CDDH-IA(2024)09
08/11/2024

STEERING COMMITTEE FOR HUMAN RIGHTS
(CDDH)

_____

DRAFTING GROUP ON HUMAN RIGHTS AND ARTIFICIAL INTELLIGENCE

(CDDH-IA)

Summary of the exchange of views with external independent experts and representatives of the Committee on Artificial Intelligence (CAI), the Committee of Experts on the Implications of Generative Artificial Intelligence on Freedom of Expression (MSI-AI), the Consultative Committee for Convention 108 (T-PD), and the Working Group on Cyberjustice and Artificial Intelligence (CEPEJ-GT-CYBERJUST) (25 September 2024)

(prepared by the Secretariat)

1. INTRODUCTION

The CDDH-IA held an exchange of views at its first meeting on the intersection of artificial intelligence (AI) and human rights, focusing on the technical, legal, and ethical implications of AI technologies, as well as on the work of various Council of Europe intergovernmental committees on human rights and AI.

The CDDH-IA held the exchange of views with:

- Marko GROBELNIK, AI Researcher & Digital Champion at the AI Lab of Slovenia's Jozef Stefan Institute
- Alberto QUINTAVALLA, Assistant Professor at Erasmus University Rotterdam,
- Miriam KULLMANN, Professor of Labour Law and Social Security Law at Utrecht University School of Law, and member of the European Committee of Social Rights
- David LESLIE, Director of Ethics and Responsible Innovation Research at The Alan Turing Institute
- Thomas SCHNEIDER, former Chair of the Committee on Artificial Intelligence (CAI) of the Council of Europe (2022–2024) and current vice-Chair of CAI
- Peter KIMPIÁN, Secretary of the Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (T-PD)
- Andrin EICHIN, Chair of the Council of Europe Committee of Experts on the Implications of Generative Artificial Intelligence on Freedom of Expression (MSI-AI)
- Jan SPOENLE, Member of the Working Group on Cyberjustice and Artificial Intelligence (CEPEJ-GT-CYBERJUST) and contributor to the European Ethical Charter on AI in judicial systems

2. SUMMARY OF DISCUSSIONS

Key points made by Marko GROBELNIK

- The evolving definition and operational framework for artificial intelligence (AI) systems reflect the increasing need for clarity and actionable terms, with recent updates from the OECD widely adopted in international policy frameworks. These definitions describe AI systems as machine-based frameworks capable of making predictions, recommendations, or decisions that influence real or virtual environments, often operating autonomously.
- Key components of AI systems include a sequence beginning with data input, progressing through model construction—often via machine learning algorithms—followed by inference and output, which subsequently influences the external environment. At the core of AI systems lies the model, a compressed, often opaque representation of data that guides the system's responses. The process of inference, whereby the AI interprets and applies the model, is crucial to the system's operational logic and capacity for decision-making.
- Recent updates to the OECD's definition incorporate key terms such as "adaptiveness" and "autonomy," reflecting the profound impact of generative AI systems and their ability to learn and adjust to new data and environments. While these developments represent advances in AI capabilities, they also introduce challenges around accountability and system

robustness, particularly where autonomous systems may act without immediate human oversight.

- AI's "black-box" nature presents significant challenges to transparency, as advanced models often function beyond the complete understanding of even expert operators. This opacity raises substantial concerns about accountability, especially as AI's speed, complexity, and autonomy may surpass human ability to supervise effectively. Addressing these challenges within AI frameworks is critical to aligning technological developments with human rights protections.
- AI systems are increasingly interacting with unknown or human-uninterpretable phenomena, as machine learning models uncover patterns and insights that extend beyond current human knowledge. While this capability represents a notable technological advancement, it emphasises the urgent need for robust policies that prevent AI systems from inadvertently undermining fundamental rights or established societal norms.

