

Artificial Intelligence and the Dilemma of Ethics

Fahad Nasser Al-Khaldi

Saudi Aramco, Information Technology, Saudi Arabia

DOI: <https://doi.org/10.5281/zenodo.6572997>

Published Date: 23-May-2022

Abstract: The article focused on the field of artificial intelligence including decision-making, and the concerns of ethical violations. It assessed the ethical concerns associated with various AI approaches and the steps that can be taken to address the challenge, identified that the major ethical concerns over AI adoption relate to individual privacy, unknown intentions of the AI developers (big corporations) and the possibility of inability to control future AI systems. The dilemma also exists especially in health care where ethics on non-disclosure of certain information seems to disagree with the need to disclose information that can be used for AI training for the benefit of the patients. Adherence to ethical standards during the development of AI is key in addressing ethical concerns. There is also a need for ethical governance and the need to ensure transparency while protecting the identifiable data.

Keywords: Artificial intelligence, ethics, ethical consideration, dilemma, privacy, trust, standards, artificial ethical agent, ethical reasoning, ethical artificial intelligence, artificial intelligence safety, machine ethics, and Robot Rights.

1. INTRODUCTION

Artificial Intelligence (AI) is a term that is used to refer to the computer systems that can execute tasks that ordinary requires human intellect such as learning and problem-solving instructions [1]. One of the AI systems is machine learning, which is capable of using data for image processing, independent recognition of patterns and predictions [2]. AI systems have become an essential part of mankind's daily life. Areas, where AI has been of immense importance to mankind, include in scheduling, the control process industrial, road safety and even for autonomous driving [3]. The use of AI has been associated with enhanced road safety [4], improved health care delivery [5], and improved efficiency and accuracy in various human activities [6].

Despite the importance of the AI to mankind, there have been fears over the misconceptions regarding the potential of the AI to be like or even exceed mankind [7]. There have been demonstrations against AI with some associating AI with decline attention to workers, reduced employment opportunities and over-dependence on machines [1]. The fast pace at which AI is outperforming humans in areas such as learning, emotion-sensing, and tacit judgment is causing concern over the potential negative effects on nations, workers, and society in general [1]. Understanding the concerns of the public over AI is important since such concerns inform the decision-making process that can hurt or enhance AI adoption and progress [8].

One of the major concerns over AI revolves around ethics [9]. In this article, the term ethics refers to the normative beliefs on how individuals should act toward one another. There are various dimensions of ethics with one being consequentialist ethics, which identify ethical acts as those that are executed upon the careful weighing of the consequences of the existing options and choosing one that has the most benefits and least harm [10]. Deontological ethics involves acting based on existing social norms while virtue ethics is where an individual act to be perceived favorably by others [10]. Major concerns have been raised concerning the morality of the AI systems that have superhuman abilities, which are feared for their existential risk to humans [1]. Ethical concerns over AI are also associated with the fact that the systems act the

guidelines, which are fed to them by algorithms written by humans [10]. Therefore, there always will be huge concerns regarding the integrity of those instruments, as they are subjected to manipulation and wrongness.

When dealing with AI, there are situations where the available choices lead to the infringement of the ethics but yet the situation demands that a choice has to be made, which results in an ethical dilemma [10]. This article is intended to study the relationship between AI and ethics and how they are influencing each other in one hand, and to explore the environmental factors, context, and best practices which can be implemented to ensure that AI instruments will not violate the code of ethics or engage in unethical conduct. Another aspect to shed the light on is the ethics involved in collecting the data used to train and create the decision model.

2. OBJECTIVE

AI systems are programmed to solve problems, assist in making decisions, and make predictions based on which decisions shall be made. However, giving the fact that these are programmable instruments, which are fed by humans, raises concerns about the abuse of ethical codes. This study is trying to help the IT manager by doing a literature review of the subject to understand the ethical risks that are associated with leveraging AI systems into the organization, assess in deciding what are the applications that can add value to the organization, and what are the limitations of these applications.

