Ethics in AI Decision Making: Mechanisms And Variables

Patil Rohini Ashok[1] and Dr. Aprana Sachin Pande[2]

Research Scholar, Department of Department of Computer Science[1]

Research Guide, Department of Department of Computer Science[2]

Sunrise University, Alwar, Rajasthan, India

Abstract*: While artificial intelligence (AI) technology has the potential to benefit society and well-being, it also poses ethical dilemmas for decision-makers in areas such as skewed data, algorithmic discrimination, and unclear accountability. In this work, we use a qualitative research approach to identify ethical risk factors of AI decision making, apply rooting theory to construct a risk-factor model of ethical risks associated with AI decision making, and explore the ways in which risks interact through system dynamics, from which risk management strategies are proposed. Our study indicates that technological ambiguity, insufficient data, and administrative errors are the main sources of ethical hazards in AI decision making. Components of risk governance may be able to successfully restrict the social risks brought on by data, algorithm, and technical hazards. We provide strategies for handling ethical risks in AI decision-making from the perspectives of development, research, and management in light of this.*

Keywords: Bias and Discrimination, Lack of Transparency, Privacy Concerns, Issues, Job Displacement, Misuse.

I. INTRODUCTION

McCarthy initially put up the idea of artificial intelligence in 1956 to explain the sentient behavior of manufactured items. AI is now extensively employed in many aspects of life, including face and fingerprint identification and virtual reality interactions. It has significantly increased our productivity and enhanced our quality of life. Intelligent judgments based on huge data have also evolved with the development of AI. The most well-known example is Google's creation of the robot AlphaGo, which ultimately prevailed against the world's best professional human go player. Whereas human experience, emotional feelings, and "limited rationality" are the foundation of conventional decision-making processes, machine learning algorithms and underlying data are the basis for AI choices. This allows AI to make assessments about how things are evolving. Artificial intelligence is seen as a process that may improve human decision-making efficiency and is playing a more and bigger role in supporting human decision-making in contemporary life [1]. Artificial intelligence search algorithms, intelligent choices, and recommendations based on public browsing behavior are the source of much of the information, advertising, sound, and images that people access through their smartphones or personal computers. Even credit assessment tools rely on AI's intelligent decisions through big data and cloud computing.

The ethical hazards associated with AI decision-making include moral and ethical dilemmas pertaining to people and society that result from mistakes made by algorithms or data, and the development of AI must take these risks' detrimental impacts into consideration. AI frequently struggles to cope with complex decision-making scenarios because tacit knowledge, such as customs, emotions, and beliefs, is difficult to fully digitize and structure. Some examples of ethical risks associated with AI decision making include the decision to choose between the lives of pedestrians and drivers in an emergency, infringement of privacy rights of people involved in “human flesh searching” based on big data technology, and incorrect decisions made by “intelligent courts” that lack human feelings. The "moral dilemma" of ethical risk, however, is whether or not human decision-making will be surpassed or even replaced by intelligent decision-making in the age of powerful AI. Concerns over AI decision-making are growing since it is unclear whether AI will undermine human control or expose people to unpredictably high societal hazards.

DOI: 10.48175/568

Understanding and addressing the ethical, legal, and social implications of AI is one of the seven strategies included in the National Strategic Plan for AI Research and Development, which the US introduced in 2016 to control the direction of AI. Another measure was the creation of a new National Science and Technology Council (NSTC) Subcommittee on Machine Learning and Artificial Intelligence [2]. The European Commission formally unveiled the White Paper on Artificial Intelligence, A European Pathway to Excellence and Trust in Brussels in 2020. It stated that ethical oversight, sustainability, and human-centered development of AI are necessary to uphold people's fundamental rights and prevent the issue of risks arising from AI decisions [3]. The European Commission also released a draft Artificial Intelligence Act in 2021, with the goals of addressing AI's dangers, creating a reliable and cohesive EU AI market, and defending EU citizens' basic rights [4]. Japan and South Korea have developed pertinent papers for robots as early as 2007, suggesting that humans should be in charge of the machines, among other things [5]. Furthermore, the establishment of AI-focused ethics committees and data ethics institutes in the UK and Japan has helped to progressively highlight the ethical concerns surrounding AI [6].

