Artificial Intelligence: Risks and opportunities

Background

While it is hard to come up with a short and comprehensive definition of artificial intelligence (AI), we could define it as the ability of technical systems to perceive their environment, deal with what they perceive, solve problems, and act to achieve specific goals. The computer receives data, processes it, and responds. While developers always set goals and define limits on possible outcomes, different AI systems have different degrees of automation and autonomy. AI is ubiquitous in our daily lives: boosting manufacturing productivity, providing public services and healthcare, designing "smart cities", connecting objects, and ensuring cybersecurity.

In recent years it has been widely acknowledged that the use of AI has had an impact on human rights. AI development and use can have a positive impact on the right to life and the right to health, as well as on the right to education (by facilitating more personalised education). At the same time, as the EU white paper on Artificial Intelligence states: "*AI entails a number of potential risks, such as opaque decision-making, gender-based or other kinds of discrimination, intrusion in our private lives or being used for criminal purposes*".

Many experts and international organisations have been reflecting on how states can ensure that AI development complies with human rights. Calls are made for systematic human rights impact assessments before, during and after the development and use of AI systems. Recommendations are made for stronger human control and oversight, and for more transparency so that individuals can make informed choices on whether or not to use AI. Effective remedies to end human rights violations are also recommended.

Research to be published by the EU Agency for Fundamental Rights will analyse real-life use cases of AI to consider its potential positive and negative impacts on fundamental rights.

UNESCO and the Council of Europe have been working on guidelines and recommendations to develop a human-rights-based approach to AI.

The European Union has designed guidelines recommending that AI development should ensure: *human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental wellbeing and accountability.*

As the EU works towards a new legal framework to regulate AI, a key issue is how to ensure that human rights remain at the core of AI development worldwide for the next decade.

Objectives

1. Present current international efforts to regulate the development and use of AI.
2. Present some of the benefits that AI can bring to human rights and reflect on how to disseminate these uses.

3. Raise awareness of the threats to human rights that AI can represent. The debate will touch upon specific aspects, such as the use of data that reinforce discrimination (for instance on ethnic and gender grounds).
4. Identify concretely how solutions to mitigate the misuse of AI can be implemented (transparency; human rights impact assessments; access to remedies and strategic litigation). Feed specific recommendations ahead of the upcoming EU legislation on AI.

Methodology

This working group will host one main thematic session (open to the general public) and two smaller interactive sessions: one open, and a closed one exclusive to human rights defenders. During the interactive sessions, the participants will discuss in smaller groups and build on their concrete experience to draw specific recommendations and lessons learned for the EU.

Main thematic session (public)

"Artificial intelligence: state of play of international efforts to regulate it"

9 December 2020, 10.00–12.30 CET

Panellists will explain current efforts to regulate AI internationally, as well as the current non-binding guidelines drafted by several international bodies. The panel discussion will provide an opportunity to discuss the upcoming AI legislation, which shortcomings it should address, and to reflect on the possibility for the regulation to have an international impact.

Interactive session I (closed)

"Addressing artificial intelligence harms: individual, collective, societal"

9 December 2020, 12.00–13.30 CET

In this interactive session, participants will collectively explore and outline how AI use may impact human rights in their own contexts. As governments, institutions and industry move to promote innovation in AI, grave concerns and major questions remain as to how the "benefits" of AI will be distributed. We will discuss how such changes will impact people, communities and society as a whole.

Use of AI systems can exacerbate surveillance, reflect and reinforce discrimination and broader societal inequalities, alter the delivery of essential public services, and disrupt the democratic process itself. For instance, AI use can reinforce discrimination in law enforcement, border management and justice systems. Due to the inherent power asymmetry between those who develop and deploy AI systems and those who are subject to them, a number of human rights issues remain. In this session, human rights defenders will

examine these consequences according to their contexts and put forward recommendations for policymakers.

Interactive session II (public)

"Fostering artificial intelligence governance: ensuring human rights compliance, ensuring justice"

10 December 2020, 16.30–18.00 CET

Following on from the main session, the second interactive session will explore specific ways to ensure a human-rights-centred governance model for AI. In response to the myriad harms AI systems can generate, in particular when deployed in sensitive areas, civil societies globally have called for comprehensive governance responses as opposed to technical fixes. This approach acknowledges that AI systems often play a role in reinforcing and amplifying existing societal problems. As such, we must also explore structural solutions that put human rights and democratic oversight at the core.

Drawing from experience across the globe, this session will explore how the harms outlined in the previous session can be cohesively tackled through AI governance frameworks that are truly centred on human and collective rights, and broader societal well-being. We will discuss a number of potential frameworks to ensure the inclusion of human rights in AI governance, including: the need for clear legal limits on harmful uses of AI ("red lines"); the need for democratic oversight and sustained models of citizen engagement; human rights impact assessments, and more.