Ethical Dilemma of Advanced Technology: AI Associated Risks

Hassanat Oluwatobi Awodipe, 280020

Abstract: Artificial Intelligence (AI) has significantly enhanced our daily lives and transformed various processes, acting as an enabling technology across diverse applications such as chatbots, algorithmic financial trading, and language translation. However, the benefits of AI come with considerable ethical and societal costs. This study reviews recent literature to illuminate some of the critical ethical dilemmas posed by advanced AI technologies. It examines challenges such as data theft, fairness, the digital divide, and the unique issues introduced by Generative AI tools like ChatGPT. The paper concludes by offering recommendations for promoting ethical AI development while discussing the inherent difficulties in implementing these solutions.

1. Introduction

The rapid progression of technology continues to redefine the way we live and work. Advanced innovations like blockchain, autonomous vehicles, and augmented reality have brought undeniable benefits to society. Various industries integrate such technologies into their operations: international business leverages big data analytics, healthcare utilizes image recognition and robotic systems, and the aerospace sector employs virtual reality (VR) for training and design. The primary driving force behind this relentless advancement is the promise of improved efficiency, enhanced capabilities, and economic growth, pushing researchers and technologists to explore new possibilities continuously.

However, these technological advancements come with significant risks. The surge in cyberattacks and data breaches has raised pressing ethical concerns. For instance, IT Governance, a global leader in cyber risk and privacy management, reported on May 2, 2024, that over 35 billion known data records had been compromised in 9,478 publicly disclosed incidents since the start of the year[1]. This marked an alarming 330% increase in data breaches and a 236% rise in the number of incidents compared to 2023, when over 8 billion records were leaked across 2,814 reported cases[2]. These statistics highlight the growing severity of data security challenges in the modern digital age.

Before the advent of advanced technology, these types of security breaches were largely non-existent. The idea of physically stealing a company's data to expose client information was impractical due to the complexity involved. A perpetrator would need to break into a company, sift through countless paper records, compile them, and then seek out a publisher willing to risk their reputation. The high likelihood of being caught and the logistical hurdles acted as deterrents. In contrast, modern technology allows attackers to carry out sophisticated data breaches remotely, bypassing the need for physical presence and minimizing the risk of exposure.

[1] identifies five major ethical dilemmas tied to emerging technologies: (i) data privacy, (ii) risks associated with artificial intelligence, (iii) creating sustainable environments, (iv) health impacts of

[1] IT Governance, 'Global Data Breaches and Cyber Attacks in 2024', https://www.itgovernance.co.uk/blog/global-data-breaches-and-cyber-attacks-in-2024

[2] IT Governance, 'List of Data Breaches and Cyber Attacks in 2023', https://www.itgovernance.co.uk/blog/list-of-data-breaches-and-cyber-attacks-in-2023

technology use, and (v) issues related to the infodemic and data weaponization. This essay will focus on exploring the risks associated with artificial intelligence (AI), discussing its ethical implications, and proposing ways to develop and maintain ethical AI systems while navigating the challenges involved.

2. Understanding AI-related Ethical Challenges

The rapid pace of technological advancement brings with it an array of ethical challenges that cannot be overlooked, especially in the realms of security, privacy, and sustainability. Consider the ethical dilemma faced by an autonomous vehicle in an unavoidable accident: should the vehicle swerve and risk the safety of its passenger, or continue on its path, potentially harming a pedestrian? Such scenarios highlight the complex moral decisions that arise when machines are programmed to simulate human decision-making [1]. Addressing these issues requires continuous learning and adaptability; however, not all individuals and organizations[3] can keep up, leading to disparities in capability and competitive advantage [2].

Cybercrime and data breaches are among the most pressing threats linked to advanced technology, posing ethical challenges for both developers and users. Users often hesitate to share personal data online due to the risk of cyberattacks, while developers must work tirelessly to fortify their systems—such as securing links in IoT networks—to safeguard user data. AI systems, which rely heavily on extensive data training to learn and make decisions autonomously, amplify these concerns. The potential for data theft, unauthorized access, denial of service, or cyber-physical attacks is ever-present. A stark example occurred in 2015 when BlackEnergy malware infiltrated the Industrial Control System (ICS) of Ukraine's power grid, causing a power outage that affected 225,000 customers for hours[4].

