8-2018

Ethical and moral issues with AI

Weiyu WANG

Keng SIAU
Singapore Management University, klsiau@smu.edu.sg

Ethical and Moral Issues with AI -- A Case Study on Healthcare Robots

Emergent Research Forum (ERF)

Weiyu Wang
Missouri University of Science and Technology
wwpmc@mst.edu

Keng Siau
Missouri University of Science and Technology
siauk@mst.edu

Abstract

AI-based technology has achieved many great things, such as facial recognition, medical diagnosis, and self-driving cars. AI promises enormous benefits for economic growth, social development, as well as human well-being and safety improvement. However, the low-level of explainability, data security, data privacy, and ethical problems of AI-based technology also pose significant risks for users, developers, and governments. As the AI advances, one critical issue is how to address the ethical and moral challenges associated with AI. This study will focus on the ethics and morality issues that may be caused by AI, and may arise because of AI. This research uses a qualitative approach and will conduct interviews with AI experts, programmers, workers, labor unions' representatives, legislators, and other stakeholders. The research focuses on two research questions: What are the perceived ethical and moral issues with AI, and how can these issues be solved or attenuated.

Keywords

Artificial Intelligence, Ethics, Morality, Ethical framework, Moral status

Introduction

Artificial Intelligence (AI) is an umbrella concept that is influenced by many disciplines, such as computer science, business, engineering, biology, psychology, mathematics, statistics, logic, philosophy, and linguistics. The complexity and capability of AI make it unique and controversial (Siau 2018). AI could be classified into weak AI and strong AI. Comparing to weak AI, which can only process specific tasks, researchers from different domains are collaborating to create strong AI (artificial general intelligence), which will be able to process multiple tasks with human-like intelligence. General AI is more controversial and caused a heated discussion because researchers are concerned that general AI will lead to superintelligence (Müller and Bostrom 2016), which could be loosely defined as "any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest" (Bostrom 2014, p.22). The conception is that the more advanced the AI is, the more risks AI will bring to humanity. For instance, AI may cause mass unemployment, make decisions that people cannot understand and control, lead to the wealth redistribution, and replace humans eventually (Siau and Wang, 2018).

Since the concept of "machine ethics" was proposed (Anderson and Anderson 2006), the ethical issues of the machine have just been discussed and debated. Comparing to the heated discussion and investment in AI technology, the consideration of AI ethics and morality is just at the budding stage. Some think that there is no rush to consider these problems since there is a long way for AI to go for it to be comparable to humans and have consciousness. But some researchers believe that ethics and morality issues must be considered early before the ethical and moral issues related to AI become importunate. Further, AI, combined with other smart technology such as robotics, is already spreading like wildfire in businesses, healthcare, and societies. For instance, IBM Watson has been used to help analyze cancer symptoms and make diagnoses. Amazon Go has realized cashier-free shopping.

Ethics is a complex, complicated, and convoluted concept. Even the definitions of ethics deserve a single paper to discuss. This paper does not aim to define its concept; instead, the objective is to review relevant literature, obtain a broader overview of what are the perceived ethical and moral issues related to AI, and

collect experts' opinions on how ethical and moral issues related to AI can be studied, analyzed, and addressed. Since AI has been applied in a wide range of fields, it is not possible to study ethical issues of AI under all situations. This work will focus on the use of intelligent robots in the healthcare field. On one hand, evidence shows that many people perceive robots to perform better than humans in some aspects of healthcare (Broadbent 2017). A survey shows that more than 80% participants accept healthcare robots for children with autism (Coeckellbergh et al. 2016). On the other hand, for healthcare directly related to the safety of human life, the potential threats of ethical issues have a more significant impact. For instance, who should be responsible for a failed surgery if human doctors and robots worked together.

Literature Review

Ethics

Ethics is a complex and comprehensive concept that research on the topic is usually focus on a single aspect. Table 1 shows some ethical frameworks studied by researchers from different domains.

