Ethical Considerations Related to Artificial Intelligence

Tomas Cahlik
http://orcid.org/0000-0003-4243-9428
Charles University. Prague, Czech Republic

ABSTRACT

The purpose of this chapter is to present ethical considerations related to AI according to the three stages of AI development: Narrow AI, General AI, and Superintelligence. For narrow AI, its impact on creative practices is discussed. There are two types of goods linked with practices, external and internal. AI can make creative practices more efficient. On the other side, it can diminish the internal goods linked with a practice, and it can diminish the enjoyment linked with the creation. For General AI, the problem of its moral status is posed with possible threats of a General AI with a moral status. For Superintelligent AI, its understanding of justice and the threats linked with different understandings of justice are discussed.

INTRODUCTION

In this article, I basically structure ethical considerations related to the artificial intelligence (AI) according to the three stages of the AI development: the Narrow AI, the General AI (AGI) and the Superintelligence.

I set this article into the background based on the scientometric analysis of the scientific literature on the border between AI and ethics. I try to suggest some future research directions that could help in minimizing the threats and realizing the opportunities of AI.

BACKGROUND

Literature review in research articles is used for showing that the research described in the article fits into research themes that are interesting for contemporary research community. Literature review in ethical considerations is used differently, just for illustration of ideas that have been published in the area of interest and for giving the background, for "opening the scene".

Looking into the Web of Science Core Collection database on May 16, 2024 and using keywords "artificial intelligence" and "ethics" in the whole timespan from 1990 to 2024, there are 4739 documents

DOI: 10.4018/406018

obtained (from that 4051 articles or conference proceedings). Development in years is as it could be expected: 2024 (till May) – 369 documents, 2023 – 1227, 2022 – 943, 2021 – 689, 2020 – 447, 2019 – 308, 2018 - 199 and then basically steadily decreasing to just a couple of publications yearly in the 90-ies.

To summarize the content of this research activity, there is a visible shift from theoretical principles to specific ethical frameworks and ethical questions in practical applications. The theoretical principles revolve around the principles of human rights and dignity. Researchers are emphasizing the importance of developing AI technologies that respect and promote human rights, that do not infringe on privacy or lead to discrimination. They usually stress principles like transparency, accountability and the promotion of fairness in AI systems. Specific ethical frameworks tackle surveillance, the manipulation of behavior and the potential for AI to make decisions that traditionally would require human judgement. There is also an ongoing debate about the opacity of AI system - how their decision-making processes can be black boxes, which are difficult for even their developers to interpret or understand. Very much discussed practical application is healthcare, where ethical dilemmas abound regarding patient privacy, the accuracy of AI diagnoses and the potential for AI to influence or even replace human decision-making in clinical settings.

Using keywords ("general artificial intelligence" OR "human level artificial intelligence") and "ethics" in the whole timespan from 1990 to 2024, there are just 7 documents, all from the timespan 2018 - 2022. Relevant for this article are two of them:

The article "Human versus Artificial Intelligence," authored by J. E. Korteling and others (Korteling et al., 2021), explores the nuanced distinctions and connections between human intelligence and AI. The authors emphasize that while human intelligence is often considered the benchmark or "real" intelligence, this perspective may overlook the breadth and potential of AI. The article challenges the anthropocentric view that human-like capabilities are the gold standard for AI, suggesting instead that many forms of intelligence can exist and that AI does not necessarily need to mirror human cognitive abilities to be effective.

The article "Accompanying Technology Development in the Human Brain Project: From Foresight to Ethics Management" by Christine Aicardi and colleagues (Aicardi et al., 2018) discusses the integration of ethics into the technological advancements within the Human Brain Project. This research emphasizes the importance of managing both the existential and immediate risks associated with the development of AI and brain-inspired technologies. One of the key points of this article is the importance of foresight methodologies for anticipating future ethical issues and societal impacts.

Using keywords "superintelligence" and "ethics" in the whole timespan from 1990 to 2024, there are 29 documents, all from the last 10 years.

We can use the article "Superintelligence as a Cause or Cure for Risks of Astronomical Suffering" by Kaj Sotala and Lukas Gloor (Sotala & Gloor, 2017) as an example. It explores the dual potential of superintelligent AI to either cause or mitigate extreme suffering risks. The authors discuss how superintelligence could lead to scenarios with severe suffering risks but also highlight how it might prevent such outcomes. They emphasize the importance of aligning superintelligent AI's goals with ethical considerations to safeguard against risks while harnessing its potential benefits. The discussion includes strategies to align AI with human values to minimize risks.

The preceding summary reveals high and steadily increasing activity on the interdisciplinary border between AI and Ethics. Furthermore, it reveals that if we compare the current discussions on ethical problems linked to the General AI or the Superintelligence with the classical texts from Karel Capek, Isaac Asimov or Michal Crichton, we do not find any deeper insights now that these classics had.