COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Strasbourg, 17 October 2018

T-PD(2018)17

**CONSULTATIVE COMMITTEE OF THE CONVENTION FOR THE PROTECTION
OF INDIVIDUALS WITH REGARD TO AUTOMATIC PROCESSING
OF PERSONAL DATA**

(Convention 108)

DRAFT GUIDELINES ON ARTIFICIAL INTELLIGENCE

Directorate General of Human Rights and Rule of Law

The present Guidelines provide a set of baseline measures which governments, AI developers, AI manufacturers, and AI service providers should follow to secure the human dignity and the human rights and fundamental freedoms of every individual, in particular with regard to personal data protection.[1]

Nothing in the present Guidelines shall be interpreted as precluding or limiting the provisions of the European Convention on Human Rights and of Convention 108 as amended ("Convention 108+")[2].

I. General guidance

1. Responsibility towards individuals and society is the corollary of any AI development, taking the safeguard of fundamental rights as an absolute pre-requisite.
2. A fundamental rights-oriented perspective is to be adopted by AI development and AI applications, in particular when AI is used in the context of decision-making processes.
3. AI development relying on personal data must be based on the principles of Convention 108+. The key elements of this approach are: proportionality of data processing, responsibility, transparency and risk management.
4. A risk-aware approach is not a barrier to innovation, but an enabler and the risks of datafication and the potentially adverse implications of data-driven solutions should thus be considered.
5. Individuals and communities should have the right to freely decide what role AI should play in analysing collective behaviour, influencing social dynamics, and in decision-making processes affecting entire groups of individuals.
6. In line with the guidance on risk assessment provided in the Guidelines on Big Data[3], a wider view of the possible outcomes of data processing should be adopted to consider the impact of data use not only on fundamental rights but also on collective social and ethical values.
7. AI development and AI applications cannot diminish or negatively affect the rights of data subjects enshrined in Convention 108+.

II. Guidance for AI developers

1. The Committee of Convention 108 actively encourages AI developers to adopt a value-oriented design of their products and services, consistent with Convention 108+ and other relevant instruments of the Council of Europe.
2. AI developers have to assess the adverse consequences of AI applications on the fundamental rights and freedoms of data subjects. Considering such consequences, precautionary approach based on risk prevention policies have to be adopted.
3. AI developers have to adopt a by-design approach to avoid potential unintentional and hidden data biases, and the risk of discrimination or negative impacts on the rights and fundamental freedoms of data subjects, in all phases of the processing, including data collection and analysis stages.
4. In developing AI applications, it is important to adopt a design paradigm that critically assesses the nature and amount of data used. Such design paradigms aim at reducing redundant or marginal data, starting with a restricted amount of training data, and then monitoring the model's accuracy as it is fed with new data. The use of synthetic data can be considered as one of the possible solutions to minimise personal data processed.

---

[1] These Guidelines derive from and are built upon the Report on Artificial Intelligence ("Artificial Intelligence and Data Protection: Challenges and Possible Remedies") available at
https://rm.coe.int/report-on-artificial-intelligence-artificial-intelligence-and-data-pro/16808e6012
[2] Amending Protocol CETS n°223 to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data.
[3] Guidelines on big data adopted by the Committee of Convention 108 in January 2017, available at
https://rm.coe.int/CoERMPublicCommnSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a.

5. The risk of de-contextualised data (i.e. ignoring the contextual information characterising the specific situations where the proposed AI-based solutions should be applied) and de-contextualised algorithmic models (i.e. using AI models originally designed for different contexts or purposes) should be adequately considered in developing AI applications.

6. Committees of experts from a range of fields, as well as independent academic institutions, should be involved in AI development to provide a valuable support in designing rights-based and socially-oriented AI and to contribute to detect potential bias. Such committees play an even more important role in areas where transparency and stakeholders' engagement are difficult to achieve, such as for instance in AI designed to be used in a judicial or law enforcement context.

7. Participatory forms of risk assessment, based on the active engagement of the groups potentially affected by AI applications, should be adopted.

8. When it is technically feasible, AI developers should design their products and services in a manner that safeguards users' freedom of choice over the use of AI and provide alternatives to AI-equipped devices and services.

9. Data subjects are entitled to know the AI applications used and the reasoning underlying AI data processing operations, including the consequences of such a reasoning.

III. Guidance for policy makers

1. Public procurement procedures could impose specific duties of transparency and prior assessment of AI systems to service providers.

2. Public trust in AI products and services could benefit from an increased AI developers' accountability and the adoption of risk assessment procedures.

3. Data protection supervisory authorities and data controllers should adopt forms of algorithm vigilance to better ensure compliance with data protection and human rights principles over the entire lifetime of AI applications.

4. Overconfidence in the reliable nature of the solutions provided by AI systems, and fears of potential liability when taking a different decision than the one suggested by AI systems risk altering the autonomy of human intervention in decision-making processes. It is thus crucial that the freedom of human decision makers not to rely on the result of the recommendations provided using AI be preserved.

5. When AI applications may significantly impact on the rights and fundamental freedoms of data subjects, data controllers should consult the supervisory authorities to seek advice to mitigate this potential adverse impact.

6. Countries having established independent bodies supervising specific sectors where AI applications operate or may operate, should strengthen the mutual cooperation between these bodies and their cooperation with data protection supervisory authorities.

7. When committees of experts are created at company level, data protection supervisory authorities should be empowered to scrutinise those committees when shortcomings in their independency, abilities or decisions affect data processing.