Key points made by Alberto QUINTAVALLA

- The relationship between AI and human rights is often viewed through specific rights such as privacy and non-discrimination, due to AI's data-driven nature and its inherent biases. Privacy risks are heightened in governmental AI applications, where personal data collection is central to functions like predictive policing, counter-terrorism, and surveillance, raising significant privacy concerns. While these uses of AI bring privacy risks, existing legal protections, such as the Council of Europe's Convention 108+ and various national frameworks, offer solid safeguards. However, as AI advances, vigilance is essential to ensure effective protection of these rights. Similarly, non-discrimination rights are challenged by AI's susceptibility to biases, particularly in systems used by public authorities, as demonstrated in cases like predictive policing and welfare fraud detection. A notable case illustrating this is the Dutch "*Toeslagenaffaire*," where tax authorities used algorithms that disproportionately flagged non-Dutch nationals in childcare benefit fraud detection, revealing how biases embedded in AI can lead to discrimination.
- A piecemeal approach to AI and human rights risks overlooking AI's broader impacts, potentially leading to unintended consequences. A systematic view, by contrast, offers advantages by recognizing the indivisibility, interdependency, and interrelatedness of all human rights. This approach ensures that rights are not considered in isolation but within a framework that acknowledges cross-impacts. A broader perspective on AI's impact is important because AI may simultaneously enhance one right while compromising another, requiring a careful balance of rights. Additionally, a systematic approach helps identify common risks across rights, providing valuable insights for policymakers to create safeguards that address and prevent human rights violations associated with AI.
- AI's influence is both positive and negative, creating a two-way relationship with rights. For instance, AI can enhance the right to a fair trial by improving access to legal resources, identifying case law, and predicting outcomes, thus making justice more accessible. However, it can also produce biased outcomes, violating fair trial rights and non-discrimination protections. AI's dual nature means it can impact multiple rights at once; for example, authoritarian uses of AI for surveillance can infringe on privacy, freedom of assembly, and expression. The Council of Europe's Framework Convention underscores this, recognizing AI's potential to advance human rights while also posing risks to human dignity and autonomy.

- This double-edged relationship forms a web of conflicting or reinforcing rights. For example, the “internet of bodies”— devices like smartwatches and digital pills that track health data — can improve the right to health by providing continuous health feedback but risks infringing on privacy rights. Individuals may want to revoke consent for data collection, yet doing so could impact their health. Additionally, while such devices enhance health rights, they may conflict with environmental rights due to their high ecological footprint, highlighting trade-offs between health and environmental protection. The complex web of AI's impact on human rights presents significant regulatory challenges. Policymakers and judges must address various human rights trade-offs arising from AI, making the proportionality principle increasingly relevant. Though sometimes contentious, this principle provides a structured way to balance competing rights, allowing for practical solutions in litigation.
- Private businesses played a central role in the AI sector being the key actors in the development stage and, sometimes, in the deployment stage. Governments may rely on non-state actors to undertake some governmental activities, including the criminal justice sector. Moreover, non-state actors perform governance activities themselves to address governance gaps, such as in the development of smart cities. It is in this context that human rights law should develop ways to deal with the role of private businesses. The indirect horizontal effect of human rights and the existing UN human rights guidelines have thus far been used to mitigate this issue. States have a positive obligation under the indirect horizontal effect of human rights to protect individuals from harmful interference by non-state actors. This responsibility calls for more effective measures to shield individuals from potential harms caused by private entities. Additionally, soft law instruments, such as corporate responsibility guidelines, have been introduced to encourage respect for human rights. However, these approaches are relatively new, and their true impact remains to be evaluated.
- The interaction between AI and human rights extends to emerging issues like environmental protection, with AI affecting the environment in both positive and negative ways. For instance, while AI can help prevent illegal activities in protected areas, it also demands significant energy and water, raising concerns over its ecological footprint. Quantifying AI's environmental impact is challenging, as there is no consensus on measuring its carbon footprint. Additionally, the Council of Europe does not recognize a standalone right to a healthy environment, complicating efforts to litigate environmental harms linked to AI. Claims must often be based on established rights like privacy, making it difficult for individuals to meet the "victim requirement." Furthermore, soft law on human rights typically targets states, not private companies, limiting its effectiveness in addressing environmental impacts from AI technologies.

