22 May 2025

EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE

(CEPEJ)

REPORT ON THE FIRST USE OF THE SELF-ASSESSMENT TOOL FOR AI SYSTEMS IN THE JUDICIAL FIELD

GENERAL ASPECTS

Document prepared by Matthieu Quiniou

I.   Introduction and background

In 2018, the European Commission for the Efficiency of Justice (hereinafter "CEPEJ") adopted its European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment (hereinafter "the CEPEJ Charter").

The CEPEJ Charter sets out five key principles that should be upheld by the judiciary in the design and use of artificial intelligence (hereinafter AI):

1) Principle of respect for fundamental rights: ensuring that AI tools and services are designed and implemented in a manner compatible with fundamental rights;
2) Principle of non-discrimination: specifically preventing the creation or reinforcement of discrimination between individuals or groups of individuals;
3) Principle of quality and security: ensuring that, in the processing of judicial decisions and judicial data, certified sources and reliable data are used, with models designed in a multidisciplinary manner within a secure technological environment;
4) Principle of transparency, neutrality, and intellectual integrity: making data processing methods accessible and understandable, and allowing for external audits;
5) Principle of user control: avoiding a prescriptive approach and enabling users to be informed actors in control of their choices.

In this context, an Assessment Tool for the Operationalization of the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment (hereinafter "the Assessment Tool") was designed and published on December 4, 2023. This tool aims to implement the CEPEJ Charter by establishing a set of verifications, key measures, and safeguards that decision-makers within judicial systems should follow when purchasing, designing, developing, implementing, and/or using AI in judicial systems and their environment, in compliance with the CEPEJ Charter.

The Assessment Tool consists of 29 questions designed to raise awareness and identify potential breaches of the CEPEJ Charter's principles by an AI system, particularly those related to human rights and the rule of law.

This tool was tested in 2024 in four European jurisdictions. These jurisdictions conducted a self-assessment of their respective AI systems, with CEPEJ providing support to clarify certain points.

Several meetings were held with various representatives of the European jurisdictions, both in groups and individually, to facilitate responses to the self-assessment questionnaires.

This report aims to provide an initial summary of the key points of these self-assessments, the content of the responses provided by the jurisdictions, and the aspects of the self-assessment questionnaire that require clarification or improvement.

II.   Summary of self-assessments and suggestions for improving the self-assessment tool

In this context, several European judicial institutions took part in this experimental phase of the self-assessment tool with different systems deployed:

- In Spain: a speech-to-text system was self-assessed with the contact point being Mr Alejandro Fernández Muñoz, Area Manager, Support Unit of the Directorate General, DG for Digital Transformation of Justice, Ministry of the Presidency, Justice and Parliamentary Relations.

- In Estonia: a court transcription system, Salme, was self-assessed with Mr Indrek Tops, Head of the Data Analysis team at the Centre of Registers and Information Systems, as the point of contact.

- In France, a system for pseudonymising court rulings as part of the Open Data obligation has been self-assessed, with Mr Edoaurd Rottier, Conseiller référendaire, head of the Open Data for court rulings unit of the Documentation Service, as the point of contact.

- In Italy, a system for automatically assigning and sorting criminal files was self-assessed, with Mr Stefano Brunetti, Funzionario tecnico di edilizia senior, Ministero della Giustizia, Corte d'Appello di Bologna, as the point of contact.

Handling the self-assessment tool did not pose any difficulties, and the purpose and wording of the questions therefore appear to be suited to the types of people designated to be in charge of the self-assessment.

Participants in the self-assessment recommended a few changes to the wording for greater clarity: rewording of question 21 (Italian self-assessment) and clarification of question 10 (Spanish self-assessment).

The following reformulations are therefore proposed for a new version of the self-assessment tool:

- Question 10: "Is the AI system likely to provide an advantage (e.g. ~~real-time~~ data processing during witness interviews) to its user(s) in legal proceedings?"

