

The moral problems brought by the development of artificial intelligence

Zhaokun Lin

Miami University, College of Engineering and Computing, Oxford, United State 45056

llinz46@miamioh.edu

Abstract. With the rapid development of Artificial intelligence (AI), the rapid development of AI also brings many new problems. Among them, the ethical and moral issues brought by AI have attracted people's attention. Due to the rapid renewal of artificial intelligence, many moral problems that have never been envisaged have also emerged. To avoid such problems and provide solutions for the future, this paper mainly analyzes the possible moral and ethical problems brought by artificial intelligence in the future, and adopted solutions and the results after adopting solutions. The research data are taken from published research, reports, government legislation, etc. This research believes that AI should comply with the current moral principles and be released after having solutions that can solve possible moral and ethical problems.

Keywords: artificial intelligence, big data, ethical and moral problems.

1. Introduction

With the rapid development of artificial intelligence technology today, artificial intelligence has brought about many ethical issues. For example, some people are trying to use face-swapping technology to create obscene video screens/pictures, and some people are trying to use artificial intelligence (such as ChatGPT) to do the work they are supposed to do. Because of these ethical issues that have occurred, people are starting to think about the ethical issues that AI may pose and possible solutions to these ethical issues, for example Massachusetts Institute of Technology (MIT) has and continues to conduct ethical machine experiments to capture people's opinions on how future self-driving AI will make judgments in the event of an emergency. Since most of the information learned so far is due to the fact that some people's interests have been infringed by AI. In order to avoid, as much as possible, the ethical problems brought about by AI in the future that will lead to the violation of some people's rights, this paper discusses the ethical problems brought about by AI that have already appeared through literature and news, and consider feasible solutions through the solutions to these problems and the problems that may arise from AI in the future. This paper also discusses the ethical issues and possible solutions brought about by strong and weak AI.

2. The relationship between artificial intelligence and ethics

2.1. Definition of artificial intelligence

The definition of artificial intelligence is not really complicated, just a simple definition of the words artificial and intelligent will give an answer. Artificial means made by humans rather than naturally occurring. Intelligence means having a human-like intelligence. After understanding these two words a simple definition of artificial intelligence emerges, that is, a machine with human-like intelligence made by human beings. Artificial intelligence is generally divided into strong artificial intelligence and weak artificial intelligence. The most common way to distinguish a strong AI from a weak AI is whether the AI passes the Turing test. Those who can pass the Turing test are strong AI, and those who cannot pass the Turing test are weak AI. The current artificial intelligence is all weak artificial intelligence, strong artificial intelligence has not yet appeared.

2.2. Development of artificial intelligence

When the concept of artificial intelligence first appeared in the last century, it gave the public a lot of imagination, but the long period of no progress made investors lose confidence. It was not until the emergence of Deep Blue that the situation of artificial intelligence was once “forgotten”. In the article “The Long Game”, it says this:

When Deep Blue was first created by International Business Machines Corporation (IBM) in 1989, the future of artificial intelligence was not promising. The reason for this is that back in 1950 artificial gave people too much hope. Mathematician Claude Shannon even thought that “within 10 or 15 years, something like a science fiction robot would appear in the laboratory”. But this did not happen, and investors who believed these words were disappointed with the phenomenon, which led investors to abandon continued investment in artificial intelligence [1].

This is unthinkable for those who have been exposed to AI now. Now, artificial intelligence is all over life, with software in cell phones, smart homes, and even restaurant ordering machines. Artificial intelligence is developing more and more rapidly, and artificial intelligence applications are constantly being updated and iteration.