Reference	Ethical Frameworks
Belmont 1979	1. Respect for subject: the right to decide whether to participate 2. Beneficence: do no harm to participants 3. Justice: fairly distribute costs and benefits of research
Mason 1986	PAPA issues — privacy, accuracy, property, and accessibility
Bentham 1996	Act utilitarianism: tally the consequences of each action first and then determine on a case by case basis whether an action is morally right or wrong Hedonistic utilitarianism: pleasure and pain are the only consequences that matter in determining whether the conduct is moral or not
Wallach 2014	Ethical principles – 1. Fairness: bias, fairness, and inclusion 2. Accountability 3. Transparency
Sinnott-Armstrong 2015	Consequentialism: engaging in action that causes more good than harm
Hursthouse and Pettigrove 2016	Virtue ethics: having ethical thoughts and ethical characters
Alexander and Moore 2016	Deontological ethics: conforming to rules, laws, and other statements of ethical duty (religious texts, industry codes of ethics, and laws)

Table 1: Examples of Ethical Frameworks

Ethical issues with AI

AI, at the present stage, is referred to Narrow AI or Weak AI. It can do well in a narrow and specific domain. The performance of narrow AI depends much on the training data and programming, which is closely related to big data and humans. The ethical issues of Narrow AI, thus, involve human's factors. "A different set of ethical issues arises when we contemplate the possibility that some future AI systems might be candidates for having moral status" (Bostrom and Yudkowsky 2014, p.5). They adopt the definition of moral status that "X has moral status = because X counts morally in its own right, it is permissible/impermissible to do things to it for its own sake." From this perspective, once AI has moral status, we should treat it not as a machine/system, but an object that has equal rights as humans.

Research about ethical issues of AI basically falls into three categories: features of AI that may give rise to ethical problems (Timmermans et al. 2010), human factors that cause ethical risks (Larson 2017), and the ways to educate AI system to be ethical (Allen et al. 2006; Anderson and Anderson 2007).

Features of AI that may give rise to ethical issues

Recent work has shown that AI is possible to "generate audio that sounds like speech to machine learning algorithms but not to humans" (Carlini and Wagner 2017). In this case, it is possible that AI could get access to personal information without the host's knowledge. If AI would be in charge of making a decision and utilize the "machine speech", then how can we control the outcomes? This kind of threat also exists in the

physical world (Kurakin et al. 2016) such as self-driving cars. AI, especially machine learning and deep learning, are not always transparent to inspection. Because of the black box that humans are not able to interpret, AI may evolve without human's monitoring and guidance. The low level of transparency also gives rise to the risks of malicious utilization.

Security and privacy are other challenges. The development of AI system relies heavily on the huge amount of data, including personal data and private data. Those data must be managed properly to prevent misuse and malicious use (Timmermans et al. 2010). To keep data safe, each action to the data should be detailed and recorded. Both the data per se and the action's record may cause privacy-related risks. It is, therefore, important to consider what should be recorded and who should take charge of the recording action, and who can have access to the data and records.

Human factors that may give rise to ethical issues

The most significant factor is human bias, such as the gender bias (Larson 2017) and race bias (Koolen and Cranenburgh 2017) that may be inherited by AI. Since AI system is still being trained by a human and using dataset made by a human, existing biases may be learned by AI systems and display in real applications. For instance, a software used to predict future criminals showed bias against a certain race (Bossmann 2016). This kind of bias comes from the training data that contains human biases. Thus, how to program and train AI systems without human biases are very important. Further, if AI gets its own sentience and sapience (Bostrom and Yudkowsky 2014), will it come up with its own biases?

Another concern is accountability. When an AI system fails at a certain assigned task, who should be responsible. This may lead to what is referred as "the problem of many hands" (Timmermans et al. 2010). When using an AI system, an undesirable consequence may be caused by the programming code, the entered data, the improper operation, or other factors. Who should be the responsible entity for the undesirable consequence, the programmer, the data owner, or the end users?

Ways to educate AI system to be ethical

Moor (2006) indicates three potential ways to transfer AI: to train AI into "implicit ethical agents", "explicit ethical agents", and "full ethical agents". Implicit ethical agents mean constraining the machine's actions to avoid unethical outcomes. Explicit ethical agents mean stating explicitly what action is allowed and what is forbidden. Full ethical agents mean machines, as humans, have consciousness, intentionality, and free will. An explicit ethical agent is currently getting the most attention and is considered to be more practical (Anderson and Anderson 2007).

Besides the above three categories, how to treat an AI system that has consciousness, moral sense, emotion, and feelings is another important consideration. For instance, is it ethical to "kill" (shut down) an AI system if it replaces human jobs or even endangers human lives? Is it ethical to deploy robots into a dangerous environment? These questions are also related to human ethics and moral values.

