The future of ethics in AI: challenges and opportunities

The future of ethics in AI: challenges and opportunities

Angelo Trotta* - Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

Marta Ziosi - Oxford Internet Institute, University of Oxford, Oxford, UK

Vincenzo Lomonaco - Department of Computer Science, University of Pisa, Pisa, Italy

Introduction

Artificial Intelligence (AI) systems that are able to learn and reason like humans are now widely used in a large number of real-world applications. Their performance has dramatically increased the degree to which we are able to address previously inaccessible tasks. However, the debate around such systems is now pivoting towards their non-functional properties and their impact on society as a whole.

In this Special Issue, we collected several works that would trigger the discussion about a more sustainable AI developmental framework that encompasses a set of principles at the crossroads of AI, Ethics, Philosophy, and Sociology.

As the range of AI capabilities expands, so does our awareness of the ethical issues related to the design, development, deployment, and use of AI systems or their application for the social good. The promise for positive change that AI represents has been challenged by several reports on ethically questionable uses of AI in contexts as varied as healthcare, education, law enforcement, recruitment, risk assessment, and more.

*Corresponding author: Angelo Trotta, angelo.trotta5@unibo.it

Key themes in AI for people: an analysis of ethics in AI

Artificial intelligence (AI) has the potential to revolutionize the way we live and work, from improving healthcare to advancing scientific research. However, as with any powerful technology, there are concerns about its impact on society and ethics. To address these concerns, the Special Issue focuses on the ethical and societal implications of AI. The published papers cover a wide range of topics. This summary will provide an overview of the topics covered in the journal, highlighting some of the key insights and debates in the field.

Ethics in AI

This Special Issue addresses various ethical concerns related to the use of AI, such as psychological targeting, empathetic AI, cultural theories, fairness and discrimination, and accountability. Several papers face ethical concerns related to the development and deployment of artificial intelligence. For instance, the consideration of the ethical implications of AI for the entire world and beyond. Here, AI needs to be developed in a way that is transparent, controllable and aligned with human values.

Similarly, there is a discussion on how AI can be designed to act ethically in everyday situations, particularly those involving empathetic interactions. The papers argued that AI systems should be designed to understand human emotions and respond appropriately. The works highlight the need for transparency and accountability in the design of AI systems in the workplace.

AI and society

The Special Issue also looks at the societal implications of AI, such as the perception of AI, ageism, and the use of AI for societal interest. It features papers that consider the impact of AI on society examining the role of AI in the geopolitical landscape and argues that global cooperation is necessary to ensure that AI benefits humanity rather than causing harm.

Furthermore, it considers how AI is changing the nature of work and argues that workers should be empowered to adapt to new technologies. The papers highlight the need for AI to be designed with the goal of enhancing human capabilities rather than replacing human workers. There is a discussion on the importance of public interest theory in shaping the development and deployment of AI. The papers argue that AI should be developed in the public interest, with a focus on promoting social welfare and minimizing harm.

Explainable AI

Several papers focus on explainable AI, which is the idea that AI systems should be transparent and understandable to humans. Here, the reader can find works about the explainability of AI systems is essential for building trust between humans and machines and the importance of timing and context in providing explanations for AI decisions. Moreover, studies propose an "explanation

space" that would allow humans to interact with AI systems and provide feedback on their decisions.

AI and healthcare

Several papers consider the role of AI in healthcare and address the ethical concerns related to bias and discrimination in healthcare AI. The papers propose a new principle of fairness that takes into account the social context in which the algorithm is used and where AI can be designed to provide individuals with the information they need to make informed decisions about their health.

Discussion and Future Work

As AI continues to evolve and become increasingly integrated into our daily lives, it is crucial that we continue to examine its ethical and societal implications. The papers published in "AI for People" represent important contributions to this ongoing conversation. However, there is much work to be done to address the complex and multifaceted issues that arise from the development and deployment of AI. Future research will need to grapple with questions of fairness, transparency, accountability, and the potential unintended consequences of AI.

Here are some potential topics for future discussions:

- **AI for People and the World:** Discusses the expansion of AI to encompass not only people, but the world and the universe as well.
- **AI in the Workplace:** Examines the impact of AI on the workplace and its potential for human adaptation, as well as the application of ethics to workplace AI.
- **Trustworthy AI:** Explores the importance of trustworthy programming for autonomous concurrent systems, as well as the role of explainability in supporting trust.
- **AI for Social Good:** Considers the use of AI for social good, aligning academic journal ratings with the United Nations Sustainable Development Goals (SDGs), and investing in AI for social good.
- **AI and the Law:** Examines the legal implications of AI, including the legal status of intelligent service robots and the use of AI in criminal justice.
- **AI and Culture:** Explores the intersection of AI and culture, such as cultural robotics, the cultural history of consciousness, and posthuman perception of AI in science fiction.
- **AI and Bias:** Investigates biases in AI, such as implicit biases and stereotypes, age discrimination and exclusion, and bias and discrimination in job advertisements.
- **AI and Technology:** Addresses various technological issues related to AI, such as developing a practical ethical methodology for integrating AI into the industry, algorithmic fairness, and the importance of transparency in AI operations.

As we continue to explore the possibilities of this technology, it is essential that we do so with a clear understanding of its limitations and ethical implications. By building on the work done so far, we can work towards a future in which AI is developed and deployed in a way that benefits all members of society.

The discussions on AI and ethics presented in this special issue demonstrate the complex interplay of different disciplines in creating a more sustainable AI development framework. Such a framework should be able to take into account moral and ethical considerations as well as other nonfunctional properties and social implications of AI technology. We have seen that addressing the ethical implications of AI goes beyond the traditional technical fields of AI and robotics and takes into consideration a much broader context of disciplines, such as ethics and philosophy, sociology and economics.

From the perspective of AI engineers and developers, it is important to embrace and practice a responsible approach to the development of AI technologies. This implies, for example, that developers should consider not only the technical aspects of their systems, but also the ethical implications of their use. In addition, developers should think about the various delegated responsibilities, such as data collection, privacy, and intellectual property, when developing or using AI technologies.

Discussion of the ethical implications of AI should not be limited to individual technology development. It should also consider the broader context of government policies and civil society initiatives. As such, it implies the need for a common framework of principles and standards that can ensure accountability, fairness and transparency in the use of AI. In addition, governments should create policies and guidelines to ensure that AI is used responsibly while enabling innovation and progress.

In conclusion, ethical considerations are a critical element in the development and use of AI technologies and must be taken into account by all stakeholders. The multidisciplinary approach of this special issue makes an important contribution to the debate by bringing different perspectives and disciplines into the conversation. We hope this will raise awareness and provide a framework for more ethical and responsible use of AI in the future.