International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 8.141
Peer Reviewed Bi-Monthly
Vol.7 No.22 March –April- 2026

The Dangers of Artificial Intelligence

Shraddha Thorat
Department of Computer Science,
Dr. D. Y. Patil Arts, Commerce and Science College, Akurdi, Pune – 44
Corresponding Author – Shraddha Thorat
DOI - 10.5281/zenodo.19345321

Abstract:

The rapid expansion of Artificial Intelligence (AI) has captured global attention, not only for its potential to revolutionize automation, productivity, and decision-making processes, but also for the profound ethical, societal, economic, and existential threats it poses. This paper delves into the multifaceted risks associated with AI, exploring how its applications and development raise urgent questions in terms of discrimination, unemployment, data security, privacy violations, and even humanity's long-term survival. By examining these concerns, we aim to highlight the critical need for governance frameworks, ethical guidelines, and robust safety protocols in the development and deployment of AI systems. AI systems, particularly those leveraging machine learning models, are increasingly recognized for perpetuating and exacerbating biases present in their training datasets. These biases can manifest in areas like hiring, criminal justice, healthcare, and lending decisions, where AI-based systems have been shown to unfairly disadvantage marginalized groups. Such discriminatory practices widen existing social inequalities, making it necessary to design algorithms with fairness and transparency as foundational principles. From an economic standpoint, AI's capacity for automation poses a significant threat to traditional job markets. The displacement of workers, particularly in industries such as manufacturing, logistics, transportation, and retail, demands more than just upskilling or reskilling initiatives; it calls for structural adjustments, including policies for income redistribution and the creation of new employment sectors. Without these proactive measures, AI could exacerbate global unemployment rates, contributing to social unrest and deepening economic divides between nations and within communities. The cybersecurity landscape has also been dramatically altered by AI, which functions both as a tool for attackers and as a target for malicious activities. AI enables the development of sophisticated cyberattacks, such as automated phishing, deepfakes, and more efficient hacking tools. Simultaneously, AI systems themselves are vulnerable to adversarial attacks, where small manipulations of the input data can result in significant changes in the output, potentially leading to catastrophic failures in critical systems like autonomous vehicles, medical diagnostics, and financial trading platforms. The dual role of AI in both strengthening and weakening cybersecurity underscores the importance of creating AI systems that are robust, secure, and resilient to manipulation. In terms of privacy, AI's ability to process vast amounts of personal data raises significant concerns. The increasing use of AI in surveillance systems, facial recognition technologies, and social media analytics threatens individuals' privacy rights and civil liberties. Such tools, when used without adequate oversight, can lead to widespread surveillance, government control, and erosion of personal freedoms. There is an urgent need to regulate how AI systems collect, process, and store data to protect individuals from privacy violations and ensure that AI technologies serve society in an ethical manner. Furthermore, long-term concerns about AI's role in society have sparked debates about its potential to become an existential threat. Superintelligent AI, often portrayed in popular media, is not just a distant hypothetical concept but a genuine scientific and philosophical debate. If AI systems surpass human intelligence and decision-making capabilities, they could act in ways that are misaligned with human values

and objectives, potentially causing irreversible harm to humanity. Even short-term misalignments, such as deploying autonomous weapons systems, could have devastating consequences if not properly regulated. In conclusion, while AI holds immense potential to drive innovation and societal advancement, it also presents a wide range of dangers that must be addressed comprehensively. This paper emphasizes the importance of ongoing research, global cooperation, and interdisciplinary dialogue in managing the risks associated with AI development. By developing governance frameworks that address issues like bias, unemployment, cybersecurity, privacy, and existential risks, society can harness AI's benefits while minimizing its dangers. The time to act is now, as the trajectory of AI's influence will shape the future of humanity in profound ways.

