/07/everyday-ai-consumer/>
- [23] W. Lehmacher, "5 reasons consumers will embrace artificial intelligence," *World Economic Forum*, Jan. 04, 2018. <https://www.weforum.org/agenda/2018/01/consumers-will-embrace-artificial-intelligence/> (accessed Oct. 20, 2023).
- [24] S. Puntoni, R. W. Reczek, M. Giesler, and S. Botti, "Consumers and Artificial intelligence: An Experiential perspective," *Journal of Marketing*, vol. 85, no. 1, pp. 131–151, Oct. 2020, doi: 10.1177/0022242920953847.
- [25] T. H. Davenport, "3 things AI can already do for your company," *Harvard Business Review*, Sep. 15, 2023. <https://hbr.org/2018/01/artificial-intelligence-for-the-real-world>
- [26] Forbes Technology Council, "15 Tech Experts Share Potential Impacts Of AI On Society," *Forbes*, Jun. 26, 2020. Accessed: Oct. 20, 2023. [Online]. Available: <https://www.forbes.com/sites/forbestechcouncil/2020/06/26/15-tech-experts-share-potential-impacts-of-ai-on-society/?sh=49b02bce3714>
- [27] T. Weitzman, "The top five ways AI is transforming business," *Forbes*, Nov. 21, 2022. [Online]. Available: <https://www.forbes.com/sites/forbesbusinesscouncil/2022/11/21/the-top-five-ways-ai-is-transforming-business/?sh=1990ebd18e7f>
- [28] Y. Hamdar, K. N. Massally, and Peiris, "Are countries ready for AI? How they can ensure ethical and

- responsible adoption | United Nations Development Programme,” *UNDP*, Apr. 25, 2023. <https://www.undp.org/blog/are-countries-ready-ai-how-they-can-ensure-ethical-and-responsible-adoption> (accessed Oct. 20, 2023).
- [29] E. Brynjolfsson, “The business of artificial intelligence,” *Harvard Business Review*, Nov. 08, 2022. <https://hbr.org/2017/07/the-business-of-artificial-intelligence>
- [30] R. L. Stevenson, *Strange case of Dr. Jekyll and Mr. Hyde*. 1886.
- [31] E. Ntoutsis *et al.*, “Bias in data-driven artificial intelligence systems—An introductory survey,” *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*, vol. 10, no. 3, Feb. 2020, doi: 10.1002/widm.1356.
- [32] K. Lum and W. Isaac, “To predict and serve?,” *Significance*, vol. 13, no. 5, pp. 14–19, Oct. 2016, doi: 10.1111/j.1740-9713.2016.00960.x.
- [33] A. Köchling and M. C. Wehner, “Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development,” *Business Research*, vol. 13, no. 3, pp. 795–848, Nov. 2020, doi: 10.1007/s40685-020-00134-w.
- [34] A. Klein, “Reducing bias in AI-based financial services | Brookings,” *Brookings*, Jul. 10, 2020. <https://www.brookings.edu/articles/reducing-bias-in-ai-based-financial-services/> (accessed Oct. 20, 2023).
- [35] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and machine learning: Limitations and Opportunities*. MIT Press, 2023.
- [36] “CODED BIAS — 7th Empire Media,” *7th Empire Media*. <https://www.7thempiremedia.com/films-codedbias>
- [37] C. Kerry, “Protecting privacy in an AI-driven world | Brookings,” *Brookings*, Feb. 10, 2020. <https://www.brookings.edu/articles/protecting-privacy-in-an-ai-driven-world/> (accessed Oct. 20, 2023).
- [38] H. Davidson, “China’s coronavirus health code apps raise concerns over privacy,” *The Guardian*, Jul. 01, 2020. [Online]. Available: <https://www.theguardian.com/world/2020/apr/01/chinas-coronavirus-health-code-apps-raise-concerns-over-privacy>
- [39] J. W. Burton and S. R. Soare, “Understanding the Strategic Implications of the Weaponization of Artificial Intelligence,” *IEEE Xplore*, Jul. 2019, doi: 10.23919/cycon.2019.8756866.
- [40] B. Dresch-Langley, “The weaponization of artificial intelligence: What the public needs to be aware of,” *Frontiers in Artificial Intelligence*, vol. 6, Mar. 2023, doi: 10.3389/frai.2023.1154184.