Theoretical Foundation

As machines, especially these intelligent machines such as home robots and healthcare robots, increase in capability and ubiquity, they will inevitably affect human lives not only physically but also ethically. At the same time, human-robot interactions will grow significantly (You and Robort 2017).

Whether the robots are regarded as moral agents affect the interactions (Sullins 2011). To be seen as real moral agents, robots have to meet three criteria: autonomy, intentionality, and responsibility (Sullins 2011). Autonomy means that machines are not under direct control of any other agent. Intentionality means that machines "act in a way that is morally harmful or beneficial and the actions are seemingly deliberate and calculated" (p.28). Responsibility means the machines fulfill some social role that carries with it some assumed responsibilities.

The notion of "having ethical status" can be separated into two associated aspects: ethical productivity, and ethical receptivity (Torrance 2011). Ethical producers are those who do or do not do their duties, such as saints and murderers. Ethical recipients are those who stand to benefit from or are harmed by the ethical

producers. From this perspective, AI and other smart machines can be both ethical producer and ethical recipients。

In the very classic trolley cases, the one who controls the trolley is the ethical producer (Allen et al. 2006). To continue to run on the current track and kill five workers or to turn to another track and kill a lone worker is a hard-ethical choice for humans. What choice would AI make? Who should be responsible for the AI's choice? The military robots that take charge in bomb disposal are ethical recipients. Is it ethical that human decide the destiny of these robots? Human ethics and morality today may not be seen as perfect by future civilizations (Bostrom and Yudkowsky 2014). One reason is that human cannot solve all the recognized ethical problems. The other reason is that human cannot recognize all the ethical problems.

"The ultimate goal of machine ethics is to create a machine that itself follows an ideal ethical principle or set of principles" (Anderson and Anderson 2007 p.15). It is theoretically easy but practically hard to formulate ethical principles for AI systems. For instance, if we program robots to always perform no harm, we should first make sure that the robots understand what is harm. This result in another problem -- what should be the ethical standards for harm? A global or universal level of ethics is needed. To put such ethics into machines, it is necessary to reduce the information asymmetries between AI programmers and ethical standards makers.

Research Questions and Procedure

As discussed earlier, AI could be an ethical producer or ethical recipient when it satisfies the three criteria indicated by Sullins (2011). Ethical and moral issues arise because of AI cannot be ignored. This research aims to study two research questions: What are the perceived ethical and moral issues with AI, and how can these issues be addressed. As a pioneering research in this area, we will conduct a case study on healthcare robots.

Since the research questions are more subjective, it is proper to utilize a qualitative approach to conduct the research (Yin 2016). An interview is an excellent way to gather insights and in-depth answers from interviewees. Also, the interview approach is more flexible and interviewers can ask follow-up questions according to the interviewee's answers to each question. Since ethical and moral issues with AI are new and complex topics and many people have different ideas about the topics, qualitative research provides the flexibility in gathering data and managing the research process, which may be lengthy and ambiguous. The target participants are physicians working with the intelligent robot, patients, healthcare robot experts and producers, and legislators. To bridge the information asymmetries among AI experts and those who do not understand AI well, AI experts and programmers will also be included. Snowball sampling will be used to find more interviewees. One-to-one interviews, as well as video interviews, will be conducted according to the location of interviewees.

To ensure the validity and reliability of research findings, a semi-structured interview will be conducted. The structured questions could guarantee the reliability of the interview, while the unstructured open-ended questions could increase the validity of the interview. Data collected will be categorized and stored.

Conclusions and Expected Contributions

Understanding and addressing ethical and moral issues related to AI is still in a very early stage. It is not a simple problem about "right or wrong", "good or bad", and "virtue and vice". It is not even a problem that can be solved by a small group of people. However, ethical and moral issues related to AI are critical and need to be discussed now. This research aims to call attention to the urgent need for various stakeholders to pay attention to the ethics and morality of AI systems. While attempting to formulate the ethical standards for AI and other advanced computing technologies, we will also understand human ethics better, improve the existing ethical principles, and improve our application of ethical principles and moral values in the AI age. Last but not least, this study will contribute to academic progress in the field by figuring out activities academia could do to help train programmers to build ethical AI and build AI ethically, as well as educate potential users of AI to treat artificial general intelligence ethically.

