

Artificial Intelligence Ethics and Cybersecurity: Overview and Prospects

Dengxu Li

School of International Education, Quanzhou University of Information Engineering, Quanzhou, Fujian, China

Keywords: Artificial Intelligence, Ethics, Cybersecurity.

Abstract: This paper summarizes the current situation and future outlook in the field of artificial intelligence ethics and cybersecurity. With the rapid development of AI technology, the industrial upgrading and social progress it brings cannot be ignored. However, it is accompanied by increasingly severe cybersecurity challenges, such as information leakage and system crashes, which seriously threaten user privacy and social stability. The article points out that although AI ethics and cybersecurity have been studied in depth respectively, the exploration of the intersection of the two is still insufficient. By reviewing relevant research results at home and abroad, this article emphasizes the importance of AI ethics in ensuring that the use of technology does not violate human rights and interests and meets ethical standards, and discusses the challenges and solutions that cybersecurity faces in the era of big data and advanced analytics. The article aims to fill the gap in the intersection research of AI ethics and cybersecurity, propose countermeasures to deal with the existing challenges and contribute to building a harmonious and stable social environment.

1 INTRODUCTION

With the rapid development of modern computer technology, Artificial Intelligence (AI) technology has become more and more mature and has penetrated into many fields such as healthcare, education, finance, etc., which has promoted the progress of society while upgrading the industry. Artificial intelligence (AI) technology in OpenAI ChatGPT application: a review of ChatGPT in writing English essay mentions that AI technology (such as ChatGPT) has been developed to facilitate human activities and productivity. By automating and intelligently handling tasks, AI can significantly increase productivity and enable humans to focus on higher-level work (Fitria, 2023).

However, while AI is becoming more and more developed, some cybersecurity issues cannot be ignored, with frequent incidents of information leakage and website crashes, which greatly threaten the privacy and security of users, and even pose a serious threat to social stability. Therefore, while researching AI, what should not be ignored is the network security problem, how to strengthen network maintenance, to ensure that the information is safe and reliable is a topic that needs to be studied in depth.

AI ethics and network security have always been two very important topics, which are related to the stability of social development and the privacy and security of users. From the current point of view, although the two topics have been studied more deeply, the intersection of the two topics has not yet been insufficient and there are large gaps. This paper aims to fill the gaps between the two, deeply analyze and study the impact and challenges of AI ethics and cybersecurity and put forward countermeasures against their vulnerabilities, so as to contribute to the construction of a harmonious and stable society.

Reviewing the research results on AI ethics and cybersecurity at home and abroad, it is found that both have been studied in depth. For example, J.E.(Hans) Korteling and G.C. van de Boer-Visschedijk discussed how to safely and reasonably develop AI systems in order to ensure that their use does not infringe upon the rights and interests of human beings and at the same time conforms to moral and ethical standards (Korteling, 2021). Further, Amin Al Ka'bi in Challenges Encountered by AI Applications in Today's Interconnected World states that in the AI/ML field, ethics and transparency are crucial as they ensure credibility, accountability, security, and interpretability, which are key to the widespread adoption and acceptance of these

technologies in today's interconnected world (Al Ka'bi, 2022). In terms of cybersecurity, Network security AIOps for online stream data monitoring demonstrates some of the current issues for cybersecurity in the field of technology, emphasizing the dynamic thresholds, deep learning, the role of SOCs, situational awareness, and the cybersecurity tools and techniques in addressing modern cybersecurity challenges (Nguyen, 2022). In Enhancing cybersecurity protocols in the era of big data and advanced analytics, it is mentioned that with the exponential growth of big data, the diversity and complexity of data is increasing, and the different formats and sources of data as well as the speed of data processing require cybersecurity protocols to have greater adaptability and flexibility (Nwobodo, 2024).

AI ethics are closely related to cybersecurity. In the future, with the development of technology and the expansion of application scenarios, people need to pay more attention to AI ethical issues and actively search for effective solutions to ensure the healthy development of AI technology and social well-being.

