



Strasbourg, 12 mai 2023

CDCJ-ADMIN-AI(2023)01 prov3

**ADMINISTRATION ET INTELLIGENCE ARTIFICIELLE
GROUPE DE TRAVAIL RESTREINT (CDCJ-ADMIN-AI)**

**PROJET DE NOTE CONCEPTUELLE CONCERNANT L'UTILISATION DE
L'INTELLIGENCE ARTIFICIELLE (IA) POUR LE MAINTIEN DE L'ORDRE,
L'ADMINISTRATION DE LA JUSTICE ET LES FRONTIERES/MIGRATIONS ET
L'UTILISATION DE L'IA OU D'AUTRES FORMES DE PRISE DE DECISION
AUTOMATISEE (ADM) A DES FINS DE COMMUNICATION**

Préparée par le groupe de travail restreint ADMIN-AI pour examen par le CDCJ

(point 5.4 du projet de l'ordre du jour de la 100^e réunion plénière du CDCJ)

Contexte

Lors de sa 99^e réunion plénière (23 au 25 novembre 2022), le CDCJ a analysé les propositions formulées par le CDCJ-ADMIN-AI sur d'éventuels travaux du CDCJ dans le domaine de l'intelligence artificielle (IA) et du droit administratif et donné au groupe de travail des orientations supplémentaires, en lui demandant d'élaborer des propositions sur a) l'usage de l'IA dans la police, la justice et la gestion des frontières et des migrations et b) l'usage de l'IA ou d'autres outils de prise de décision automatisée (ADM) à des fins de communication, pour examen par le CDCJ lors de sa 100^e réunion plénière en mai 2023.

Les membres du CDCJ-ADMIN-AI ont proposé que les problèmes soulevés par l'usage de l'IA ou de l'ADM à des fins de communication soient abordés lors de la mise à jour du manuel *L'Administration et vous*. La présente note conceptuelle développe la proposition, déjà présentée au CDCJ par le CDCJ-ADMIN-AI, de s'intéresser à l'usage de l'IA dans la police, la justice et la gestion des frontières et des migrations ; d'autre part, elle explique comment les questions d'usage de l'IA ou de l'ADM à des fins de communication pourraient trouver leur place dans la mise à jour du manuel.

La note conceptuelle couvre les aspects suivants :

1. usage de l'IA dans la police, la justice et la gestion des frontières et des migrations :
 - a) introduction, brève explication sur le choix du sujet
 - b) risques/problèmes pouvant être créés par le déploiement de l'IA et de l'ADM
 - c) exemples pratiques d'usage de l'IA dans la police, la justice et la gestion des frontières et des migrations
 - d) propositions d'activités futures du CDCJ dans ce domaine ;
2. modalités d'intégration des enjeux liés à l'usage de l'IA ou de l'ADM à des fins de communication dans le manuel *L'Administration et vous*.

1. Usage de l'IA dans la police¹, la justice et la gestion des frontières et des migrations

a) Introduction : pourquoi traiter ce sujet ?

1. Le déploiement de l'IA et l'ADM (pour *automated decision-making*, prise de décision automatisée) ne cesse de prendre de l'ampleur, et un certain nombre de tâches auparavant confiées à des êtres humains sont désormais – et seront plus encore à l'avenir – accomplies par des machines. Aussi, dans un avenir proche, il n'est plus du tout impensable que les usagers des services publics soient davantage en contact avec des machines qu'avec des êtres humains.
2. L'IA et l'ADM sont actuellement déployées dans différents domaines du secteur public, y compris dans le fonctionnement quotidien des forces de l'ordre et de l'administration de la justice.
3. Contrairement aux êtres humains, les machines ne peuvent ni rendre de comptes sur leurs actions ou inactions, ni être sanctionnées pour leurs erreurs. Seules les personnes physiques qui les manient ou les contrôlent peuvent être tenues pour responsables. Il convient donc de mettre en place des garanties supplémentaires lorsque l'IA et l'ADM sont utilisées dans le secteur public. Cette affirmation est d'autant plus légitime lorsque l'IA est déployée dans la police, la justice ou la gestion des frontières et des migrations, car le déséquilibre entre l'autorité publique et les personnes concernées y est particulièrement

¹ La notion de "police" aux fins de la présente note conceptuelle ne doit pas être comprise comme couvrant les aspects de droit pénal de l'application de la loi, mais la gestion des données à l'aide de l'IA et de l'ADM.

marqué. Ces garanties supplémentaires peuvent prendre la forme, entre autres possibilités, d'études d'impact et d'évaluations *a posteriori* visant à assurer la transparence et la responsabilisation, à réduire et prévenir les risques et à remédier en temps utile à tout manquement.

4. Cette suggestion de travaux futurs pour le CDCJ comprend trois volets : les pratiques policières, l'administration de la justice et la gestion ou le contrôle des frontières et des migrations. Chacune de ces sous-catégories peut être traitée à part. Cependant, et contrairement aux autres sujets proposés, elles ont un point commun : dans ces trois domaines, les actions officielles reposent régulièrement sur un fort déséquilibre entre l'autorité publique et les personnes concernées, créant un risque accru d'atteinte à leurs droits. Le risque proportionnellement plus élevé de violations des droits et l'absence de politiques, normes ou lignes directrices communes à l'attention des autorités nationales ne sont que deux motifs parmi tous ceux qui justifient l'adoption de mesures supplémentaires pour protéger les droits de l'homme dans ce domaine.

b) Quels sont les risques et problèmes associés au déploiement de l'IA et de l'ADM ?

