Meeting Report

COMMISSIONER FOR HUMAN RIGHTS
COMMISSAIRE AUX DROITS DE L'HOMME

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

THE HUMAN LINE: SAFEGUARDING RIGHTS AND DEMOCRACY IN THE AI ERA

PREPARED BY THE COUNCIL OF EUROPE COMMISSIONER FOR HUMAN RIGHTS

Strasbourg 20 October 2025

CommHR(2025)57
Original version

CONTENTS

On 27 and 28 May 2025, the Council of Europe Commissioner for Human Rights brought together a group of experts working on or engaging with artificial intelligence (AI) to discuss the risks and opportunities of emerging AI technologies, and how to ensure a human-centred approach to AI governance. The interdisciplinary composition of the group enabled different perspectives to be brought to the discussion. The Commissioner chaired the meeting. The agenda can be found in Annex I.

The consultation aimed to understand current trends in AI technological development and the associated risks and opportunities for human rights. It also aimed to understand the role of regulation in ensuring safeguards for human rights in the design, development, and deployment of AI systems. The consultation centred on two main themes:

Theme 1: Emerging AI technologies. risks and opportunities for human rights

Theme 2: Embedding human rights in AI governance.

This report provides a non-exhaustive overview of the consultation's main points and conclusions in the form of a chairperson's summary. A list of participants is provided in Annex II to the report.

DISCUSSION AND CONCLUSIONS

Theme 1: Emerging AI technologies: risks and opportunities for human rights

1. The rapid development of AI requires continued engagement from policymakers. 2025 marked an inflection point, with widespread discussions about AI-powered technology. In such a fast-paced context, it is imperative to ensure that the design, development and deployment of AI takes a human-centred approach to foster opportunities and prevent the risks it poses to human rights, democracy and the rule of law.

2. The opaque use of AI technology poses a fundamental challenge for two key reasons. Firstly, individuals are often unaware of its deployment. For example, surveillance cameras with face recognition capabilities used in public assemblies interfere with individuals' right to privacy. Secondly, individuals may be aware of the use of AI technology but be unaware of its discriminatory design. For example, biased data fed into AI systems can result in a violation of the right to non-discrimination. The absence of explicability can lead to a lack of contestability, which in turn can hinder access to justice and the right to an effective remedy. In any case, greater transparency in AI systems would contribute to better protection of human rights.

3. Generative AI (AGI) has shifted the focus from the automation of tasks to the autonomy of machines, raising concerns about the potential absence of human control. Fully autonomous or agentic AI, which is subject to little or no oversight, could have profound adverse social consequences. A concerning trend is the potential use of agentic AI in military and security contexts. In this respect, human oversight is paramount to avoid gross human rights violations, including loss of life. Furthermore, individuals should always have the right to challenge decisions made by machines. Similarly, when it comes to content moderation, the discretionary power to identify harmful content may require human intervention to prevent the amplification of harm through the use of AI technology.

4. There is a significant risk associated with operating large language model (LLM)-powered technology, especially when it acts as a substitute for humans in discussions or educational settings. Such technology is increasingly producing so-called “careless speech” that we could describe as a type of hallucination, whereby the information received by a user is subtly incorrect, incomplete or biased towards a particular viewpoint. In other words, inaccurate information is repackaged and presented to individuals in a clearer, more confident version, which can subsequently be consumed uncritically. This phenomenon requires specific domain expertise for its detection. The AI-powered amplification of immaterial degradation of information quality could contribute to a reduction in the plurality of ideas and opinions.

5. Similarly, the use of generative AI in education can have a long-term impact on society. The way history and truth are presented can be a powerful tool for shaping a collective identity, or for fostering distinct clusters of opinions that may become disengaged from one another. If misused, such technology can rapidly amplify problems that can, in turn, impact the human rights of targeted groups, such as migrants or minorities. Similarly, the use of LLM-powered technology, particularly in an educational context, can lead to concerning levels of skill degradation, whereby individuals become reliant on such technology to make everyday decisions. This results in their inability to think, judge, discern and reason independently. All of the above also has implications for how information is controlled centrally by those who design, develop and deploy such AI systems.