2 OVERVIEW OF AI TECHNOLOGY DEPARTMENT

AI is a cross-discipline based on computer science and technology and cross-fertilized by multiple disciplines such as computing, philosophy, psychology, etc. It is a new technological science that researches and develops theories, methods, technologies, and application systems used to simulate, extend, and expand human intelligence. The research in this field covers a wide range of aspects such as robotics, language recognition, image recognition, machine learning, deep learning, natural language processing and expert systems. AI can simulate the information process of human consciousness and thinking, and although it is not human intelligence, it is able to think like a human being and may even exceed human intelligence in some aspects. As discussed in A Scholarly Definition of AI: Advancing AI as a Conceptual Framework in Communication Research, the authors define AI as non-human machines or artificial entities with the tangible ability to perform, solve tasks, communicate, interact, and act logically in the real world, as occurs in biological humans (Gil de Zúñiga, 2024). They also stated that AI is mainly defined as computational technology capable of achieving specific tasks,

imitating human behavior, cognitive heuristics and intelligence.

The Dartmouth Conference held in 1956 marked the emergence of AI as an independent field of study. An important branch of early AI was semiotics, whose main research directions included natural language processing and logical reasoning, etc. The 1960s and 1970s were the early years of the development of AI. In 1961, John McCarthy introduced the term "artificial intelligence", which meant that the research on AI was gradually deepened, but due to the limitations of technological development and computing power, the development of AI was relatively slow. 1980s, connectionism and neural networks became an emerging field of research, and in 1986, the back propagation algorithm opened the way to the research and development of neural networks, and after this, the research direction of AI gradually shifted to machine learning and deep learning. 21st century to the present is the beginning of deep learning. The beginning of the 21st century to the present is the rise of deep learning, which began to rise in 2006 with the proposal of deep confidence networks. Deep learning analyzes and learns the special features of data and their deeper meanings through multi-level neurons, and is widely used in natural language recognition, image recognition, speech recognition, etc. In 2016, AlphaGo defeated Go world champion Lee Sedol, demonstrating the potential of AI and reinforcement learning technology in the field of AI. This was followed by the birth of ChatGPT, which was introduced in 2022. ChatGPT, as a powerful language model developed by OpenAI, has strong language processing capabilities and the ability to converse with human beings, making it a new milestone in AI. The introduction of ChatGPT marks a major breakthrough for AI in the field of science and technology, and also signals that AI will gradually integrate into all aspects of human life.

The technical classification of AI mainly includes: machine learning, deep learning, natural language processing, computer vision, robot vision, robotics, biometrics, portrait recognition, image recognition, text recognition, aerospace applications, information processing and many other types.

An important application of AI in the field of cyber security is intrusion detection. The system is able to automatically identify and analyze anomalies in network traffic through robotic algorithms, thus enabling timely detection and blocking of virus intrusions and network attacks. For example, the malicious code analysis and detection system based on machine learning can automatically identify and

extract the characteristics of malicious code and analyze them, so that network attacks such as Trojans and worms can be effectively identified.

AI can also predict cyber attacks that have not yet occurred by analyzing historical data and network behavior. This predictive ability can help enterprises and organizations deploy preventive measures in advance, which can greatly reduce the losses caused by cyber attacks. Another important task in cybersecurity is malicious domain name detection. By introducing data enhancement techniques and some neural network models, the system is able to automatically detect malicious domain names, thus protecting network security.

Currently, AI has infiltrated into people's daily life as well as all walks of life, and some potential cybersecurity issues are gradually emerging as AI becomes more and more prevalent.

The AI technology in vehicles such as automobile's automatic driving, unmanned vehicles, drones and other means of transportation, although not directly related to cybersecurity, the communication network it uses may become the target of cyber virus attacks. Once the automatic driving system is invaded by a virus, it may lead to serious consequences such as leakage of driving data and loss of control of the vehicle, seriously threatening privacy and personal safety. Applications of AI in financial services include investment management, risk prediction, stock prediction, and so on. AI improves the efficiency and accuracy of financial services. As demonstrated by LONGBIN CAO in AI in Finance: Challenges, Techniques and Opportunities, AI can assist financial institutions in risk assessment and monitoring, identifying potential dangers through complex models, thus dramatically improving risk management. AI algorithms can quickly analyze large amounts of transaction data and identify abnormal transaction behavior, thus effectively preventing financial crimes (CAO, 2022).