5. Dans une large mesure, l'usage de systèmes d'IA et d'ADM pose, dans la police, la justice et la gestion des frontières et des migrations, les mêmes problèmes et risques que dans d'autres contextes, que ce soit dans l'administration ou dans d'autres domaines comme la santé, l'éducation ou la consommation. Ils peuvent être globalement décrits comme suit :
 - **Biais et discrimination** : la discrimination consiste à traiter différemment, pour des motifs illégitimes, des personnes se trouvant dans une situation comparable, avec pour effet de les défavoriser. La discrimination peut être directe ou indirecte. Elle est directe lorsqu'un acte juridique est immédiatement lié à une caractéristique sensible, sans justification adéquate et suffisante. Elle est indirecte lorsqu'une réglementation neutre sur le plan formel, c'est-à-dire ne prévoyant rien qui désavantage clairement les membres de groupes spécialement protégés, a des effets qui créent une inégalité de traitement dans la pratique. Les systèmes d'IA créent différentes sources de discrimination². La technologie adopte souvent, de manière implicite ou explicite, les partis pris déjà présents dans la société. Ces derniers peuvent relever de phénomènes sociaux ou des positions personnelles d'individus (clients, programmeurs, etc.) qui exercent une forte influence sur la conception des systèmes. La discrimination résulte alors d'un biais préexistant dans les données d'entraînement, représentatif de l'état actuel de la société. Par exemple, un algorithme qui évalue le risque de crédit va pénaliser les femmes au profit des hommes car historiquement, les premières ont eu moins accès au crédit que les seconds³. Dans le cas de systèmes de prévision policière, la mise sous surveillance excessive de certaines communautés ou zones géographiques par le passé risque de se voir renforcée et perpétuée si le système algorithmique repose sur les données d'archive de la police⁴. Les biais peuvent aussi naître d'aspects techniques ; c'est le cas, par exemple, lorsqu'un système de reconnaissance faciale fonctionne mieux sur les visages à la peau claire que sur ceux à la peau sombre parce que ces derniers sont sous-représentés dans les jeux de

² Kordzadeh Nima/Ghasemaghaei Maryam (2022), Algorithmic bias: review, synthesis, and future research directions, *European Journal of Information Systems*, vol. 31 n° 3, pp. 388–409.

³ Richardson Sharon (2022), Exposing the many biases in machine learning, *Business Information Review*, vol. 39 n° 3, pp. 82–89.

⁴ Alikhademi Kiana et al. (2022), A review of predictive policing from the perspective of fairness, *Artificial Intelligence and Law*, vol. 30, pp. 1–17.

données⁵. Le fait qu'un groupe de population soit sous-représenté dans les données d'entraînement suscite une discrimination. Cependant, certaines formes de discrimination n'apparaissent qu'au moment de l'application de l'IA, en particulier quand le système continue d'apprendre en fonctionnant. Les biais se développent généralement après la mise en service du système et sont mis en lumière par des changements dans les conceptions sociales, la démographie ou les valeurs culturelles. En outre, la discrimination peut résulter de corrélations avec ce qu'on appelle des *proxies*, ou données indirectes. Il s'agit de caractéristiques apparemment innocentes mais qui peuvent être étroitement corrélées à des motifs de discrimination interdits. Il peut y avoir un lien, par exemple, entre le lieu de résidence et l'appartenance ethnique ou le statut social⁶, ce qui crée une impression de causalité alors qu'il n'y en a pas.

- **Erreurs invisibles et fausses corrélations** : beaucoup de systèmes d'IA et d'ADM emploient un type de raisonnement probabiliste qui identifie des liens de corrélation avec un certain degré de certitude, livrant des résultats qui ne sont ni vraiment solides, ni fondés sur une relation de cause à effet et sont donc souvent inexacts, voire totalement faux. Cette limite de l'IA actuelle – qui devrait s'atténuer à l'avenir avec l'amélioration des systèmes – est aggravée par le caractère de « boîte noire » des systèmes d'IA et d'ADM, qui empêche leurs utilisateurs de vérifier l'exactitude du résultat ou de comprendre le raisonnement qui sous-tend la décision. Plusieurs exemples illustrent le danger que représentent des systèmes d'IA et d'ADM inexacts dans le secteur de la protection sociale ; ainsi, des bénéficiaires d'allocations ont vu leur situation financière et personnelle ruinée par de fausses allégations de fraude⁷.
- **Vie privée et banalisation de la surveillance** : les systèmes d'IA et d'ADM reposent sur d'immenses quantités de données qui sont souvent de médiocre qualité, non représentatives ou de provenance inconnue, et leur déploiement incite à collecter et à conserver encore plus de données pour continuer à entraîner les modèles. Certains systèmes d'IA, en particulier les systèmes embarqués (robots ou drones) ou la reconnaissance faciale, généralisent ouvertement la surveillance dans la sphère publique et privée et finissent ainsi par banaliser cette surveillance.
- **Autonomie et dignité humaines** : les systèmes d'IA embarqués, comme les robots et les drones, bousculent les attitudes, interactions et relations humaines, et nous n'avons pas fini de cerner tous leurs effets sur la notion de dignité. Étant donné que ces nouvelles technologies effectuent des tâches auparavant réservées aux humains, il y a un risque que les êtres humains se comportent conformément aux exigences des systèmes automatiques ou qu'ils disposent d'un moindre éventail de choix autonomes du fait que les options auront été prédéfinies / recommandées par un système d'ADM. L'introduction de systèmes d'IA dans des structures humaines a pour effet de modifier ces structures et de les rendre non pas *plus* mais *moins efficaces*.