6. Beyond education, personalised or targeted information through algorithms can contribute to the creation of separate informational spaces. This creates two main problems. Firstly, it isolates individuals from one another. Secondly, it potentially makes individuals more susceptible to manipulation and diminishes their critical thinking skills. Recent research attests to the impact of increased isolation and decreased social interaction on individuals' cognitive systems and resilience. In this context, using chatbots and algorithms to amplify disinformation could push individuals towards extremism. These elements affect not only individuals, but also the way our democratic societies are organised to safeguard human rights, including freedom of expression and access to information, which are vital pillars of any democratic society.

7. According to the Harvard Business Review, the number one use of generative AI in 2025 is for companionship and therapy purposes. Companion AI chatbots are said to be having a positive effect on the so-called “crisis of loneliness and isolation” and these are being designed and marketed by companies in highly anthropomorphic and personified ways. However, it should be noted that the long-term effects of using such technology may exacerbate human isolation further and contribute to the breakdown of the social fabric of our societies. In other words, such technology introduces structural social distancing. While it is common for humans to ascribe human features to new things, anthropomorphising AI technology carries serious risks and implications regarding what humans expect from it. This technology could exploit those in vulnerable situations, such as those dealing with death and loss, who may develop a strong emotional attachment to it. The elderly and children are particularly at risk in this regard, as research shows that these groups are highly susceptible to algorithms and addiction. Children are adversely affected at an early stage of development, most notably when they develop expectations around human relationships and social skills.

8. AI technology in general, and LLMs in particular, uses vast amounts of data, including personal data. This data originates from what individuals voluntarily and involuntarily post on the internet. LLM-powered technology is increasingly being used by individuals and the public sector and deploys LLM-based multimodal data aggregation and prediction with the help of

advanced simulation techniques. Recent research demonstrates that these systems could enable the 360-degree profiling of individuals, resulting in so-called “data cages” in which individuals' personal data is aggregated and categorised to provide detailed insights into human behaviour at scale. These “data cages”, which also facilitate collective profiling, enable LLM-powered technology to predict what individuals might wish to purchase, view or react to, thereby assisting in the targeted delivery of content with assumedly outstanding accuracy. Thus, this technology poses risks to human dignity and autonomy (e.g. disempowering individuals to make choices), as well as to human rights, particularly the right to private life (e.g. tracking individuals' personal and sensitive data). Understanding how data is compressed within an LLM can provide a pathway for governing its outputs.

9. AI technology has the potential to promote human rights further in the public sector, for example by promoting faster inclusion and access to healthcare and education. However, the deployment of AI systems in the public sector under a cost-cutting narrative poses significant risks, as the economic incentive to rationalise expenses can harm the human rights of individuals in vulnerable socioeconomic conditions. For instance, governments use AI technology on a large scale to detect fraud within social welfare systems, but this is rarely done to increase access to benefits. Therefore, the problem lies not in the technology itself, but in its use through the lens of securitisation – where individuals in vulnerable socioeconomic situations are treated as suspects by default, rather than the technology being used to identify ways to improve access to rights. Currently, it is difficult to identify this type of harm as it is not always immediately tangible; however, it often becomes apparent when the accumulation of harm becomes systemic. Increased dependence on AI systems requires bias control to be embedded in the data used for machine learning, as well as cultural awareness and replicability across cultures, in the design phase to ensure respect for the principles of equality and non-discrimination.

10. There is already an internationally agreed framework to protect human rights. It is therefore important to insist on due diligence and accountability, and to take a human-centric approach to AI. After all, it is always a human who develops the machine, the algorithms and the coding. Individuals also provide the data used for machine learning and should ultimately be responsible for the consequences of AI systems. Additionally, educating the general public on how to consume information and be critical of technology's outputs could preserve the integrity of the information system as a whole. Increased AI literacy is therefore of paramount importance, without detracting from the conversation over the liability of providers for harm caused by AI systems.

