Guidance on the ethical use of Artificial Intelligence (AI) in research

DCU Research Ethics Committee

August 2024

As the new EU AI act came into force on 1st August 2024, the DCU REC aims to give some guidance to DCU researchers on the ethical use of AI in research. This guidance will be temporary for two reasons: 1) AI is developing fast and our understanding of its potential benefits and harms will change over time, and 2) legal obligations may change over time in line with the act. In the meantime the REC offers guidance on a theoretical and a practical level, based on useful work done by others (see list of sources), as detailed below.

As to the theoretical level, the REC endorses the ethics guidelines for trustworthy AI, as advanced by the Independent High Level Expert Group on AI. Their thinking about trustworthy AI rests on "… four ethical principles, rooted in fundamental rights, which must be respected in order to ensure that AI systems are developed, deployed and used in a trustworthy manner." (EC 2019, 11) These are the principle of *respect for human autonomy*, the principle of *prevention of harm*, the principle of *fairness*, and the principle of *explicability* (EC 2019, 12-13). Based on these principles and following the approach taken by this expert group, trustworthy AI then has the following seven requirements:

1. Human agency and oversight
   "AI systems should support human autonomy and decision-making, as prescribed by the principle of *respect for human autonomy*. This requires that AI systems should both act as enablers to a democratic, flourishing and equitable society by supporting the user's agency and foster fundamental rights, and allow for human oversight." (EC 2019, 15)

2. Technical robustness and safety
   "A crucial component of achieving Trustworthy AI is technical robustness, which is closely linked to the *principle of prevention of harm*. Technical robustness requires that AI systems be developed with a preventative approach to risks and in a manner such that they reliably behave as intended while minimising unintentional and unexpected harm, and preventing unacceptable harm. This should also apply to potential changes in their operating environment or the presence of other agents (human and artificial) that may interact with the system in an adversarial manner. In addition, the physical and mental integrity of humans should be ensured." (EC 2019, 16)

3. Privacy and data governance

   "Closely linked to the *principle of prevention of harm* is privacy, a fundamental right particularly affected by AI systems. Prevention of harm to privacy also necessitates adequate data governance that covers the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy." (EC 2019, 17)

4. Transparency

   "This requirement is closely linked with the *principle of explicability* and encompasses transparency of elements relevant to an AI system: the data, the system and the business models." (EC 2019, 18)

5. Diversity, non-discrimination and fairness

   "In order to achieve Trustworthy AI, we must enable inclusion and diversity throughout the entire AI system's life cycle. Besides the consideration and involvement of all affected stakeholders throughout the process, this also entails ensuring equal access through inclusive design processes as well as equal treatment. This requirement is closely linked with *the principle of fairness*." (EC 2019, 18)

6. Societal and environmental well-being

   "In line with the *principles of fairness* and *prevention of harm*, the broader society, other sentient beings and the environment should be also considered as stakeholders throughout the AI system's life cycle. Sustainability and ecological responsibility of AI systems should be encouraged, and research should be fostered into AI solutions addressing areas of global concern, such as for instance the Sustainable Development Goals. Ideally, AI systems should be used to benefit all human beings, including future generations." (EC 2019, 19)

7. Accountability

   "The requirement of accountability complements the above requirements, and is closely linked to the *principle of fairness*. It necessitates that mechanisms be put in place to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use." (EC 2019, 19)

As to the practical level, the Independent High Level Expert Group on AI published a Self-Assessment for Trustworthy Artificial Intelligence to assist developers and users of AI in the implementation of the above seven requirements. If your project entails the use of AI such that ethical issues are raised, please complete the Self-Assessment checklist available on the REC website.

Useful sources

- European Commission. High-level Expert Group on Artificial Intelligence (AI HLEG). *Ethics Guidelines for Trustworthy AI*. Brussels, Belgium: European Commission; 2019.
- European Commission. High-level Expert Group on Artificial Intelligence (AI HLEG). *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment.* Brussels, Belgium: European Commission; 2020.
- Responsible AI at Stanford. Enabling innovation through AI best practices
- EU Grants: How to complete your ethics self-assessment: V2.0 – 13.07.2021
- https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/guidelines-responsible-use-generative-ai-research-developed-european-research-area-forum-2024-03-20_en