COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAI)

OUTLINE OF HUDERIA RISK AND IMPACT ASSESSMENT METHODOLOGY

Document prepared by the CAI Secretariat

www.coe.int/cai

Introduction

In line with the recommendations of the CAHAI (see paragraphs 45 to 53 of Possible elements of a legal framework on artificial intelligence, based on Council of Europe's standards on human rights, democracy and the rule of law, adopted by the CAHAI in December of 2021), the Secretariat is working on a methodology providing clear, concrete and objective criteria for identifying such artificial intelligence systems or combined technologies based on such systems that are likely to pose significant levels of risk to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law. The methodology would ensure a uniform approach towards identification, analysis and evaluation of risk and assessment of impact of such systems in relation to the enjoyment of human rights, the functioning of democracy and the observance of rule of law.

It could to a large extent be based on the work of the CAHAI Policy Development Group (PDG), specifically its Sub-Working Group 1 on human rights impact assessment and ideas set out in the "Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems" document prepared by The Alan Turing Institute at the request of the Secretariat of the Council of Europe's Ad hoc Committee on Artificial Intelligence (CAHAI).

The model should be based on the assumption that the domestic authorities are usually better placed to make the relevant regulatory choices, taking into account the political, economic, social, technological and other types of contexts, and that they should enjoy a certain margin of appreciation in this sphere. In this view, the role of the methodology is to assist the domestic authorities in identifying such artificial intelligence systems or combined technologies based on such systems that are likely to pose significant levels of risk to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law and help them manage the related risks.

Another equally important aspect of the methodology is to ensure seamless compatibility of our approach with the existing compliance practices followed by the industry, to be achieved through, inter alia, consultations with the relevant stakeholders representing both large and small and medium sized enterprises.

It is important to note that a legally binding instrument should also require future Parties to adopt appropriate measures, in particular regarding the manner in which the competent domestic authorities would supervise and control the application of the methodology by the relevant actors as well as the training of national supervisory authorities, artificial intelligence providers and artificial intelligence users, ensuring that the relevant actors are capable of applying it.

General approach

The methodology should be designed to be as "algorithm neutral" and practice-based as possible so that it can remain maximally future proof and inclusive of different AI applications. The model will, however, need to stay responsive to the development of novel AI innovations and use-cases, and should be seen as dynamic and in need of regular revisitation and re-evaluation. Hence, the choice of putting it in the Appendix to the future legal instrument, which should facilitate the revision of the methodology in the future.

Key elements of the methodology

The methodology will likely be composed of a number of clearly articulated and interrelated processes and instruments/steps:

1) A context-based risk analysis (COBRA) provides an initial indication of the context-based risks that an AI system could pose to human rights, democracy and the rule of law. Its main purpose is to identify whether an artificial intelligence system, in view of the context of its deployment and future use as well as other relevant factors, is at all likely to pose significant levels of risk to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law.

   This part will serve as a threshold test and consist of a series of specific questions based on the relevant case-law of the European Court of Human Rights for human rights and the specialised research for democracy and the rule of law (both conducted by the team of experts with the assistance of the Secretariat), the answers to which will give a fairly clear indication as to whether the remaining steps of the methodology should be triggered. The artificial intelligence systems which have not been identified as such will fall outside the scope of the methodology and will not require further assessment.

2) Provided that an artificially intelligence system has been identified as likely to pose significant levels of risk to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law, and depending on the exact level of risk identified, at the next stage the methodology would require some form of stakeholder engagement process to help project teams identify stakeholder salience and to facilitate proportionate stakeholder involvement and input throughout the project workflow.

   It is important to underline that the basic idea behind this stage is simply to improve the quality of the risk and impact assessment by amplifying the perspective of the actors whose rights and interests could be potentially at stake. In any event, the level of such involvement should be proportionate to the level of risk and other relevant factors. The methodology will define certain basic principles whilst remaining quite flexible and, in the end, it would be up to the Parties to define the exact modalities of this part of the process.

3) The core of the methodology would be the actual Human Rights, Democracy, and the Rule of Law Impact Assessment (HUDERIA) which will consist essentially of an obligation to address:

(a) a certain number of specific sociotechnical questions regarding the artificial intelligence system under examination and
(b) questions reflecting the considerations that are specific to human rights, democracy and the rule of law based on the standards of the Council of Europe.

   Both sets of questions are currently under development by a team of experts (with the assistance of the Secretariat).

   The process of answering these questions essentially contextualises and corroborates potential harms which have been previously identified, enables the discovery of further harms through the integration of stakeholder perspectives, makes possible the collaborative assessment of the severity of potential adverse impacts identified, facilitates the co-design of an impact mitigation plan, sets up access to remedy, and establishes monitoring and impact eventual re-assessment protocols.

   At final stages (4) and (5) the methodology would require the composition of an impact mitigation plan as well as, if appropriate, mechanisms providing access to remedy(ies) and lastly contain a requirement that the carrying out of such assessments would have an iterative character. This means that it should be repeatable at regular intervals

within the lifecycle of the artificial intelligence system in question, making sure that the relevant intervening changes in both the context and the system itself are properly identified, understood, taken account of and dealt with.

***