CEPEJ-GT-QUAL(2025)6rev2

22 mai 2025

COMMISSION EUROPÉENNE POUR L'EFFICACITÉ DE LA JUSTICE

(CEPEJ)

RAPPORT PREMIERE UTILISATION DE L'OUTIL D'AUTO-EVALUATION DES SYSTEMES D'IA DANS LE DOMAINE JUDICIAIRE

ASPECTS GENERAUX

Document préparé par Matthieu Quiniou

I. Introduction et contexte

En 2018, la Commission européenne pour l'efficacité de la justice (ci-après « CEPEJ ») a adopté sa Charte éthique européenne sur l'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement (ci-après « la Charte de la CEPEJ »).

La Charte de la CEPEJ énonce cinq principes clés qui devraient être respectés par le pouvoir judiciaire dans la conception et l'utilisation de l'intelligence artificielle (ci-après IA) :

1) principe de respect des droits fondamentaux : assurer une conception et une mise en œuvre des outils et services d'IA qui soient compatibles avec les droits fondamentaux ;

2) principe de non-discrimination : prévenir spécifiquement la création ou le renforcement de discriminations entre individus ou groupes d'individus ;

3) principe de qualité et de sécurité : en ce qui concerne le traitement des décisions juridictionnelles et des données judiciaires, utiliser des sources certifiées et des données intangibles, avec des modèles conçus de manière multi-disciplinaire, dans un environnement technologique sécurisé ;

4) principe de transparence, de neutralité et d'intégrité intellectuelle : rendre accessibles et compréhensibles les méthodes de traitement des données, autoriser des audits externes;

5) principe de maîtrise par l'utilisateur : bannir une approche prescriptive et permettre à l'usager d'être un acteur éclairé et maître de ses choix.

Dans ce contexte, un Outil d'évaluation pour l'opérationnalisation de la Charte éthique européenne sur l'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement (ci-après "l'Outil d'évaluation") a été conçu et publié le 4 décembre 2023. Cet outil vise à opérationnaliser ladite Charte de la CEPEJ en prévoyant un ensemble de vérifications, de mesures clés et de garanties que les décideurs au sein des systèmes judiciaires devraient suivre lorsqu'ils achètent, conçoivent, développent, mettent en œuvre et/ou utilisent l'IA dans les systèmes judiciaires et leur environnement, en conformité avec la Charte de la CEPEJ.

L'outil d'évaluation se compose de 29 questions conçues pour sensibiliser et identifier les atteintes potentielles d'un système d'IA aux principes de la Charte de la CEPEJ, et en particulier celles relatives aux droits humains et à l'État de droit.

Cet outil a été expérimenté en 2024 auprès de quatre juridictions européennes. Ces juridictions ont mené une auto-évaluation de leurs systèmes d'IA respectifs en étant accompagnées par la CEPEJ pour clarifier certains points.

Plusieurs réunions ont été réalisées avec différents représentants des juridictions européennes concernées en groupe et individuellement pour faciliter les réponses aux questionnaires d'auto-évaluation.

Le présent rapport vise à dresser une première synthèse des principaux points du déroulement de ces auto-évaluation, du contenu des réponses apportées par les juridictions et des points à clarifier ou améliorer dans le questionnaire d'auto-évaluation.

II. Synthèse des auto-évaluations et propositions d'amélioration de l'outil d'auto-évaluation

Dans ce contexte plusieurs institutions judiciaires européennes ont participé à cette phrase expérimentale de l'outil d'auto-évaluation avec différents systèmes déployés :

- En Espagne : un système de textualisation (speech-to-text) a été auto-évalué avec comme point de contact M. Alejandro Fernández Muñoz, responsable de secteur, unité de support de la Direction Générale, DG pour la Transformation Numérique de la Justice, Ministère de la Présidence, de la Justice et des Relations parlementaires.

- En Estonie : un système de transcription des audiences, Salme a été auto-évalué avec comme point de contact M. Indrek Tops, Responsable de l'équipe d'analyse des données au Centre des registres et des systèmes d'information, en tant que personne de contact.

- En France, un système de pseudonymisation des décisions de justice dans le cadre de l'obligation de l'Open data a été auto-évalué avec comme point de contact M. Edoaurd Rottier, conseiller référendaire, chef du pôle Open Data des décisions de justice du Service de Documentation, en tant que personne de contact,

- En Italie, un système d'affectation et de tri automatique des dossiers en matière criminelle a été auto-évalué avec comme point de contact M. Stefano Brunetti, Funzionario tecnico di edilizia senior, Ministero della Giustizia, Corte d'Appello di Bologna, en tant que personne de contact.