6. Bien que ces inconvénients soient courants dans les systèmes d'IA et d'ADM, les torts que ces systèmes peuvent causer dans les sous-domaines spécifiques que sont la police, la justice et la gestion des frontières et des migrations sont particulièrement graves, car ils portent – entre autres – sur le droit de chacun à un procès équitable, à la liberté de circulation et à la protection contre l'arbitraire et la privation de liberté. De ce fait, malgré

⁵ Bacchini Fabio/Lorussi Ludovica (2019), Race, again: how face recognition technology reinforces racial discrimination, *Journal of Information, Communication and Ethics in Society*, vol. 17 n° 3, pp. 321–335.

⁶ Prince Anya E.R./Schwarcz Daniel (2019), Proxy Discrimination in the Age of Artificial Intelligence and Big Data, *Iowa Law Review*, vol. 105 n° 1–5, pp. 1257–1318.

⁷ Heikkilä Melissa (2022), Dutch Scandal serves as a warning for Europe over risks of using algorithms, Politico, <https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms>

les aspects en apparence avantageux des applications de l'IA et de l'ADM dans ces domaines (rapidité et moindre coût), le seuil à atteindre pour qu'elles soient mises en œuvre en conformité avec les droits de l'homme est beaucoup plus élevé.

c) Exemples pratiques d'usage de l'IA dans la police, la justice et la gestion des frontières et des migrations

Police et autres services répressifs

7. Plusieurs types de systèmes d'IA ou d'ADM sont utilisés dans le travail de la police ou des services répressifs. Il s'agit par exemple des systèmes suivants :

- Détection des infractions : le traitement de vastes volumes de données prélevées en ligne (« aspiration de données ») sert parfois à détecter des propos haineux, des menaces terroristes ou des contenus pédopornographiques (tâche principalement confiée à des plateformes privées en vertu de la loi d'octobre 2022 sur les services numériques), ainsi que les cas de fraude fiscale et de fraude aux prestations sociales.
- En 2019, par exemple, on a appris que les services fiscaux néerlandais avaient utilisé un algorithme apprenant pour créer des profils de risque et détecter les fraudes aux allocations familiales. Sur la base des indicateurs de risque de ce système, des familles se sont vu infliger des amendes pour de simples soupçons de fraude⁸.
- Les autorités fiscales françaises utilisent l'IA pour lutter contre la fraude fiscale depuis 2014. Une loi adoptée en 2020 les a même autorisées à utiliser l'analyse massive de données sur les réseaux sociaux (*data mining*) pour détecter les fraudes. Depuis 2022, la France emploie aussi l'intelligence artificielle pour repérer les piscines non déclarées – plus de 20 000 à ce jour⁹.

8. En Estonie, le Bureau des registres et informations agricoles fait analyser les prises de vues satellitaires de l'Agence spatiale européenne par un système d'IA, nommé SATIKAS, pour s'assurer que les bénéficiaires d'aides ont fauché leurs prairies avant la date limite. Cette vérification s'étend à tous les bénéficiaires. Cependant, il n'y a pas de sanction automatique, même si les personnes qui n'ont pas encore procédé au fauchage sont contactées par un agent¹⁰. En ce qui concerne les prévisions policières, les systèmes d'IA peuvent être utilisés par la police et par d'autres services répressifs compétents pour prédire et/ou prévenir des infractions pénales. Un logiciel d'IA est capable d'analyser de vastes quantités de données. Il peut donc servir à repérer des tendances, des types de comportements et d'autres corrélations de manière beaucoup plus rapide et précise que ne le pourrait un être humain. Différents pays du monde pratiquent les prévisions policières par IA, appliquées à des personnes ou à des territoires. Le pronostic porte sur la dangerosité – ou sur la mise en danger – d'une personne ou sur la probabilité, vu les caractéristiques d'un territoire, que des actes criminels soient commis. La question posée est donc, dans le premier cas : « Qui pourrait devenir dangereux ou être mis en danger ? », et dans le second : « Où tel ou tel danger pourrait-il se présenter ? ».

9. De tels systèmes sont utilisés aux États-Unis d'Amérique. En voici quelques exemples :

⁸ Heikkilä Melissa (2022), Dutch Scandal serves as a warning for Europe over risks of using algorithms, Politico, <https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/>

⁹ Euronews avec l'AFP (2022), France uses artificial intelligence to detect more than 20,000 undeclared swimming pools, Euronews, <https://www.euronews.com/my-europe/2022/08/30/france-uses-artificial-intelligence-to-detect-more-than-20000-undeclared-swimming-pools>

¹⁰ Association européenne des entreprises de télédétection (2021), A Case Study Grassland Monitoring in Estonia, https://ears.org/sebs/wp-content/uploads/2021/05/Grassland-Monitoring-in-Estonia_vfinal.pdf

- *Strategic Subject List*, à Chicago (en vigueur depuis 2020) : système de prévision policière appliqué aux personnes. Les informations concernant les contacts sociaux servent à calculer le risque qu'une personne fasse partie, par exemple, d'un gang criminel¹¹.
- *Predpol*, dans plusieurs villes dont Los Angeles, Chicago, Seattle et Boston : système conçu pour prédire le moment et le lieu de menaces potentielles et prévoir les patrouilles de police appropriées¹².
- *Risk Terrain Modelling (RTM)* : outil d'analyse géospatiale des infractions, conçu pour examiner les facteurs de risque environnementaux associés aux infractions pénales et pour identifier les domaines dans lesquels l'influence de ces facteurs est liée à une vulnérabilité aux comportements délictueux¹³.
- *HunchLab*, utilisé dans plusieurs administrations policières, par exemple à Chicago et Philadelphie : système intégrant plusieurs paramètres tels que le taux de criminalité, les récidives sur le même territoire, la situation socio-économique, les facteurs temporels et les événements sociétaux. Les informations sont traitées par un algorithme d'apprentissage automatique et régulièrement mises à jour. HunchLab « apprend » en surveillant les données sur la criminalité et crée, sur cette base, des prédictions ensuite utilisées pour suggérer l'affectation des patrouilles. Pour offrir un support sur le terrain, la plateforme HunchLab est accessible sur les appareils mobiles, ce qui permet aux agents de visualiser en temps réel les zones de criminalité potentielles¹⁴.

