

SETTING UP AN ETHICAL FRAMEWORK AS A FIRST STEP TO COMPREHENSIVE REGULATION OF ARTIFICIAL INTELLIGENCE TOOLS IN THE JUSTICE SYSTEM

There is an urgent need for a binding and comprehensive legal instrument which regulates the use of artificial intelligence in the field of Justice. The European Ethical Charter on the use of artificial intelligence in judicial systems and their environment provides a non-binding approach to the subject and contains five fundamental principles which must govern any artificial intelligence software used in courts or by legal professionals. The use of artificial intelligence in Law and Justice may be highly beneficial, but the perils it also entails encourage the establishment of a more robust legal framework to prevent any compromise of fundamental rights.

João Arsénio de Oliveira*



* João Arsénio de Oliveira is the President of the Working Group on Quality of Justice (CEPEJ) and works at the Ministry of Justice of Portugal.

The World lacks a comprehensive regulation on artificial intelligence. The judicial legal systems (international, national, regional, and local) lack a comprehensive and binding regulation on artificial intelligence applicable to Courts and the judicial environment, as a whole.

The Council of Europe, however, keeps striving to establish guidelines for the use of artificial intelligence in all fields of knowledge, as well as for the judicial systems, with the main aim (which is in line with its own goals) of avoiding any kind of infringement of the rights enshrined in the European Convention of Human Rights and, therefore, preserving the Rule of Law and Democracy. Although it is a regional international organization, its pioneer work serves as a model for other continents.

The Council of Europe decided to take a step-by-step approach. It started by establishing an Ethical Charter on the use of artificial intelligence in judicial systems and their environment¹, while, at the same time, but at a slower pace, set up a group to establish a binding instrument on artificial intelligence, which will also have provisions applicable to Courts and Law. This group was recently replaced by a new formation, whose main mission will be to draft an international instrument aimed at setting the binding rules for artificial intelligence in Europe (and, most probably, beyond).

In fact, the Council of Europe is the international organization most concerned with artificial intelligence and its implications in society, although other organizations, such as the United Nations and OECD, are also working on the subject.

One of the bodies of the Council of Europe, the European Commission for the Efficiency of Justice (CEPEJ), has been working on the issue for some time. One of its most apparent outcomes is the already alluded European Ethical Charter on Artificial Intelligence in the Legal System and its Environment. In development of the ethical framework of the Member States of the Council of Europe regarding the use of the set of sciences, theories, and techniques designed to mimic human cognition by machines, the Ethical Charter forms the cornerstone of its design.²

The Charter is intended for public and private stakeholders responsible for the design and deployment of artificial intelligence tools and services that involve the processing of judicial decisions and data (machine learning or any other methods deriving from data science).³ It also concerns public decision-makers in charge of

¹ Which can be found on <https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment> (February 2022).

² João Arsénio de OLIVEIRA, « La Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires, » in *Revue Experts* n° 141, (Décembre 2018), p. 30.

³ CEPEJ, *The European Ethical Charter of the CEPEJ*, (2019): p. 5.

<https://experts-institute.eu/en/europe-of-justice/cepej-en/the-european-ethical-charter-of-the-cepej/>

the legislative or regulatory framework, of the development, audit or use of such tools and services.⁴

The Charter consists of five major principles and four appendices. A detailed study of the use of artificial intelligence in judicial systems is presented in the first appendix. In the second one, using artificial intelligence for its intended uses is encouraged, but only for those requiring considerable methodological precautions. This is a set of uses that should be considered after further research. An appendix contains a list of uses that should be regarded with the greatest caution; a glossary forms the third appendix and the final annex contains a checklist for integrating the Charter's principles into processing methods.

“The principle of quality and security requires that automatic learning can be performed on certified originals and that the integrity of these data is ensured at all stages of processing.”

The five principles enshrined in the Charter are:

1. The Principle of Respect for Fundamental Rights: ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights;
2. The Principle of Non-Discrimination: specifically prevent the development or intensification of any forms of discrimination between individuals or groups of individuals;
3. The Principle of Quality and Security: process judicial decisions and data using certified sources and intangible data derived from models designed in a multidisciplinary manner, within a technologically secure environment;
4. The Principle of Transparency, Impartiality and Fairness: make data processing methods accessible and understandable, authorize external audits;
5. The Principle «Under User Control»: preclude a prescriptive approach and ensure that users are informed actors and in control of their choices.

⁴ CEPEJ, (2019).

The first principle stipulates that, human rights must be respected from the inception of any artificial intelligence instrument (human rights by design). It intends to make sure that all artificial intelligence solutions respect the rights foreseen in the European Convention of Human Rights, as well as other relevant Council of Europe instruments.

Within the *acquis* of fundamental rights foreseen in the European Convention of Human Rights, the CEPEJ considered necessary to highpoint the principle of non-discrimination, since it is extremely susceptible to be breached by artificial intelligence tools in the field of Justice (mainly in criminal matters), as is better explained below.