Key conclusions:

- The lack of a human-centric approach to the design, development and deployment of AI can pose a high risk to democracies and human rights, particularly for those in vulnerable situations.
- Data used for training AI systems can be biased, and the use of algorithms can lead to discriminatory practices; meaningful transparency is therefore required.
- There are some fields in which the adverse consequences of using AI can be significant, such as the military and security sectors. Human oversight and individual accountability remain key to ensuring its safe use.
- Large language model-powered technology can amplify disinformation, including through so-called “careless speech”, which can lead to the degradation of the informational ecosystem. If unchecked, access to quality, pluralistic information is at risk.

- Without education and awareness around AI, users are rendered vulnerable to manipulation by information providers, including through generative AI systems, which can further polarise societies. Efforts must be made to promote digital trust and literacy among end users in information consumption.
- Anthropomorphising AI technology can make users more vulnerable and increase their isolation in the long term, which can exacerbate the degradation of the social fabric in societies. Children and the elderly are particularly vulnerable to the negative impacts of social chatbots.
- Data, including personal data, gathered by AI systems and processed by advanced simulation techniques could be used for in-depth profiling of individuals, which could result in risks to their human dignity, autonomy, and human rights.
- The use of AI in the public sector offers opportunities to advance human rights and foster the inclusion of marginalised communities and can contribute to promote equality and non-discrimination and enhance access to rights.

Theme 2: Embedding Human Rights in AI Governance

11. Embedding human rights in AI governance requires a multistakeholder approach and international collaboration. This can only be effective if it is based on a shared understanding and meaningful dialogue. Globally agreed standards could ensure coherence, although the implementation of these principles may differ at a local level. Nevertheless, such standards must take due consideration of children and individuals in situations of vulnerability, who have limited political agency and economic influence. In this respect, it is paramount to embed safety by design in national regulations and to promote digital literacy and resilience, including among children. Other important principles for building responsible AI governance include ensuring transparency and accountability, investing in research and disaggregated data, and fostering cross-border collaboration. To safeguard human rights and democracies, regulatory responses must move beyond a purely reactive approach. An effective regulatory framework should include robust structural auditing mechanisms to assess how algorithms used by major platforms shape and influence public discourse.

12. The current regulatory framework in Europe shows great promise. The Council of Europe Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law takes a flexible approach that allows it to adapt to the specific needs and context of each ratifying country. Under the European Union (EU) Artificial Intelligence Act (AI Act), providers are specifically required to conduct conformity assessments and establish a quality management system, as well as maintain appropriate documentation. Consequently, deployers must operate AI systems in accordance with the providers' instructions and ensure effective human oversight throughout their use. The requirement for certain deployers to conduct human rights impact assessments will be crucial for protecting them, as will the obligations concerning the establishment of a risk management system and the implementation of data governance frameworks. In terms of human rights oversight, national human rights institutions and data protection authorities remain key actors within their respective mandates to access information and requesting documentation related to AI systems. The establishment of an Advisory Forum on AI within the EU is considered an example of good practice. The Forum serves as a platform that ensures diverse perspectives from industry, civil society, and academia are included in discussions on AI governance.

13. Recent research indicates that the primary challenge is no longer the formulation of new regulations for governing AI, but rather the interpretation and effective implementation of those already established. For example, a significant number of companies fail to report the

systematic implementation of risk mitigation AI measures. Companies require support to interpret and implement regulations in practical terms, accompanied by clear guidance, since a one-size-fits-all approach is not sufficient across diverse sectors. Furthermore, governance discussions often fail to address the entire AI ecosystem, including issues such as access to and equity in data, and the management of data and AI infrastructure.

14. Investment in more research is required to explore the associated risks and mitigation measures relating to human rights. In this context, the most notable risks to human rights posed by emerging technologies, such as artificial general intelligence (AGI) — which remains vaguely defined, but generally refers to human-like intelligence — are: the loss of human control over the system; widespread AI-enabled hacking; and the concentration of power. Other risks associated with AI agents, which are defined as autonomous systems that sense and act upon their environment to achieve goals, include ethical questions about their decision-making capabilities, as well as a lack of transparency and explainability, both of which are closely related to the principle of responsibility. Focusing on fears of societal extinction posed by AGI distracts from the real challenges and shifts attention away from present-day discussions. Individuals will always play a role in technological development, and policymakers must ensure they remain an active and responsible part of the process. The focus should be on the purpose of the technology. The current regulatory instruments in Europe are sufficiently robust to address any technology that could be classified as AGI.