10. Des systèmes similaires sont aujourd'hui testés ou déployés en Europe. En voici deux exemples :

- *Burglary Predictor*, à Utrecht (Pays-Bas) : outil s'appuyant sur des algorithmes d'apprentissage automatique pour analyser des données et repérer, semaine par semaine, où se situent les plus forts risques de cambriolages. Sont ainsi créés des modèles capables de prédire la probabilité de cambriolages dans des lieux spécifiques. Les facteurs pris en compte sont par exemple les conditions météorologiques, les dates des jours fériés et les événements particuliers, ainsi que le profil sociodémographique des différents quartiers¹⁵.
- *AIDA* : algorithme utilisé par la police des West Midlands pour prédire la probabilité d'infractions pénales dans des zones spécifiques¹⁶.

Administration de la justice

11. Dans le cadre l'administration de la justice, les systèmes d'IA et d'ADM poursuivent différentes fins (liste non exhaustive¹⁷) :

- traiter de grandes quantités de données (en vue d'établir les faits, par exemple en identifiant des suspects ou en détectant des corrélations, et d'identifier les règles juridiques applicables à un différend) ;

¹¹ Portail de données de Chicago (2020), Strategic Subject List – Historical, <https://data.cityofchicago.org/Public-Safety/Strategic-Subject-List-Historical/4aki-r3np>

¹² Site web de Predpol : <https://www.predpol.com>

¹³ Site web de *Risk Terrain Modeling* sur l'analyse spatiale des infractions : <https://www.riskterrainmodeling.com>

¹⁴ Chammah, Maurice (2016), Does Predictive Policing Lead to More Police in Black Communities?, The Marshall Project, <https://www.themarshallproject.org/tag/hunchlab>

¹⁵ Prévision des cambriolages pour la commune d'Utrecht, Xomnia, <https://www.xomnia.com/burglary-prediction-for-the-municipality-of-utrecht/>.

¹⁶ Site web du projet AIDA : <https://www.project-aida.eu/index.php/about-aida>.

¹⁷ Xu Zichun, Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities, *Applied Artificial Intelligence*, vol. 36 n° 1.

- traiter des données et générer des contenus (anonymisation, traduction, rédaction de décisions¹⁸) ;
- automatiser des procédures (relations avec les parties à un litige et organisation interne des tribunaux, par exemple répartition des dossiers entre les juges).

12. Pour diverses raisons dont en particulier l'inacceptabilité sociale, le remplacement du juge par une machine afin de permettre l'automatisation complète des jugements est une pratique balbutiante et très rarement expérimentée. En revanche, l'aide juridique par IA est plus développée. En voici quelques exemples :

- Dans le système judiciaire américain, COMPAS est utilisé pour aider les juges à prendre des décisions sur certaines questions. COMPAS est un système d'analyse prédictive personnelle par IA qui évalue le risque de récidive d'un défendeur, le risque de comportement violent et celui de non-comparution en l'absence de détention préventive. Ce logiciel, qui illustre la pratique de la condamnation fondée sur des preuves, consiste à décider des peines et de leurs modalités en fonction d'un score de risque de récidive calculé à partir de 137 paramètres différents. Il a été montré que cet algorithme discriminait les populations afro-américaines, alors même qu'il n'utilise aucun critère ethnique. Le croisement des données, comprenant le lieu de résidence (qui peut révéler l'appartenance ethnique), a suscité une discrimination indirecte. Situation aggravante, le système repose sur des données biaisées, celles de décisions de justice qui reflètent les stigmatisations sociales dans un pays où les Afro-américains sont déjà condamnés à des peines en moyenne 20 % plus longues que les personnes de la catégorie « Caucasiens¹⁹ ».
- En Amérique du Sud, un autre outil, baptisé Prométéa, a été développé au sein du ministère public de Buenos Aires pour recommander des solutions aux litiges concernant l'attribution de logements sociaux ou de prestations sociales.
- En Colombie, tout justiciable qui considère que l'action ou l'inaction des pouvoirs publics menace ses droits fondamentaux peut déposer un recours en protection des droits (*tutela*), qui est automatiquement transmis à la Cour constitutionnelle. La Cour utilise un outil nommé PretorIA pour filtrer et analyser ce type de recours.
- En matière de justice pénale, l'outil *HART* (Harm Assessment Risk Tool), mis en œuvre à Durham (Royaume-Uni), est un programme d'apprentissage automatique utilisé par des agents du système carcéral pour évaluer le risque de récidive des détenus (fort, moyen ou faible²⁰).
- Dans le domaine du traitement des données, le système *DataJust* utilisé en France développe un algorithme extrait de manière automatique et exploite les données contenues dans les décisions de justice portant sur l'indemnisation des préjudices corporels. Il s'agit plus précisément de recenser les montants demandés et offerts par les parties aux instances, les évaluations proposées dans le cadre de procédures de règlement amiable des litiges et les montants alloués aux victimes par les juridictions²¹.
- En Estonie, les tribunaux utilisent une application d'IA nommée Salme. Il s'agit d'un programme d'enregistrement des audiences incluant un logiciel de reconnaissance

¹⁸ Niiler, Eric (2019), Can AI Be a Fair Judge in Court? Estonia Thinks So, Wired, <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/>.