Key points made by Miriam KULLMANN

- Public authorities and private entities increasingly use automated decision-making systems, applying them to tasks like granting social benefits, hiring, and assessing creditworthiness. While these systems can improve efficiency, it cannot be ignored that automating decision-making, even if at some point in this process a human is involved, can violate economic and social rights as for automated decision-making processes personal data is being processed. Relevant personal data are, e.g., a person’s age, gender, religion or belief, political views, trade union membership, place of living and/or work, family and/or marital status, the income situation and whether or not the person is in need of social security or social benefits, or an individual’s health condition. These

systems are limited by the quality of their data and the algorithms guiding them, often leading to adverse outcomes for marginalised groups.

- Digitalisation has transformed recruitment, with automated systems widely used to target and screen job candidates, though risks of bias persist. These systems also monitor productivity, influencing promotion, pay, or dismissal. In social protection, automated decision-making is applied in benefits administration, as seen in the Dutch childcare benefits scandal (*kinderopvangtoeslagaffaire*), where algorithms flagged individuals based on profiles like dual nationality, leading to wrongful penalties. Similar issues arise in Denmark, where extensive data is cross-checked to detect welfare fraud, raising concerns about fairness. Automation also affects access to housing and healthcare, with algorithms determining eligibility for mortgages, social housing, and even healthcare treatments. When trained on biased data, these systems risk harmful generalizations and can overlook individual needs.

- The ESC enshrines several rights relevant to employment, impacted by the rise of automated decision-making. Art 1 guarantees the right to work, including non-discrimination (linked with Art 20, the right to equal treatment in employment), Art 8 (protection of maternity), Art 15 (employment of persons with disabilities), Art 23 (social protection of elderly workers), and Art 27 (equal treatment for workers with family responsibilities). Discrimination, based on characteristics like gender or disability, is prohibited in recruitment, employment conditions, and promotions, while Art 18 ensures foreign nationals have equal access to work. Automated decision-making systems affecting factors such as working hours, monitored productivity, or remuneration must respect limitations that protect health, safety, and non-discrimination (Art 4 and Art 3). Where AI is used for job performance tracking or dismissals, notice and valid reasons are required (Art 24). Workers must have a safe environment, including rights like digital disconnection, preventing penalization for not working outside normal hours. Art 5 of the ESC guarantees freedom of association, which may be impacted if AI systems factor in union membership, possibly deterring workers from joining unions. Art 6 ensures collective bargaining rights, encouraging representation in discussions on automated systems' workplace roles. This is closely linked to rights on workplace consultation, conditions, and dignity (Arts 21, 22, and 26).

- Concerning health and social security, the ESC guarantees the right to social security for employed and self-employed individuals and their dependents under Article 12 (1), covering areas like medical care, unemployment, family, and maternity benefits. Social security access must be free from discrimination as outlined in Articles 20 and E, ensuring equal treatment in aspects like benefit calculation and eligibility. Public authorities using automated decision-making for social security or medical assistance (Article 13 (1)) must ensure these systems are non-discriminatory, as eligibility should only depend on need. Access to social welfare services similarly requires non-discriminatory practices (Articles 14 and E).

- Concerning housing, under Article 31 of the ESC, everyone has the right to adequate housing, with special attention given to vulnerable groups such as low-income individuals, single parents, and persons with disabilities. When automated decision-making systems are used by public authorities or private entities to allocate housing or mortgages, access must be non-discriminatory, as required by Article E. States are also obligated to prevent and reduce homelessness, ensuring that automated systems do not exclude or discriminate against individuals at risk of losing housing.