La prise en main de l'outil d'auto-évaluation n'a pas posé de difficulté, l'objet et les formulations des questions paraissent donc adaptés aux types d'interlocuteurs désignés pour être en charge de l'auto-évaluation.

Les participants à l'auto-évaluation ont conseillé quelques modifications de formulation pour plus de clarté : reformulation de la question 21 (auto-évaluation italienne) et clarification de la question 10 (auto-évaluation espagnole).

Les reformulations suivantes sont ainsi proposées pour une nouvelle version de l'outil d'auto-évaluation :

- Question 10 : « Le système d'IA est-il susceptible d'apporter un avantage (par exemple, traitement de données ~~en temps réel~~ lors d'entretiens avec des témoins) à son ou ses utilisateur(s) dans le cadre d'une procédure judiciaire ? »

~~G1.~~ Question 21 : Le code source est-il auditable *(techniquement auditable, absence de secret limitant l'auditabilité…)* ? ~~Un secret commercial est-il susceptible d'entraver l'auditabilité du système d'IA ?~~

Les quatre systèmes auto-évalués à ce stade sont des systèmes à risque limité appartenant à trois catégories (triage, pseudonymisation, transcription) des neuf catégories d'IA judiciaires listées par le Centre de Ressources sur la Cyberjustice et l'IA de la CEPEJ. Il paraîtrait intéressant de suivre l'utilisation de l'outil d'auto-évaluation pour chacune des catégories d'IA judiciaire, tout particulièrement des systèmes d'IA dans les domaines les plus à risque comme l'aide à la décision, la prévision des résultats des litiges ou la résolution automatisée des litiges. Cette approche systématique permettrait de s'assurer de l'intérêt de l'outil d'auto-évaluation notamment pour les systèmes d'IA pour lesquels l'évaluation paraît impérative en raison des risques qu'ils présentent pour les droits fondamentaux. Il convient de noter que plusieurs participants aux auto-évaluations ont indiqué que certaines questions étaient surtout pertinentes pour les systèmes à haut risque (et non pas dans leur cas) ce qui tend à confirmer l'intérêt de cet outil d'auto-évaluation pour les systèmes d'IA à haut risque qui n'ont pas été évalués dans le cadre de ce pilotage par la CEPEJ. Il conviendrait certainement d'inclure de nouvelles questions dédiées aux systèmes à faible risque et éventuellement de dissocier plus nettement le questionnaire en fonction du risque et/ou du type d'usage.

Il convient également d'indiquer que les auto-évaluations réalisées dans le cadre de ce travail de pilotage ont été réalisées avec des systèmes d'IA déjà implémentés ou en cours d'implémentation et non au stade de la finalisation du cahier des charges ou du choix d'un logiciel préexistant, comme suggéré pour l'utilisation de l'outil d'auto-évaluation. Il en résulte que les participants à cette auto-évaluation avaient déjà réalisé des audits plus ou moins approfondis de leurs systèmes. Pour autant, d'après les représentants des systèmes participants à cette auto-évaluation, l'outil d'auto-évaluation a été utile en complément des évaluations déjà réalisées antérieurement et pourrait leur être utile à l'avenir pour disposer d'un outil standardisé d'évaluation des systèmes d'IA dans le domaine judiciaire. L'outil d'auto-évaluation apparaît donc comme un outil pouvant être utilisé aux différentes étapes de la production et de l'implémentation d'un système d'IA dans le domaine judiciaire.

Par ailleurs, dans le cadre des échanges avec le bureau consultatif sur l'intelligence artificielle (AIAB) de la CEPEJ, il a été évoqué l'intérêt de travailler sur des représentations graphiques pour faciliter la réalisation de l'auto-évaluation ainsi que sur la création d'un formulaire dynamique en ligne pour améliorer l'expérience de réponse au questionnaire (en excluant automatiquement les questions non pertinentes en fonction notamment des réponses précédentes et du type de système d'IA analysé).

La création de passerelles entre cet outil d'auto-évaluation et les travaux qui ont été menés par le CAI du Conseil de l'Europe avec la méthode HUDERIA a également été présenté par l'AIAB de la CEPEJ comme une perspective à explorer.

