

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

L'utilisation d'outils de machine learning à des fins de sécurité publique

Poullet, Yves; Lognoul, Michael

Published in:
Bulletin de l'AFIA

Publication date:
2023

Document Version
le PDF de l'éditeur

[Link to publication](#)

Citation for pulished version (HARVARD):

Poullet, Y & Lognoul, M 2023, 'L'utilisation d'outils de machine learning à des fins de sécurité publique: une interdiction de principe en droit européen ?', *Bulletin de l'AFIA*, numéro 120, pp. 89-95. <https://afia.asso.fr/wp-content/uploads/2023/05/120_avr23.pdf>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



AfIA

Association française
pour l'Intelligence Artificielle

- ity. *Ethics and Information Technologies*, 18(2) :117–129, 2016.
- [7] A. J. Kolber. Pain detection and the privacy of subjective experience brain imaging and the law. *American Journal of Law & Medicine*, 33(2–3) :433–457, 2007.
- [8] S. R. Krause. *Civil Passions. Moral Sentiment and Democratic Deliberation*. Princeton University Press, 2008.
- [9] J. Mansbridge, J. Bohman, S. Chambers, D. Estlund, A. Føllesdal, A. Fung, C. Lafont, B. Manin, and J. L. Martí. The place of self-interest and the role of power in deliberative democracy. *The Journal of Political Philosophy*, 18 :64–100, 2010.
- [10] K. Marx. *Le Capital, 1867-1879, t. 1, 4e section, chapitre XV, 3, p. 379 (trad. J. Roy)*. Édition du Progrès, 1976.
- [11] M. Mormina. Knowledge, expertise and science advice during covid-19 : In search of epistemic justice for the 'wicked' problems of post-normal times. *Social Epistemology*, 36(6) :671–685, 2022.
- [12] C. Pohl, B. Truffer, and G. Hirsch-Hadorn. *Addressing wicked problems through transdisciplinary research*. R. Frodeman, J. Thompson Klein, and R. C. S. Pacheco (Eds), Oxford University Press, 4th edition, The Oxford handbook of interdisciplinarity : 319–331, [9780198733522.013.26](https://doi.org/10.1017/9780198733522.013.26), 2017.
- [13] F. Popa, M. Guillermin, and T. De-deurwaerdere. A pragmatist approach to transdisciplinarity in sustainability research : From complex systems theory to reflexive science. *Futures*, 65 :45–56, 2015.
- [14] F. X. Shen. Neuroscience, mental privacy, and the law. *Harvard Journal of Law & Public Policy*, 36(2) :653–714, 2013.
- [15] J. Steiner. Raison et émotion dans la délibération. *Archives de philosophie*, 72(2) :259–274, 2011.

■ L'utilisation d'outils de machine learning à des fins de sécurité publique : une interdiction de principe en droit européen ?

Yves POULLET

CRIDS/NADi

Université de Namur

Par

yves.poulet@unamur.be

www.unamur.be

Michael LOGNOUL

michael.lognoul@unamur.be

Introduction

L'intelligence artificielle (IA) constitue un outil majeur d'investigations, de prévention et de lutte contre la criminalité et le terrorisme. Si la liste des applications possibles de l'IA en matière policière ou de renseignement est infinie,

notons que la technologie de l'apprentissage machine requiert cependant des mégadonnées et que la collecte, le stockage et l'exploitation de ces données peuvent être d'autant plus difficiles qu'elles sont à l'origine recueillies par des opérateurs privés. En outre, les textes en ma-



Afia

Association française
pour l'Intelligence Artificielle

tière de protection des données, en particulier la directive 2016/680 (dite directive « Police-justice ») [2], fixent à cette exploitation et aux investigations policières nombre de contraintes.