Introduction:

AI is already deeply embedded in many aspects of our daily lives, yet public awareness of its dangers is still limited. Most discussions tend to focus on the positive aspects of AI, like increased efficiency and convenience, but the risks, which can have a lasting and global impact, are often overlooked. The goal of this research is to provide a balanced perspective on the topic and raise awareness of the dangers AI presents**.** Artificial Intelligence (AI) is rapidly changing the way humans interact with technology, businesses, and each other. From self-driving cars to virtual assistants like Siri and Alexa, AI is integrated into various aspects of daily life. Despite its numerous benefits, such as improving efficiency and automating tasks, AI also brings a host of dangers that could impact societies negatively if not managed properly. This research paper aims to explore the various risks associated with AI and understand why these dangers must be addressed sooner rather than later. Through examining key risks such as job displacement, ethical concerns, privacy issues, and even existential risks, this paper will argue for the responsible development and regulation of AI technologies to safeguard human welfare. The rapid advancement of AI technologies has the potential to bring about revolutionary changes to society, but with those benefits come serious risks that must be understood and managed. AI is both beneficial and dangerous. It is not inherently harmful, but its misuse or lack of control can create significant risks for humans **Why Choose the Topic "Dangers of AI"?** Artificial Intelligence (AI) is rapidly transforming every part of society, from healthcare to education, transportation, and business. While AI presents numerous benefits, it also brings significant risks.

Choosing the topic "Dangers of AI" is important because: **Growing Presence**: AI is integrated into many everyday applications, such as virtual assistants, social media algorithms, and self- driving cars. It's essential to understand its potential dangers as its influence expands.

Ethical Concerns:

The widespread use of AI raises ethical issues like bias, surveillance, and lack of accountability. These concerns affect privacy, job security, and personal freedoms. **Societal Impact:** As AI becomes more autonomous, understanding how it may impact society—positively and negatively—is crucial to preventing harmful consequences. **Technological Control:** Advanced AI systems can eventually operate without much human intervention. If not properly regulated or controlled, they can lead to unintended actions with serious consequences.

Key Points on Dangers of AI:

1. Job Displacement and Economic Disruption:

Overview: The advent of AI and automation poses a significant threat to millions of jobs, especially within industries such as manufacturing, retail, and transportation.

Example: In the automotive sector, the introduction of autonomous vehicles may render traditional driving roles obsolete, affecting taxi drivers, truck drivers, and delivery workers.

Dangers: The swift pace of automation could result in extensive unemployment and exacerbate economic disparities. Individuals displaced from their jobs may find it challenging to secure new employment, potentially leading to social unrest.

Solution: It is imperative for governments and industries to prioritize the reskilling and upskilling of the workforce. Policies should be designed to stimulate job creation in fields that demand human creativity and problem-solving abilities, such as healthcare, education, and the creative arts.

2. Bias and Discrimination:

Overview: AI systems derive their learning from the datasets they are exposed to. If these datasets contain biases related to gender, race, or social status, the AI will adopt and perpetuate these biases. **Example:** Recruitment algorithms powered by AI may show a preference for male candidates due to historical trends indicating a higher number of men in leadership roles.

Dangers: The potential for AI to reinforce systemic discrimination poses a risk of deepening inequalities in areas such as hiring practices, lending, and law enforcement.

Solution: It is essential for developers to focus on creating algorithms that are transparent, accountable, and equitable. Implementing regular audits, utilizing diverse datasets, and adopting inclusive programming methodologies can significantly mitigate bias.

3. Privacy Invasion:

Overview: AI technologies have the capability to analyze and process extensive amounts of personal information, which can lead to violations of privacy. AI-driven surveillance systems can monitor individual activities in real-time, often without obtaining consent.

Example: Facial recognition technology is increasingly employed by governments and corporations to surveil public areas, raising significant concerns regarding mass surveillance.

Dangers: The extensive collection of personal data may result in misuse, cyber threats, or the potential for authoritarian control over individuals.

Solution: It is crucial to establish robust regulations and frameworks that protect individual privacy rights while ensuring responsible data usage.

4. Autonomous Weapons and Warfare:

Overview: Artificial intelligence can facilitate the development of autonomous weaponry, which are systems capable of identifying and engaging targets independently of human operators**.**

Example: AI-enhanced drones may be deployed in military operations, autonomously determining targets for attack with minimal human oversight.

Dangers: The deployment of autonomous weapons poses risks of malfunction, unintended harm, and potential misuse by malicious entities for warfare or terrorist activities. Their integration into combat scenarios raises significant ethical questions regarding accountability and the decision-making process. **Solution:** It is imperative that international agreements and regulations ban the deployment of fully autonomous weapons. Maintaining human oversight in military AI systems is essential to prevent unnecessary conflicts and accidents.