- [41] D. Harwell, "Faked Pelosi videos, slowed to make her appear drunk, spread across social media," *Washington Post*, May 24, 2019. [Online]. Available: <https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media/>
- [42] M. Sadiq, "Real v fake: debunking the 'drunk' Nancy Pelosi footage - video," *The Guardian*, Feb. 26, 2020. <https://www.theguardian.com/us-news/video/2019/may/24/real-v-fake-debunking-the-drunk-nancy-pelosi-footage-video>
- [43] G. Shao, "What 'deepfakes' are and how they may be dangerous," *CNBC*, Jan. 17, 2020. [Online]. Available: <https://www.cnn.com/2019/10/14/what-is-deepfake-and-how-it-might-be-dangerous.html>
- [44] B. Cheatham, K. Javanmardian, and H. Samandari, "Confronting the risks of artificial intelligence," *McKinsey & Company*, Apr. 2019, [Online]. Available: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/confronting-the-risks-of-artificial-intelligence>
- [45] S. Russell, D. Dewey, and M. Tegmark, "Research priorities for robust and beneficial artificial intelligence," *Ai Magazine*, vol. 36, no. 4, pp. 105–114, Dec. 2015, doi: 10.1609/aimag.v36i4.2577.
- [46] Mila, "Missing links in AI governance," *UNESCO*, Aug. 2023, [Online]. Available: <https://www.unesco.org/en/articles/missing-links-ai-governance>
- [47] Future of Life Institute, "Asilomar AI Principles - Future of Life Institute," *Future of Life Institute*, Oct. 19, 2020. <https://futureoflife.org/person/asilomar-ai-principles/>
- [48] M. Cheng, "What AI mavens can learn from a 1975 genetics conference," *Quartz*, May 03, 2023. [Online]. Available: <https://qz.com/what-ai-mavens-can-learn-from-a-1975-genetics-conferenc-1850305170>
- [49] F. Heylighen, "A cognitive-systemic reconstruction of maslow's theory of self-actualization," *Systems Research and Behavioral Science*, vol. 37, no. 1, pp. 39–58, Jan. 1992, doi: 10.1002/bs.3830370105.
- [50] A. H. Maslow, *A theory of human motivation*. Simon and Schuster, 2013.
- [51] A. Turing, "I.—COMPUTING MACHINERY AND INTELLIGENCE," *Mind*, vol. LIX, no. 236, pp. 433–460, Oct. 1950, doi: 10.1093/mind/lix.236.433.
- [52] B. J. Copeland and D. Proudfoot, "The computer, artificial intelligence, and the Turing Test," in *Springer eBooks*, 2004, pp. 317–351. doi: 10.1007/978-3-662-05642-4\_13.
- [53] S. Harnad, "The Turing Test is not a trick," *SIGART Newsletter*, vol. 3, no. 4, pp. 9–10, Oct. 1992, doi: 10.1145/141420.141422.