Beyond data breaches, the data used to train AI models raises critical questions about bias and fairness. Personal experiences highlight this issue; for example, as a citizen of a country popularly known for *Yahoo yahoo*[5]. I have encountered discrimination in online banking due to a generalized association with fraud. In 2022, Bper Bank in Italy imposed restrictive measures on Nigerian customers, preventing mobile app transfers and necessitating in-person bank visits with added fees and delays. While such measures aim to prevent fraud, they are unfair to honest individuals and reflect the challenge of teaching AI to differentiate between legitimate users and potential threats without bias.

AI's capability extends to a range of tasks, including repetitive and routine operations, which raises concerns about job displacement [2]. Chatbots and automated ordering systems are now commonplace, taking on roles once filled by human workers. While AI development can create new specialized job opportunities, these positions often require advanced education, training, and access to costly resources, widening the digital divide. Those unable to afford or access such resources risk marginalization, reinforcing socioeconomic inequalities.

The recent release of tools like ChatGPT has sparked significant debate in the education sector [3]. Students can utilize AI for writing assignments and conducting analyses, adding to the plagiarism problem that institutions already must deal with. In response, some researchers have developed models to detect AI-generated text, and educators are learning to identify language patterns typical of AI. A notable

[3] Organisations include public (government) and private organisations.

[4] Christina Miller, https://www.industrialcybersecuritypulse.com/threats-vulnerabilities/throwback-attack-blackenergy-attacks-the-ukrainian-power-grid/

[5] Yahoo yahoo is a broad term for cybercime in Nigeria, mostly perpetuated by the youths and involves defrauding people abroad. More information here: https://dailypost.ng/2023/11/09/yahoo-yahoo-concerns-as-nigerian-higher-institutions-witness-surge-in-internet-scams-among-students/

incident occurred on X (formerly Twitter), where an American user tweeted that a cold email he received was AI-generated because it contained the word ‘delve’.[6] This sparked a debate among users, especially those familiar with the British educational system, who argued that the word was common in formal writing. Such exchanges reveal that individuals may feel pressured to avoid certain words or phrases to distinguish their writing from AI outputs and prove their authenticity.

AI’s integration into social networks has further ethical implications. Facebook, originally launched for Harvard students to connect and share photos, has evolved into a platform serving over 2 billion active monthly users with services such as Marketplace, Gaming, Facebook Watch, and more[7]. This expansion, driven by advanced AI and VR, has exposed Facebook to significant ethical scrutiny. Analyst David Lauer described Facebook’s algorithms as a “black box,” lacking transparency [4]. He argued that these algorithms often amplify political extremism, misinformation, and harmful content to drive user engagement. The core issue lies not in the AI filters themselves but in Facebook’s business model, which prioritizes virality and user interaction over accuracy. Consequently, users are more likely to encounter content with higher engagement, even if it perpetuates falsehoods.

This overview is not exhaustive, but it underscores the breadth of ethical issues surrounding AI. While leveraging AI can enhance problem-solving capabilities, it is essential to balance technological progress with ethical considerations to ensure responsible usage and development.

3. Achieving Ethical AI

AI ethics has become a major focus for research, particularly in Europe, where it is closely tied to the principle of data protection as a fundamental human right [5]. The development and evaluation of ethical AI can be guided by several key principles: transparency, privacy, fairness, accountability, safety, accuracy, and respect for human values [1, 5, 6]. These principles align with existing regulations such as the General Data Protection Regulation (GDPR) and other digital safety laws. One significant challenge in the pursuit of ethical AI lies in designing AI systems that consistently adhere to these ethical standards [1]. Currently, many AI products prioritise commercial profitability over societal well-being, complicating efforts to establish and enforce ethical norms [6].

The dynamic nature of AI further complicates the establishment of ethical guidelines. The technology evolves rapidly, making it difficult to create static regulations that remain relevant over time. To address these challenges, the European Union introduced the Artificial Intelligence Act, which advocates for an "ethics-by-design" approach to AI development [7]. This approach aims to ensure that AI is created with human rights, social equity, and environmental sustainability in mind. Under the Act, existing AI systems will need to be updated or, in some cases, banned if they do not comply with these standards. For instance, the use of CCTV footage to build facial recognition databases is prohibited to protect individual privacy and prevent misuse[8].