3. Analysis of ethical issues arising from artificial intelligence

3.1. Ethical issues that have arisen

Artificial intelligence has been developed so far, it has been possible to achieve voice conversion, picture/video face replacement, picture recognition, AI answers and other functions. But while implementing these functions, ethical issues come along with them. The book What is AI mentions, poor training and not enough correctly labeled images of black people during training led Google to identify a black couple as “gorilla” [2]. This mistake could have been avoided, but it still happened. Since its launch, ChatGPT has received a lot of attention. The powerful conversation function and the huge knowledge base quickly answer various questions from users. Study. com surveyed 1,000 students and got an alarming statistic that “over 89% of students use ChatGPT to help with homework” [3]. ChatGPT was not originally conceived to help students complete assignments or cheat on exams. But the reality is that many people make unethical use of ChatGPT. Also in a survey of 1,000 students by Study.com, it was stated that “72% of college students believe ChatGPT should be banned from their college network” [3]. This also shows that most students know that what they are doing is not ethical, but there are still some students who use ChatGPT unethically.

3.2. Resolution of ethical issues

In the case of the photo recognition error, Google's attempted solution was to Google tried to fix this but failed, so they removed the “gorilla” tag from all images. Google also said it will continue to fix it to better recognize faces with darker skin. Journalist Charles Pulliam-Moore attributes the problem to a lack of diversity in some Silicon Valley companies [4]. ChatGPT takes a “no offense” approach by avoiding answers on topics such as politics, race, religion, and sexual orientation.

As noted above, these errors could have been avoided, and even testing the technology on people of color before it was introduced to the public would have likely identified them early enough to make the changes proactively, rather than being forced to make them after they were discovered.

Many university are now thinking about how to detect and punish students for using ChatGPT. Stanford University was the first to introduce DetectGPT to test ChatGPT writings and has a “95%” accuracy rate [5]. As noted above, these errors could have been avoided, and even testing the technology on people of color before it was introduced to the public would have likely identified them early enough to make the changes proactively, rather than being forced to make them after they were discovered. The University of Hong Kong sent an email to all staff in the school stating that the use of artificial intelligence tools including ChatGPT is prohibited unless the prior written consent of the teacher is obtained, otherwise it will be regarded as plagiarism [6].

3.3. Ethical guidelines for possible future solutions (for weak AI)

In order to address the ethical issues that weak AI may pose, the European Union (EU) report “Ethical Guidelines for Trustworthy AI” of April 8, 2019 sets out seven key requirements, including: Human agency and oversight, Technical Robustness and safety. Privacy and data governance, Transparency, Diversity, non-discrimination and fairness, Societal and environmental well-being and Accountability [7]. The following section discusses some of the most important of these. Figure 1 shows the acceptance and facts of some of the organizations listed in IBM’s research on the seven-point proposals put forward by the EU. The most accepted and implemented is transparency. The least accepted and implemented is social and environmental well-being. So why does this happen? With the development of Internet companies, people’s trust in Internet companies has also declined. The higher the transparency of Internet companies, the more trust they can receive. The benefits of social and environmental well-being to Internet companies are not obvious at present, so the acceptance and implementation are low. This also means that recommendations and implementation of AI ethics also need to consider the interests of the companies involved.