- [54] M. E. Koltko-Rivera, "Rediscovering the later version of Maslow's hierarchy of needs: Self-Transcendence and opportunities for theory, research, and unification," *Review of General Psychology*, vol. 10, no. 4, pp. 302–317, Dec. 2006, doi: 10.1037/1089-2680.10.4.302.
- [55] A. H. Maslow, *Motivation and personality*. New York : Harper & Row, 1970.
- [56] R. Kurzweil, *The singularity is near: When Humans Transcend Biology*. Penguin, 2005.
- [57] V. Vinge, "The coming technological singularity," 1993. <https://philpapers.org/rec/VINTCT>
- [58] *Transcendence*, (Apr. 18, 2014).
- [59] J. McCarthy, M. Minsky, N. Rochester, and C. E. Shannon, "A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955," *Ai Magazine*, vol. 27, no. 4, p. 12, Dec. 2006, doi: 10.1609/aimag.v27i4.1904.
- [60] Future of Life Institute, "AI Principles - Future of Life Institute," *Future of Life Institute*, Jun. 08, 2023. <https://futureoflife.org/open-letter/ai-principles/>
- [61] H. Stapf-Fine, U. Bartosch, S. Bauberger, and A. Sülzen, "Policy Paper on the Asilomar Principles on Artificial Intelligence," *ResearchGate*, Dec. 2018, [Online]. Available: [https://www.researchgate.net/publication/329963051\\_Policy\\_Paper\\_on\\_the\\_Asilomar\\_Principles\\_on\\_Artificial\\_Intelligence](https://www.researchgate.net/publication/329963051_Policy_Paper_on_the_Asilomar_Principles_on_Artificial_Intelligence)
- [62] S. Sarangi and P. Sharma, *Artificial intelligence*. 2018. doi: 10.4324/9780429461002.
- [63] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *Journal of Business Research*, vol. 104, pp. 333–339, Nov. 2019, doi: 10.1016/j.jbusres.2019.07.039.
- [64] F. Jiang *et al.*, "Artificial intelligence in healthcare: past, present and future," *Stroke and Vascular Neurology*, vol. 2, no. 4, pp. 230–243, Jun. 2017, doi: 10.1136/svn-2017-000101.
- [65] S. E. Dilsizian and E. L. Siegel, "Artificial Intelligence in medicine and cardiac imaging: Harnessing big data and advanced computing to provide personalized medical diagnosis and treatment," *Current Cardiology Reports*, vol. 16, no. 1, Dec. 2013, doi: 10.1007/s11886-013-0441-8.
- [66] H. Kim, E. Y. Kim, I. Lee, B. Bae, M. Park, and H. Nam, "Artificial Intelligence in Drug Discovery: A Comprehensive review of data-driven and machine learning approaches," *Biotechnology and Bioprocess Engineering*, vol. 25, no. 6, pp. 895–930, Dec. 2020, doi: 10.1007/s12257-020-0049-y.
- [67] J. Vamathevan *et al.*, "Applications of machine learning in drug discovery and development," *Nature Reviews Drug Discovery*, vol. 18, no. 6, pp. 463–477, Apr. 2019, doi: 10.1038/s41573-019-0024-5.

- [68] D. Nahavandi, R. Alizadehsani, A. Khosravi, and U. R. Acharya, "Application of artificial intelligence in wearable devices: Opportunities and challenges," *Computer Methods and Programs in Biomedicine*, vol. 213, p. 106541, Jan. 2022, doi: 10.1016/j.cmpb.2021.106541.
- [69] E. B. Boukherouaa *et al.*, *Powering the Digital Economy: Opportunities and risks of Artificial intelligence in finance*. International Monetary Fund, 2021.
- [70] T. Miziołek, "Employing artificial intelligence in investment management," in *Routledge eBooks*, 2021, pp. 161–174. doi: 10.4324/9781003095354-9.
- [71] S. Kiderlin, "Banking in the metaverse? A.I. could be about to change the way you manage your money," *CNBC*, May 02, 2023. [Online]. Available: <https://www.cnbc.com/2023/05/02/banking-in-the-metaverse-ai-could-change-money-management.html>
- [72] L. Wewege and M. C. Thomsett, *The digital banking revolution*. 2019. doi: 10.1515/9781547401598.
- [73] D. J. Fagnant and K. M. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," *Transportation Research Part A: Policy and Practice*, vol. 77, pp. 167–181, Jul. 2015, doi: 10.1016/j.tra.2015.04.003.
- [74] B. Schoettle, "A survey of public opinion about autonomous and self-driving vehicles in the U.S., the U.K., and Australia," Jul. 01, 2014. <https://deepblue.lib.umich.edu/handle/2027.42/108384>
- [75] R. L. Abduljabbar, H. Dia, S. Liyanage, and S. A. Bagloee, "Applications of Artificial Intelligence in Transport: An Overview," *Sustainability*, vol. 11, no. 1, p. 189, Jan. 2019, doi: 10.3390/su11010189.
- [76] V. D. V. T. & J. Nick, "Artificial Intelligence in Education: Can AI bring the full potential of personalized learning to education?," *ideas.repec.org*, 2019, [Online]. Available: <https://ideas.repec.org/p/zbw/itse19/205222.html>
- [77] E. Brynjolfsson and A. McAfee, *The second machine age: work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company, 2014.
- [78] J. Bughin *et al.*, "Skill shift: Automation and the future of the workforce," *McKinsey & Company*, May 23, 2018. <https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce> (accessed Oct. 20, 2023).
- [79] M. R. Frank *et al.*, "Toward understanding the impact of artificial intelligence on labor," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 14, pp. 6531–6539, Mar. 2019, doi: 10.1073/pnas.1900949116.
- [80] J. Manyika, K. Sneider, and McKinsey Global Institute, "AI, automation, and the future of work: Ten things to solve for," *McKinsey & Company*, Jun. 01, 2018. <https://www.mckinsey.com/featured->