2.1. Theme 1: Relationship between AI and ethics:

This theme focused on how AI and ethics influence each other as presented in the existing literature. The most prominent subthemes that emerged from this theme include concerns over AI's ethics and the existing ethical dilemmas

2.1.1 Subtheme 1: Concerns over AI's ethics

[11] noted deliberate malpractice committed by system controllers such as the issue of black boxes, which are indicated to be inscrutable, therefore cannot be audited or controlled. The article also reported the influence of big companies (especially those in the US) on AI governance and ethics and possible conflict of interest such as the recent concerns over the misuse of AI by Cambridge Analytica and Facebook. [12] argued that AI mirrors human biases since they depend on instructions created by humans. Although AI systems can be good in addressing challenges in health care they can also lead to racial profiling as seen in other fields a possible bias that arises from the data used to train the AI machines. Another concern noted by [12] is the intention for design of the AI such as the AIs that they have been fundamentally created to cheat for example the Volkswagen case that enabled the cars to pass emission tests and Uber's software that allowed detection of law enforcers and therefore circumventing regulations. [13] also raised concerns over the intention of creators (individuals and large corporations). While [14] raised concerns over privacy, consent, and bias that affect the data used in training of the AI systems. [15] mentioned that AI technologies such as instant facial recognition could be used by governments in oppressing their opposition.

2.1.2 Subtheme 2: Ethical dilemma associated with the use of AI

The use of AI in healthcare causes ethical dilemmas to health care providers. [12] noted that for healthcare providers it is unethical to violate the Hippocratic ethics that calls for physicians to withhold certain information from the medical record, which means such data may not be available for machine learning. But such action also means that the patients whose information is withheld cannot benefit from machine-learning analyses and they become disadvantaged.

[16] also argued that AI facilitates better knowledge-gathering and can be used to make predictions concerning an individual's behavior and preferences. However, the AI can be used to illegally target individuals. When used in warfare, [17] noted that questions on how to create a target in autonomous war while adhering to noncombatant immunity need to be answered.

[18] highlighted the issue of proprietary when robots are producing and consuming art and literature. In other words, if expert systems use big data and machine learning to contribute to human knowledge; would the people who owned the data and copy rights (in some cases) be financially or otherwise rewarded?

[19] discussed the ethical issue in employing lethal autonomous weapons systems (LAWS) in warfare. Giving a robot the full autonomy to search and eliminate a "threat" is not something that is particularly against international humanitarian laws, although the criteria in these laws are subjective. Different countries are such as Japan and Germany showed

concern regarding developing such weapons, however, US, UK and Israel see no need for a ban on developing such weapons as they have internal compliance reviews in-place.

2.2. Theme 2: Approaches to enhance adherence to ethics by AI systems:

2.2.1 Subtheme 1: Standards

[20] Advocate for machine ethics where the AI is trained to adhere to set ethical standards. They argue that the solution to the ethical dilemmas faced with the use of AI lies in the development of systems that can use inductive programming techniques to identify the ethical principles that the machines need to adhere to. The machines that do not adhere to the ethical principles should not be tolerated. [11] also proposed a complementary adoption of relevant laws/standards and technology to address ethical concerns in AI. [21] noted that for the development of built-in functional morality to ensure ethically adequate behavior of the robots there is a need for ethical considerations during design and the need for policies and standards. [22] also noted that the solution to the issue of ethical issues associated with AI rest on the development of the required laws, standards, and use of adequate data in the development of ethical algorithms.

2.2.2 Subtheme 2: Depending on AI devices, and the over trust phenomenon

It also emerged from the review that the solution to the ethical concerns and dilemmas faced in AI lies in developing smart devices that are capable of monitoring and addressing ethical violations by AI. [23] argued that to determine and address ethical issues associated with AI, there is a need to depend on other AI devices that are capable of identifying ethical adherence and ensure no violation of ethical laws and standards by other AI systems.