Key points made by David LESLIE

- The rapid growth of AI has reshaped the digital environment, embedding AI-driven systems into both public infrastructure and the private sphere. This transformation enables algorithmic systems to influence personal identity and societal interactions through targeted decision-making and content curation. Practices like attention-mining, relevance-ranking, and trend-predicting direct user behaviour and affect how individuals engage with information, raising significant concerns for human rights. Such technologies, by steering digital experiences, may undermine individual autonomy, privacy, and freedom of thought, as they increasingly control access to and interpretation of information in the public sphere.
- AI's pervasive role in behavioural manipulation extends to the workplace, where algorithmic management systems monitor employee productivity, satisfaction, and interactions. These systems, while aimed at optimizing efficiency, may erode personal privacy and alter behaviour due to constant surveillance. The lack of interpersonal trust and solidarity induced by such monitoring compromises freedom of expression and assembly, essential for collaborative work environments. In labour, education, and criminal justice, these technologies can impose coercive surveillance that limits open communication and deters democratic agency and participation.
- Generative AI technologies intensify these risks by enabling the large-scale production of disinformation, synthetic content, and deep fakes, which can degrade the quality and reliability of information in the digital public sphere. This influx of manipulated content threatens the social trust foundational to democratic societies, as it becomes increasingly difficult for individuals to discern fact from fabrication. The pollution of public discourse with AI-generated misinformation poses a serious threat to the integrity of democratic engagement and civic responsibility.
- Together, these developments highlight a critical ethical challenge: AI's influence on human rights, democracy, and the rule of law is extensive, and its unchecked expansion risks undermining the very foundations of these institutions. Recognising the transformative potential of AI, society must actively shape its deployment through democratic governance and ethical foresight, ensuring that technological advancement aligns with the values of justice, fairness, and respect for human rights.

Key points made by Thomas SCHNEIDER

- AI, as a disruptive technology, parallels past technological shifts, like the transition from physical to cognitive automation. Unlike physical engines, AI operates with dematerialized data that can be reproduced and deployed globally, necessitating adaptive governance. Schneider underscored that the convention aims not to replace existing standards but to complement them, focusing on upholding human rights, democracy, and rule of law while promoting innovation. The approach seeks to address AI's unique attributes through a high-level, future-proof framework that remains adaptable to technological advance.

- The scope of the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (Framework Convention) extends to both public and private sectors, with flexibility for national application but holds all actors accountable for protecting rights. Limited exemptions apply, such as national security and certain research activities. The Framework Convention applies a differentiated approach, weighing risks based on context and impact severity, aiming to ensure regulatory measures are proportional to the technology's potential impact on rights.

- The Framework Convention stipulates core principles including human dignity, individual autonomy, transparency and oversight, accountability and responsibility, equality, and privacy. Procedural safeguards mandate documentation to enable individuals to contest AI-driven decisions, aligning with a commitment to transparency and access to remedies. On future regulatory needs, the conventional legal frameworks may be too slow to address rapidly evolving AI technologies. More agile, adaptable forms of governance suited to 21st-century technology is needed. The Framework Convention, which now awaits ratification, marks a significant step, but the need for ongoing dialogue, technical norms, and international cooperation to create a shared vision for AI that aligns with global human rights and democratic standards is to be emphasised.

Key points made by Andrin EICHIN

- The Committee of Experts on the impacts of generative artificial intelligence for freedom of expression (MSI-AI) is preparing Draft Guidelines on the implications of generative artificial intelligence for freedom of expression. The work started in 2024, with the first meeting aimed at exploring generative AI's unique challenges. A first draft of the Guidelines is anticipated by the end of 2024, with the final document scheduled for completion by 2025.

- The process includes identifying whether and how generative AI uniquely impacts freedom of expression compared to other technologies, assessing if these impacts require immediate intervention, and determining the necessary conditions to safeguard freedom of expression in a society increasingly shaped by generative AI. To guide this analysis, the Committee employs a "tech-stack" approach, examining risks at each stage of generative AI's lifecycle—foundational model training, model fine-tuning, and user interaction.

- The MSI-AI identified that the analysis should include both the risks and benefits of generative AI. Generative AI offers advantages, such as personalized learning support, access to mental health resources through avatars, and tools to counter disinformation. However, it also presents risks, including biased outputs, "hallucinations" (incorrect information generation), and the potential for manipulation via hyper-personalized content. These issues directly impact freedom of expression, highlighting the need for safeguards, particularly transparency in AI-generated content. Such measures are crucial to prevent misuse in political contexts, where tools like voice cloning and manipulated images could unduly influence public opinion.