- insights/future-of-work/ai-automation-and-the-future-of-work-ten-things-to-solve-for (accessed Oct. 20, 2023).
- [81] S. Ransbotham, D. Kiron, P. Gerbert, and M. Reeves, “Reshaping business with artificial intelligence: Closing the gap between ambition and action - ProQuest,” *MIT Sloan Management Review*, no. Vol. 59, Iss. 1, 2017.
- [82] K. Aliche *et al.*, “Succeeding in the AI supply-chain revolution,” *McKinsey & Company*, Apr. 30, 2021. <https://www.mckinsey.com/industries/metals-and-mining/our-insights/succeeding-in-the-ai-supply-chain-revolution> (accessed Oct. 20, 2023).
- [83] K. Butner and IBM Institute for Business Value, “AI is reshaping the supply chain,” *IBM*, Jun. 01, 2017. <https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/cognitivesupplychain> (accessed Oct. 20, 2023).
- [84] S. Mullainathan and J. Spiess, “Machine Learning: an Applied Econometric approach,” *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 87–106, May 2017, doi: 10.1257/jep.31.2.87.
- [85] European Parliament, “Artificial intelligence: threats and opportunities | News | European Parliament,” Jun. 20, 2023. <https://www.europarl.europa.eu/news/en/headlines/society/20200918STO87404/artificial-intelligence-threats-and-opportunities> (accessed Oct. 20, 2023).
- [86] C. Bibby, J. Gordon, G. Schuler, and E. Stein, “The big reset: Data-driven marketing in the next normal,” *McKinsey & Company*, Mar. 2021, [Online]. Available: <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-big-reset-data-driven-marketing-in-the-next-normal>
- [87] G. Allen and T. Chan, “Artificial Intelligence and National Security | Belfer Center for Science and International Affairs,” *Belfer Center for Science and International Affairs*, Jul. 2017. <https://www.belfercenter.org/publication/artificial-intelligence-and-national-security> (accessed Oct. 20, 2023).
- [88] V. M. Hudson, “Introduction,” in *Artificial Intelligence And International Politics*, 1st ed., Routledge, pp. 1–8. doi: 10.4324/9780429033575-1.
- [89] A. Echterhölter, J. Schröter, and A. Sudmann, “How is Artificial Intelligence Changing Science? Research in the Era of Learning Algorithms.,” *MediArXiv*, pp. 1–24, Jun. 2021, doi: 10.33767/osf.io/28pnx.
- [90] C. B. Frey and M. A. Osborne, “The future of employment: How susceptible are jobs to computerisation?,” *Technological Forecasting and Social Change*, vol. 114, pp. 254–280, Jan. 2017,

doi: 10.1016/j.techfore.2016.08.019.

- [91] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, “Dissecting racial bias in an algorithm used to manage the health of populations,” *Science*, vol. 366, no. 6464, pp. 447–453, Oct. 2019, doi: 10.1126/science.aax2342.
- [92] S. Barocas and A. D. Selbst, “Big data’s disparate impact,” *California Law Review*, vol. 104, no. 3, p. 671, Jan. 2016, doi: 10.15779/z38bg31.
- [93] C. C. Miller, “When Algorithms Discriminate,” *The New York Times*, Jul. 09, 2015. [Online]. Available: <https://www.nytimes.com/2015/07/10/upshot/when-algorithms-discriminate.html>
- [94] J. Buolamwini, “Gender Shades: Intersectional accuracy Disparities in commercial gender classification,” *PMLR*, Jan. 21, 2018. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- [95] B. Cowgill, “The Impact of Algorithms on Judicial Discretion: Evidence from Regression Discontinuities,” 2018. <https://www.semanticscholar.org/paper/The-Impact-of-Algorithms-on-Judicial-Discretion-%3A-Cowgill/cdd4484708af448831eeb3d76a7c2d1e5b0a4ff2>
- [96] A. Lambrecht and C. Tucker, “Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads,” *Social Science Research Network*, Jan. 2016, doi: 10.2139/ssrn.2852260.
- [97] R. P. Bartlett, “Consumer-Lending discrimination in the era of FinTech \*,” 2018. [https://www.semanticscholar.org/paper/Consumer-Lending-Discrimination-in-the-Era-of-\\*-Bartlett/c9c4fd3dbae4a33f7f985a184d9340be3b8db59](https://www.semanticscholar.org/paper/Consumer-Lending-Discrimination-in-the-Era-of-*-Bartlett/c9c4fd3dbae4a33f7f985a184d9340be3b8db59)
- [98] A. Datta, M. C. Tschantz, and A. Datta, “Automated Experiments on AD Privacy Settings: A tale of opacity, choice, and discrimination,” *arXiv (Cornell University)*, Aug. 2014, doi: 10.48550/arxiv.1408.6491.
- [99] L. Sweeney, “Discrimination in online ad delivery,” *Social Science Research Network*, Jan. 2013, doi: 10.2139/ssrn.2208240.
- [100] S. U. Noble, *Algorithms of oppression: How Search Engines Reinforce Racism*. NYU Press, 2018.
- [101] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, “Machine Bias,” *ProPublica*, May 23, 2016. Accessed: Oct. 21, 2023. [Online]. Available: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- [102] J. Dressel and H. Farid, “The accuracy, fairness, and limits of predicting recidivism,” *Science Advances*, vol. 4, no. 1, Jan. 2018, doi: 10.1126/sciadv.aao5580.