¹⁹ <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

²⁰ Oswald Marion et al. (2018), Algorithmic risk assessment policing models : lessons from the Durham HART model and "Experimental" proportionality, *Information & Communications Technology Law*, vol. 27 n° 2, pp. 223–250, Barnes (2022), AI can predict reoffending; university finds, Durham Constabulary, <https://www.durham.police.uk/News/News-Articles/2022/January/AI-can-predict-reoffending-university-study-finds.aspx>; Fair Trials (2022), FOI reveals over 12,000 people profiled by flawed Durham police predictive AI tool, <https://www.fairtrials.org/articles/news/foi-reveals-over-12000-people-profiled-by-flawed-durham-police-predictive-ai-tool/>.

²¹ (2020), Traitement automatisé de données à caractère personnel, Justice.fr, <https://www.justice.fr/donnees-personnelles/datajust>

vocale. Conçu pour tenir compte du vocabulaire spécifique employé, il fournit une transcription des propos tenus ainsi qu'un enregistrement audio. Le système permet de modifier en temps réel la transcription par reconnaissance vocale et de distinguer les voix des différentes personnes présentes à l'audience²².

Gestion des frontières et des migrations

13. La tendance à la montée des systèmes d'IA et d'ADM dans le secteur public s'observe aussi dans le domaine du contrôle des frontières et des migrations.

14. Les images captées par des caméras installées dans des avions, drones ou hélicoptères permettent de détecter les cas de franchissement illégal des frontières. Par rapport aux contrôles humains, ces systèmes d'IA présentent l'avantage de pouvoir analyser en continu un plus important volume de données, y compris la nuit, dans des conditions météorologiques difficiles et pour un plus faible coût.

- En Estonie, la Direction de la police et des gardes-frontières a acquis, avec un soutien financier de l'Union européenne, neuf drones ELIX-XL en vue de surveiller la situation quotidienne à la frontière orientale du pays et de réagir aux incidents, notamment ceux nécessitant des secours. Ces drones font partie d'un projet global de renforcement des frontières destiné à sécuriser les frontières extérieures de plusieurs États²³.
- Deux années de suite (2021 et 2022), un aéronef sans pilote (UAV) de l'Agence européenne pour la sécurité maritime (AESM) a été déployé pour quelques mois en Estonie pour aider à surveiller la frontière maritime, à effectuer des opérations de recherches et de secours et à détecter la pollution marine²⁴.

15. L'IA et l'ADM sont aussi utilisées sur des aspects relatifs à l'asile. Par exemple :

- En Allemagne, l'Office fédéral pour les migrations et les réfugiés, autorité compétente dans ces domaines au niveau national, a mené des projets pilotes consistant à appliquer des outils numériques, dont la reconnaissance faciale et linguistique, pour aider à vérifier l'identité des demandeurs d'asile à leur arrivée dans le pays²⁵.
- iBorderCtrl est un système expérimental d'intelligence artificielle (IA) financé depuis 2016 par l'Union européenne et utilisé en au moins trois lieux de franchissement des frontières de l'espace Schengen (Grèce, Hongrie et Lettonie). Il évalue la « fiabilité » de la personne qui souhaite entrer sur le territoire de l'Union européenne : il s'agit d'un « détecteur de mensonge par reconnaissance faciale » qui redirige ensuite l'intéressé vers une file d'attente rapide ou, au contraire, vers un contrôle approfondi. Le système est capable de détecter, sur la base de 38 « micromouvements », les déclarations supposées inexactes d'une personne. Ces « micromouvements » incluent par exemple l'inclinaison de la tête ou les mouvements oculaires. Le garde-frontière virtuel pose des questions au voyageur (nom, pays d'origine, durée du séjour, raison du voyage, etc.). Les personnes dont le système estime qu'elles ont répondu honnêtement reçoivent un

²² Oyetunde, Blessing (2022), Introducing Salme; Estonian courts' speech recognition assistant, e-Estonia, <https://e-estonia.com/introducing-salme-estonian-courts-speech-recognition-assistant/>

²³ LETA/TBT (2018), Estonia's police authority to showcase drones purchased for guarding eastern border, ThenBalticTimes, https://www.baltictimes.com/estonia_s_police_authority_to_showcase_drones_purchased_for_guarding_eastern_border/

²⁴ (2022), Drones launched from Saaremaa monitor pollution and ships in Baltic Sea, News, <https://news.err.ee/1608743026/drones-launched-from-saaremaa-monitor-pollution-and-ships-in-baltic-sea>

²⁵ Forster, Madeleine (2022), Refugee protection in the artificial intelligence era, Chatham House, <https://www.chathamhouse.org/2022/09/refugee-protection-artificial-intelligence-era/2-near-future-ai-and-asylum>.

code qui les autorise à franchir la frontière, tandis que les autres personnes sont orientées vers des gardes-frontières.