Annexes :

1 - Italie
2 - Espagne
3 - Estonie
4 - France

Annexe 1 : Italie

I. Synthèse de l'auto-évaluation

En Italie, GIADA 2 est utilisé comme outil automatique d'attribution des affaires pénales. Il a été développé par un prestataire de services sur la base des spécifications de l'institution judiciaire.

GIADA 2 s'entraîne avec un ensemble de données provenant directement du Système de Gestion des Affaires Judiciaires (SICP – RegeWEB). Le flux de données va directement du système de gestion des affaires au système d'IA, et seuls les utilisateurs enregistrés peuvent mettre à jour les données utilisées. Chaque action (connexion, consultation et modification des données, etc.) est enregistrée et peut être examinée en cas de besoin. L'annotation des données et des jetons a été réalisée par des professionnels du droit et de l'éthique.

Le système d'IA ne collecte pas de données sensibles et n'autorise ni ne facilite le profilage des juges dans le processus de triage.

L'ensemble de données et le modèle ont été audités (audit d'explicabilité du modèle, audit des données, test A/B des résultats) par l'équipe pluridisciplinaire de l'institution judiciaire et par des tiers afin d'identifier d'éventuels biais. Aucun biais n'a été détecté lors des audits.

Des mécanismes de surveillance ont été mis en place pour garantir un environnement sécurisé : « Les systèmes informatiques sont hébergés et exécutés dans un environnement sûr et sécurisé, répartis dans différentes salles de serveurs à travers le pays, séparées du WEB par un proxy central ».

GIADA 2 est auditable, et une formation ainsi qu'une documentation ont été mises à disposition des utilisateurs (FAQ, guide utilisateur détaillé et test de prise en main).

II. Remarques liées à l'outil d'auto-évaluation

Il a été mentionné que le système d'IA soumis à l'auto-évaluation est un système de triage, qui est intrinsèquement moins susceptible de créer des risques significatifs qu'un système d'aide à la décision ou de prise de décision automatisée. Certaines questions de l'auto-évaluation n'étaient pas adaptées à ce type de système à faible risque.

Annexe 2 : Espagne

I. Synthèse de l’auto-évaluation

En Espagne, un système de reconnaissance vocale (« textualisation ») a été développé en interne par le ministère de la Présidence, de la Justice et des Relations parlementaires, à partir d'une version affinée des modèles Whisper et Pyannote, disponibles dans Hugging Face sous licence MIT.

Des mécanismes de surveillance ont été mis en place pour sécuriser les données. Les données sont fournies dans des systèmes de fichiers en lecture seule. Le modèle, l’ensemble de données et les machines d’entraînement sont situés sur un réseau à accès restreint.

Le système d’IA peut fournir des résultats erronés, mais cela est clairement expliqué à l’utilisateur, et des recommandations sont suggérées. Des tutoriels vidéo, un guide utilisateur détaillé et un ensemble complet de supports de formation sont accessibles afin de former les utilisateurs.

II. Remarques liées à l’outil d’auto-évaluation

Il a été mentionné que le système d'IA soumis à l'auto-évaluation est un système de reconnaissance vocale vers texte, qui est intrinsèquement moins susceptible de créer des risques significatifs qu'un système d'aide à la décision ou de prise de décision automatisée. Certaines questions de l'auto-évaluation n'étaient pas adaptées à ce type de système à faible risque.

Annexe 3 : Estonie

I. Synthèse de l’auto-évaluation

En Estonie, Salme est utilisé comme solution de traitement du langage naturel pour l'enregistrement audio des audiences judiciaires, la transcription et la génération des comptes rendus de séance. Il a été développé par un prestataire de services sur la base des spécifications de l'institution judiciaire.

En cas de dysfonctionnement de l'IA, cette tâche de transcription doit être réalisée par un humain.

Afin de minimiser le risque de vulnérabilités liées aux données personnelles, le répondant a déclaré : « De nombreuses solutions de sécurité, aucun prestataire tiers. »

Le modèle d'entraînement (ainsi que le modèle pré-entraîné tiers) n'a pas été audité pour identifier d'éventuels biais. Toutefois, le répondant a identifié un biais potentiel dans la compréhension de certains dialectes, dû à un échantillonnage et à des données d'entraînement non représentatives, biais atténué grâce à l'utilisation d'ensembles de données plus riches.