China’s State Council said in the 2017 growth Plan for a New Generation of Artificial Intelligence that AI is undergoing fast growth and that, in order to guarantee its safe and healthy development, rigorous attention should be given to its risk problems [7]. In 2018, General Secretary Xi Jinping emphasized that preventing potential risks in AI development is essential to the technology's healthy development while presiding over a group study on the state and trends of AI development organized by the CPC Central Committee's Political Bureau. A new-generation AI governance professional committee was created by China's new-generation AI development plan in 2019. This committee is in charge of all things related to AI, including the creation of normative governance and research on ethical codes. The aforementioned papers on AI and committees show that decision-making using AI has drawn interest from all across the globe, and that research into the ethical issues raised by AI is essential to the advancement of both AI technology and humankind.

This study examines the ethical risks and dimensions associated with AI decision making. Using rooted theory and system dynamics, we analyze the mechanisms of action between the hazards. In order to guarantee the sound and long-term growth of AI, this study aims to provide references for the scientific avoidance, accurate reaction, and prompt resolution of the ethical hazards associated with AI decision making.

This is how the rest of the paper is structured. Section 2 reviews pertinent literature; Section 3 identifies and analyzes ethical risks associated with AI decision-making using rooting theory; Section 4 analyzes the mechanisms of action of these risks using system dynamics and conducts simulation experiments; and Section 5 offers the paper's discussion and conclusions.

II. LITERATURE REVIEWS

There is no standard book or theoretical framework pertaining to ethics, but it serves as a benchmark for assessing the interaction between humans and nature as a moral need and norm [8]. In order to promote co-development between intelligent technology, humans, and nature, the ethics of AI serve as a guide for technical advancement and recognized ethical principles that do not contradict with human interests [9]. Academics are becoming more aware of the hazards associated with technology as it advances. The one that has drawn the most attention from academics is the question of the ethical dangers connected to AI decision-making. Robots were the subject of the first ethical studies on AI decision-making [10], which raised concerns among academics about whether machine learning will eventually replace or exceed human decision-making abilities and take into account significant ethical hazards including the erosion of human dignity and existential crises. "Robots killing people" is inevitable due to the absence of human emotions, robots' incapacity to recognize emotions and make complicated judgments, and the inadequate rules and regulations pertaining to ethics.

In order to create the optimal plan, artificial intelligence decision-making is dependent on a restricted amount of data, programs, pertinent algorithms, and other input circumstances. But uncertainty is a part of technology itself. When combined with incomplete data, decisions devoid of human emotions are prone to error and can significantly change even human decisions, which can lead to ethical risks like invasions of privacy, endangering human life, and undermining social justice. These uncertainties are a significant source of ethical risks. In order to effectively prevent and protect against these risks and to enable intelligent decision making to develop in a strong direction, the study of the ethical risks of artificial intelligence decision making entails defining the ethical risks brought on by the uncertainty of technology and the uncertainty of human complex emotional decision making. Two main sources of risk are technology unpredictability and human limited rationality, which might pose ethical problems in AI decision making [11]. The biggest causes of technical danger, from a technological standpoint, are technological loss of control, misuse, and abuse [12]. Particular sources of ethical risk include intelligent algorithms, program design, and other technologies used throughout the AI decision-making process [13]. Humans are the primary source of risk creation from the standpoint of human limited rationality because programming and data importation samples in intelligent decision making involve human decisions [14]. The complex interactions between technology, humans, society, and nature are the source of ethical risks in AI decision making.