- Question 21: Is the source code auditable *(technically auditable, no secrets limiting auditability, etc.)*? ~~Is a trade secret likely to hinder the auditability of the AI system?~~

The four self-assessed systems at this stage are limited-risk systems belonging to three categories (triage, pseudonymisation, transcription) of the nine categories of judicial AI listed by the CEPEJ's Cyberjustice and AI Resource Centre. It would be useful to observe the use of the self-assessment tool for each of the categories of judicial AI, especially AI systems in the highest-risk areas such as decision support, litigation outcome prediction or automated dispute resolution. This systematic approach would make it possible to ensure the relevance of the self-assessment tool, particularly for AI systems for which the assessment seems imperative because of the risks they present to fundamental rights. It should be noted that several participants in the self-assessments indicated that some questions were particularly relevant to high-risk systems (and not to their case), which tends to confirm the interest to apply the self-assessment tool for high-risk AI systems that were not assessed as part of this piloting by the CEPEJ. It would certainly be advisable to include new questions dedicated to

low-risk systems and possibly to separate the questionnaire more clearly according to risk and/or type of use.

It should also be noted that the self-assessments carried out as part of this pilot work were carried out with AI systems already implemented or in the process of being implemented, and not at the stage of finalising the specifications or choosing pre-existing software, as suggested for the use of the self-assessment tool. As a result, the participants in this self-assessment had already carried out more or less in-depth audits of their systems. However, according to the representatives of the systems participating in this self-assessment, the self-assessment tool was useful as a complement to the assessments already carried out previously and could be useful to them in the future as a standardised tool for assessing AI systems in the judicial domain. The self-assessment tool therefore appears to be a tool that can be used at the various stages of the production and implementation of an AI system in the judicial field.

Furthermore, in the context of discussions with the Artificial Intelligence Advisory Board (AIAB) of the CEPEJ, the value of working with graphic representations to facilitate the self-assessment process was raised, as well as the creation of a dynamic online form to improve the questionnaire response experience (automatically excluding irrelevant questions based on previous answers and the type of AI system analysed).

The creation of bridges between this self-assessment tool and the work carried out by the CAI of the Council of Europe with the HUDERIA method was also presented by the CEPEJ's AIAB as a perspective to be explored.

Appendices :

1 - Italy
2 - Spain
3 - Estonia
4 - France

Appendix 1: Italy

I.   Self-evaluation summary

In Italy, GIADA 2 is used as an automatic tool for assigning criminal cases. It was developed by a service provider based on the specifications of the judicial institution.

GIADA 2 is trained with a set of data taken directly from the Court Case Management System (SICP - RegeWEB). The data flow goes directly from the case management system to the AI system, and only registered users can update the data used. Every action (logging in, consulting and modifying data, etc.) is recorded and can be reviewed if necessary. The annotation of data and tokens was carried out by legal and ethics professionals.

The AI system does not collect sensitive data, nor does it authorize or facilitate the profiling of judges in the triage process.

The dataset and model were audited (model explicability audit, data audit, A/B test of results) by the multi-disciplinary team of the judicial institution and by third parties in order to identify possible biases. No biases were detected during the audits.

Monitoring mechanisms have been put in place to guarantee a secure environment: "IT systems are hosted and run in a safe and secure environment, distributed across different server rooms across the country, separated from the WEB by a central proxy".

GIADA 2 is auditable, and training and documentation have been provided for users (FAQ, detailed user guide and test run).

II.   Notes on the self-assessment tool

It was mentioned that the AI system submitted for self-assessment is a triage system, which is inherently less likely to create significant risks than a decision support or automated decision-making system. Some of the self-assessment questions were not suitable for this type of low-risk system.

Appendix 2: Spain

I.   Self-evaluation summary

In Spain, a voice-to-text recognition system ("textualization") was developed in-house by the Ministry of the Presidency, Justice and Parliamentary Relations, using a fine-tuned version of the Whisper and Pyannote models, available in Hugging Face under MIT license.

Monitoring mechanisms are in place to ensure data security. Data is provided in read-only file systems. The model, data set and drive machines are located on a restricted-access network.

The AI system may provide erroneous results, but this is clearly explained to the user, and recommendations are suggested. Video tutorials, a detailed user guide and a comprehensive set of training materials are available to help train users.