5. Loss of Human Autonomy:

Overview: As artificial intelligence technologies advance, there exists a concern that individuals may become excessively dependent on AI for making decisions.

Example: AI-based personal assistants might take charge of decisions related to finances, healthcare, or daily routines, often with minimal human involvement.

Dangers: Such reliance could result in a

reduction of personal autonomy, where individuals find themselves unable to make independent choices. This dependency may also stifle creativity and critical thinking skills.

Solution: AI should serve as an aid to enhance human decision-making rather than replace it. Mechanisms must be established that empower individuals to override AI-generated decisions, ensuring they maintain control over significant life choices.

6. Security Risks and AI Hacking:

Overview: AI systems, especially those used in critical infrastructures like power grids, healthcare, or financial systems, could be targeted by hackers.

Example: Hackers could exploit vulnerabilities in AI algorithms to manipulate stock markets, shut down power plants, or steal sensitive information.

Dangers: AI systems are not immune to attacks, and when they control essential services, a breach could have catastrophic consequences.

Solution: AI security protocols must evolve alongside AI technologies. Regular security assessments, encryption, and the development of AI that can self-detect and defend against attacks are essential.

7. Unemployment and Inequality:

Overview: The development and implementation of AI technologies often benefit large corporations and wealthy nations, while poorer regions and lower- income individuals may struggle to compete.

Example: Countries with advanced AI industries, like the U.S. and China, could dominate global markets, while developing nations fall further behind economically.

Dangers: This disparity could increase global inequality, leaving large portions of the world's population with fewer job opportunities and lower access to education and technology.

Solution: Global cooperation is required to share the benefits of AI. Investments in education, technology transfer, and collaborative development can help bridge the gap between wealthy and developing nations.

Literature Review:

Artificial Intelligence (AI) is one of the most transformative technologies of the 21st century, revolutionizing industries and reshaping how humans interact with machines. However, alongside its benefits, AI presents significant risks and challenges. These dangers span ethical, social, and technical domains, ranging from immediate threats, such as job displacement, to more speculative existential risks, like loss of human control over superintelligent systems. This literature review examines key areas of concern in AI safety, ethics, and its potential long-term dangers, as discussed in academic and popular literature.

1. Bias and Discrimination:

There is a considerable concern regarding bias within the AI algorithms since it is claimed that for the efficiency of the AI systems, the training data sets which are utilized in order to train the systems are supposed to be clean. When businesses deploy the AI systems in areas such as law, recruitment, and even healthcare, various studies indicate that training datasets which are used contain biased data which in turn yield discriminatory outcomes. For example, when examining the decision making algorithm process using structures, Barocas and selbst (2016) argue that the process practices or decision-making frameworks poses a risk for perpetuating racism, sexism and classism among marginalized groups. In a related context, noble (2018) in her work Algorithms of Oppression argues that social search engines and other AI based system have a tendency to reinforce the existing social bias rather than oppose them. The danger exists in the ability of AIs to increase the already existing inequalities and reinforce the systemic

discrimination further, with insufficient controls or mechanisms for redress ongoing or later.

2. Job Displacement and Economic Impact:

One of the most intriguing questions related to the artificial intelligence and automation processes in the economy is how far AI can displace jobs. Frey and Osman (2013) estimate that 47% of employment positions within the country (United States) will be at risk, due to automation somewhere in the next few decades. The social economic factors associated with such chaotic employment structural change is very high and critical for employed individuals at low level action repetitive type of labor.

3. The Integration of Autonomous:

Weapons in Military Operations:

President Barack Obama put the world on high alert when he characterized artificial intelligence as potentially "the most dangerous invention ever." humanoid robots that are fully automated are indeed an ethical worry within military advancement. It is suggested that the autonomous military system, also commonly referred to as a 'killer robot,' is one that is capable of executing any lethal action without human control; this elicits important questions such as who is liable, the limit of humanity in battle, and the possibility of overexaggeration (Meneguello, 2019).

Modern non-governmental organizations like Human Rights Watch or several other agencies raised their voices concerning the future of automated weapons; however, there is still nothing concrete about the 2021 legislation. Such discussions tell the reader that, where there is no effective controls, these systems may or can cause catastrophic effects, such as accidents and other forms of violation of international laws on war and conflicts (Sharkey, 2018).