[6] Paul Graham on X, https://x.com/paulg/status/1777030573220933716?s=46&t=OFHqNeIZo_eqKekLWzLezA
[7] History.com Editors, ‘Facebook Launches’, https://www.history.com/this-day-in-history/facebook-launches-mark-zuckerberg
[8] European Parliament, ‘Artificial Intelligence Act: MEPs adopt landmark law’, https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law

Complementing this, the EU Data Act empowers individuals and businesses to control how data generated by their connected products is used. This regulation enhances data sovereignty, ensuring that data handling aligns with privacy and ethical standards[9].

3.1. Privacy and Transparency

The volume and variety of data being collected make this a significant issue in data processing. The GDPR and other national data protection laws have outlined steps to be taken when processing private data including mitigation measures when a data breach occurs. In [5], experts encouraged researchers to apply a privacy-by-design approach to their projects such as encryption and hashing, and using data-protection-focused service providers and storage platforms. Pseudonymisation and anonymisation of personal data are encouraged while data minimisation is emphasized. In addition, the law places obligations on the *data controller* – who determines the purpose and means of processing data – and the *data processor* – who processes data on behalf of the controller.

Informed consent to processing also helps with achieving privacy, accuracy and respect for human values [5, 7]. In this case, users are informed on how their data will be used and what it will be used for, and they reserve the right to agree, disagree or withdraw after agreeing. They are also allowed to edit or update their data. However, data processors must be cautious not to present this information in highly technical terms or downplay the operations involved. [7] equally emphasises the need to pay attention to human well-being and the impact data processing can have on people's intimate lives even after giving consent.

Transparency plays a critical role in fostering trust. To this end, some organizations make their application code publicly accessible on platforms like GitHub. GitHub facilitates the storage, tracking, and collaborative development of software, enabling interested parties to review the code, seek clarifications, or contribute to the project. A notable example is Google DeepMind's open-source research on using AI to predict the progression of retinal diseases[10]. By making such projects publicly available, organizations enhance transparency and encourage public engagement with their technology.

3.2. Interpretable and Explainable AI

The most effective AI methods are often complex, leading to challenges in understanding how specific outcomes are achieved [8]. These methods are commonly referred to as *black-box* approaches due to their opaque nature. With rising demands for ethics and transparency, and the implementation of legal frameworks such as the GDPR, the interpretability of AI has become an important focus. The ambiguity inherent in AI models, particularly in high-stakes areas like healthcare, limits their application and trust [9].

For improved predictive accuracy, there is an increase in model complexity. However, by integrating *Explainable AI* (XAI) techniques, scientists can interpret and communicate how models produce their results [8]. This advancement is crucial as it allows for the identification of potential errors, biases, or the use of irrelevant data features [9]. A practical example in action is IBM Watson for Oncology, an AI software that aids in diagnosing cancer and recommending treatment plans. This tool not only ranks treatment options but also provides supporting evidence for each recommendation, drawing on medical

[9] Cyber Risk GmbH, 'The European Data Act', https://www.eu-data-act.com/#:~:text=The%20data%20act%20will%20give,objects%2C%20machines%2C%20and%20devices

[10] Jason Yim, Reena Chopra, Jeffrey De Fauw, Joseph Ledsam, 'Using AI to predict retinal disease progression', https://deepmind.google/discover/blog/using-ai-to-predict-retinal-disease-progression/

literature and expert-curated guidelines, such as those from Memorial Sloan Kettering[11]. Such transparency helps oncologists trust and verify the AI's recommendations.

According to [9], interpretability methods can be categorized based on their algorithms. *Model-specific* interpretability is limited to particular predictive models, while *model-agnostic* methods can be applied across various models . Additionally, these methods may bed by their scope. If an interpretation pertains to a single data instance, it is *local*; if it covers a broader set of instances, it is *global*. Using this taxonomy, interpretability techniques are divided into four main categories: methods for explaining complex black-box models, techniques for developing transparent *white-box* models, methods aimed at reducing discrimination and enhancing fairness in machine learning, and tools for assessing the sensitivity of model predictions. This range ensures that, regardless of complexity, there is a way to demystify every type of AI model.