II.   Notes on the self-assessment tool

It was mentioned that the AI system submitted for self-assessment is a speech-to-text recognition system, which is inherently less likely to create significant risks than a decision support or automated decision-making system. Some of the self-assessment questions were not suitable for this type of low-risk system.

Appendix 3: Estonia

I.   Self-evaluation summary

In Estonia, Salme is used as a natural language processing solution for audio recording of court hearings, transcription and generation of court reports. It was developed by a service provider based on the specifications of the judicial institution.

In the event of an AI malfunction, this transcription task must be performed by a human.

To minimize the risk of personal data vulnerabilities, the respondent said, "Many security solutions, no third-party providers."

The training model (as well as the third-party pre-trained model) was not audited for potential bias. However, the respondent did identify a potential bias in the understanding of some dialects, due to unrepresentative sampling and training data, a bias mitigated by the use of richer datasets.

To monitor the mechanisms, the respondent emphasized human assessment, the use of recent models trained in-house, and the absence of third-party access.

Documentation (FAQ and detailed user guide) has been created to accompany the use of the AI system.

II.   Notes on the self-assessment tool

It was mentioned that the AI system submitted for self-assessment is a transcription system, which is inherently less likely to create significant risks than a decision support or automated decision-making system. Some of the self-assessment questions were not suitable for this type of low-risk system.

Appendix 4: France

I.   Self-evaluation summary

The Cour de cassation wanted to test the self-evaluation tool on a tool for pseudonymizing court rulings before releasing them in open data, developed in-house and hosted on internal servers, in order to meet the regulatory requirements stipulating the principle of open data for court rulings.

According to the questionnaire, this AI system is trained by annotators and is designed to meet the pseudonymization needs of the 3 to 5 million decisions that will be transmitted to the Cour de cassation every year once the open data rollout is complete

The AI system handles sensitive personal data or trade secrets, and risk mitigation systems have been put in place, using the Réseau privé virtuel justice (RPVJ) which integrates the servers of the Cour de cassation. The AI system in place is designed to limit the exposure and export of sensitive data to external servers in the context of open data.

The AI system was audited at various levels: source code, model explanation and data.

Various biases were identified, in particular the more delicate and therefore less systematic pseudonymization of foreign localities compared to French localities, which could discriminate against people living abroad. This bias was corrected thanks to the expertise of the S.D.E.R.'s human annotators, engineers and magistrates, by modifying the model to take greater account of document structure and less of the typology of locality addresses.

The AI model is based on a third-party model, the CamemBERT model developed by a French public research organization, INRIA. The training data for this model are expert-generated and taken from the French sub-corpus of the OSCAR multilingual corpus

In addition, the model was refined on the basis of official and unadulterated court decisions contained in the databases of the Cour de cassation. The data was annotated by interdisciplinary professionals, including legal and ethics experts.

Mechanisms have been put in place to control data integrity from collection to AI model training, through the use of business applications managed by the court clerk's office, the guarantor of the procedure. In addition, commercial court decisions transmitted since December 2024 are electronically signed via a trusted third party before being transmitted and centralized within GIE Infogreffe, which feeds the API exposed by the Cour de cassation.

The model has been designed to guarantee transparency, neutrality and data integrity: the code is auditable, the data is verifiable, and the weighting criteria are clear and allow the model to be audited.

In addition, users are clearly informed that an AI system is being used, and are provided with a detailed user guide and dedicated training. As the AI system is liable to produce erroneous pseudonymizations, in addition to the recommendations, the Innovation Lab has developed a questioning algorithm to alert annotating agents to cases where the results of the pseudonymization model are less reliable.

II.   Notes on the self-assessment tool

We noted the convergence between the questions posed in the questionnaire and those already asked by the departments in charge of creating and deploying the AI system.

It was pointed out that the AI system undergoing self-assessment is a pseudonymization system, and therefore by its very nature less likely to create significant risks than a system providing automated decision support or decision-making. Some of the self-assessment questions were not adapted to this type of limited-risk system.