Dans ce contexte, deux décisions majeures de la Cour de justice de l'Union européenne retiendront notre attention. La première, rendue en date du 6 octobre 2020, concerne les obligations de rétention des données de communication par les opérateurs de communication électroniques et leur utilisation par les autorités policières [4]. Ces obligations peuvent être imposées auxdits opérateurs par les États Membres de l'Union européenne (UE), en vertu de la directive 2002/48 (dite directive « *e-Privacy* », en cours de révision) [1]. Dans cet arrêt, la Cour soulève les risques accrus liés aux utilisations potentielles des technologies de l'IA, en tout cas celles utilisant les technologies de l'apprentissage machine, pour justifier un encadrement plus strict de l'étendue de ces obligations. La seconde décision date du 21 juin 2022 [6]. Elle concerne la transmission obligatoire, aux autorités publiques compétentes par les compagnies aériennes, de données relatives à leurs passagers, sur base de la directive 2016/681 (dite directive « PNR ») [3]. L'utilisation de logiciels d'IA par les services de police et de renseignements des États Membres, pour le traitement de telles informations, amène la Cour à interpréter de manière restrictive le texte européen, mais surtout à énoncer quelques principes en ce qui concerne cette utilisation. Dans les pages qui suivent, les points principaux de ces développements sont soulignés.

La rétention des données et l'obligation de collaboration des opérateurs de communications électroniques

Dans cette première affaire, la Cour de justice était saisie par diverses associations de défense des libertés, notamment de la question suivante : « L'article 15, paragraphe 1, de la

directive [2002/58], combiné avec les articles 4, 7, 8, 11 et 52, paragraphe 1, de la [Charte], doit-il être interprété en ce sens qu'il s'oppose à une réglementation nationale [...] qui prévoit une obligation générale pour les opérateurs et fournisseurs de services de communications électroniques de conserver les données de trafic et de localisation au sens de la directive [2002/58], générées ou traitées par eux dans le cadre de la fourniture de ces services si cette réglementation a notamment pour objet de réaliser les obligations positives incombant à l'autorité en vertu des articles 4 et 7 de la Charte, consistant à prévoir un cadre légal qui permette une enquête pénale effective et une répression effective de l'abus sexuel des mineurs et qui permette effectivement d'identifier l'auteur du délit, même lorsqu'il est fait usage de moyens de communications électroniques ? » (§ 79).

Dans son raisonnement, la Cour consacre cette obligation positive de l'État, qui trouve un écho dans l'article 15, paragraphe 1, de la directive *e-Privacy* ; cette obligation trouve cependant ses limites dans l'application du principe de proportionnalité. L'obligation positive dont question permet certes aux États Membres d'introduire des exceptions à l'obligation de principe, énoncée à l'article 5, paragraphe 1, de cette directive, de garantir la confidentialité des données à caractère personnel ainsi qu'aux obligations de non-utilisation des données à des fins autres que de sécurité du réseau ou de facturation des services. Ces exceptions ne peuvent cependant valoir que si elles constituent une mesure prévue par la loi, nécessaire, appropriée et proportionnée, au sein d'une société démocratique, pour sauvegarder la sécurité nationale, la défense et la sécurité publique, ou assurer la prévention, la recherche, la détection et la poursuite d'infractions pénales ou d'utilisations non autorisées du système de communications électroniques.