- [103] J. Burrell, “How the machine ‘thinks’: Understanding opacity in machine learning algorithms,” *Big Data & Society*, vol. 3, no. 1, p. 205395171562251, Jan. 2016, doi: 10.1177/2053951715622512.
- [104] F. A. Pasquale, *The Black Box Society*. 2015. doi: 10.4159/harvard.9780674736061.
- [105] C. S. Nam, J.-Y. Jung, and S. Lee, *Human-Centered Artificial Intelligence: Research and Applications*. Elsevier, 2022.
- [106] R. Calo, “Artificial Intelligence Policy: a roadmap,” *Social Science Research Network*, Jan. 2017, doi: 10.2139/ssrn.3015350.
- [107] M. Scherer, “Regulating Artificial intelligence Systems: Risks, challenges, competencies, and strategies,” *Social Science Research Network*, Jan. 2015, doi: 10.2139/ssrn.2609777.
- [108] N. B. Jones, “Computer science: The learning machines,” *Nature*, vol. 505, no. 7482, pp. 146–148, Jan. 2014, doi: 10.1038/505146a.
- [109] F. Bu, N. Wang, B. Jiang, and H. Liang, “‘Privacy by Design’ implementation: Information system engineers’ perspective,” *International Journal of Information Management*, vol. 53, p. 102124, Aug. 2020, doi: 10.1016/j.ijinfomgt.2020.102124.
- [110] A. Cavoukian, “Staying one step ahead of the GDPR: Embed privacy and security by design,” *ideas.repec.org*, 2018, [Online]. Available: <https://ideas.repec.org/a/aza/csj000/y2018v2i2p173-180.html>
- [111] J. P. Carlin and G. M. Graff, *Dawn of the Code War: America’s Battle Against Russia, China, and the Rising Global Cyber Threat*. PublicAffairs, 2018.
- [112] P. Scharre, *Army of None: Autonomous weapons and the future of war*. W. W. Norton & Company, 2018.
- [113] N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, USA, 2014.
- [114] M. Chui, J. Manyika, and M. Miremadi, “Where machines could replace humans—and where they can’t (yet),” *McKinsey & Company*, Jul. 2016, [Online]. Available: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet>
- [115] K.-F. Lee, *AI superpowers: China, Silicon Valley, and the New World Order*. Houghton Mifflin, 2018.
- [116] K. D. Acemoglu and P. Restrepo, “Automation and New Tasks: How technology displaces and reinstates labor,” *Journal of Economic Perspectives*, vol. 33, no. 2, pp. 3–30, May 2019, doi:

10.1257/jep.33.2.3.