Solutions technologiques

Reconnaissance biométrique

16. La police et d'autres services répressifs compétents utilisent des logiciels de reconnaissance biométrique pour identifier des individus à partir de caractéristiques telles que le visage, la voix ou la démarche. De tels logiciels épargnent aux agents la vérification manuelle des documents d'identité dans différentes bases de données. En plus d'enregistrer une image, la plupart de ces applications permettent aussi de collecter et de traiter des données biométriques et peuvent ainsi identifier les personnes. L'un de leurs usages les plus répandus est l'identification des délinquants (par exemple : système TAJ (Traitement des antécédents judiciaire) en France²⁶, Interpol).
17. Outre leur utilisation à des fins d'identification des personnes, les technologies de reconnaissance biométrique permettent également d'authentifier l'identité d'une personne par comparaison entre les traits de son visage et ceux stockés dans une base de données. Captée par une caméra numérique (création d'un modèle biométrique), l'image du visage d'une personne est ensuite traitée en vue d'établir des données clés comme la distance entre les yeux, la position du nez ou la forme de la bouche (extraction de caractéristiques). Ces caractéristiques sont comparées, au moyen d'un algorithme, à celles enregistrées dans la base de données. En fonction du score de similitude obtenu, la personne peut être considérée comme dûment authentifiée. Par exemple, l'authentification par reconnaissance faciale est utilisée lors des contrôles de sécurité dans certains aéroports (contrôle automatisé aux frontières – ABC, pour *automated border control*²⁷).
18. Il est possible de déployer la reconnaissance biométrique à des fins d'identification en tant qu'instrument de surveillance, y compris de masse, dans les espaces publics. Le contrôle de la population au moyen du traitement de données biométriques autorise en effet une surveillance massive. L'utilisation de tels systèmes dans les espaces publics peut non seulement violer le droit à la vie privée, mais aussi avoir un effet tétanisant sur la jouissance d'autres droits humains comme la liberté d'expression ou d'association, puisqu'elle décourage la participation à des rassemblements ou manifestations. Ces systèmes peuvent aussi produire des effets discriminatoires, en affectant de manière disproportionnée les populations déjà exposées à une discrimination.
19. La reconnaissance biométrique en temps réel est utilisée ponctuellement en Europe pour surveiller des zones spécifiques et géographiquement limitées. Les visages sont scannés, puis comparés à des listes de surveillance établies par la police ou les tribunaux.
- Par exemple, la Police métropolitaine de Londres utilise des caméras de reconnaissance faciale en direct pour améliorer la capacité de ses agents à identifier les suspects²⁸.

²⁶ Traitement d'antécédents judiciaires (Taj), site web Service-Public.fr : <https://www.service-public.fr/particuliers/vosdroits/F32727>

²⁷ Commission européenne, Migrations et affaires intérieures, Glossaire sur le contrôle automatisé aux frontières (ABC, automated border control) : https://home-affairs.ec.europa.eu/networks/european-migration-network-emn/emn-asylum-and-migration-glossary/glossary/automated-border-control-abc_en

²⁸ UK Metropolitan Police Facial Recognition Technology Advice and Information : <https://www.met.police.uk/advice/advice-and-information/fr/facial-recognition>; Dodd Vikram (2020), Met police to begin using live facial recognition cameras in London, *The Guardian*, <https://www.theguardian.com/technology/2020/jan/24/met-police-begin-using-live-facial-recognition-cameras>.

- En France, la reconnaissance faciale a été expérimentée lors du carnaval de Nice en 2019. Les personnes volontaires pour participer à l'expérimentation avaient donné leur consentement préalable²⁹.

20. Il est à noter que les usages en temps réel non seulement permettent une surveillance (de masse), mais ils entraînent aussi des risques accrus de violation des droits humains. Il est souvent inutile de savoir si une personne pourrait être identifiée en temps réel ou quelques jours plus tard.

Suivi à distance

21. Les drones servent de plus en plus à des fins de surveillance et de patrouille. Ils permettent d'inspecter des structures difficiles à atteindre depuis le sol. Ainsi, des chercheurs de l'Université du Maryland et de l'Université de Zurich ont équipé un drone de caméras et d'un sonar pour qu'il puisse détecter et éviter les objets qui croisent sa trajectoire. Les caméras en question n'utilisent pas nécessairement la reconnaissance biométrique. Elles peuvent cependant détecter des mouvements de foule, la chute d'une personne, la présence d'une arme à feu dans une main ou même compter les personnes portant des masques dans l'espace public, par exemple. Ces drones peuvent être utilisés dans les environnements à haut risque.

- En Grèce, par exemple, des drones ont été expérimentés pour vérifier les déclarations au fisc dans les régions touristiques. À l'aide des données collectées par les drones, les autorités ont pu déterminer le nombre de participants aux excursions en bateau et le comparer aux reçus pris en compte par les contribuables dans leurs déclarations³⁰.
- En Estonie, l'Académie de défense développe actuellement un drone en vue d'une utilisation par les services de police et de secours pour rechercher des personnes disparues en terrain boisé ou accidenté. L'IA traite des données visuelles pour détecter des objets. Fin 2022, une entreprise estonienne a annoncé avoir commencé à produire des drones conçus pour les services de secours, embarquant une IA capable de détecter un feu de forêt ou la présence d'une personne³¹.
- En France, des drones ont été utilisés pendant le confinement, en 2020, pour surveiller l'espace public et, ultérieurement, des manifestations³².
- L'Estonie n'a pas appliqué de confinement total, mais la règle dite « 2+2 » (autorisation de se déplacer par deux avec une distance de deux mètres avec les autres) et une interdiction des rassemblements. Des drones équipés de haut-parleurs ont également été utilisés pour rappeler l'interdiction aux personnes concernées³³.

Robots

22. De plus en plus, des robots sont utilisés à des fins de surveillance, tant dans des zones à risque faible qu'à risque élevé. Ils patrouillent dans des centres commerciaux et surveillent des réseaux électriques et autres lieux sensibles. Les robots peuvent atteindre des zones qui sont inaccessibles ou peu adaptées aux humains. De plus, ils peuvent assurer une

²⁹ Jasserand Catherine (2022), Experiments with Facial Recognition Technologies in Public Spaces: In Search of an EU Governance Framework. SSRN, <https://ssrn.com/abstract=4204452>

³⁰ De Hoon, Iven. Greece uses drones to find tax fraud, No More Tax, <https://nomoretax.eu/greece-drones-tax-fraud/>.