3 EXISTING ISSUES IN ARTIFICIAL INTELLIGENCE ETHICS

With the rapid development of AI technology, it is increasingly being used in many areas of life. This has a positive and positive impact on service quality and productivity. However, as technology advances, so do the privacy risks associated with the collection and use of user data. The development of AI technology, especially complex algorithms like deep learning,

relies on the training and optimization of large amounts of data, much of which contains sensitive information such as location trajectories, behavioral preferences and personally identifiable information. The lack of clear guidelines for data use and strict user permission processes during the data collection phase may increase the risk of data leakage and misuse.

The risk of data leakage mainly stems from technical vulnerabilities, human errors or malicious attacks. CYBERSECURITY RISK mentions cases such as SolarWinds, in which a cyberattack not only affected the direct target, but also spread to many client organizations, including some large U.S. federal agencies. This suggests that data breaches or cyberattacks can trigger a chain reaction with indirect effects on the wider economy (Florackis, 2023).

Data misuse raises more complex ethical issues as it involves social justice and equity in addition to privacy violations. Certain companies may use user data for inappropriate marketing, tailored pushes, and discriminatory pricing, thus negatively impacting consumer rights. More importantly, the use of data for algorithmic decision-making can lead to "algorithmic discrimination" and exacerbate social inequality if the training data is insufficient or the algorithm is biased. For example, gender bias in historical data may lead employment algorithms to discriminate against female applicants, while regional disparities may lead banking algorithms to refuse to serve people in a particular region.

To solve this problem, a combination of ethical, technological and legislative elements is necessary. Legally, relevant laws and regulations must be updated to strengthen the regulation of data collection, processing, transmission and storage, and to clarify the parameters of data use and the responsible parties. In order to address the difficulties posed by data flows in the context of globalization, efforts should be made to harmonize and unify international data protection legislation. In order to prevent the misuse or leakage of user data, we should improve data security protection at the technical level by implementing encryption technology and using measures such as anonymous processing. Auditing and traceability procedures should also be established in order to ensure the legality and compliance of data use. Companies should adhere to the values of fairness, openness and respect when using data, while reinforcing social responsibility and ethical obligations. At the same time, they should improve their personal privacy protection skills and raise public awareness of data protection, so that they can work together to create a thriving data ecosystem.

In the field of machine learning, although the algorithms themselves are not subjectively biased, they may unintentionally reflect social and cultural biases and discrimination when processing data. This phenomenon has become one of the hotspots of AI ethics and fairness research. The bias and discrimination of machine learning algorithms mainly stems from the following aspects.

The process of developing machine learning algorithms begins with data collection and processing, which usually involves bias aka data discrepancy. Since training data is usually created by people, social and cultural biases are inevitable. If historical data from a hiring system shows gender bias, i.e., the percentage of successful female applicants is significantly lower than that of male applicants, then an algorithm trained on that data may replicate that bias in hiring decisions, thereby treating female applicants unfairly. In addition, an important factor contributing to bias is the underrepresentation of the dataset. When the dataset is not fully representative of all groups, the algorithm may not be able to correctly learn the characteristics of the minority group, leading to biased decision making.

Next is the design and training of the algorithm. When dealing with difficult problems, algorithm designers may unintentionally bias or discriminate, especially by simplifying assumptions. In addition, the optimization goal during algorithm training is usually to maximize some kind of accuracy rather than to maximize justice. This may cause algorithms to prioritize high accuracy over specific demographics. For example, in an image recognition task, if there are few examples of a certain image type (e.g., dark-skinned people) in the training dataset, the algorithm may not be able to correctly learn the features of that type of image, leading to recognition bias.