Smart devices that are capable of addressing the violation of ethics by AI devices need to be based on better-trained algorithms. [24] noted that smart devices that can overcome the human level limitation and subjectivities are those that trained using data from a large number of humans. Some of the designs that can be used to ensure the development of smart devices capable of addressing AI's ethics include the Value-Sensitive Design, which facilitates the development of AI that behaves in a perceptibly ethical manner

In recent years, people have developed the misconception of robotic “overtrust”, where they either overestimate the effectiveness of AI instruments or ignore the chances that AIs might involve in any ethical violation

2.2.3 Subtheme 3: Governance

Ethical governance is important in ensuring adherence to ethical standards by AI systems. [25] noted that there is a need for a strong institutional framework and principled leadership. A code of ethical conduct should be published; value ethical governance. [26] also argued that government intervention is needed to ensure adherence to and enforcement of ethics in the design of AI systems. Ethical governance should start with the individual and institutions that create [25]. Government policies that ensure ethics is part of AI curriculum and collaboration between the policymakers, ethic committees, and ethical AI technologists should also be prioritized [10].

2.2.4 Algorithm Transparency and data anonymity

One of the important steps in addressing ethical concerns over AI is by enhancing transparency and protecting personal data. [11] advocates for increased explainability and interpretability of the algorithms, which are aimed at enhancing the algorithmic transparency and fairness. [11] also indicates that the incorporation of ethical auditing of the inputs and outputs of algorithms enables the identification of possible bias and harms. The dilemma on data sharing in health care such as for radiologists can be addressed by adopting robust anonymization where identifiable data is not provided for AI.

2.2.5 Subtheme 5: The case for an Ethical Black Box

One way of gaining the public trust over the AI instruments is to equip those machines with continuous data recording, similar to the black box in flights, which can be called an ethical black box. They claim that the ethical black box will help in monitoring and understanding why and how an AI instrument was involved in an ethical violation, or why a robot, in driverless cars, caused an accident.

2.3. Theme 3: Theme 3: Challenges in enhancing the AI's ethics:

This theme focuses on are various challenges faced in addressing AI's ethics. One of the challenges is the lack of proper definition of terms related to ethics such as fairness and transparency especially when applied to machines. Attempts to

define using machine language risk narrow definition for the otherwise complex terms [11]. Another area of concern is the fact that there are no standardized ethical codes; hence what is acceptable in one region, might not be normal in another one. Therefore, the AI instruments will always require human supervision which contrasts the reason why those tools were developed. One more challenge is the view that AI can outperform human's decision-making on what is moral, and the possibility of humans being enslaved by machines [27]. [28] also note the challenge in adherence to ethics by AI systems when empathy and care are involved.

3. CONCLUSION

Despite the highlighted concerns that reviewed studies have provided in-depth insights into AI's ethics. Major ethical concerns over AI adoption relate to individual privacy, unknown intentions of the AI developers (big corporations) and the possibility of inability to control future AI systems. The dilemma also exists especially in health care where ethics on non-disclosure of certain information seems to collide with the need to disclose information that can be used for AI training for the benefit of the patients. According to this review, addressing the ethical concerns that are associated with AI requires the adherence to ethical standards during the development of AI. There is also a need for ethical governance and the need to ensure transparency while protecting the identifiable data. However, the development of super-intelligent machines that are capable of determining what is ethical and addressing ethical concerns seems to escalate the ethical concerns over AI. [28] Therefore argues AI should never be given the mandate to decide what is ethical.

To make decisions on matters of ethical importance then a fair and robust system for translating the patchwork of human values into moral guidance must be developed. The intractability of our moral disputes and the magnitude of our moral errors can both be minimized by compromising with a decision framework which emphasizes the respective priorities of different moral systems in accordance with their plausibility to ensure proper conduct by AI instruments—people will need to employ other AI systems

