WRITEN BY

JURAJ PODROUŽEK
Lead and Researcher
Ethics and Human Values in Technology
Kempelen Institute of Intelligent Technologies

ADRIÁN GAVORNÍK
Research Intern
Ethics and Human Values in Technology
Kempelen Institute of Intelligent Technologies

THE PLACE FOR ETHICS IN AI

AI has become an indispensable part of digital innovation across industries. As AI adoption progresses rapidly, it is important to consider AI as not purely technical but also as sociotechnical systems. Paying attention to ethical and societal issues can ensure that AI systems will be developed and implemented in a way that benefits society and avoids possible harm.

FROM PRINCIPLES TO PRACTICES

Considering AI regulation, Europe has the ambition to become the global leader in establishing harmonized rules and common standards for AI systems. In 2021 the European Commission proposed The AI Act – the first horizontal regulation of AI systems. The AI Act adopts the risk-based approach and categorizes AI systems into categories based on their impacts on health, safety, and fundamental rights. AI applications posing an unacceptable risk, such as government-imposed credit scoring algorithms should be banned. High-risk AI applications such as algorithms used in education, public services, or safety components of vehicles are subject to specific legal requirements. Lastly, AI applications not explicitly banned or listed as high-risk are largely left unregulated.

However, there are still a couple of years of negotiation and work ahead of us until the AI Act comes into force. So far, the AI Act aims to outline the high-level picture of the balance between protecting the health, safety, and fundamental rights of individuals while reaping the benefits of innovation. At the same time, the accompanying technical standards are currently being drafted.

Meanwhile, there is a plethora of guidelines and frameworks for building ethical and trustworthy AI systems as a form of soft regulation. These guidelines were developed not only by experts from academia or industry but also by international organizations such as OECD or UNESCO. Unfortunately, it is not always clear how the ethical principles and values from guidelines like transparency, autonomy, or diversity should be put into practice. Therefore, many companies often struggle to implement ethical guidelines and frameworks effectively, which can lead to unintended harm to people and unwanted ethical dilemmas for their employees.

Fortunately, in recent years there has been a shift in AI ethics from generating guidelines consisting only of high-minded moral principles to the question of how to operationalize these principles into practice. Organizations and their AI teams who are willing to put some effort into building ethical and trustworthy AI systems can now apply specific tools to test their systems regarding specific areas, such as explainability or fairness. There are also some general assessment lists available to assess AI systems according to the requirements of ethical and trustworthy AI and to provide recommendations on how to meet high ethical standards.

However, we have to keep in mind that thorough application of such tools and assessment lists might require cooperation with relevant experts from the field of AI ethics. Sometimes, participation of various stakeholders and affected groups in the assessment process is also necessary.

> “There are still a couple of years of negotiation and work ahead of us until the AI Act comes into force.

Underestimating the need for proper expertise on trustworthy AI within organizations can lead to considerable risks of superficial adoption of ethical principles into practice. Most often we are witnessing “ethics washing” - the practice of pretending to be ethical through the implementation of superficial measures. Another problematic practice is ethics shopping which includes mixing and cherry-picking the ethical principles, guidelines, or other tools, in order to advocate some pre-existing behaviors, and hence justify them a posteriori, instead of implementing or improving new behaviors.

WHY SHOULD WE PAY ATTENTION TO AI ETHICS?

If the adoption of ethical principles in the development and deployment of AI systems is not an easy and straightforward task, why should companies put an effort into their operationalization? There is no single answer to that question. Firstly, it can help to mitigate the risk of unintended consequences or ethical dilemmas arising from the development and use of AI systems. Aligning AI systems with generally accepted principles can help to ensure that the harm done to individuals or groups will be minimized. This, in turn, can help to build the trust and confidence of society in AI systems, which is essential for their widespread adoption.

Secondly, considering ethics during development phases can help to improve the performance and accuracy of AI systems. Ethical considerations such as fairness and non-discrimination can help to ensure that AI systems are developed in a manner that is inclusive and does not perpetuate biases. This in advance can help to provide satisfactory outcomes for everyone, not only some preselected or prominent groups. For example, researchers have demonstrated that fairness-aware algorithms used in loan approval systems have reduced disparity among applicants with various sensitive attributes such as race or gender, while maintaining equality of opportunity among all applicants.

Finally, taking AI ethics seriously can help companies to enhance their reputation and build brand value. Ethical and societal considerations are becoming increasingly important to consumers, who are looking for companies that are socially responsible and committed to ethical practices. By demonstrating a true commitment to ethical AI development and use, companies can differentiate themselves in the market and build long-term customer loyalty.

DELIVERING ETHICAL AI

By understanding the benefits of utilizing AI ethics tools and frameworks, an increasing number of companies and AI teams have decided to assess their development and use of AI systems from an ethical standpoint. By proactive mitigation of the ethical and societal risks, they want to ensure that their AI systems are developed and used in a manner that benefits society and does not cause unwanted harm. At the same time, they will also improve the performance and accuracy of their AI systems and strengthen the company’s reputation and brand value.