15. While human oversight is a legal prerequisite under existing AI regulations, there remains a lack of ethical and responsible oversight, particularly in the military and security sectors. This has led to situations in which AI systems are deployed in military operations under the supervision of individuals who trust the technology more than they trust their own judgement or expertise. Consequently, critical decisions involving physical harm to humans may go unchallenged. Policymakers should ensure that AI system design criteria include human responsibility and oversight by individuals with adequate moral judgement and discernment and capability to challenge those systems' decision-making. There is an increasing need to put pressure on policymakers to introduce clear guardrails for high-risk and potentially harmful AI use cases in these sectors. As technologies deployed in a military or security context continue to evolve and improve, and as their societal impact grows, proactive regulatory measures will be essential to safeguarding human rights and public trust.

16. Regarding the debate on whether regulation stifles innovation, a recent counterexample can be found in China, where the world's most regulated LLM, DeepSeek, was developed. The Chinese authorities have developed detailed standards on what algorithms should encompass, including how training data should be stored, albeit without due consideration for human rights. Risk management frameworks that take human rights into account, such as those established in Europe, can actively support and enhance innovation. For example, requirements to document and test AI tools prior to market release benefit both the economy and human rights by helping to mitigate the risks associated with autonomous online updates to these tools online that could cause harm to individuals later on. Investors must recognise the economic value of integrating human rights considerations alongside commercial viability and market fit in venture capital decisions. Ultimately, respecting the human rights of end users contributes to effective AI investment.

17. A preliminary step towards effective AI governance is having a clear understanding of what is at stake in terms of both technology and human rights. Currently, there is a significant policy-level gap in grasping the nuances of technology, such as the potential of AI for public good

and the potential harm it could cause. Without this foundational knowledge, policymakers struggle to strike the right balance when developing smart, effective, and simplified regulation. They must be able to identify where to strategically implement safeguards to protect human rights, as well as understand the impact of regulation on individuals, society and companies. This requires not just theoretical knowledge, but also practical examples that illustrate the issues. Similarly, companies should be able to measure and understand the limits of the technologies they develop and deploy. For example, while technology can offer quick solutions to social problems such as loneliness, it is inherently limited in its capacity to fully resolve them. To measure the real-life impact of such technologies, companies should be encouraged to invest in research and engineering to develop better measurement tools.

18. There is a growing need for clearer regulations regarding the responsibilities of AI providers in helping users to interpret content. This could include requirements to disclose sources and present a diverse set of references reflecting a range of perspectives, including conflicting ones. This could encourage critical engagement and reduce users' vulnerability to manipulation. Additionally, a framework should be established to assign liability for downstream harms caused by general-purpose AI technology, such as life assistants. Certain categories of prompts may need to be limited, particularly those involving sensitive data, such as medical advice, where the use of generative AI systems is especially problematic. In such cases, providers should be held responsible for the consequences of the outputs generated by their systems.

19. Standardisation in technology must be articulated in a way that conveys meaningful parameters to all stakeholders involved in its design, development, and deployment, including innovators, developers, operators, and maintainers. This can be achieved horizontally by examining overarching principles and translating them into applicable rules. However, the vertical dimension, comprising the translation of standards within specific sectors, is also essential. This sector-specific approach enables relevant industries to participate actively in shaping the standards, fostering a sense of ownership and responsibility in aligning them with real-world needs and practices. The end goal is to create meaningful regulations that serve society rather than focusing exclusively on minimising harm.

20. Policy discussions must encompass the principle of individual responsibility for the environmental footprint of AI systems, addressing the entire AI supply chain and paying particular attention to how each stage may affect human rights. Discussions should focus on concrete processes involved in building AI systems from the bottom, i.e. the raw materials required to build chips, the vast energy consumption in data storage and management, and the societal and environmental impacts, particularly on underrepresented or excluded communities.