3.3. Cybersecurity

Cybersecurity encompasses strategies and measures designed to safeguard digital systems from malicious attacks. AI-based products often feature characteristics such as connectivity, autonomy, alterability, location tracking, and user intimacy [7]. As a result, breaches involving these systems can have severe repercussions. To combat these risks, both individuals and organizations need to proactively identify potential cyber threats and establish preventive measures. Routine security evaluations to detect system vulnerabilities, robust data encryption, user authentication protocols, and comprehensive employee training are all essential components of a strong cybersecurity framework [2].

From an ethical standpoint, the role of hacking must also be considered [1]. While malicious hackers—categorized as black hat and red hat hackers—pose clear threats, there are also ethical hackers, or white hat hackers, who contribute positively to cybersecurity. These professionals conduct threat intelligence and penetration testing with the company's permission, adhering to both organizational policies and national laws. In contrast, gray hat hackers perform similar activities without explicit consent, which, while potentially beneficial, breaches ethical and legal boundaries. By employing ethical hacking practices, organizations can better understand their vulnerabilities and prepare responses for potential threats. If a cyberattack occurs, it is critical for organizations to analyze the breach and use their insights to reinforce existing security measures.

Nevertheless, even the most sophisticated cybersecurity strategies are ineffective if system users lack awareness. The BlackEnergy malware attack in Ukraine serves as a case study demonstrating this point. Attackers gained entry through phishing emails containing malicious MS Office attachments, which were sent to employees in the Industrial Control System (ICS) and energy sectors. These emails appeared to originate from the Ukrainian Parliament, making the recipients more susceptible to opening them. This incident underscores the importance of training employees on best practices for cybersecurity, such as recognizing phishing attempts, and educating consumers on taking personal responsibility for their online choices [6].

3.4. AI Regulation and Compliance

While ethical frameworks provide valuable guidance for organizations developing AI products, formal policies set by higher authorities are essential for ensuring compliance and accountability. A significant

[11] IBM sales manual, https://www.ibm.com/docs/en/announcements/watson-oncology?region=CAN

concern in this area is *ethics washing*, where companies publicly showcase an interest in ethical practices to avoid stringent regulations or to enhance their reputation without genuine commitment [6]. This deceptive approach is aimed at gaining public trust while prioritizing the company's interests. Thus, it is crucial that when organizations pledge adherence to ethical codes or regulations, regulatory bodies actively monitor their compliance to prevent such practices.

To strengthen accountability, regulatory agencies should implement methods to measure and publicly report the ethical performance of AI systems. Offering incentives for organizations with high ethical standards could encourage companies to prioritize ethics in their operations. These rewards should be designed to outweigh the costs associated with integrating ethical practices, thereby fostering a competitive advantage and promoting consumer trust [6].

The presence of dedicated ethics departments within organizations can further ensure that ethical considerations remain central to product development. Since many AI engineers and software developers have limited training in ethics, having an in-house team focused on ethical review facilitates compliance. For instance, Microsoft has established the *Office of Responsible AI* and the *AETHER* committee (AI, Ethics, and Effects in Engineering and Research) to oversee its ethical initiatives [6]. Smaller organizations could adopt a similar approach by employing an ethics officer to work closely with development teams and review AI outputs.

The European Parliamentary Research Service (EPRS) has suggested establishing an EU observatory to monitor the convergence of digital technologies. AI is increasingly integrated with other technologies, such as Internet of Things (IoT) devices, compounding the risks associated with each individual component. This observatory would assess the ethical, legal, and social issues (ELSI) related to such convergences and recommend appropriate regulatory measures [7].

Regulating AI also involves scrutinizing the purpose of future AI products. Some applications may be inherently unethical due to their capabilities. For instance, in 2017, the United Nations Institute for Disarmament Research (UNIDIR) published a study examining the implications of *Autonomous Weapon Systems* (AWS) [10]. This research emerged amidst debates at the Convention on Certain Conventional Weapons (CCW) over whether to adopt lethal AWS. Although there is no consensus among member states, over 100 countries[12] and the UN Secretary-General have called for a binding international agreement to prohibit such weapons[13].