Studies on the Ethical Risk Management of AI-Powered Decision Making

Many researchers have proposed risk governance, primarily via top-down and bottom-up governance methods, in response to the ethical issues that AI may bring. In order to force robots to make judgments and behave in accordance with moral and ethical standards, the top-down method entails creating a framework with moral and ethical awareness. Kant's categorical imperative [17], the three laws of robotics [16], Amoff's moral calculus [15], and general moral philosophical content are a few examples. In terms of governance measures, risks in the decision-making process may be avoided by creating governance systems [20], matching ethical risk governance framework guidelines [19], and a list of principles for emerging technologies [18]. However, human emotions are complicated and impacted by a wide range of values, social norms, etc., which cannot be generalized by simple rules. In addition, all ethics and regulations contain flaws. Because of this, creating intelligent decision-making systems via a top-down methodology is very challenging. Under a bottom-up governance model, a computer repeatedly mimics human behavior and emotions in order to develop a system of moral judgments that is akin to human thought patterns. This process is known as machine learning. The most well-known example is autonomous driving technology; nevertheless, humans' incorrect understanding of the laws may instill undesirable habits in robots, increasing the potential of danger and even making decision-making more difficult. In terms of technological or moral ethics, neither bottom-up nor top-down methods to governance can imbue computers with human-like moral consciousness or cognitive abilities. It is especially crucial to strengthen the potential ethical review and legal implications of the AI decision-making process [22], as well as to govern the ethical risks of AI decision making [23]. Studies have shown that people are not opposed to the implementation of new technologies, and that people's fear of AI decision making is primarily based on a mistrust of government [21].

Artificial intelligence (AI) has generally advanced to a point where many facets of human existence have benefited from its intelligent decision-making [24], including social management [27], ecology [26], medicine [25], and ecology [24]. Fewer research have been able to compile the hazards and risk creation processes of AI ethical decision making, as well as look into the links between the risks. This is because, as a technology, AI's intelligent decision making inherently has related ethical concerns. In order to identify and categorize the risk aspects of AI ethical decision making, including risk sources and risk repercussions, we use a qualitative research technique of anchored theory in this work. In order to effectively support ethical decision-making and lessen the negative effects of AI ethics, we build a conceptual model and a feedback model of risk factors through system dynamics to explore the formation mechanism of AI ethical risks and analyze the causes of risks from multiple perspectives and in an all-around way.

III. RESEARCH METHODS

Evolutionary reasoning is a prevalent technique in qualitative research, which examines social phenomena via human and action-based analysis. 28]. One popular technique in qualitative research is rooting theory. According to Juliet (2015), P48–49, "Rooting theory" is a fact-based theory of induction and conceptualization of unstructured material based on data collecting and interviews. 29]. It is a bottom-up simulation research approach, and the theory that is produced must be standardized via an ongoing development and improvement process. Thus, rooting theory is broken down into seven parts, as seen in

The following are the seven steps: establishing the study question gathering and compiling data free coding spinning code theory creation, theory saturation testing, and selective coding. In roots theory, gathering data and level-by-level coding are the two most crucial processes. It is possible to generalize complicated data and create a comprehensive, standardized theoretical model by using the three-level coding procedure.

Data Collection and Collation

The field of artificial intelligence has seen an explosion in study since the notion was first introduced. Since roots theory is a qualitative research approach, it needs a lot of data to be validated. We adhered to the "everything is data" tenet in our research and went back to the source material. In order to browse and gather various information related to the research topic and to obtain secondary data, we used official websites, authoritative news media websites, Baidu, Zhihu, the China Knowledge Network, and relevant literature-reading websites in Chinese, in addition to Google, Yahoo, Twitter, and other websites. The transcripts that were retrieved include reports and viewpoints pertaining to AI ethics in addition to literary works.

The China National Knowledge Infrastructure was the primary consideration for choosing Chinese literature. The screening criteria that we used were the Chinese Science Citation Database and the Chinese Social Sciences Citation Index. The topic of "ethical risk of artificial intelligence" was used to generate 84 publications in total, whereas the subject of "ethical risk of artificial intelligence decision making" yielded one article. Furthermore, further items from People's Daily, Guangming Daily, and other publications were filtered out of Baidu's engine by using the term "ethical risks of AI decision making." Since there isn't much study on the subject, what is known about the ethical dangers of AI decision making has to be gleaned from publications about such issues. Elsevier's full-text journal database was used to pick English literature, and the issue "ethical risks of AI decision making" yielded 587 articles in 2022. This indicates that other nations are more concerned about the problems associated with AI decision making. Nonetheless, Flynn [30] discovered that there were only 4 to 49 publications on rooted theory; as a result, Flynn thought that a sample size of around 20 could ensure the theory's reasonableness.