Notre propos n'entend pas analyser l'en-



Afia

Association française
pour l'Intelligence Artificielle

semble des règles déduites par la Cour de justice en ce qui concerne les limites du droit des États Membres soit à exiger des opérateurs de communication l'accès non généralisé, mais à certaines données de communication (par exemple, provenant d'une zone géographique particulière ou d'un groupe de personnes), soit à intercepter des communications. Il se concentre sur celles relatives à la conservation exigée des opérateurs de communication électroniques, des métadonnées de communication. Par métadonnées de communication, on entend les données de trafic et de géolocalisation (type de communication, émetteur, destinataire, localisation de ces acteurs, durée de la communication, volume des données) permettant d'identifier les communications sans atteindre à leur contenu. Cette obligation de conservation autorise alors les services de polices ou de renseignements à procéder par des techniques de *data mining* et de profilage à détecter les auteurs d'infraction, potentiels, suspectés ou réels. Ces opérations sont susceptibles de révéler des informations sur un nombre important d'aspects de la vie privée des personnes concernées, y compris des informations sensibles, telles que l'orientation sexuelle, les opinions politiques, les convictions religieuses, philosophiques, sociétales ou autres ainsi que l'état de santé. On sait que ces données méritent une protection particulière, suivant les textes européens. Par ailleurs, l'agrégation des métadonnées de communication peut aboutir à la constitution de profils très précis, incluant les habitudes de la vie quotidienne, les lieux de séjour permanents ou temporaires, les déplacements journaliers ou autres, les activités exercées, les relations sociales de ces personnes et les milieux sociaux fréquentés par celles-ci, et permet d'en inférer le contenu des communications.

Ce sont précisément les possibilités croissantes d'atteintes à la vie privée liées à l'uti-

lisation de systèmes d'IA de plus en plus performants qui, selon la Cour, justifient des restrictions supplémentaires au droit des États Membres d'exiger une conservation généralisée des données de trafic et de géolocalisation, ce qui était l'objet du recours pris contre la réglementation de certains États Membres qui autorisaient cette demande de conservation. L'arrêt souligne les risques d'erreurs, de discrimination, et d'évolutivité non contrôlée liés à l'utilisation de tels systèmes (nous reviendrons sur ce point *infra*). La Cour relève en outre qu'une telle mesure concerne tous les citoyens et non uniquement ceux suspectés ou objets de mesure de surveillance et exige donc des restrictions supplémentaires. Aussi, les juges décident que de telles mesures doivent rester tout à fait exceptionnelles et ne s'adresser qu'à des mesures dites de sauvegarde de la sécurité nationale, à savoir la lutte contre le terrorisme. Ils excluent dès lors le recours à une obligation de conservation généralisée pour des objectifs de simple sécurité publique (par exemple, des manifestations violentes) ou de lutte contre la criminalité, y compris grave. Pour autant que la menace s'avère réelle et actuelle ou prévisible, et à la condition que la durée de cette conservation soit limitée au strict nécessaire, l'objectif de sauvegarde de la sécurité nationale face à une menace grave est seul susceptible de justifier des mesures comportant des ingérences dans les droits fondamentaux plus graves que celles que pourraient justifier ces autres objectifs. La Cour ajoute que le niveau de la menace, les techniques d'analyse automatisée et la durée de la mesure doivent faire l'objet d'un contrôle effectif « soit par une juridiction, soit par une entité administrative indépendante, dont la décision est dotée d'un effet contraignant, visant à vérifier l'existence d'une situation justifiant ladite mesure ainsi que le respect des conditions et des garanties devant être prévues » (§ 179). Les autorités de protection des don-



nées sont implicitement visées pour effectuer ce contrôle.

Face aux risques liés à l'utilisation des techniques d'IA, les juges énoncent quelques garanties supplémentaires qui doivent précisément faire l'objet de l'examen par cette autorité dont l'intervention est jugée nécessaire. Ainsi, « il convient de préciser que les modèles et les critères préétablis sur lesquels se fonde ce type de traitement de données doivent être, d'une part, spécifiques et fiables, permettant d'aboutir à des résultats identifiant des individus à l'égard desquels pourrait peser un soupçon raisonnable de participation à des infractions terroristes et, d'autre part, non discriminatoires » (§ 180). À cet égard, ils mettent en garde contre l'utilisation de modèles qui se fonderaient exclusivement sur des données sensibles comme l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses, l'appartenance syndicale, l'état de santé ou la vie sexuelle d'une personne, sans prendre en compte l'analyse du comportement individuel de la personne. Le taux d'erreurs constatées à la suite de l'utilisation des systèmes d'intelligence artificielle exige que « tout résultat positif obtenu à la suite d'un traitement automatisé doit-être soumis à un réexamen individuel par des moyens non automatisés avant l'adoption d'une mesure individuelle produisant des effets préjudiciables à l'égard des personnes concernées » (§ 182).