The principle of quality and security requires that automatic learning can be performed on certified originals and that the integrity of these data is ensured at all stages of processing. The principle also encourages a multidisciplinary approach, urging the designers of machine learning models to rely on the expertise of the relevant justice system professionals and researchers in the fields of Law and social sciences. Moreover, it requires that data must come from certified sources and that it should not be modified until it has been used.

The principle of transparency, impartiality and fairness indicates that it is necessary to find a balance between the intellectual property of the processing methods and the need for transparency (access to the design process), impartiality (absence of bias), fairness and intellectual integrity (prioritizing the interests of Justice).

Last but not least, the principle ‘under user control’ emphasizes the importance of the human being and affirms that all artificial intelligence tools should focus on the person above all in the field of justice. Corollaries of this principle determine that, legal professionals should, at any moment, be able to review judicial decisions taken by IT and that the subjects of artificial intelligence decisions should be conferred special information rights. At the same time, the knowledge and understanding of AI should be promoted in government institutions, independent oversight bodies, national human rights structures, the judiciary and law enforcement, with addition to the general public.⁵

CEPEJ’s Plenary Meeting approved a revised roadmap in December 2021 for ensuring an appropriate follow-up to the CEPEJ Ethical Charter, which established (again) five key elements aimed at increased application of the Charter.

⁵ Commissioner for Human Rights, *Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights* (Recommendation) (Strasbourg: Council of Europe, May 2019): p. 14.

The first one comprises developing more practical guidance to legal professionals and IT developers on how to apply the five principles laid down in the CEPEJ Charter, through the detailed operationalization of the five principles. This means developing a detailed description of what needs to be checked and how to ensure compliance with the respective principle.

The second idea consists of creating a pilot project on CEPEJ Charter conformity assessment. It engages the operationalization of the checklist attached to the Charter on one artificial intelligence program, through a test case led by the CEPEJ secretariat.

“Even though artificial intelligence is promising (but also a threat) to all fields of law, it seems to pose a greater threat to criminal and criminal procedure law in particular.”

Furthermore, the CEPEJ has decided to create a five member Artificial Intelligence Advisory Board (AIAB) to monitor the emergence of artificial intelligence applications in the justice system, to discuss issues, and to propose new ways to protect fundamental rights when using artificial intelligence. The Advisory Board will meet virtually every trimester.

The Advisory Board will also be in charge of feeding the Resource Centre on Artificial Intelligence, which will be a publicly accessible database containing reliable and exhaustive overview of the existing artificial intelligence programs in the judicial area.

The final vector of the roadmap relies on the need to increase the awareness of the CEPEJ Charter, by means of concrete training activities.

Even though artificial intelligence is promising (but also a threat) to all fields of law, it seems to pose a greater threat to criminal and criminal procedure law in particular. Several experiences have already highlighted how dangerous automatic interventions can be in the penal realm, and how aggressive they can be towards human rights. By itself, this justifies establishing a binding legal instrument applicable to the use of artificial intelligence in the judicial process.

Not long ago, a report from ProPublica⁶ described that an artificial intelligence tool used in courtrooms across the United States to predict future crimes, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), was biased against black defendants.⁷

Often, the problem relies on the data used to feed the artificial intelligence machine, as it can only proceed if supplied with data. If the data fed to the machine already comprises biases, the machine and the algorithm on which it relies will only perpetuate the discriminatory behaviour and, therefore, contribute to aggravate the discrimination.

According to a recent article, researchers in China have developed a machine, using artificial intelligence that can charge people with crimes. This artificial intelligence «prosecutor» can file a charge with more than 97 percent accuracy, based on a verbal description of the case, according to the same source.⁸ The lack of human intervention is scary and, as one Prosecutor interviewed by the same reporter has stated an AI prosecutor could file a charge based only on its previous experience. It could not foresee the public reaction to a case in a changing social environment. AI may help detect a mistake, but it cannot replace humans in making a decision.

Fundamental rights such as freedom of expression and assembly, right to privacy, as well as non-discrimination are at high risk of being violated by artificial intelligence tools. For this reason, we thought it fundamental to have a binding (hopefully global) instrument that regulates the use of artificial intelligence in the justice system.

⁶ ProPublica is, according to its own words “An Independent, Non-Profit Newsroom that Produces Investigative Journalism with Moral Force,” <https://www.propublica.org>

⁷ Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, “Machine Bias,” *ProPublica*, 23 May 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

⁸ Stephen Chen, “Chinese Scientists Develop AI ‘Prosecutor’ that can press its own Charges,” *South China Morning Post*, 26 December 2021, <https://www.scmp.com/news/china/science/article/3160997/chinese-scientists-develop-ai-prosecutor-can-press-its-own>