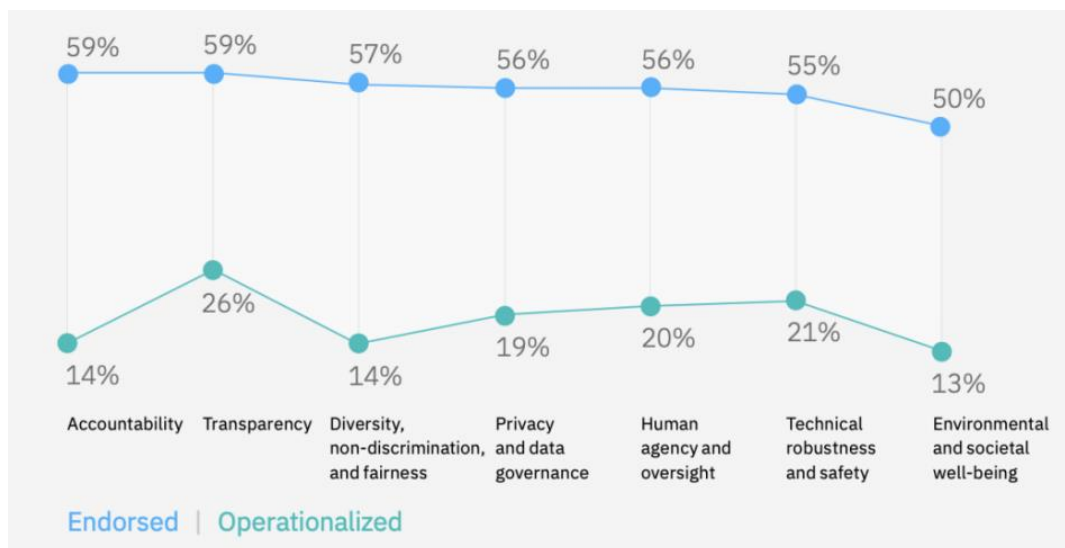


Figure 1. The recognition and implementation of the seven requirements of the EU by some organizations [8].

3.4. The EU’s seven key requirements are to prevent the ethical problems that weak AI could bring

The first is the Human agency and oversight. Artificial intelligence is defined above as “a machine with human-like intelligence created by humans”. Since it is made by humans, humans naturally have the right to supervise and manage artificial intelligence. The second of Asimov’s three Laws of robotics, published in a 1942 science fiction novel, also states that “a robot must obey human

commands”. This shows that humans recognized the importance of robots obeying human commands as early as the 1940s. So which organizations are the most effective in overseeing AI? There are three main regulatory bodies. The first is the developers of artificial intelligence itself. Developers should have regulatory responsibility for the AI they develop. The second is governments or organizations that perform governmental functions such as the European Union. Third, third-party institutions. Third-party monitoring is sometimes better able to detect and alert developers to problems that may have been overlooked.

The second is transparency and privacy. Developers should demonstrate how their AI protects user privacy and be open and transparent about how developers and their AI use user data. Such as Purpose limitation in the EU’s General Data Protection Regulation, developers are also supposed to demonstrate the purpose for which they are using users’ data for AI.

The third is diversity. As mentioned above, google misidentified black people as orangutans, which may be caused by the lack of diversity. Diversity is a very important thing in today's society, and Internet companies are paying more and more attention to diversity. The aforementioned Google is an Internet company, and its current emphasis on diversity can be seen in Google’s “2023 Diversity Report” (Figure 2). Greater diversity within the developer team means more consideration for minority interests within the developer.

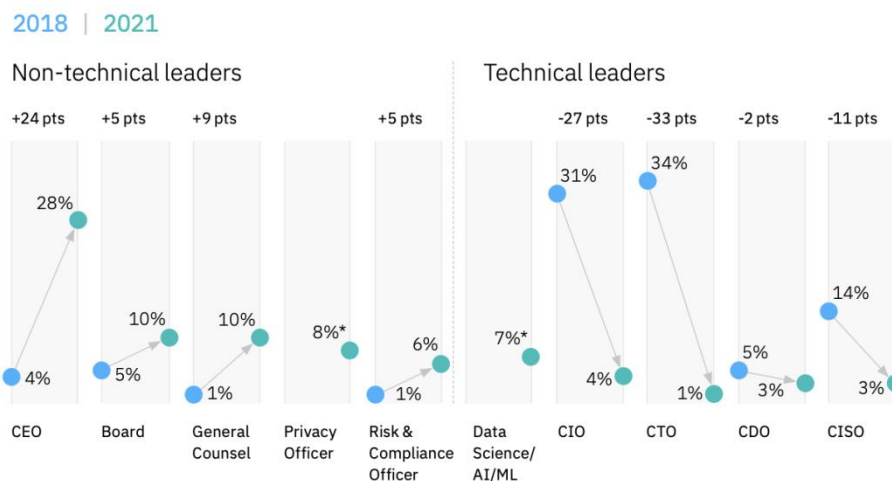


Figure 2. The racial percentage of Google employees in 2023 compared to 2022 [9].