4. Privacy and Surveillance:

The use of AI for surveillance purposes presents significant concerns regarding privacy. Zuboff (2019), in her book The Age of Surveillance Capitalism, outlines how AI and big data analytics enable companies and governments to gather unprecedented amounts of personal information. This information can be used for targeted advertising, political manipulation, and even predictive policing, raising concerns about individual autonomy and freedom. The growing use of facial recognition technology, as highlighted by Garvie, Bedoya, and Frankle (2016), exemplifies this threat. While governments and private companies often tout these systems as enhancing security, their deployment often occurs without adequate oversight or consent, leading to concerns about mass surveillance and the erosion of civil liberties.

5. Existential Risks and Superintelligence:

The long-term dangers of AI, particularly the potential development of superintelligent systems, are widely discussed in both academic and popular literature. Scholars like Bostrom (2014) argue that the development of artificial general intelligence (AGI)— systems capable of outperforming humans in all intellectual tasks— could pose an existential risk to humanity if not properly controlled. A super intelligent AI could develop goals misaligned with human interests, leading to catastrophic outcomes. The "control problem," as described by Yudkowsky (2008), refers to the difficulty of ensuring that an AGI remains aligned with human values as it rapidly evolves. If such a system were to act autonomously, it could prioritize its objectives over human survival, potentially leading to human extinction.

6. Ethical and Governance Challenges:

As AI continues to evolve, its governance has become a major focus of academic inquiry. The literature highlights the need for robust regulatory frameworks to manage the risks associated with AI while fostering innovation.

Floridi and Cowls (2019) discuss the ethical implications of AI in their paper on AI for social good, stressing the importance of transparency, accountability, and fairness in AI systems. However, there is widespread agreement that current regulatory frameworks are insufficient to address the rapid pace of AI development. The global nature of AI also presents challenges, as different countries adopt varying standards for AI ethics and governance. Without international cooperation and standards, the risks of AI misuse or malfunction could be exacerbated (Brynjolfsson & McAfee, 2017).

Hypothesis:

As AI systems become more advanced and widely integrated into various sectors of society, they pose significant risks that could lead to unintended harmful outcomes, including ethical violations, economic disruption, loss of privacy, and even existential threats, unless carefully regulated, aligned with human values, and designed with robust safety protocols.

Key Assumptions in this Hypothesis:

1. Advancement of AI Systems: AI will continue to evolve, becoming increasingly autonomous and capable of decision-making in complex environments.

2. Lack of Proper Oversight: Without adequate regulation, there is a risk that AI will be deployed in ways that prioritize efficiency or profit over ethical considerations. Misalignment with Human Values: AI systems might not be able to fully grasp or implement human values, leading to actions that could cause harm or exacerbate existing societal issues.

3. Unintended Consequences: AI, especially in the form of machine learning systems, may produce unexpected results due to the complexity of real-world environments and their inability to fully anticipate all possible outcomes.

4. Increase in Bias and Discrimination: AI systems may lead to decisions that exacerbate societal inequalities or make discriminatory judgments due to biased training data. Economic Disruption: The widespread adoption of AI in industries will lead to job displacement, particularly in sectors relying on routine tasks, increasing unemployment rates.

5. Cybersecurity Threats: AI will be used to enhance cyberattacks, making it harder for traditional security measures to detect and defend against these threats.

6. Privacy Infringements: AI-powered surveillance systems will increase the potential for privacy violations as they become more integrated into public and private monitoring systems.

7. Long-term Existential Risk: The development of superintelligent AI that lacks value alignment with human objectives could lead to decisions or actions that threaten human existence. Testing this hypothesis would involve examining real-world AI deployment, observing ethical, social, and economic impacts, and assessing the regulatory and safety frameworks in place.

Research Methodology:

To explore the dangers of AI, a multi-faceted research methodology can be used, incorporating qualitative, quantitative, and experimental approaches.

Below is an outline of a possible methodology:

1. Literature Review:

Objective: To review existing research, case studies, and theories related to the risks and dangers of AI. **Approach:** Analyze peer-reviewed papers, books, and reports on AI safety, ethics, job displacement, bias, and existential risks. Focus on historical case studies (e.g., bias in facial recognition, job displacement in manufacturing, cybersecurity breaches).

