COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Strasbourg, 30 November 2018

T-PD(2018)17rev

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

AI-based applications[1] are providing new and valuable solutions to tackle needs and address challenges in a variety of fields, from smart homes to smart cities, from the industrial sector to healthcare and defence, supporting evidence-based and inclusive policies. As may be the case with other technological innovations, these applications may have adverse consequences for individuals and society. In order to prevent and avoid this, the Parties to Convention 108 will ensure and enable an AI development focused on the safeguard of human rights and fundamental freedoms.

The present Guidelines to this end 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.[2]

Nothing in the present Guidelines shall be interpreted as precluding or limiting the provisions of the European Convention on Human Right and of Convention 108. These Guidelines also take into account the new safeguards of the modernised Convention 108 (more commonly referred to as "Convention 108+")[3].

I. General guidance

1. The safeguard of human rights and fundamental freedoms, and in particular human dignity and the right to the protection of personal data, is an absolute pre-requisite in developing and adopting AI applications that may have effects on individuals and society.
2. AI development and AI applications which affect individuals and society must be fundamental rights-oriented. This is even more important when AI applications are used in the context of decision-making processes.
3. AI development relying on personal data should be based on the principles of Convention 108+. The key elements of this approach are: lawfulness, fairness, purpose specification, proportionality of data processing, responsibility and demonstration of compliance, transparency, and risk management.
4. A risk-aware approach, focused on the potential risks of data-driven applications, is a necessary element of responsible innovation in the field of AI.
5. In line with the guidance on risk assessment provided in the Guidelines on Big Data adopted by the Committee of Convention 108 in 2017, a wider view of the possible outcomes of data processing should be adopted to consider the impact of data use not only on human rights and fundamental freedoms but also on collective social and ethical values.

[1] A definition of AI is available here: https://www.coe.int/en/web/human-rights-rule-of-law/artificial-intelligence/glossary.

[2] These Guidelines follow and build on the Report on Artificial Intelligence ("Artificial Intelligence and Data Protection: Challenges and Possible Remedies") available at *****

[3] Amending Protocol CETS n°223 to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data.

6. AI development and AI applications must be done in full respect of the rights of data subjects.
7. AI applications must allow users to have control over the purposes of data processing and related effects both on a collective and individual level.

II. Guidance for developers, manufacturers and service providers

1. AI developers, manufacturers and service providers should adopt a value-oriented approach in the design of their products and services, consistent with the principles of Convention 108+ and other relevant instruments of the Council of Europe.
2. AI developers, manufacturers and service providers have to assess the adverse consequences of AI applications on human rights and fundamental freedoms of data subjects, and considering these effects to adopt a precautionary approach based on risk prevention policies.
3. In all phases of the processing, including data collection and analysis stages, AI developers, manufacturers and service providers shall adopt a human rights by-design approach to avoid potential unintentional and hidden data biases, and the risk of discrimination or negative impacts on the human rights and fundamental freedoms of data subjects.
4. In developing AI applications, developers shall critically assess the nature and amount of data used, 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[4] can be considered as one of the possible solutions to minimise the amount of personal data processed by AI applications.
5. The risk of de-contextualised data[5] and de-contextualised algorithmic models[6] should be adequately considered in developing and using AI applications.
6. AI developers, manufacturers and service providers may be supported by independent committees of experts from a range of fields, as well as independent academic institutions, which can contribute to design rights-based and ethically- and socially-oriented AI applications, and to detect potential bias. Such committees play an even more important role in all areas where the respect for human rights and fundamental freedoms is crucial and where transparency and stakeholders' engagement can be more difficult to achieve due to competing interests and rights, such as in the fields of predictive justice, crime detection or predictive policing.

[4] Synthetic data are generated from a data model built from real data. They should be representative of the original real data. See the definition of synthetic data in OECD. 'Glossary of Statistical Terms'. 2007. http://ec.europa.eu/eurostat/ramon/coded_files/OECD_glossary_stat_terms.pdf ("An approach to confidentiality where instead of disseminating real data, synthetic data that have been generated from one or more population models are released").

[5] This is the risk of ignoring contextual information characterising the specific situations in which the proposed AI-based solutions should be used.

[6] This happens when AI models, originally designed for a specific application, are used in a different context or for different purposes.

7. Participatory forms of risk assessment, based on the active engagement of the individuals and groups potentially affected by AI applications, should be considered.
8. In order to enhance users' trust, AI developers, manufacturers and service providers are encouraged to design their products and services in a manner that safeguards users' freedom of choice over the use of AI and provides alternatives to AI-equipped devices and services.
9. Data subjects are entitled to be informed appropriately when they are interacting with an AI application, and to know the AI applications used and the reasoning underlying AI data processing operations, including the consequences of such a reasoning.
10. Individuals should be entitled to object to technologies having an impact on their opinions and personal development.

III. Guidance for legislators and policy makers

1. Without prejudice to secrecy safeguarded by law, public procurement procedures should impose to service providers specific duties of transparency, prior assessment of the impact of data processing on human rights and fundamental freedoms, and vigilance on the potential adverse effects and consequences of AI applications (hereinafter algorithm vigilance[7]).
2. Trust in AI products and services could be enhanced by respect for the principle of accountability, the adoption of risk assessment procedures and the application of other suitable measures, such as codes of conduct and certification mechanisms.
3. 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. When AI applications may significantly impact on the human rights and fundamental freedoms of data subjects, supervisory authorities should be encouraged to develop algorithm vigilance programmes, which promote accountability of all the relevant stakeholders throughout the entire life cycle of these applications.
4. Overreliance in the solutions provided by AI applications, and fears of potential liability when taking a decision other than the one suggested by AI applications risk altering the autonomy of human intervention in decision-making processes. The autonomy of human intervention in decision-making processes and the freedom of human decision makers not to rely on the result of the recommendations provided using AI shall therefore be preserved.

[7] On the notion of algorithmic vigilance, see also 40th International Conference of Data Protection and Privacy Commissioners, Declaration on Ethics and Data Protection in Artificial Intelligence, Tuesday 23rd October 2018, Brussels, guiding principle no. 2. See also the Report on Artificial Intelligence (fn 1), Section II.4

5. Supervisory authorities should be consulted by controllers, when AI applications may significantly impact on the human rights and fundamental freedoms of data subjects.
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. Appropriate mechanisms should be put in place to ensure the independence of the committees of experts mentioned in Section II.6.
8. Individuals, groups, and stakeholders should be informed and actively involved in the debate on what role AI should play in shaping social dynamics, and in decision-making processes affecting them.
9. Policy makers should invest resources in digital literacy and education to increase data subjects' awareness and understanding of AI applications and their effects. They should also encourage professional training for AI developers to raise awareness and understanding of the potential effects of AI on individuals and society.