4 AI ETHICS FROM A CYBER SECURITY PERSPECTIVE

In the field of network security, although the introduction of AI technology has brought great innovation to the defense mechanism, it also faces many difficulties and serious challenges. First, the endless growth of complex threats is the first difficulty that AI defense systems must address. These include ransomware, phishing, and traditional virus propagation, as well as advanced persistent threats (apt) and distributed denial of service (DDoS) attacks, which pose more diverse and subtle threats to

cyberspace security. Traditional defense strategies no longer seem to be effective in the face of ever-innovating attacks. While AI systems are able to detect and react automatically, maintaining effective and accurate judgment in a complex and dynamic attack environment has become a major challenge.

In addition, vulnerability exploitation is another major issue facing AI in cybersecurity. As software systems become more complex, new vulnerabilities are often discovered. Attackers can often take advantage of these weaknesses to launch attacks quickly. AI systems have the ability to monitor network traffic and detect abnormal behavior in real time, but the protection efficiency of AI systems may be seriously affected by the presence of unknown security vulnerabilities or new attack techniques. Therefore, how to improve the defense capability of AI systems in terms of vulnerability exploitation has become an urgent problem.

In order to cope with the ethical challenges of AI in the field of cybersecurity, we need to strengthen the ethical construction of AI from multiple dimensions. First, the key to ensure the healthy development of AI is to improve laws and regulations. In order to control the use of AI in cybersecurity, clarify the roles and responsibilities of all parties, and maintain the progress of AI technology, the government should formulate clear laws and regulations. Changwu Huang in *An Overview of AI Ethics* mentions that discussing the ethics of AI will inevitably involve the legal considerations (Huang, 2022). At the same time, it is necessary to establish a strict regulatory mechanism to crack down on illegal behaviors and maintain the security and order of cyberspace.

Second, more regulation is needed to ensure that AI ethics are properly maintained. As mentioned in *Ethical framework for AI and digital technologies* with the popularization of AI and DT, a series of ethical issues have surfaced, such as privacy leakage and algorithmic bias. It is the responsibility of the government to establish and improve regulatory mechanisms to ensure that the application of AI meets ethical standards and prevents potential social risks (Ashok, 2022).

In addition to establishing joint forces and collaborative regulatory mechanisms across sectors and domains, the government should strengthen the regulation of AI technologies. At the same time, businesses and non-governmental organizations are urged to participate in the supervision and utilize their technology, resources, and other advantages to jointly promote the application of AI ethics.

In addition, enhancing technology safety is the cornerstone of safeguarding AI ethics. Security and

privacy protection need to be considered at the R&D stage of technology to ensure that AI systems strictly comply with applicable laws, regulations and ethical norms when handling sensitive information. At the same time, technological innovation should be strengthened to continuously improve the defense and self-learning capabilities of AI systems in order to cope with increasingly complex cybersecurity challenges.

5 CHALLENGES AND PROSPECTS

With the rapid development of AI technology, its application in the field of network security is increasingly promising. In the future, AI technology will show the following key development trends in the field of network security. AI technology will further enhance the intelligence level of network security defense, and through machine learning, deep learning and other algorithms, it will realize the intelligent identification, warning and response to network attacks. This trend will greatly improve the efficiency and accuracy of cybersecurity defense, but at the same time, it also puts higher requirements on the accuracy and transparency of the algorithms. As the technology matures, AI will be able to handle cybersecurity incidents more autonomously and automate emergency response. This will shorten the time from detection to response and reduce human error, but it may also trigger ethical discussions about the interpretability of machine decision-making processes and attribution of responsibility. Cybersecurity threats are constantly evolving, requiring AI technologies to have the ability to continuously learn and evolve. Future cybersecurity systems will be able to continuously learn new attack patterns and automatically update defense strategies, but this self-evolutionary process may also bring new ethical challenges, such as algorithmic bias and uncontrollability.

As the application of AI in cybersecurity deepens, the transparency and interpretability of algorithms becomes an urgent issue. Users and society need to understand how algorithms make decisions to ensure their fairness and reasonableness. Lack of transparency may lead to public distrust of the technology and even ethical controversies.

The application of AI in cybersecurity inevitably involves the collection, processing and analysis of large amounts of data. How to protect personal privacy and data security has become an important

ethical issue. Strict data protection mechanisms need to be established to ensure the security and privacy of data during collection, storage, transmission and use.