Outcome: Identify key areas of concern and

common themes, such as AI bias, privacy violations, job loss, and potential long-term risks.

2. Case Study Analysis:

Objective: Examine real-world instances where AI systems have resulted in harmful or unintended outcomes.

Approach: Select well-documented examples (e.g., AI bias in recruitment tools, autonomous weapons, AI- enhanced surveillance leading to privacy breaches). Investigate the circumstances, causes, and impacts of these cases.

Outcome: Detailed analysis of how and why AI can go wrong, identifying patterns and contributing factors.

3. Simulation and Modeling:

Objective: Model potential future scenarios where AI poses significant dangers, particularly in complex systems like healthcare, security, or autonomous vehicles.

Approach: Use simulations to model AI decision- making in high-stakes scenarios (e.g., autonomous vehicles navigating urban environments, AI in financial markets). Model adversarial attacks on AI systems (e.g., cybersecurity vulnerabilities).

Outcome: Quantitative data showing the likelihood and impact of different risks, allowing for a better understanding of how AI systems might fail or be exploited.

4. Quantitative Data Analysis:

Objective: Analyze data on the social and economic impacts of AI to quantify risks such as job displacement, economic inequality, or cybersecurity incidents.

Approach: Collect data from industries undergoing AI-driven automation (e.g., manufacturing, logistics, customer service) to measure job displacement. Analyze trends in cybersecurity breaches and AI- driven attacks.

Outcome: Empirical evidence that helps quantify the scale of potential dangers such as economic disruption or increased cyber vulnerabilities.

5. Policy Analysis:

Objective: Evaluate current AI regulations and policies to determine their effectiveness in mitigating risks.

Approach: Examine AI-related policies across various countries and industries, focusing on ethics, data privacy, and safety. Assess how well these regulations are enforced and whether they adequately address key dangers (e.g., bias, privacy).

Outcome: Insights into regulatory gaps and areas where policy intervention is needed to mitigate AI risks.

Results:

This study employs a quantitative methodology to address the research objectives. The sampling process was conducted effectively, utilizing questionnaires as the primary data collection tool. These questionnaires featured closed-ended questions and incorporated visual aids such as pie charts. A survey was carried out targeting individuals within the age range of 18 to 30, with the sample predominantly drawn from the surrounding region. A total of 100 questionnaires were completed and returned enthusiastically by the participants, who answered a series of 14 questions. When asked about their primary concerns regarding the dangers of artificial intelligence (AI), 30% of respondents indicated **privacy violations** as their top concern, followed by **job displacement** at 25%, **bias and discrimination in AI systems** at 15%, and **misuse of AI for misinformation** at 25%. A small fraction (5%) mentioned **other concerns**, such as the potential loss of human connection and over-reliance on AI technologies. This data highlights that the majority of the audience perceives AI-related privacy issues and job displacement as the most pressing dangers, emphasizing the need for further discussion and regulation in these areas. As part of our research,

a survey was conducted to explore public perceptions of the potential dangers of AI-powered systems. When respondents were asked why they believe AI can be dangerous in terms of privacy, a significant **62.5%** identified privacy violations as their primary concern. This reflects the growing apprehension about the risks associated with AI's reliance on personal data.

Our survey findings reveal that **41.3%** of respondents have experienced or know someone who has experienced job loss due to technological advancements or economic changes. This data reflects the widespread impact of automation and market shifts, particularly in sectors where routine and repetitive tasks are common. As shown in Figure 2, these results underscore the importance of addressing job displacement through targeted policies, such as reskilling programs and social safety nets, to mitigate the negative effects of technological progress.

Our survey results indicate that **55%** of respondents reported a noticeable impact of AI on their professional or personal lives. This finding highlights the pervasive role AI technologies now play in shaping individual experiences, both at work and at home. Figure 3 shows that a majority of respondents— 55%—acknowledge AI's impact, while the remaining 45% either reported no significant impact or were uncertain about its influence.

The survey results revealed that **52.5%** of respondents reported experiencing negative effects of AI in their daily lives. This indicates that a slight majority of participants have encountered challenges associated with AI technologies, suggesting a need for a closer examination of its impact. Figure 4 illustrates the survey findings, showing that **52.5%** of participants reported negative effects of AI in their day-to-day activities, while **47.5%** indicated they had not experienced such effects.