- The Draft Guidelines also aim to establish clear responsibilities and oversight mechanisms for both public and private actors, stressing the importance of transparency, fairness, and accountability across all stages of Generative AI development and deployment. They also intend to provide practical guidance for users, focusing on raising public awareness of Generative AI's potential impact on freedom of expression and equipping individuals with tools to manage these risks. This approach reflects a broader commitment to protecting democratic values and human rights as AI technologies continue to advance.

Key points made by Peter KIMPIAN

- Convention 108 is the only legally binding multilateral agreement on privacy and data protection, covering both public and private sectors with 55 member states and additional observers globally. The convention provides a framework designed to safeguard individual rights to privacy and data protection, aiming to protect human dignity in the digital age. It incorporates data protection principles such as necessity, proportionality, transparency, and the right of individuals to understand and contest decisions made by automated data processing, a key aspect also in regulating AI's societal impact.
- Cross-border data flows present significant challenges, with a pressing need for harmonized international standards to prevent circumvention of data protection laws, especially given the prevalence of AI-driven data sharing. Cooperation among data protection authorities is essential for upholding the goals of Convention 108(+) in cross-border contexts. Under the treaty, such cooperation is mandatory, promoting a coordinated approach to data protection that aligns with global data governance standards.
- The 2019 Guidelines on Artificial Intelligence and Data Protection, elaborated by the Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (TP-D), emphasise AI policies that safeguard privacy, prevent discrimination, and uphold democratic values. The guidelines are structured in a way that they offer guidance for policymakers, as well as AI developers, manufacturers and service providers. These guidelines advocate for privacy-by-design in AI development, integrating robust data protection to minimize risks related to automated decision-making and profiling. Transparency is also emphasised, with clear explanations of AI-driven decisions encouraged to foster public trust and enable democratic participation.
- Another important tool is Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems which aims to prevent abuses of these systems and ensures effective protection for human rights, the rule of law, and democracy. It provides guidance on state obligations and corporate responsibilities, focusing on rights such as fair trial, privacy and data protection, freedom of thought, conscience, and religion, freedoms of expression and assembly, equal treatment, and economic and social rights. Updating an earlier recommendation to align with the modernized Convention 108+, it highlights concerns around profiling, transparency, and the need to protect vulnerable groups. The recommendation also calls for a legally binding "privacy by design" approach.

Key points made by Jan SPOENLE

- The Commission for the Efficiency of Justice (CEPEJ) aims to enhance justice system efficiency and functioning and enable better implementation of international legal instruments. CEPEJ's work supports the use of technology to streamline judicial processes, improve legal training, and manage court administration, aligning with fairness and accessibility in justice. Over time, the close connection between environmental protection and social rights has become increasingly apparent, with the degradation of the environment significantly affecting various social rights, such as the right to health and the right to safe and healthy working conditions.

- In response to the ethical and legal challenges posed by AI in judicial systems, CEPEJ adopted the European Ethical Charter on the Use of AI in Judicial Systems in 2018. The Charter addresses concerns about AI biases, as well as fears about AI predicting court decisions and the “black box” nature of algorithms. It is the first European instrument to outline five core principles for AI use in judicial contexts: (1) respect for fundamental rights, ensuring AI tools align with human rights standards; (2) prevention of discrimination; (3) quality and security, mandating certified, multidisciplinary data processing; (4) transparency, impartiality and fairness, requiring clear, auditable AI processes; and (5) user control, ensuring that users retain are informed actors and in control of their choices.
- To facilitate practical application, CEPEJ introduced an assessment tool in 2023. This tool aims to complement the Framework Convention, and helps judicial bodies operationalise the Ethical Charter’s principles by offering 29 guiding questions on ethical AI use. The questions are intended for judicial officials and IT managers to evaluate potential risks, such as profiling judges or jurors, before implementing AI systems in the judiciary.
- Further supporting ethical AI use in justice, CEPEJ established a permanent working group on Cyberjustice and AI, an AI advisory board, and the European Cyberjustice Network, which offers cross-border collaboration and webinars among judicial experts. CEPEJ also improved its monitoring tools with a new ICT chapter in its evaluation framework and launched a cyberjustice resource centre to track AI applications in European judicial systems. In 2023, it issued a guidance note on generative AI for judicial professionals, providing insights into AI’s responsible use in work contexts.