In automated emergency response and intelligent defense systems, it becomes difficult to determine the attribution of responsibility when the system makes mistakes or causes problems. Clear accountability mechanisms need to be established to ensure that those responsible can be held accountable when problems arise, while protecting technological innovation and development.

Given the potential impact of the ethical issues mentioned above, it is particularly important to establish a comprehensive and systematic code of ethics for AI. This will not only help regulate the R&D and application behavior of AI technology and reduce ethical risks, but also help enhance public trust and acceptance of the technology.

In order to promote the construction of AI ethical norms, the construction of international cooperation mechanisms is crucial. Different countries and regions have different views and positions on AI ethics, and through international cooperation, mutual understanding and consensus can be enhanced, and ethical norms and standards of universal applicability can be jointly formulated. At the same time, international cooperation can also promote the exchange and sharing of technology and promote the healthy development of AI technology.

Therefore, this paper suggests strengthening international dialog and cooperation to jointly promote the construction of ethical norms for AI. Through the formulation of clear ethical principles, the establishment of an effective regulatory mechanism, and the strengthening of ethical review in the research, development and application of technology, we can ensure that the research, development and application of AI technology meets ethical requirements and promotes the sustainable development of society.

6 CONCLUSIONS

This paper focuses on the ethical and cybersecurity issues facing the rapid development of AI technology. The paper first outlines the application of AI technology in the fields of healthcare, education, finance, etc. and its positive impact on the society, and then points out the threats to user privacy and security and social stability posed by cybersecurity issues such as information leakage and website crashes. Then, the paper deeply analyzes the close connection between AI ethics and cybersecurity, and

reviews domestic and international research results in this field, such as the ethical discussion on the development of AI systems and the importance of ethics and transparency in the AI/ML field. In addition, the paper introduces some current key issues and technical solutions in the field of cybersecurity, such as the application of machine learning and deep learning in cybersecurity monitoring. Finally, the paper looks into the future direction of AI ethics and cybersecurity, emphasizing the importance of paying attention to ethical issues and actively finding solutions to ensure the healthy development of AI technology and social stability.

REFERENCES

- Fitria, T. N., 2023. Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. In *ELT Forum: Journal of English Language Teaching* (Vol. 12, No. 1, pp. 44-58).
- Korteling, J. E., Visschedijk G. C., Blankendaal, R. A. M., Boonekamp, R. C. & Eikelboom, A. R., 2021. Human-versus Artificial Intelligence. *Frontiers in Artificial Intelligence*. 622364-622364.
- Al Ka'bi, A., 2022. Challenges encountered by artificial intelligence applications in today's interconnected world. In *Proceedings of the 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)* (pp. 1-6). Prague, Czech Republic. doi: 10.1109/ICECET55527.2022.9873465
- Nguyen, G., Dlugolinsky, S., Tran, V., & López García, Á., 2024. Network security AIOps for online stream data monitoring. *Neural Computing and Applications*, 1-25.
- Nwobodo, L. K., Nwaimo, C. S., & Adegbola, A. E., 2024. Enhancing cybersecurity protocols in the era of big data and advanced analytics. *GSC Advanced Research and Reviews*, 19(3), 203-214.
- Gil de Zúñiga, H., Goyanes, M., & Durotoye, T., 2024. A scholarly definition of artificial intelligence (AI): advancing AI as a conceptual framework in communication research. *Political communication*, 41(2), 317-334.
- Cao, L., 2022. Ai in finance: challenges, techniques, and opportunities. *ACM Computing Surveys (CSUR)*, 55(3), 1-38.
- Florackis, C., Louca, C., Michaely, R., & Weber, M., 2023. Cybersecurity risk. *The Review of Financial Studies*, 36(1), 351-407.
- Huang, C., Zhang, Z., Mao, B., & Yao, X., 2022. An overview of artificial intelligence ethics. *IEEE Transactions on Artificial Intelligence*, 4(4), 799-819.
- Ashok, M., Madan, R., Joha, A., & Sivarajah, U., 2022. Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information Management*, 62, 10243.