4. Conclusion

This paper has explored the ethical challenges associated with AI technology and assessed approaches for achieving ethical AI practices. Despite existing measures, it remains uncertain whether cybercrime can be entirely eradicated or AI ethics fully sustained. Factors such as the rapid pace of technological advancements, diverse cultural and social norms, and the limited cyber literacy of many end-users contribute to the complexity of maintaining ethical AI practices. Furthermore, the misconception that enforcing ethical standards hinders innovation in AI development presents an additional obstacle [1].

[12] UNODA Meetings, 'Statement by Stop Killer Robots', https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2024)/SKR-Statement_GGE-March.pdf

[13] Mary, 'Stop Killer Robots', https://www.stopkillerrobots.org/news/unban/

Cybercriminals often evade prosecution, which can serve as an encouragement to potential offenders. This is partly due to the transnational nature of cybercrime, where perpetrators are often located in different jurisdictions than their victims. Jurisdictional limits complicate prosecution, prompting calls for subjective territorial jurisdiction to prosecute cybercriminals where they reside, with transnational jurisdiction reserved for exceptional cases [11]. The *Budapest Convention on Cybercrime*, already adopted by EU countries and some non-EU nations, should be ratified globally to strengthen international cooperation in combating cybercrime [11].

Ultimately, ensuring AI ethics is a continuous and adaptive process that necessitates the involvement of all technology stakeholders. Governments play a crucial role by implementing laws such as the GDPR, the EU AI Act, and the EU Data Act, which regulate data collection and AI development. Meanwhile, some organizations are actively working to build secure, inclusive systems and comply with regulations. For their part, individuals should exercise caution in their online activities and support organizations that demonstrate a commitment to ethical practices and fairness.

References

[1] L. L. Dhirani, N. Mukhtiar, B. S. Chowdhry, and T. Newe, “Ethical Dilemmas and Privacy Issues in Emerging Technologies: A Review,” Feb. 01, 2023, *MDPI*. doi: 10.3390/s23031151.

[2] M. Thulani Bongani Makhanya, “The Implications for Risk Management in the Era of Technological Advancements,” in *The Future of Risk Management [Working Title]*, IntechOpen, 2024. doi: 10.5772/intechopen.1003899.

[3] D. Baidoo-Anu and L. Owusu Ansah, “Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning,” *Journal of AI*, vol. 52, no. 7, pp. 52–62, Dec. 2023, doi: 10.61969/jai.1337500.

[4] D. Lauer, “Facebook’s ethical failures are not accidental; they are part of the business model,” *AI and Ethics*, vol. 1, no. 4, pp. 395–403, Nov. 2021, doi: 10.1007/s43681-021-00068-x.

[5] B. Hayes and A. Kuyumdzhieva, “Ethics and data protection,” 2021. Accessed: Jul. 15, 2024. [Online]. Available: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf

[6] A. Schoentgen and L. Wilkinson, “Ethical issues in digital technologies,” Gothenburg: International Telecommunications Society, 2021. [Online]. Available: https://hdl.handle.net/10419/238052

[7] I. Poel, T. Wildt, I. Oosterlaken, and J. Hoven, “Ethical and societal challenges of the approaching technological storm,” 2022. Accessed: Jul. 11, 2024. [Online]. Available: https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)729543

[8] A. Holzinger, R. Goebel, R. Fong, T. Moon, · Klaus-Robert Müller, and W. Samek, “xxAI-Beyond Explainable AI,” 2020. [Online]. Available: https://link.springer.com/bookseries/1244

[9] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, “Explainable AI: A review of machine learning interpretability methods,” Jan. 01, 2021, *MDPI AG*. doi: 10.3390/e23010018.

[10] D. Atkinson *et al.*, “The Weaponization of Increasingly Autonomous Technologies: Autonomous Weapon Systems and Cyber Operations,” 2017. Accessed: Jul. 15, 2024. [Online]. Available: https://unidir.org/publication/the-weaponization-of-increasingly-autonomous-technologies-autonomous-weapon-systems-and-cyber-operations/

[11] P. Arnell and B. Faturoti, “The prosecution of cybercrime–why transnational and extraterritorial jurisdiction should be resisted,” *International Review of Law, Computers and Technology*, vol. 37, no. 1, pp. 29–51, 2023, doi: 10.1080/13600869.2022.2061888.