NVivo is used in this work to compile the literature that has been screened. NVivo is a potent software program for qualitative analysis that can import and compile many forms of data. After being chosen at random, two thirds of the text data were uploaded into NVivo for extensive data mining and compilation. Since the article's content required analysis of the ethical risks associated with AI decision-making, we sorted through the literature using NVivo's word frequency analysis and manual coding functions to form the initial concept. The coding was then compiled using the coding classification method for spindle coding and selective

Open Coding

The process of arranging and condensing lengthy passages of gathered text into definitions in the form of ideas is known as "open coding." There are three processes involved in open coding. Labeling textual phrases is the initial stage in the process called tagging. The tagged ideas are further examined, condensed, and simplified in the second phase, conceptualization, where keywords are removed to create a tentative concept. Scoping is the third phase, when ideas are further developed and distilled into concepts at a more in-depth level. For instance, the original record entry "the introduction of discrimination or bias into the decision-making process by an algorithm for human reasons" served as the foundation for the idea of "human-caused discrimination." The original remarks are not included in this work due to space restrictions. In this work, NVivo was used to annotate the sentences. Then, via discussion and analysis, codes with identical or comparable semantic meanings were integrated to produce 126 basic ideas. There were 22 initial categories as a result of the combination of the initial concepts based on the expansion and analysis of each concept's meanings in the research context, as demonstrated by machine self-learning, discriminatory algorithm design, non-discriminatory algorithm design, data bias, prejudice, discrimination, and user equality.

Hacking, obedient data behavior, biased data omissions, unstable hardware, holes in data management, inadequate data security, voice and picture recognition, smart homes, adequate data, and misleading information

Theoretical Models and Selective Coding

The process of removing the essential categories from the major categories is known as selective coding. Through the core categories, the key categories are greatly reduced and connected to build a coherent narrative that culminates in a theoretical model.

Causal Construction

Plotting the causality and flow diagrams of the AI ethical decision risk system requires first determining the important variables in a complex system, which is a prerequisite for employing system dynamics for modeling and simulation. Based on the findings of the impact connection diagram and the rooting theory, the 26 variables in this research were plotted into two causality diagrams, which show the trend of risk change after risk governance and the shift in ethical risk causes in the ungoverned condition.
Analysis of Risk Subsystem Causality

System Flow Diagram

The fundamental structures of a system dynamics model may be reflected in causality diagrams and system feedback loops. Although they are qualitative evaluations of the system model, they are unable to reveal the characteristics of the system's variables or the quantitative connections among them. System flow diagrams were used to investigate and evaluate the linkages in more detail.

Model Assumptions and Equations

The study's variables were categorized as level, rate, auxiliary, and constant variables in accordance with the conceptual paradigm of ethical risk in AI decision making. The relevant data are simulated values since the emphasis of this work was on the evolutionary trend of risk and the state impact of risk under the condition of governance. Based on how much risk is discussed in pertinent literature, including Zhang Tao [34], Lo Piano [36], the Artificial Intelligence Development Report (2018–2019), and the Ethical and Moral Standards for Artificial Intelligence published by the Defense Innovation Board under the US Department of Defense, data values were assigned to the variables.

Simulation and Testing

One cannot compare experimental data with real data since there is a dearth of empirical study on the ethics of ethical risk in AI decision making. As a consequence, by repeatedly tweaking the formulas and sensitivity testing for the evolution of risk before and after governance, respectively, we were able to get more accurate findings. Utilizing the vensim PLE simulation tool, we ran simulated operations with the following parameters: timestep = 0.125, unit of time = month, beginning time = 0, and end time = 6. To acquire the modifications in the risk subsystem and the governance subsystem, we changed the parameter values of the ethical risk variables associated with artificial intelligence decision making. As an example, Figure 9 displays the risk development level before to governance, while Figure 10 displays the risk level after governance.