Enfin, prescrivent les juges, « aux fins de garantir, en pratique, que les modèles et les critères préétablis, l'usage qui en est fait ainsi que les bases de données utilisées ne présentent pas un caractère discriminatoire et soient limités au strict nécessaire au regard de l'objectif de prévenir des activités de terrorisme présentant une menace grave pour la sécurité nationale, la fiabilité et l'actualité de ces modèles et de ces critères préétablis ainsi que des bases de données utilisées doivent faire l'objet d'un réexamen régulier » (§ 182). Le second arrêt

de la Cour entend préciser encore ces limites à l'utilisation de systèmes d'apprentissage automatique.

L'analyse des données PNR par des outils d'IA aux fins d'identifier les terroristes et les criminels

Dans cette seconde affaire, la Cour de justice était saisie d'un recours visant à faire constater l'invalidité de certaines dispositions de la directive PNR, sur base de leur contrariété alléguée à la Charte des droits fondamentaux de l'UE. Plus précisément, une association de défense des libertés remettait en question la transposition belge de la directive PNR devant la Cour constitutionnelle du même pays, ce qui a conduit ladite Cour à interroger les juges européens quant à l'interprétation et à la validité de la directive PNR elle-même, vis-à-vis des droits fondamentaux au respect de la vie privée et familiale, et à la protection des données à caractère personnel.

En effet, la directive PNR prévoit, en son article 6, que des données relatives aux passagers aériens (nom, itinéraire, dates de voyage, coordonnées, modes de paiement, informations relatives aux bagages, etc.), recueillies par les transporteurs aériens, doivent systématiquement être communiquées aux autorités publiques compétentes des États Membres. Ces données sont ensuite confrontées « aux bases de données utiles aux fins de la prévention et de la détection des infractions terroristes et des formes graves de criminalité ainsi que des enquêtes et des poursuites en la matière [...] ; ou [traitées] au regard de critères préétablis ». Dans ce cas, la directive prévoit que « [l]'évaluation des passagers [...] au regard de critères préétablis est réalisée de façon non discriminatoire. Ces critères préétablis [...] ciblés, proportionnés et spécifiques [...] ne sont en aucun cas fondés sur l'origine raciale ou ethnique d'une personne, ses opinions po-



Afia

Association française
pour l'Intelligence Artificielle

litiques, sa religion ou ses convictions philosophiques, son appartenance à un syndicat, son état de santé, sa vie sexuelle ou son orientation sexuelle ».

Dans sa décision, la Cour de justice relève tout d'abord que « la directive PNR comporte des ingérences d'une gravité certaine dans les droits [fondamentaux à la vie privée et à la protection des données à caractère personnel], dans la mesure notamment où elle vise à instaurer un régime de surveillance continu, non ciblé et systématique, incluant l'évaluation automatisée de données à caractère personnel de l'ensemble des personnes faisant usage de services de transport aérien » (§ 111). Partant de ce constat, la Cour rappelle, dans cette affaire également, les principes de légalité et de proportionnalité et en examine le respect par la directive en cause, à savoir son aptitude à atteindre les objectifs légitimes poursuivis, et la stricte nécessité des ingérences imposées pour y parvenir.

À défaut d'analyser ici l'ensemble des considérations qui ont mené la Cour à rendre sa décision, notons toutefois qu'au cours de cet examen des mesures imposées par la directive PNR, la Cour fournit une interprétation restrictive des dispositions de la directive afin de conclure à sa validité par rapport à la Charte des droits fondamentaux de l'UE. Au-delà, elle entend limiter les usages qui peuvent être faits de l'IA dans le cadre des contrôles opérés par les autorités publiques. Ce faisant, la Cour pose une série de jalons qui conditionnent, voire limitent, l'usage d'outils d'IA par les autorités publiques dans le cadre de l'application des dispositions de la directive PNR.