The survey results revealed that **60%** of respondents expressed concerns about AI tracking their online behavior. This significant majority indicates widespread apprehension about how personal data is collected, stored, and used by AI systems in educational and other online settings. Figure 5 depicts the survey findings, illustrating that **60%** of participants are concerned about AI tracking their online behavior, while the remaining **40%** either expressed no concern or were indifferent. Our survey findings indicate that **47.5%** of respondents recognize AI as a significant cybersecurity risk. This concern is rooted in the potential for AI-driven cyberattacks, vulnerabilities in AI models, data breaches, and risks associated with autonomous systems. As illustrated in Figure 6, this reflects widespread awareness of AI's potential to be both a tool for enhancing and a target for cyberattacks. Addressing these risks will be essential in ensuring the safe deployment of AI technologies. While the remaining **52.5%** either identified other risks or did not view AI as a major threat.

Our survey found that **78.8%** of respondents believe AI in gaming has the potential to make games addictive for young users. This concern stems from AI's ability to personalize gameplay, optimize rewards, and increase player engagement through behavioral analytics, all of which could intensify gaming addiction. As shown in Figure 7 presents the survey results, indicating that **78.8%** of respondents are concerned that AI in gaming could contribute to addiction, while **21.2%** did not share this concern or were uncertain about the issue.

Our survey findings indicate that **73.8%** of respondents are concerned that AI tools may misuse or leak personal data of young users without their knowledge. This concern is grounded in fears about unauthorized data collection, sharing with third parties, and the lack of transparency surrounding AI- driven systems.

As illustrated in Figure 8, these findings stress the importance of ensuring robust privacy protections and clear consent processes when using AI technologies with young users.

Our survey revealed that **80%** of respondents believe AI-generated deepfakes have the potential to spread misinformation that might mislead young people. This concern stems from the ability of deepfakes to convincingly alter reality, making it harder for young audiences to distinguish between fact and fiction. As shown in Figure 9, these findings highlight the need for greater awareness and safeguards to protect young users from the dangers of AI-driven misinformation. Figure 9 illustrates the survey findings, showing that **80%** of respondents believe AI-generated deepfakes can spread misinformation that may mislead young people, while **20%** did not share this concern or were uncertain about the issue.

Conclusion:

Artificial Intelligence (AI) can do amazing things, but it also comes with risks we must take seriously. These include issues like job loss, invasion of privacy, biased decisions, and the potential for misuse. If we don't manage AI carefully, it could cause harm instead of helping society. To avoid these dangers, we need clear rules, ethical guidelines, and cooperation between governments, companies, and researchers. By being responsible and cautious, we can enjoy the benefits of AI while keeping its risks under control. Artificial Intelligence (AI) can change the world and make our lives better, but it also comes with risks we need to take seriously. These risks include people losing their jobs as machines take over tasks, private data being misused, unfair or biased decisions made by AI, and AI being used for harmful things like cyberattacks or weapons. AI is developing so quickly that laws and rules might not be able to keep up, creating problems that could be taken advantage of. To handle these risks, we need clear rules and strong ethical standards. We should invest in research to make AI safer and ensure it works in ways people can understand. It's also important to teach people about the risks and benefits of AI so everyone, including leaders and governments, can make better decisions. By working together around the world, we can use AI to improve lives while keeping its dangers under control, creating a future where technology helps humanity instead of harming it.

References:

1. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mane, D. (2016). *Concrete problems in AI safety.* arXiv preprint arXiv:1606.06565.
2. Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies.* Oxford University Press. 3. Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation.* Future of Humanity Institute.
3. Hendrycks, D., Carlini, N., Schulman, J., Steinhardt, J., & Amodei, D. (2023). *An overview of catastrophic AI risks.* arXiv preprint arXiv:2306.12001.
4. Müller, V. C., & Bostrom, N. (2016). *Future progress in artificial intelligence: A survey of expert opinion.* In V. C. Müller (Ed.), *Fundamental issues of artificial intelligence* (pp. 555–572). Springer.
5. Ord, T. (2020). *The precipice: Existential risk and the future of humanity.* Bloomsbury Publishing.
6. Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control.* Penguin Press.