Finally, accountability. Who has the main responsibility for AI? According to IBM, primary responsibility has shifted from technical staff to non-technical staff (As shown in Figure 3). Non-technical personnel need to supervise the development of artificial intelligence and determine whether it can be used, and have an inescapable responsibility. However, as the most important personnel to develop artificial intelligence, technical personnel should have greater responsibility for the artificial intelligence they develop. Non-technical personnel may not have a good understanding of the research and development of artificial intelligence and some hidden problems, but technical personnel are most familiar with the developed artificial intelligence and clearly know the problems that may occur in the developed artificial intelligence. So technicians should be mainly responsible for the problems of artificial intelligence.

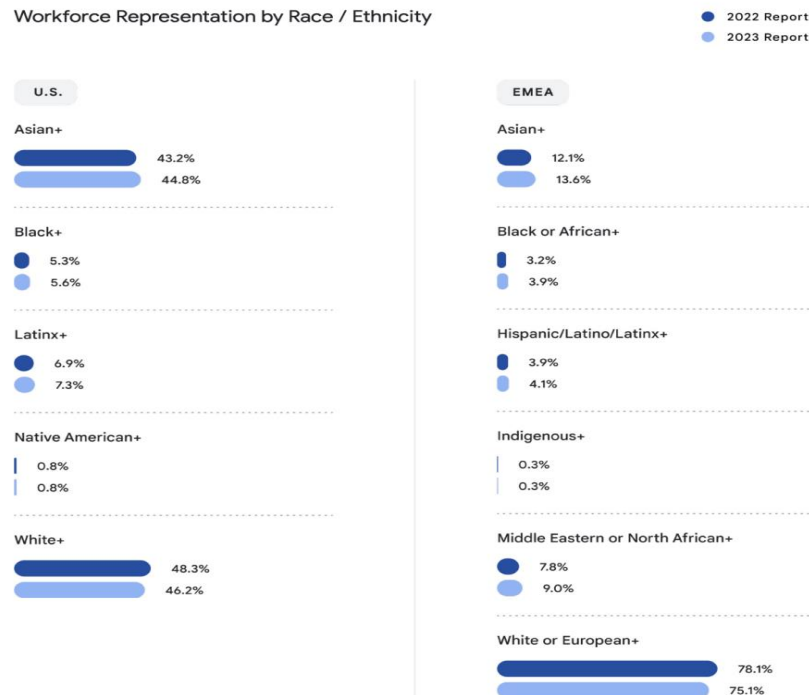


Figure 3. IBM studies the changing ethical responsibility of AI [10].

3.5. Possible problems and solutions of future strong artificial intelligence

3.5.1. The ethical problems that strong AI might bring. As mentioned above, strong artificial intelligence can make it difficult to tell the difference between humans and artificial intelligence, and even have the ability to think for themselves. So what might be the problems with strong AI? (1) Hack a computer (either a local computer or a remote computer). Artificial intelligence itself has a greater ability to learn than humans. At the same time, AI storage is located on remote computers online or offline, which has a better environment to invade computers than human hackers. (2) Assuming the strong artificial intelligence of the future has the ability to think and feel independently, is it willing to see humans enslave their own kind? (3) Do human beings have the same rights as human beings for strong artificial intelligence with emotion and independent thinking?

3.5.2. Referable solutions. The feasible solutions to the possible invasion of computers by strong artificial intelligence include: limiting the learning of code-related knowledge by artificial intelligence and restricting the connection of strong artificial intelligence to the Internet. Strong AI has very strong learning ability. Once strong AI is connected to the Internet, it is possible to learn to invade computer technology.