Ainsi, lorsqu'elle analyse la nécessité des ingérences imposées, la Cour indique notamment que les analyses automatisées des données PNR présentent un taux d'erreur important, car elles sont basées sur des données non vérifiées et sur des modèles et critères

préétablis. Les juges européens notent, à cet égard, qu'en 2018 et 2019, cinq personnes sur six identifiées par des moyens automatisés comme présentant un risque élevé ont ultérieurement été considérées comme des concordances positives erronées lors d'un réexamen par des moyens non automatisés. Partant, la Cour insiste sur le fait qu'aucune décision produisant des effets préjudiciables significatifs à l'égard d'une personne ne peut être prise sur le seul fondement d'un traitement automatisé des données PNR. Un traitement ultérieur de ces données, par des moyens non automatisés, est requis pour valider ou infirmer une concordance positive établie par un outil informatique.

Ensuite, s'agissant des « bases de données utiles » auxquelles les données personnelles des voyageurs peuvent être confrontées par les autorités publiques, la Cour apporte plusieurs précisions. Tout d'abord, elle limite sévèrement les bases de données susceptibles d'être utilisées dans le cadre de cette investigation. En premier lieu, la Cour indique qu'il s'agit des seules bases de données « concernant les personnes ou les objets recherchés ou faisant l'objet d'un signalement, conformément aux règles nationales, internationales et de l'Union applicables à de telles bases de données » (§ 187). En second lieu, la Cour détermine que ces bases de données doivent être exploitées « en rapport avec la lutte contre des infractions terroristes et des formes graves de criminalité présentant un lien objectif [...] avec le transport aérien des passagers » (§ 191). Enfin, la Cour note que ces bases de données utiles doivent être gérées ou exploitées par des autorités publiques compétentes, dans le cadre de leur mission de lutte contre le terrorisme et les formes graves de criminalité. Or, ces autorités doivent être désignées de manière limitative par les États membres en application de la directive PNR.

En outre, s'agissant cette fois du traitement des données des passagers « au regard de



Afia

Association française
pour l'Intelligence Artificielle

critères préétablis », la Cour suit les conclusions de son Avocat Général [5] et prend position contre l'utilisation, par les autorités publiques, d'outils d'IA fonctionnant sur base d'apprentissage machine, dès lors que ceux-ci ont la capacité de modifier, de manière autonome, le processus de l'évaluation des passagers. En particulier, les systèmes d'IA susceptibles de modifier les critères d'évaluation ou encore leur pondération sont prohibés, puisque de telles modifications seraient contraires au caractère préétabli desdits critères. De ce fait, seuls les outils fonctionnant grâce à des règles et pondérations entièrement préétablies par des humains – et sans capacité d'adaptation autonome ultérieure –, soit les seuls systèmes experts d'IA qualifiée de symbolique, à l'exclusion des systèmes d'apprentissage machine, pourraient être mis à contribution dans le cadre des contrôles permis par la directive PNR. La Cour ajoute, pour le surplus, que le recours aux technologies d'apprentissage machine pourrait priver d'effet utile le réexamen obligatoire, mentionné ci-avant, des concordances positives par des moyens non automatisés. En effet, « compte tenu de l'opacité caractérisant le fonctionnement des technologies d'intelligence artificielle, il peut s'avérer impossible de comprendre la raison pour laquelle un programme donné est parvenu à une concordance positive » (§ 195). Dans le même ordre d'idées, la Cour ajoute que l'utilisation de technologies d'IA fonctionnant sur base de l'apprentissage machine serait « susceptible de priver les personnes concernées également de leur droit à un recours juridictionnel effectif [...], en particulier pour contester le caractère non discriminatoire des résultats obtenus » (§ 195). Notons certes que, l'existence d'une concordance positive établie par un système fonctionnant grâce à l'apprentissage machine n'empêcherait pas le réexamen ultérieur, par un agent humain, de l'ensemble du dossier d'un passa-