Discussion

- One member of the CDDH-IA raised a concern about the evolving OECD definition of AI systems and the “black box” problem I. Marko GROBELNIK, in his response, elaborated that the definition does not touch on the explainability of AI models. On AI's interpretability challenges, it was explained that certain AI phenomena are observable yet not fully understood. This issue, where AI operates in its own "language" beyond human comprehension, underscores the need for nuanced regulatory frameworks that can account for unknown or unpredictable outcomes of advanced AI.
- Questions were also posed about the challenge to identify core areas concerning economic and social rights. Miriam KULLMANN explained that besides identifying the material scope of the rights, a way forward could be to identify main thematic areas and try to bring different human rights together rather than highlighting separate.
- A question was posed on the potential role of the right to development in the context of AI. Alberto QUINTAVALLA highlighted the role of businesses, and the issue of intellectual property (IP) rights and open-source access, questioning to what extent IP protections could be adapted to promote broader technological benefits. One suggested approach was the use of compulsory licensing to balance international obligations with development goals. The response emphasised the importance of careful balancing through, for example, the principle of proportionality, where the right to development, and possibly also the right to scientific progress be weighed alongside other competing rights such as the emerging right to a healthy environment. Finally, the discussion underscored the need for flexible regulatory approaches to address the diverse applications and impacts of AI across regions.
- The discussion also covered AI's impact on children's rights, highlighting ongoing initiatives to make digital platforms safer for minors. David LESLIE noted the Council of Europe’s work

on examining how social media apps and intensive early screen time can affect children's development. He emphasised the importance of grounding age-appropriate design in human rights principles, stressing that the broader, cumulative effects of AI on children should be assessed within a human rights framework.

- A CDDH-IA member raised a question about AI's potential to interpret and understand the language it generates and whether AI could engage in misuse independently, without human intervention. In response, Marko GROBELNIK explained that machines perceive the world as a web of interconnected concepts. He noted that while it may be possible to decode AI's "language" to some degree, doing so requires significant computational resources and remains imperfect. David LESLIE cautioned against anthropomorphizing AI. He emphasised a "deflationary" perspective, arguing that large-scale AI systems are fundamentally statistical tools that process patterns in massive datasets, including syntactic structures that resemble human language. However, he argued, there is nothing mysterious about this capability; it is simply a result of processing vast amounts of data. These systems, he asserted, are quantitative tools, lacking independent language, self-awareness, or consciousness, and should not be viewed as autonomous entities capable of intentional misuse.
- A question was raised about the potential relationship between the CDDH Handbook on AI and Human Rights and the Human Rights, Democracy, and the Rule of Law Impact Assessment (HUDERIA), particularly concerning their target audiences. In response, an expert emphasised that defining the target audience is crucial. For HUDERIA, one of the main challenges is determining the appropriate language, technical depth, and placement of key messages. If the Handbook aims to guide non-lawyers in understanding human rights implications of AI, it should align with the HUDERIA substantively but ensure accessible language. In discussing the added value of the Handbook relative to previous work by the Ad hoc Committee on Artificial Intelligence (CAHAI) and the Committee on Artificial Intelligence (CAI), Thomas SCHNEIDER suggested that, like the Guide to Human Rights for Internet Users, the Handbook could help readers understand their rights in the context of AI and outline also available remedies. Identifying rights alone is insufficient; the Handbook should put emphasis on accessible remedies and address the challenge that paths to redress remain vague.