REFERENCES

Anderson, M., and Anderson, S., eds. 2006. "Special Issue on Machine Ethics," *IEEE Intelligent Systems* 21(4) (July/August).
Anderson, M., and Anderson, S. L. 2007. "Machine ethics: Creating an ethical intelligent agent," *AI Magazine*, 28(4), 15-26.
Alexander, L., and Moore, M. 2016. "Deontological ethics". *In Edward N. Zalta, editor*, The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, Winter 2016 edition.
Allen, C., Wallach, W., and Smit, I. 2006. "Why machine ethics?". *IEEE Intelligent Systems*, 21(4), 12-17.
Belmont. 1979. "The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research". Retrieved from https://videocast.nih.gov/ethical_principles_and_guidelines.pdf.
Bentham, J. 1996. "The collected works of Jeremy Bentham: An introduction to the principles of morals and legislation". *Clarendon Press*.
Bossmann, J. 2016. "Top 9 ethical issues in artificial intelligence. World Economic Forum," Retrieved from https://www.weforum.org/ethical-issues-in-AI.
Bostrom, N. 2014. "Superintelligence: Paths, dangers, strategies". *OUP Oxford*. Ch. 2.
Bostrom, N., and Yudkowsky, E. 2014. "The ethics of artificial intelligence," *The Cambridge handbook of artificial intelligence*, 316-334.
Broadbent, E. 2017. "Interactions with robots: The truths we reveal about ourselves," *Annual review of psychology*, 68, 627-652.
Carlini, N., and Wagner, D. 2017. "Towards evaluating the robustness of neural networks," *In Security and Privacy (SP)*, 2017 IEEE Symposium, pp. 39-57.
Coeckelbergh M, Pop C, Simut R, Peca A, Pintea S, et al. 2016. "A survey of expectations about the role of robots in robot-assisted therapy for children with ASD: ethical acceptability, trust, sociability, appearance, and attachment," *Sci. Eng*. Ethics 22:47–65
Hursthouse, R., and Pettigrove, G. 2016. "Virtue ethics". *In Edward N. Zalta, editor*, The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, Winter 2016 edition.
Koolen, C., and van Cranenburgh, A. 2017. "These are not the Stereotypes You are Looking For: Bias and Fairness in Authorial Gender Attribution". *In Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*. pp. 12-22.
Kurakin, A., Goodfellow, I., and Bengio, S. 2016. "Adversarial examples in the physical world". *arXiv preprint arXiv*:1607.02533.
Larson, B. N. 2017. "Gender as a variable in natural-language processing: Ethical considerations".
Mason, R. O. 1986. "Four ethical issues of the information age," *Mis Quarterly*, 5-12.
Moor, J.H. 2006. "The nature, importance, and difficulty of machine ethics," *IEEE intelligent systems,* 21(4), pp.18-21.
Müller, V. C., and Bostrom, N. 2016. "Future progress in artificial intelligence: A survey of expert opinion". *In Fundamental issues of artificial intelligenc*e, pp. 553-570, Springer International Publishing.
Siau, K. 2018. "Education in the Age of Artificial Intelligence: How will Technology Shape Learning?" *The Global Analyst,* Vol. 7, No. 3, pp. 22-24.
Siau, K., and Wang, W. 2018. "Building Trust in Artificial Intelligence, Machine Learning, and Robotics," *Cutter Business Technology Journal*, Vol. 31, No. 2, pp. 47-53.
Sinnott-Armstrong, W. 2015. "Consequentialism,*" In Edward N. Zalta, editor*, The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, Winter 2015 edition.
Sullins, J. P. 2011. "When is a robot a moral agent," *Machine ethics,* 151-160.
Timmermans, J., Stahl, B. C., Ikonen, V., and Bozdag, E. 2010. "The ethics of cloud computing: A conceptual review," *Cloud Computing Technology and Science*. IEEE Second International Conference. pp. 614-620.
Torrance, S. 2011. "Machine ethics and the idea of a more-than-human moral world," *Machine Ethics,* op. cit, 115-137.
Veruggio, G. 2011. "Roboethics roadmap". *In in EURON Roboethics Atelier*.
Wallach, H. 2014. "Big data, machine learning, and the social sciences: Fairness, accountability, and transparency," Retrieved from https://medium.com/big-data-machine-learning-and-the-social-sciences.
Yin, R. 2016. "Qualitative Research from Start to Finish". *The Guilford Press*. Second Edition.
You, S., Ye, T., and Robert, L. 2017. "Team Potency and Ethnic Diversity in Embodied Physical Action (EPA)" *Robot-Supported Dyadic Teams. AIS.*