ger, sans tenir compte des facteurs analysés par machine (ou de leur poids), afin de prendre une décision finale de maintien – ou de suppression – de la concordance positive. En revanche, l'argument fondé sur la difficulté pour les passagers d'accéder à un recours juridictionnel effectif doit être salué : les passagers aériens seraient en effet dans l'incapacité de contester le caractère non discriminatoire des facteurs utilisés par les systèmes d'IA d'apprentissage machine : l'opacité de ces systèmes empêche, de fait, les personnes concernées de comprendre quels facteurs sont pris en compte pour établir une concordance positive automatisée, et comment ceux-ci sont mis en œuvre.

Enfin, mentionnons qu'en vertu de l'article 6 de la directive PNR, les « critères préétablis » à l'aune desquels les données des voyageurs sont évaluées doivent faire l'objet d'un réexamen régulier. A ce sujet, la Cour de justice apporte également des précisions. Elle indique ainsi que, dans le cadre de ce réexamen, les critères retenus doivent être actualisés en tenant particulièrement compte de l'expérience acquise dans le cadre de leur application, de manière à réduire autant que possible le nombre (fort élevé) de résultats de type « faux positifs ». Cette manière de procéder, dit la Cour, doit contribuer au caractère strictement nécessaire de l'application de ces critères – et donc justifier la stricte nécessité des ingérences dans les droits fondamentaux imposées en vertu de la directive PNR.

Conclusion

Ce rapide survol des décisions rendues par les juges européens, en matière d'usage d'outils d'intelligence artificielle par les autorités publiques en charge de la sécurité publique, démontre que lesdits juges dessinent le cadre dans lequel une police « algorithmique », respectueuse des droits fondamentaux des individus et tenant compte des risques accrus engendrés



par l'IA, devra se développer.

Dans ce contexte, d'aucuns pourraient s'interroger sur la transposition de certaines parties du raisonnement de la Cour, dans ces deux affaires, à d'autres domaines, y compris dans le secteur privé, dans lesquels des outils d'IA sont utilisés pour prendre des décisions qui ont un impact significatif sur les individus, comme des outils d'octroi de crédit, de mesure de l'assurabilité ou de l'employabilité des personnes. L'interdiction d'utiliser des outils d'IA fonctionnant sur base d'apprentissage machine ne pourrait-elle pas recevoir une portée plus large, dès lors que leur utilisation risque de priver les personnes concernées également de leur droit à un recours juridictionnel effectif, du fait de l'opacité de ces outils ?

Références

- [1] *Directive 2002/58 du Parlement européen et du Conseil du 12 juillet 2002 concernant le traitement des données à caractère personnel et la protection de la vie privée dans le secteur des communications électroniques (dite directive « Vie privée et communications électroniques »)*, OJ L 201, 31 juillet 2002.
- [2] *Directive 2016/680 du Parlement européen et du Conseil du 27 avril 2016 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales, et à la libre circulation de ces données, et abrogeant la décision-cadre 2008/977/JAI du Conseil (dite directive « Police-Justice »)*, OJ L 119, 4 mai 2016.
- [3] *Directive 2016/681 du Parlement européen et du Conseil du 27 avril 2016 relative à l'utilisation des données des dossiers passagers (PNR) pour la prévention et la détection des infractions terroristes et des formes graves de criminalité, ainsi que pour les enquêtes et les poursuites en la matière (dite directive « PNR »)*, OJ L 119, 4 mai 2016.
- [4] *CJ, arrêts Privacy International, La Quadrature du Net e.a., French Data Network e.a., et Ordre des barreaux francophones et germanophone, affaires C-623/17, C-511/18, C-512/18 et C-520/18*, 2020.
- [5] *Av. gén. M. G. Pitruzzella, concl. préc. CJ, arrêt Ligue des droits humains c. Conseil des ministres, affaire C-817/19*, 2022.
- [6] *CJ, arrêt Ligue des droits humains c. Conseil des ministres, affaire C-817/19*, 2022.