³¹ Site web de Krattworks : <https://www.krattworks.com/>

³² La Quadrature Du Net (2020), France: First victory against police drones, EDRI, <https://edri.org/our-work/france-first-victory-against-police-drones/> ; <https://www.reuters.com/article/us-health-coronavirus-france-drones-idUSKBN29J15Z>

³³ (2020), Police take to drones to enforce 2+2 coronavirus rule, News, <https://news.err.ee/1075942/police-take-to-drones-to-enforce-2-2-coronavirus-rule>

surveillance 24 heures sur 24 et 7 jours sur 7, de manière plus efficace et en réduisant les risques pour la sécurité du personnel. Voici quelques exemples :

- États-Unis : des robots sont utilisés à des fins de sécurité dans les aéroports et les centres commerciaux. À San Francisco, la police utilise déjà des robots contrôlés à distance, qui sont envoyés en reconnaissance en cas d'alerte à la bombe par exemple, ou pour vérifier que des lieux sont sûrs avant une intervention. Dans le même esprit, l'idée est d'aller plus loin en équipant ces robots de charges explosives pour qu'ils puissent détruire des barricades et diriger la police vers un suspect armé ou dangereux (décembre 2022³⁴).
- Japon : des robots effectuent des patrouilles dans les gares ferroviaires. À Tokyo, il est prévu d'installer dans la gare de Takanawa Gateway six types de robots capables par exemple de guider les passagers, de nettoyer les locaux ou d'assurer des tâches de sécurité³⁵. De tels robots sont déjà courants dans les gares japonaises.

d) Propositions d'activités futures pour le CDCJ

23. Compte tenu de ce qui précède, le CDCJ-ADMIN-AI est convaincu que l'usage croissant de l'intelligence artificielle (IA) et des outils de décision automatique (ADM) utilisant l'IA dans la police, la justice et la gestion des frontières et des migrations appelle à élaborer des politiques, des normes ou des lignes directrices communes à l'attention des autorités nationales.
24. Le CDCJ-ADMIN-AI considère que le CDCJ, en tant qu'organe intergouvernemental normatif du Conseil de l'Europe compétent dans le domaine du droit public et privé, pourrait agir dans le cadre de son mandat pour aborder les défis posés par le déploiement de l'IA et de l'ADM dans la police, la justice et la gestion des frontières et des migrations. Le CDCJ-ADMIN-AI confirme qu'à sa connaissance aucune autre organisation ne prévoit de travaux dans ce domaine spécifique. Cela rend les éventuelles initiatives du CDCJ encore plus pertinentes.
25. Les actions du CDCJ pourraient aider à garantir, dans la police, la justice et la gestion des frontières et des migrations, des usages de l'IA et de l'ADM responsables et pleinement respectueux de droits fondamentaux comme la liberté personnelle, l'autodétermination informationnelle et l'interdiction de la discrimination.
26. Le CDCJ-ADMIN-AI estime que l'élaboration d'une recommandation sur l'usage de l'IA et de l'ADM dans la police, la justice et la gestion des frontières et des migrations constituerait un moyen approprié de relever les défis dans ce domaine, tels que résumés dans la présente note conceptuelle.

2. Modalités d'intégration des problèmes soulevés par l'usage de l'intelligence artificielle (IA) ou de la prise de décision automatisée (ADM) à des fins de communication dans la mise à jour du manuel *L'Administration et vous*

27. Lors de sa 99^e réunion plénière (23 au 25 novembre 2022), le CDCJ a demandé au CDCJ-ADMIN-AI de développer deux de ses propositions de travaux futurs dans le domaine de l'IA et du droit administratif. Concernant les travaux sur l'usage de l'IA ou de l'ADM à des

³⁴ Labeyrie, Isabelle (2022), États-Unis : à San Francisco, les futurs robots-tueurs de la police inquiètent la population, Franceinfo, https://www.francetvinfo.fr/replay-radio/le-monde-est-a-nous/etats-unis-a-san-francisco-les-futurs-robots-tueurs-de-la-police-inquietent-la-population_5491851.html

³⁵ Nahao, Riho (2020), Robots roam Tokyo's newest train station to patrol and sweep, Nikkei Asia, <https://asia.nikkei.com/Business/Transportation/Robots-roam-Tokyo-s-newest-train-station-to-patrol-and-sweep>

fins de communication, le Comité a évoqué deux possibilités : aborder le sujet à l'occasion de la mise à jour du manuel *L'Administration et vous* ou le traiter à part.