Key conclusions:

- Effective AI governance requires a multistakeholder approach to establish and implement globally aligned standards while taking into consideration local specificities and protect groups in situation of vulnerability, including children.
- In Europe, the current regulatory framework on AI is sufficient to address the risks to human rights posed by emerging AI technologies. National human rights institutions remain key actors in conducting human rights oversight of AI systems within their respective mandates.
- Responsible oversight in the military and security sectors should be implemented by ensuring human responsibility and insisting on human oversight by individuals with moral discernment.

- Incorporating human rights considerations into risk management frameworks for technology enhances innovation and leads to effective AI investment.
- Policymakers and companies should strengthen their practical understanding of AI technologies, including their purpose and the risks and opportunities they pose to human rights, by investing in research and measurement tools. Further efforts are needed to guide companies towards achieving this goal, alongside the effective implementation of regulation.
- Clear rules should be established to define the responsibilities of AI providers in facilitating users' interpretation of online content. A liability framework for harms caused by general-purpose AI systems in sensitive areas should be established.
- The principle of individual responsibility for the environmental and societal impact of AI systems must be taken into account by policymakers when developing regulatory frameworks.

Annex I

PROGRAMME

NAVIGATING THE FUTURE: HUMAN RIGHTS IN THE FACE OF EMERGING AI TECHNOLOGIES

Expert Consultation

Paris and online, 27th and 28th May 2025

Venue: Meeting Room 3, Council of Europe's Paris Office
55 Avenue Kléber, 75116 Paris

27 May 2025 - EMERGING AI TECHNOLOGIES: RISKS AND OPPORTUNITIES FOR HUMAN RIGHTS

14:00 - 14:05 **Welcome and introduction by Michael O'Flaherty, Commissioner for Human Rights**

14:05 - 14:15 **The work on AI and human rights by the Council of Europe and the Office of the Commissioner**
Hristijan Koneski, Adviser to the Commissioner for Human Rights

14:15 - 14:45 **Tour de table**
Presentation of participants and their line of work in a nutshell

14:45 - 15:45 **Session I:** Mapping main issues – **emerging AI technologies: risks and opportunities for human rights**

Presentation by Prof. Oreste Pollicino

Presentation by Dr Murielle Popa-Fabre

Presentation by Dr Brent Mittlestadt

Discussion

15:45 - 16:00 ***Coffee break***

16:00 - 17:20 **Continuation of discussion**

Discussion

17:20 - 17:30 **Conclusions of the day**
Michael O'Flaherty, Commissioner for Human Rights

17:30 *End of day one*

******

28 May 2025 - AI GOVERNANCE

09:00 - 10:30 **Session II: Embedding human rights in AI governance**

09:00 - 09:10 **Models of AI governance**
Sandra Veloy Mateu, Adviser to the Commissioner for Human Rights

9:10 – 10:30 *Presentation by Ms Samira Gazzane*

Presentation by Dr David Reichel

Presentation by Prof. Ali Ghazi Hessami

Discussion

10:30 - 11:00 ***Coffee break***

11:00 - 12:10 **Continuation of discussion**

12:10 - 12:30 **Concluding remarks by Michael O'Flaherty, Commissioner for Human Rights**

12:30 *End of day two*

Annex II

List of Participants

Bryson, Joanna	Hertie School
Gazzane, Samira	World Economic Forum
Hasselberger, William	Catholic University of Portugal
Hessami, Ali	Vega Systems
Mittelstadt, Brent	Oxford Internet Institute
Pollicino, Oreste	Bocconi University
Popa-Fabre, Murielle	Independent
Reichel, David	EU Fundamental Rights Agency
Świerczyński, Marek	University of Cardinal Stefan Wyszynski
Tesfaye, Hiwot	Microsoft
Bishop Tighe, Paul	Holy See
Xenidis, Raphaële	Sciences Po Law School

Council of Europe

O'Flaherty, Michael	Commissioner for Human Rights
Havula-Lorenzini, Anna	Assistant
Koneski, Hristijan	Adviser to the Commissioner
Veloy Mateu, Sandra	Adviser to the Commissioner