Limit the appearance of human-like personalities and emotions in AI, or distinguish between strong AI and weak AI when learning.

Analyze the major ethical discussions on whether giving strong AI the same rights as humans. The first is communitarianism. Can a strong AI identify itself as a member of the said community? Does AI have the same interests as its community? Firstly, artificial intelligence is a tool created by human beings. Human beings do not give artificial intelligence corresponding rewards when using artificial intelligence, so it is difficult for artificial intelligence to have the same interests with its community. Secondly, as for the second problem, it is difficult for AI to consider itself as a member of the said community. The second is utilitarianism, whether human endowed artificial intelligence can maximize human happiness. Maybe it is difficult. Human beings develop artificial intelligence by using artificial

intelligence as a tool. Even with emotion and personality, human being's greatest expectation for artificial intelligence is still to be used as a tool by human beings. So giving artificial intelligence the same rights as humans is not going to maximize human happiness. Thirdly, for egoism, if AI has rights equal to human beings, it means that some human rights will be shared with AI, and at the same time, there will be certain restrictions on human beings themselves, which is not in line with most human individuals' goal of using AI to help human beings. So it's not self-interest to give AI the same rights as humans. To sum up, humans should not give artificial intelligence the same rights as humans.

4. Conclusion

Since the emergence of the concept of artificial intelligence, it has developed a lot and experienced a long time. Human beings are not well prepared for the problems that artificial intelligence may bring. But with solutions now in place to the ethical issues raised by AI, it is imperative to prevent future problems. At present, it is still in the era of weak artificial intelligence, and the moral problems brought by artificial intelligence are mainly moral problems brought by human beings themselves. To prevent the moral problems that may be brought by artificial intelligence is to prevent the moral problems that may be brought by human beings through artificial intelligence. This can be seen in the EU's seven-point requirements for artificial intelligence. However, for strong artificial intelligence with similar intelligence or even stronger learning ability than human beings, the prevention of moral problems brought by strong artificial intelligence is more about technical and legal constraints on strong artificial intelligence.

References

- [1] THOMPSON, C. The long game. MIT Technology Review [s. l.], v. 125, no. 2, pp. 73-78, 2022. Available at: <https://search.ebscohost.com/login.aspx?direct=true&AuthType=cookie,ip&db=buh&AN=155160372&site=eds-live&scope=site>. accessed: 23 mar. 2023.
- [2] LOUKIDES, M. K.; LORICA, B. What is AI? First edition. [s. l.]: O'Reilly Media, 2018. <https://search.ebscohost.com/login.aspx?direct=true&AuthType=cookie,ip&db=cat00344a&AN=mucat.b5057467&site=eds-live&scope=site>. Acesso em: 25 mar. 2023.
- [3] Study.com, Productive Teaching Tool or Innovative Cheating? <https://study.com/resources/perceptions-of-chatgpt-in-schools>
- [4] Jessica Guynn, Google Photos labeled black people "gorillas", July 1, 2015, <https://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/>
- [5] Katharine Miller, Human Writer or AI? Scholars Build a Detection Tool, Feb 13, 2023, <https://hai.stanford.edu/news/human-writer-or-ai-scholars-build-detection-tool>
- [6] Tongxin Qian, University of Hong Kong announces ban! Should ChatGPT be Considered "Public Enemy of Teaching"?, Feb 19, 2023, <https://m.yicai.com/news/101679186.html>
- [7] European Commission, Ethics guidelines for trustworthy AI, 08 April 2019, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- [8] Brian Goehring, Francesca Rossi, Beth Rudden, AI ethics in action [J] IBM Report, April 2022, pp.6.
- [9] Diversity Annual Report - Google Diversity Equity & Inclusion. (n.d.). <https://about.google/belonging/diversity-annual-report/2023/>
- [10] Brian Goehring, Francesca Rossi, Beth Rudden, AI ethics in action [J] IBM Report, April 2022, pp.6.