**A Literature Review on Artificial Intelligence Foreseeable Risks and Ethics**

***Full name: Chi Nguyen Chau***

***SID: 22156675***

## Introduction

Artificial Intelligence (AI) has been developing sharply and affecting on various disciplines including business, healthcare, manufacturing, education, agriculture, defence and universal. It can be applied to enhance the productivity, reduce costs, automate the repetitive and monotonous tasks, support medical procedures and complex data analysis like identifying fraudulent, finding shortest route, recommending systems, generating text summarization, object detection and so on (Sarangi & Sharma, 2018). However, what left behind those marvellous and astonishing advancements and what should be put more attention to contemporarily along the journey of development of AI. This report will review these aspects focusing on potential risks of AI applications, AI ethics frameworks and then draw some conclusions.

## The potential risks of AI

Similar to the advents of new machines in the Industrial Revolution, the AI technologies have potential to change how the world develops. Beside positive effects, many facets of human well-being and society are in risk like unemployment, transparency, safety, privacy, fairness and accountability.

With the application of AI technology, mass of unemployment can happen both in low-skilled and high-skilled jobs, monotonous jobs and creative jobs. For examples, manual labour can be replaced by robots in manufacturing; AI assistants can be used in creating contents and customer services instead of hiring staffs; or when AI technologies developed to a certain extent, hiring juniors for some positions may become considerably inefficient in case AI can cover their tasks well. Even for the C-level jobs, corporations can consider whether they should open or reduce the cost for those positions when the AI’s ability of decision-making will be improved more and more (Sarangi & Sharma, 2018).

Transparency is another risk of AI applications which is about the understanding of how the algorithms work or make decisions (Sarangi & Sharma, 2018; Kazim & Koshiyama, 2021). The release of deep-learning algorithms which mimic how the human brain works advocated a main part on creating more complex and human-like applications. Unfortunately, the same as our shortage of understanding of human brain, it is difficult or even impossible to comprehend what these applications learnt. And the problem without comprehending is that how we can trust their decisions or re-educate them if they make decisions which are harmful to human well-beings.

Safety and privacy are also considerable aspects. The demand of a tremendous amount of data including sensitive and private information for training AI algorithms which might be collected and used unacknowledged raised a warning in violating human privacy rights (Sarangi & Sharma, 2018; Kazim & Koshiyama, 2021). When the AI applications which are getting more and more vital and powerful in real life are attacked or used maliciously such as in war or illegal actions, it can lead to the more and more consequences. For example, Facebook has been ordered by Australian Federal Court to pay $10 million for misleading consumers about use of their data in 2023 (ACCC, 2023) or Tiktok was sued for “collecting children’s personal information without sufficient warning, transparency or the necessary consent required by law, and without children or parents knowing what is being done with that information” in 2021 (BBC News, 2021).

Fairness is still a big problem related to gender, nationality, race, income, etc. in current world. The action of finding patterns in historical data of AI algorithms can cause the bias when make decisions which treat discriminatorily among individuals or groups (Sarangi & Sharma, 2018; Kazim & Koshiyama, 2021). Besides that, for each action conducted by AI systems, argument is also made about the accountability for it. For example, when an autonomous car causes a fatal crash (BBC News, 2020), who will be responsible for that: the pedestrian, the driver who is in the car which are in the self-driving mode, the manufacturer or the developer.

Because of the appearance of these potential risks, it is necessary to have ethical standards or codes of conduct which can put the way AI applications are developed controllable. Otherwise, some horrible and unmeasurable consequences might be happening in the future like the bombings of Hiroshima and Nagasaki in case of nuclear energy invention.

## The AI ethics frameworks

Some national AI ethics frameworks have been released as guidelines for the process of designing, developing, deploying, implementing and maintaining AI applications around the world such as Australia’s AI Ethics Framework and Ethics Guidelines for Trustworthy AI set up by the European Commission. These frameworks are based on the fundamental human rights and aligned with the available laws and regulations in those countries as well as international law in order to ensure that the below main requirements should be adhered:

* *The human rights should be respected such as privacy, safety, freedom and fairness:* the AI systems should support human autonomy, allow human oversight, prevent malicious uses which can be harmful human well-beings and be protected against vulnerabilities; the privacy and data protection must be guaranteed; data collection must be under acknowledgment of direct and indirect relevant parties; avoidance of unfair bias must be conducted throughout the entire AI system’s life cycle (Department of Industry, Science and Resources, 2019; European Commission, 2019).
* *The transparency should be guaranteed*: each process in the life cycle should be traceable and documented to the best possible standards; the decisions should be able to be explained and the trade-off between explainability and accuracy should be prior to the former whenever the AI system has significant impacts on human lives; users should know clearly when interacting with AI systems (Department of Industry, Science and Resources, 2019; European Commission, 2019).
* *The accountability should be explicitly acknowledged*: the processes should be assessed and audited; negative impacts should be assessed, reported and documented; whenever tensions arise between requirements, ethical decisions should be prior to, reasoned and properly documented; having accessible mechanism for adequate redress (Department of Industry, Science and Resources, 2019; European Commission, 2019).

In addition, some tech giants such as Google, Meta, IBM, Microsoft, Amazon and Apple have been articulated their AI ethics principles which covered whole or a part of the requirements mentioned so far, mostly including fairness, privacy and transparency.

## Conclusion

Alongside the development of AI applications, several potential risks must be addressed including unemployment, transparency, safety, privacy, fairness and accountability. Even though many AI ethics frameworks have been developed to mitigate these risks, they still need to be tested and refined in real-world scenarios to ensure they are effective and appropriate. Additionally, further research is essential to explore more foreseeable consequences of AI’s fast advancement as well as solutions for facing challenges.

**REFERENCE**

ACCC (2023, Jul 26). *$20m penalty for Meta companies for conduct liable to mislead consumers about use of their data*. ACCC. https://www.accc.gov.au/media-release/20m-penalty-for-meta-companies-for-conduct-liable-to-mislead-consumers-about-use-of-their-data

BBC News (2020, Sep 17). *Uber's self-driving operator charged over fatal crash*. BBC. https://www.bbc.com/news/technology-54175359

BBC News (2021, Apr 21). *TikTok sued for billions over use of children's data*. BBC. https://www.bbc.com/news/technology-56815480

Department of Industry, Science and Resources. (2019, November). *Australia’s AI Ethics Principles*. Retrieved from https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles

European Commission. (2019, April). *Ethics guidelines for trustworthy AI*. Retrieved from <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, *2*(9).

Sarangi, S., & Sharma, P. (2018). *Artificial intelligence: evolution, ethics and public policy*. Routledge India.