A systematic exercise tool helps researchers ponder the ethical implications of AI

Leikas, Jaana

2022-10

A systematic exercise tool helps researchers ponder the ethical implications of AI

by Jaana Leikas (VTT Technical Research Centre of Finland Ltd) and Patrik Floréen (University of Helsinki, Finnish Centre for Artificial Intelligence)

In addition to many clear benefits, AI technology can also bring risks related to the misuse of technology and for example increased inequalities. There is thus an increased demand to apply the principles of human-centricity and ethics in the development of AI. Members of the Ethics Advisory Board of the Finnish Centre for Artificial Intelligence (FCAI) have developed an Ethics Exercise Tool to help AI researchers consider and discuss the ethical implications of their research and strengthen ethical understanding within their team.

Although AI is a generic technology, ethics of AI can be understood as a contextual phenomenon. Understanding the context means that the dynamic effects of AI should also be understood. For example, contextual relevance is indicated by the fact that biased data can be almost irrelevant in one context (if it is not used for any activity) and extremely relevant in another.

Problems related to social contexts and human behaviour are central in AI, placing emphasis on value questions such as the status of citizens, democracy, and fairness. These issues are often multidimensional, even ambiguous, and require an analytical approach and new tools to operationalise and integrate them into the context of AI.

Many public, private and non-governmental organizations and expert groups have provided views and guidelines for the design of ethically and socially acceptable AI. For example, The European Commission's independent, high-level expert group on artificial intelligence (AI HLEG) states in Ethics Guidelines for Trustworthy AI [L1] as the ethical purpose of the document to "ensure respect for fundamental rights, principles and values when developing, deploying and using AI". One of the most recent one, the UNESCO Recommendation [L2], starts from the premise that there are ethical issues at every stage of the life cycle of AI systems. This holistic view of the ethical development of AI is reflected in the fact that the ethical consideration is embedded in the practical activities of people by including - not only AI systems and their development - but also the issues related to their use. The Recommendation refers to all actors involved in at least one stage of the AI system life cycle as "AI actors". This includes both natural and legal persons, such as researchers, programmers, engineers, data scientists, and end-users, as well as businesses, universities, public authorities, and private entities.

Proactive assessment of the impact of AI is necessary when discussing ethics. This requires insight into both short-term and long-term challenges. For this purpose, the Finnish Centre for Artificial

Intelligence (FCAI) has launched an Ethics Exercise Tool [L3] to help AI researchers identify, explicate, and generally work with ethical issues related to their work as well as understand the ethical and social implications of their research. The tool developed by FCAI ethics experts follows the principles of responsible research and innovation. The tool is a frame for systematic thinking of futures that are and are not desired. It helps in avoiding pitfalls and in guiding research toward the greater common good. The perspective from which it is examined varies from individual projects to entire research programmes, and it is to be used iteratively in the planning, implementation, and communication of research. It comprises various thematic questions that researchers should consider in their teams in the different phases of a project.

The first, general goal of the tool is to link the results of the research to a broader societal context through ethical consideration. The second immediate goal of the tool is to help researchers identify and resolve ethical and societal problems related to their own work; consideration of the effects is also required by an increasing number of publication forums and funding instruments of research.

Stimulating debate is important

The tool is based on the assumption that ethical consideration cannot be outsourced, because of its dependence on context: answering ethical questions about the problem in question is the responsibility of researchers working with the problem. Therefore, researchers need to awaken to ethics and try to sufficiently understand the field from which ethical questions emerge. A basic understanding of ethics is starting to be essential also for students of artificial intelligence.
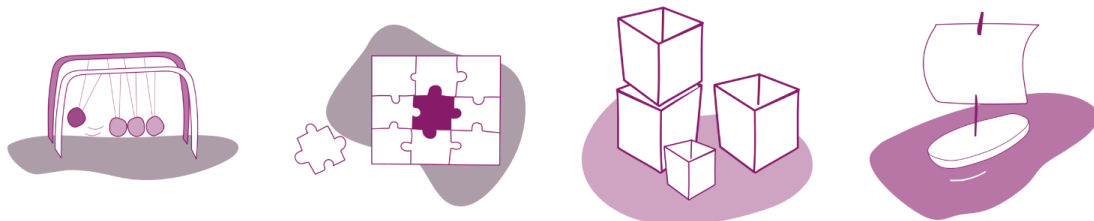

Figure 1: The FCAI Ethics Exercise Tool includes reflections on anticipation, reflexivity, inclusion, and responsiveness.

The idea of the FCAI Ethics Exercise Tool is based on four complementary points of view. *Anticipation*, involving careful consideration of the sought-after and accidental consequences of the research and innovation activities. *Reflexivity* of one's own actions, in which possible assumptions and commitments are considered, which might affect the research work. *Inclusion*, in which the relevant stakeholders are invited in dialogue on the desired and unwanted consequences of the research, and predisposition and *responsiveness*: seeking the right direction for the research in accordance with the visions and values that have emerged through the processes that have been highlighted.

Using the tool may aid in fostering ethics at critical phases in the project life cycle. It should be used iteratively, where ethical issues are systematically discussed before initiating and after the project. Reflection on ethics should thus include reflection on the ethical values and choices in respect to design decisions, as well as the possible impacts of research outcomes on a societal level.

Like understanding the ethics of AI, developing an ethics exercise tool also requires a continuous dialogue between researchers. FCAI therefore wants to iteratively develop and improve the tool and thus continuously seeks feedback from researchers on its use.

Links:

[L1]: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[L2]: https://en.unesco.org/artificial-intelligence/ethics

[L3]: https://fcai.fi/ethics-exercise

Reference:

[1]: Sigfrids, A., Nieminen, M., Leikas, J., & Pikkuaho, P. (2022). How should public administrations support ethical and sustainable development and use of Artificial Intelligence? A systematic review of proposals for developing governance of AI. Frontiers in Artificial Intelligence. https://doi.org/10.3389/fhumd.2022.858108

Please contact:

Jaana Leikas

VTT Technical Research Centre of Finland Ltd., Finland

jaana.leikas@vtt.fi