The findings of the simulation demonstrate that algorithmic risk and data risk eventually reach an unmanageable condition in the absence of government intervention, leading to a steadily rising and uncontrollable rate of societal risk. But after the governance condition was added, both algorithmic risk and data risk rate significantly decreased in the later stages as the degree of risk governance increased, even though there was not a greater effect in the early stages (likely because governance measures were not given high priority). As a result, there was less social risk overall, and the issue of ethical issues related to AI decision making was more effectively managed.

IV. CONCLUSION

As much as artificial intelligence (AI) brings ease, we also need to minimize any ethical hazards it may pose. In this research, we developed a conceptual model of risk factors and used a rooted theory method to identify and categorize the ethical risk aspects of AI decision making. Additionally, a system dynamics feedback model of risk variables was built to investigate the creation process of AI's ethical hazards. In order to effectively support ethical decision-making, lessen the unethical effects of AI, and ensure the long-term, responsible development of AI, we examined the causes of risk from a variety of angles. This analysis raised the bar for national science and technology development governance. The following are this paper's primary conclusions and revelations.

DOI: 10.48175/568

REFERENCES

[1]. Crompton, L. The decision-point-dilemma: Yet another problem of responsibility in human-AI interaction. J. Responsible Technol.

[2]. 2021, 7–8, 100013.

[3]. Yu, C.l.; Hu, W.L.; Liu, Y. US Releases New "The National Artifical Intelligence Research and Development Strategic Plan".

[4]. Secrecy Sci. Technol. 2019, 9, 35–37.

[5]. Wang, X.F. EU Releases "Artificial Intelligence White Paper: On Artificial Intelligence—A European approach to excellence and trust". Scitech China 2020, 6, 98–101.

[6]. Zhongguancun Institute of Internet Finance. Read More|The European Commission's Proposal for a 2021 Artificial Intelligence Act; Zhongguancun Institute of Internet Finance: Beijing, China, 2021.

[7]. Dayang.com—Guangzhou Daily. Korea to Develop World's First Code of Ethics for Robots; Guangzhou Daily: Guangzhou, China, 2007.

[8]. Sadie think tank. [Quick Comment], Foreign Countries Conduct Ethical and Moral Research on Artificial Intelligence at Multiple Levels. Available online: https://xueqiu.com/4162984112/135453621 (accessed on 20 August 2022).

[9]. State Council. Notice of the State Council on the Issuance of the Development Plan for a New Generation of Artificial Intelligence; State Council: Beijing, China, 20 July 2017.

[10]. Jiang, J. The main purpose and principles of Artificial Intelligence ethics under the perspective of risk. Inf. Commun. Technol. Policy 2019, 6, 13–16.

[11]. Susan, F. Ethics of Al: Benefits and risks of artificial intelligence systems. Interesting Engineering. Available online: https:// baslangicnoktasi.org/en/ethics-of-ai-benefits-and-risks-of-artificial-intelligence-systems/ (accessed on 20 August 2022).

[12]. Turing, A.M. Computing Machinery and Intelligence. In Parsing Turing Test; Springer: Dordrecht, The Netherlands, 2007; pp. 23–65.

[13]. Yan, K.R. Risk of Artificial Intelligence and its Avoidance Path. J. Shanghai Norm. Univ. Philos. Soc. Sci. Ed. 2018, 47, 40–47.

[14]. Chen, X.P. The Target, Tasks, and Implementation of Artificial Intelligence Ethics: Six Issues and the Rationale behind Them.

[15]. Philos. Res. 2020, 9, 79–87+107+129.

[16]. Marabelli, M.; Newell, N.; Handunge, V. The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges. J. Strateg. Inf. Syst. 2021, 30, 101683.

[17]. Arkin, R.C. Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture—Part 1: Motivation and Philosophy. In Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, Amsterdam, The Netherlands, 12–15 March 2008; pp. 121–128.

[18]. Zhao, Z.Y.; Xu, F.; Gao, F.; Li, F.; Hou, H.M.; Li, M.W. Understandings of the Ethical Risks of Artificial Intelligence. China Soft Sci. 2021, 6, 1–12.