REFERENCES

- [1] Wright, S. A., & Schultz, A. E. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons*, 61(6), 823-832.
- [2] King, B. F. (2018). Artificial intelligence and radiology: what will the future hold?. *Journal of the American College of Radiology*, 15(3), 501-503
- [3] Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, 105, 105-120.
- [4] Jiménez, F., Naranjo, J. E., Anaya, J. J., García, F., Ponz, A., & Armingol, J. M. (2016). Advanced driver assistance system for road environments to improve safety and efficiency. *Transportation research procedia*, 14, 2245-2254.
- [5] Weber, G. M. (2018). Using artificial intelligence in an intelligent way to improve efficiency of a heart failure care team. *Journal of cardiac failure*, 24(6), 363-364.
- [6] Jamal, A., & Syahputra, R. (2016). Heat Exchanger Control Based on Artificial Intelligence Approach. *International Journal of Applied Engineering Research (IJAER)*, 11(16), 9063-9069.
- [7] Eisen, L. (2019). Digital continuous care: Future of artificial intelligence-based healthcare. *Digital Medicine*, 5(2), 49.
- [8] Fast, E., & Horvitz, E. (2017, February). Long-term trends in the public perception of artificial intelligence. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- [9] Lugmayr, A., Tsang, J., Williams, T., Lim, C. X., Teo, Y. Y., & Farmer, M. (2018). Children Road Safety Training with Augmented Reality (AR). *International SERIES on Information Systems and Management in Creative eMedia (CreMedia)*, (2017/2), 40-42.
- [10] Yu, H., Shen, Z., Miao, C., Leung, C., Lesser, V. R., & Yang, Q. (2018). Building ethics into artificial intelligence. *arXiv preprint arXiv:1812.02953*. Retrieved from <https://arxiv.org/pdf/1812.02953.pdf>
- [11] Wright, S. A., & Schultz, A. E. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons*, 61(6), 823-832.

- [12] Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. *The New England journal of medicine*, 378(11), 981
- [13] Arruda, A. (2016). An ethical obligation to use artificial intelligence: An examination of the use of artificial intelligence in law and the model rules of professional responsibility. *American Journal of Trial Advocacy*, 40, 443.
- [14] Yuste, R., Goering, S., Bi, G., Carmena, J. M., Carter, A., Fins, J. J., ... & Kellmeyer, P. (2017). Four ethical priorities for neurotechnologies and AI. *Nature News*, 551(7679), 159.
- [15] Feldstein, S. (2019). The Road to Digital Unfreedom: How Artificial Intelligence is Reshaping Repression. *Journal of Democracy*, 30(1), 40-52.
- [16] Bryson, J., & Winfield, A. (2017). Standardizing ethical design for artificial intelligence and autonomous systems. *Computer*, 50(5), 116-119
- [17] Roff, H. M. (2014). The strategic robot problem: Lethal autonomous weapons in war. *Journal of Military Ethics*, 13(3), 211-227
- [18] Schafer, B., Komuves, D., Zatarain, J. M. N., & Diver, L. (2015). A fourth law of robotics? Copyright and the law and ethics of machine co-production. *Artificial Intelligence and Law*, 23(3), 217-240.
- [19] Russell, S., Hauert, S., Altman, R., & Veloso, M. (2015). Ethics of artificial intelligence. *Nature*, 521(7553), 415-416.
- [20] Anderson, M., & Anderson, S. L. (2018). GenEth: a general ethical dilemma analyzer. *Paladyn, Journal of Behavioral Robotics*, 9(1), 337-357.
- [21] Crnkovic, G. D., & Çürüklü, B. (2012). Robots: ethical by design. *Ethics and Information Technology*, 14(1), 61-71.
- [22] Hibbard, B. (2014). Ethical artificial intelligence. arXiv preprint arXiv:1411.1373.
- [23] Etzioni, A., & Etzioni, O. (2016). AI assisted ethics. *Ethics and Information Technology*, 18(2), 149-156.
- [24] Conitzer, V., Sinnott-Armstrong, W., Borg, J. S., Deng, Y., & Kramer, M. (2017, February). Moral decision making frameworks for artificial intelligence. In *Thirty-first aaai conference on artificial intelligence*. 4831-4835
- [25] Winfield, A. F. T., & Jirotko, M. (2017). The Case for an Ethical Black Box. *Towards Autonomous Robotic Systems Lecture Notes in Computer Science*, 262–273. doi: 10.1007/978-3-319-64107-2_21
- [26] Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harv. JL & Tech.*, 29, 353.
- [27] Davis, E. (2015). Ethical guidelines for a superintelligence. *Artificial Intelligence*, 220, 121-124.
- [28] Pavaloiu, A., & Kose, U. (2017). Ethical artificial intelligence-an open question. arXiv preprint arXiv:1706.03021.