- [117] V. Eubanks, *Automating inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, 2018.
- [118] C. O'Neil, *Weapons of math destruction: How Big Data Increases Inequality and Threatens Democracy*. Penguin UK, 2016.
- [119] S. Zuboff, *The age of surveillance capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books, 2019.
- [120] S. Zheng *et al.*, "The AI Economist: Improving Equality and Productivity with AI-Driven Tax Policies," *arXiv (Cornell University)*, Apr. 2020, [Online]. Available: <https://arxiv.org/pdf/2004.13332.pdf>
- [121] K. Yeung, "'Hypernudge': Big Data as a mode of regulation by design," *Information, Communication & Society*, vol. 20, no. 1, pp. 118–136, May 2016, doi: 10.1080/1369118x.2016.1186713.
- [122] B. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data & Society*, vol. 3, no. 2, p. 205395171667967, Dec. 2016, doi: 10.1177/2053951716679679.
- [123] B. Mittelstadt and L. Floridi, "The Ethics of Big Data: Current and foreseeable issues in biomedical contexts," *Science and Engineering Ethics*, vol. 22, no. 2, pp. 303–341, May 2015, doi: 10.1007/s11948-015-9652-2.
- [124] C. Sandvig, "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms," 2014. <https://www.semanticscholar.org/paper/Auditing-Algorithms-%3A-Research-Methods-for-on-Sandvig-Hamilton/b7227cbd34766655dea10d0437ab10df3a127396>
- [125] B. Cowgill, "Bias and productivity in humans and machines," *Social Science Research Network*, Jan. 2019, doi: 10.2139/ssrn.3584916.
- [126] C. L. P. Chen and C. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," *Information Sciences*, vol. 275, pp. 314–347, Aug. 2014, doi: 10.1016/j.ins.2014.01.015.
- [127] A. H. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, Apr. 2015, doi: 10.1016/j.ijinfomgt.2014.10.007.
- [128] K. Crawford, "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms," Oct. 01, 2013. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2325784](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2325784)

- [129] S. Hajian and J. Domingo-Ferrer, “A Methodology for direct and indirect discrimination Prevention in Data Mining,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 7, pp. 1445–1459, Jul. 2013, doi: 10.1109/tkde.2012.72.
- [130] J. A. Kroll, “Accountable algorithms,” *Penn Carey Law: Legal Scholarship Repository*. [https://scholarship.law.upenn.edu/penn\\_law\\_review/vol165/iss3/3](https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3)
- [131] B. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, “The ethics of algorithms: Mapping the debate,” *Big Data & Society*, vol. 3, no. 2, p. 205395171667967, Dec. 2016, doi: 10.1177/2053951716679679.
- [132] O. Osoba, “Keeping artificial intelligence accountable to humans,” *RAND*, Aug. 2018, [Online]. Available: <https://www.rand.org/blog/2018/08/keeping-artificial-intelligence-accountable-to-humans.html>
- [133] M. Brundage *et al.*, “The Malicious Use of Artificial intelligence: Forecasting, Prevention, and Mitigation,” *arXiv (Cornell University)*, Feb. 2018, doi: 10.17863/cam.22520.
- [134] M. Ziewitz, “Governing algorithms,” *Science, Technology, & Human Values*, vol. 41, no. 1, pp. 3–16, Sep. 2015, doi: 10.1177/0162243915608948.
- [135] A. Chouldechova, “Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments,” *Big Data*, vol. 5, no. 2, pp. 153–163, Jun. 2017, doi: 10.1089/big.2016.0047.
- [136] A. Jobin and M. Ienca, “The global landscape of AI ethics guidelines,” *Nature Machine Intelligence*, vol. 1, no. 9, pp. 389–399, Sep. 2019, doi: 10.1038/s42256-019-0088-2.
- [137] L. Floridi *et al.*, “AI4People—An Ethical Framework for a Good AI Society: Opportunities, risks, principles, and recommendations,” *Minds and Machines*, vol. 28, no. 4, pp. 689–707, Nov. 2018, doi: 10.1007/s11023-018-9482-5.
- [138] M. Tegmark, *Life 3.0: Being human in the age of Artificial Intelligence*. 2017. [Online]. Available: <https://dl.acm.org/citation.cfm?id=3169322>
- [139] L. Floridi, *The Fourth Revolution: How the infosphere is reshaping human reality*. 2014. [Online]. Available: <http://ci.nii.ac.jp/ncid/BB16630552>
- [140] L. Floridi and J. W. Sanders, “On the Morality of Artificial Agents,” *Minds and Machines*, vol. 14, no. 3, pp. 349–379, Aug. 2004, doi: 10.1023/b:mind.0000035461.63578.9d.
- [141] J. Powles and H. Hodson, “Google DeepMind and healthcare in an age of algorithms,” *Health and Technology*, vol. 7, no. 4, pp. 351–367, Mar. 2017, doi: 10.1007/s12553-017-0179-1.