28. À la suite des discussions de la réunion plénière, le CDCJ-ADMIN-AI a examiné ces deux options. De l'avis des membres du CDCJ-ADMIN-AI, les questions soulevées par l'usage de l'IA ou de l'ADM à des fins de communication sont, dans une large mesure, comparables à celles que pose le déploiement de ces technologies dans l'administration publique en général. Par conséquent, le CDCJ-ADMIN-AI a décidé de proposer au CDCJ d'aborder les enjeux relatifs à l'usage de l'IA ou de l'ADM à des fins de communication lors de la mise à jour du Manuel. La présente partie de la note conceptuelle explique de quelle manière les enjeux en question pourraient être abordés dans ce cadre.
29. Les systèmes d'IA d'usage général qui génèrent des contenus, y compris les grands modèles de langage (ci-après : LLM, *Large Language Models*) installés dans des *chatbots* (ci-après : « robots génératifs ») ont considérablement progressé, suscitant un débat sur la manière et le moment de les intégrer aux différents services utilisés par les acteurs économiques et l'administration publique.
30. Ces « robots génératifs » d'apparition récente diffèrent, sur le fond comme au plan technique, de ceux qu'on rencontrait jusqu'alors sur les sites de commerce en ligne et sur certaines applications administratives. Là où la plupart des robots conversationnels n'étaient que des arbres décisionnels automatisés, livrant des résultats précis sur la base de paramètres stricts (et donc avec des capacités limitées), les robots génératifs sont intégrés à quelques LLM, vastes systèmes d'IA qui puisent dans d'impressionnants volumes de données collectées en ligne et s'entraînent à l'aune de millions de paramètres. Ces robots conversationnels sont conçus pour imiter la parole humaine et générer des réponses réalistes, évoquant une véritable conversation.
31. Les robots conversationnels génératifs sont des outils expérimentaux. Leurs concepteurs n'en comprennent pas encore totalement le fonctionnement, et moins encore les personnes qui les déploient et les utilisent. De tels robots ne marquent sans doute que le début d'une vague d'outils génératifs qui seront conçus, dans les années à venir, sur la base des LLM. Ils impliquent toute une série de risques, d'inquiétudes et d'effets indésirables. Parce que ces outils s'entraînent à partir de données comprenant, entre autres, des forums de discussion en ligne, des œuvres couvertes par le droit d'auteur, des sites d'actualité et un large éventail d'autres contenus, leurs résultats sont imprévisibles, occasionnellement insultants ou nuisibles, souvent imprécis, parfois trompeurs (les réponses générées par le robot sont plausibles mais décousues, fausses ou dépourvues de sens, aussi appelé « hallucination »). Des mécanismes techniques permettant d'affiner les robots conversationnels génératifs sont en cours d'élaboration, et les progrès technologiques sont tels que de nombreuses inquiétudes pourraient ne plus être d'actualité dans quelques mois.
32. Sur le long terme, les difficultés posées par les robots conversationnels génératifs sont plus fondamentales et difficiles à quantifier. De tels robots pourraient égaler ou mieux imiter les performances humaines sur une série de tâches, dont les résumés de recherches ou la synthèse de documents de politique générale. Cependant, il leur manque aussi des qualités humaines : étude attentive, analyse et libre arbitre, mais aussi transparence, vue d'ensemble et capacité à rendre des comptes. Par ailleurs, les robots conversationnels génératifs pourraient si bien imiter une conversation humaine que leurs interlocuteurs auront du mal à comprendre qu'ils parlent à une IA. Ces robots pourraient également être manipulés par des acteurs extérieurs.
33. Dans les mois et années à venir, la tendance ira inévitablement à l'intégration des robots conversationnels et autres outils génératifs dans l'administration publique. Par exemple, un juge colombien a déjà admis avoir utilisé un robot conversationnel pour prendre une

décision. Il est crucial que ces outils soient considérés comme étant radicalement différents des outils technologiques qui ont été utilisés jusqu'alors et que la prudence prévale concernant leur intégration dans l'administration.

34. Si les robots conversationnels fondés sur l'IA présentent plusieurs avantages, leur utilisation soulève aussi de nouveaux défis pour l'administration publique, pour ceux qui les déploient et pour les bénéficiaires de leurs services. Face à ces défis, et en sa qualité d'organe intergouvernemental normatif du Conseil de l'Europe, le CDCJ pourrait offrir un soutien juridique supplémentaire.
35. Sur plusieurs points, le déploiement des robots conversationnels soulève des questions très similaires à celles que pose l'usage de l'IA et de l'ADM dans l'administration publique en général. Plus spécifiquement, les enjeux liés au déploiement des IA conversationnelles ont particulièrement trait aux principes suivants :
- principe 2 (**égalité de traitement**) : face à un même type de demande, les robots conversationnels devraient donner les mêmes informations ;
 - principe 3 (**objectivité et impartialité**) : les robots ne devraient pas « agir » de façon biaisée ;
 - principe 5 (**sécurité juridique**) : il faut assurer la fiabilité des informations fournies par le robot conversationnel génératif pour conseiller les citoyens ;
 - principe 6 (**transparence**) : les citoyens qui échangent avec l'administration publique par le biais d'un outil numérique doivent être informés que l'outil en question est un robot conversationnel génératif n'impliquant aucune intervention humaine ;
 - principe 7 (**respect de la vie privée et protection des données à caractère personnel**) : les robots conversationnels génératifs peuvent aussi traiter des données personnelles ;
 - principe 10 (**droit d'être entendu**) : les robots conversationnels génératifs pourraient être utilisés lorsque les citoyens ont la possibilité d'exprimer leur point de vue avant qu'une décision administrative soit prise et lorsque les limites d'une telle approche ont été définies.
 - principe 13 (**forme et notification des décisions administratives**)
36. Étant donné que le Manuel est en cours de révision et que les principes évoqués ci-dessus sont très également très pertinents pour les robots conversationnels génératifs, le CDCJ-ADMIN-AI propose d'intégrer la question des robots conversationnels à la mise à jour du Manuel. Il conviendra également d'examiner attentivement si l'usage de robots conversationnels pour communiquer des informations ou résoudre des problèmes administratifs appelle l'élaboration d'un nouveau principe, sachant que le fait de priver totalement une personne de la possibilité de communiquer avec des fonctionnaires dans le cadre de processus administratifs peut poser un problème du point de vue de la dignité humaine³⁶.
37. Compte tenu de ces considérations, le CDCJ-ADMIN-AI propose de tenir compte, dans sa mise à jour du Manuel, des questions soulevées par le recours à des robots conversationnels génératifs pour communiquer des informations ou résoudre des problèmes administratifs.

³⁶ (2022) Loi portant amendements de la loi sur les procédures administratives et d'autres lois connexes 634 SE, <https://www.riigikogu.ee/tegevus/eelnoud/eelnou/21f6df90-a333-413a-a533-ebbf7e9deebe/Haldusmenetluse+seaduse+muutmise+ja+sellega+seonduvalt+teiste+seaduste+muutmise+seadus>.