Pour surveiller les mécanismes, le répondant a mis l'accent sur l'évaluation humaine, l'utilisation de modèles récents entraînés en interne et l'absence d'accès par des tiers.

Une documentation (FAQ et guide utilisateur détaillé) a été créée pour accompagner l'utilisation du système d'IA.

II. Remarques liées à l'outil d'auto-évaluation

Il a été mentionné que le système d'IA soumis à l'auto-évaluation est un système de transcription, qui est intrinsèquement moins susceptible de créer des risques significatifs qu'un système d'aide à la décision ou de prise de décision automatisée. Certaines questions de l'auto-évaluation n'étaient pas adaptées à ce type de système à faible risque.

Annexe 4 : France

I. Synthèse de l'auto-évaluation

La Cour de cassation a souhaité expérimenter l'outil d'auto-évalution sur un outil de pseudonymisation des décisions de justice avant diffusion en open data, développé en interne et hébergé sur des serveurs internes, pour répondre aux exigences réglementaires prévoyant le principe d'open data des décisions de justice.

Il ressort du questionnaire que ce système d'IA est entrainé par des annotateurs et doit permettre de faire face aux besoins de pseudonymisation des 3 à 5 millions de décisions qui seront transmises chaque année à la Cour de cassation au terme du déploiement de l'open data.

Le système d'IA traite des données personnelles sensibles ou des secrets commerciaux et des systèmes d'atténuation des risques ont été mis en place, avec l'utilisation du Réseau privé virtuel justice (RPVJ) qui intègre les serveurs de la Cour de cassation. Le système d'IA mis en place a pour vocation de limiter l'exposition et l'exportation de données sensibles vers des serveurs externes dans le cadre de l'open data.

Le système d'IA a été audité à différents niveaux : code source, explication du modèle et donnée.

Différents biais ont pu être identifiés, notamment sur la pseudonymisation plus délicate et donc moins systématique des localités étrangères par rapport aux localités françaises, susceptible de créer une discrimination à l'égard des personnes résidant à l'étranger. Ce biais a été corrigé grâce à l'expertise des annotateurs humains, ingénieurs et magistrats du S.D.E.R en modifiant le modèle pour qu'il tienne compte davantage de la structure du document et moins de la typologie des adresses des localités.

Le modèle d'IA s'appuie sur un modèle tiers, le modèle CamemBERT développé par un organisme de recherche public français, l'INRIA. Les données d'entraînement de ce modèle sont expertisables et issues du sous-corpus français du corpus multilingue OSCAR.

En outre, le modèle a été affiné à partir des décisions de justice officielles et intègres contenues sur les bases de données de la Cour de cassation. Les données ont été annotées par des professionnels interdisciplinaires, notamment des professionnels du droit et de l'éthique.

Des mécanismes de contrôle de l'intégrité des données de la collecte jusqu'à l'entraînement du modèle d'IA ont été mises en place, à travers l'utilisation d'applicatifs métiers gérés par le greffe des juridictions, garant de la procédure. De plus les décisions des tribunaux de commerce transmises depuis décembre 2024 sont signées électroniquement via un tiers de confiance avant d'être transmises et centralisées au sein du GIE Infogreffe qui alimente l'API exposée par la Cour de cassation.

Le modèle a été conçu pour garantir transparence, neutralité et intégrité de la donnée : le code est auditable, les données sont vérifiables, les critères de pondérations sont clairs et permettent un audit du modèle.

Par ailleurs, l'utilisateur est clairement informé qu'un système d'IA est utilisé et dispose d'un guide d'utilisation détaillé de l'outil et de formations dédiées. Le système d'IA étant susceptible de produire des pseudonymisations erronées, en plus des recommandations, un algorithme de mise en doute a été développé par le laboratoire d'innovation afin de signaler aux agents annotateurs les cas dans lesquels il existerait une moindre fiabilité des résultats du modèle de pseudonymisation.

II. Remarques liées à l'outil d'auto-évaluation

Il a été constaté la convergence entre les questions posées dans le questionnaire et celles que s'était déjà posée les services en charge de la création et du déploiement du système d'IA.

Il a été évoqué le fait que le système d'IA objet de l'auto-évaluation est un système de pseudonymisation, donc un système par nature moins susceptible de créer des risques importants qu'un système portant sur de l'aide à la décision ou de la décision automatisé. Certaines questions de l'auto-évaluation n'étaient pas adaptées à ce type de système à risque limité.