[19]. Leibniz, G.W. Notes on Analysis: Past Master; Oxford University: Oxford, UK, 1984.

[20]. Anderson, S.L. Asimov's "three laws of robotics" and machine metaethics. Sci. Fict. Philos. Time Travel Superintelligence 2016, 22, 290–307.

[21]. Joachim, B.; Elisa, O. Towards a unified list of ethical principles for emerging technologies. An analysis of four European reports on molecular biotechnology and artificial intelligence. Sustain. Futures 2022, 4, 100086.

[22]. Bernd, W.; Wirtz, J.C.; Weyerer, I.K. Governance of artificial intelligence: A risk and guideline-based integrative framework. Gov. Inf. Q. 2022, 101685.

[23]. Bonnefon, J.F.; Shariff, A.; Rahwan, L. The social dilemma of autonomous vehicles. Science 2016, 352, 1573–1576.

[24]. Johann, C.B.; Kaneko, S. Is Society Ready for AI Ethical Decision Making? Lessons from a Study on Autonomous Cars. J. Behav. Exp. Econ. 2022, 98, 101881.

DOI: 10.48175/568

[25]. Cartolovni, A.; Tomicic, A.; Mosler, E.L. Ethical, legal, and social considerations of AI-based medical decision-support tools: A scoping review. Int. J. Med. Inf. 2022, 161, 104738.

[26]. Chen, L.; Wang, B.C.; Huang, S.H.; Zhang, J.Y.; Guo, R.; Lu, J.Q. Artificial Intelligence Ethics Guidelines and Governance System: Current Status and Strategic Suggestions. Sci. Technol. Manag. Res. 2021, 41, 193–200.

[27]. Weinmann, M.; Schneider, C.; vom Brocke, J. Digital Nudging. Bus. Inf. Syst. Eng. 2016, 58, 433–436.

[28]. Jian, G. Artificial Intelligence in Healthcare and Medicine: Promises, Ethical Challenges and Governance. Chin. Med. Sci. J. 2019,

[29]. 34, 76–83.

[30]. Stahl, B.C. Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence. Int. J. Inf. Manag. 2022, 62, 102441.

[31]. Galaz, V.; Centeno, M.A. Artificial intelligence, systemic risks, and sustainability. Technol. Soc. 2021, 67, 101741.

[32]. Catherine, M.; Gretchen, B.R. Designing Qualitative Research: Guidance throughout an Effective Research Program; Chongqing University Publisher: Chongqing, China, 2019.

[33]. Juliet, M.C.; Anselm, L.S. Procedures and Methods for the Formation of a Rooted Theory Based on Qualitative Research; Chongqing University Publisher: Chongqing, China, 2015.

[34]. Flynn, S.V.; Korcuska, J.S. Grounded theory research design: An investigation into practices and procedures. Couns. Outcome Res. Eval. 2018, 9, 102–116.

[35]. Li, X.; Su, D.Y. On the Ethical Risk Representation of Artificial Intelligence. J. Chang. Univ. Sci. Technol. Soc. Sci. 2020, 35, 13–17.

[36]. Tan, J.S.; Yang, J.W. The Ethical Risk of Artificial Intelligence and Its Cooperative Governance. Chin. Public Adm. 2019, 10, 46–47.

[37]. Zhang, Z.X.; Zhang, J.Y.; Tan, T.N. Analysis and countermeasures of ethical problems in artificial intelligence. Bull. Chin. Acad. Sci. 2021, 36, 1270–1277.

[38]. Zhang, T.; Ma, H. System Dynamics Research on the Influencing Factors of Data Security in Artificial Intelligence. Inf. Res. 2021, 3, 1–10.

[39]. Zhu, B.Z.; Tang, J.J.; Jiang, M.X.; Wang, P. Simulation and regulation of carbon market risk based on system dynamics. Syst. Eng. Theory Pract. 2022, 42, 1859–1872.

[40]. Lo Piano, S. Ethical principles in machine learning and artificial intelligence: A case from the field and possible ways forward.

[41]. Humanit. Soc. Sci. Commun. 2020, 7, 9.