Zeitschrift: Kriminologie / Schweizerische Arbeitsgruppe für Kriminologie SAK = Criminologie / Groupe Suisse de Criminologie GSC = Criminologia / Gruppo Svizzero di Criminologia GSC

Herausgeber: Schweizerische Arbeitsgruppe für Kriminologie

Band: 38 (2021)

Artikel: Intelligence artificielle et justice pénale : état des lieux

Autor: Gillérion, Philippe

DOI: https://doi.org/10.5169/seals-1051602

Intelligence artificielle et justice pénale : état des lieux

Philippe Gillérion*

Table des matières

Résumé

Dans cette contribution, l'auteur propose un aperçu de différents cas de figure où les systèmes d'intelligence artificielle sont susceptibles d'être utilisés dans le cadre de l'administration de la justice pénale, en en soulignant les risques et le cadre juridique qui pourrait, dans un proche avenir, entourer leur adoption à l'aune de la proposition de règlement sur l'intelligence artificielle adoptée le 21 avril 2021 par la Commission européenne.

Zusammenfassung

In diesem Beitrag gibt der Autor einen Überblick über verschiedene Szenarien, in denen Systeme der künstlichen Intelligenz möglicherweise im Bereich der Strafjustiz eingesetzt werden. Er hebt deren Risiken hervor und beschreibt den rechtlichen Rahmen, der ihre Anwendung in naher Zukunft auf Grundlage des Richtlinienvorschlags der Europäischen Kommission zur

* Professeur à l'Université de Lausanne, avocat en l'étude Wilhelm Gilliéron Avocats SA, à Lausanne. Cette contribution se base sur la présentation donnée au Congrès annuel du Groupe Suisse de Criminologie au mois de septembre 2020 à Interlaken. Elle a été remaniée pour tenir compte, en particulier, de la proposition de règlement sur l'intelligence artificielle adoptée le 21 avril 2021 par la Commission européenne.

künstlichen Intelligenz, die am 21. April 2021 verabschiedet wurde, bestimmen könnte.

20 mai 2018. Un homme prénommé Yu se rend au stade de sports de Jiaxing, une ville située dans la province du Zhejiang, à l'est de la Chine, pour écouter un concert donné par Jacky Cheung. Au moment de passer le contrôle de sécurité, il est interpellé par la police après avoir été identifié par des caméras de reconnaissance faciale pour un vol de pommes de terre ayant eu lieu trois années plus tôt .

Avec plus de 600 millions de caméras de surveillance prévues en 2020 capables de reconnaissance faciale, un budget annuel de 60 milliards de dollars pour encourager les innovations dans le domaine de l'intelligence artificielle et l'introduction d'un score de comportement social (*social credit scoring*, SCS), la Chine est en passe de faire de la dystopie orwellienne une réalité.

Nul doute que l'intelligence artificielle est susceptible de modifier profondément les rapports entre l'Etat et les citoyens suivant l'usage que l'on souhaite en faire. La Chine en donne un exemple, peu enviable.

La présente contribution s'efforce d'offrir un aperçu des enjeux susceptibles d'être rencontrés lors du recours à des systèmes d'intelligence artificielle, dans le cadre d'un rapport particulier entre l'Etat et les citoyens, à savoir celui de l'administration de la justice pénale, à l'aune de trois volets :

- Les dispositifs de surveillance (1.) ;
- Les outils de prévisibilité en matière criminelle (2.) ;
- Le recours à de tels systèmes dans l'enceinte des tribunaux (3.).

1. Les dispositifs de surveillance

Le recours à des dispositifs de surveillance fonctionnant en tout ou partie sur la base de systèmes d'intelligence artificielle (dont la reconnaissance faciale est l'exemple type) présente, à mon sens, deux préoccupations majeures :

- Le rôle croissant joué par des acteurs privés (1.1) ;
- Le risque de dévier vers une société de surveillance (1.2).

1.1 Le rôle croissant joué par des acteurs privés

12 juillet 2019. Chandler, Arizona. Un homme se réveille en sursaut après que sa caméra de surveillance *Ring* a détecté un mouvement suspect dans le périmètre de sa maison et déclenché une alarme sur son téléphone portable. Grâce à l'application *Neighbors*, téléchargée par près de 20 000 résidents de Chandler

et permettant d'équiper les appareil de caméras *Ring* directement accessibles aux forces de l'ordre, les voisins et la police sont informés de l'indésirable présence, et l'indésirable rapidement arrêté[1].

Ring. Si ce nom n'est encore guère évocateur sous nos latitudes, il l'est davantage outre-Atlantique. Ayant flairé le potentiel de la technologie et la possibilité de renforcer ainsi ses liens avec les autorités étatiques, Amazon acquiert, au mois de mars 2018, la société pour un peu plus de 1 milliard de dollars[2]. Au seul mois de décembre 2019, plus de 400 000 caméras *Ring* auraient été vendues[3].

Suite à l'interpellation émise le 10 octobre 2019 par le sénateur Edward Markey, qui s'interroge sur la manière dont les vidéos, prises par des particuliers au moyen de l'application *Neighbors*, sont partagées avec la police, Amazon répond que chaque requête émise en ligne sur l'application doit être individuelle et porter un numéro d'affaire, sans qu'aucune preuve ne soit en revanche exigée ; *Ring* produit alors jusqu'à 12 heures de vidéos, enregistrées en 45 jours[4]. Quand bien même nul ne saurait se voir imposer de permettre un tel visionnage, on peut sans peine imaginer qu'un utilisateur recevant une invitation de la police à partager sa vidéo pour l'aider à résoudre un délit aura quelque peine à refuser un tel visionnage. Aucune statistique ne semble être disponible sur cette question. Dans le prolongement de cet échange, le *Washington Post* soulignait, dans son édition du 19 novembre 2019, que les prises de vidéo communiquées à la police pouvaient être conservées par cette dernière sans limite de temps[5].

Le 20 mars 2020 – soit quelques mois plus tard –, la *BBC* révélait que toutes les interactions avec les caméras *Ring*, les logs comprenant le moment de l'utilisation, le nombre d'occurrences et la durée de ces utilisations étaient conservés par Amazon[6] ; autant de données porteuses d'informations permettant à Amazon d'en savoir toujours davantage sur ses consommateurs. A l'heure du *Big Data* et des développements majeurs que connaît l'intelligence artificielle, on comprend aisément la valeur intrinsèque que l'agrégation de ces informations peut représenter pour une société comme Amazon.

---

1 ‹https://kvoa.com/news/2019/08/08/amazon-is-developing-high-tech-surveillance-tools-for-an-eager-customer-americas-police/› (consulté le 31.5.2021).

2 ‹https://www.businessinsider.com/why-amazon-acquired-ring-2018-3 ?IR=T› (consulté le 3.5.2021).

3 ‹https://www.vox.com/recode/2020/1/21/21070402/amazon-ring-sales-jumpshot-data› (consulté le 31.5.2021).

4 ‹https://www.markey.senate.gov/imo/media/doc/Response %20Letter_Ring_Senator %20Markey %2011.01.2019.pdf› (consulté le 31.5.2021).

5 ‹https://www.washingtonpost.com/technology/2019/11/19/police-can-keep-ring-camera-video-forever-share-with-whomever-theyd-like-company-tells-senator/› (consulté le 31 mai 2021).

6 ‹https://www.bbc.com/news/technology-51709247› (consulté le 31.5.2021).

Au mois de janvier 2021, Amazon avait signé des partenariats avec plus de 2000 forces de police et de pompiers, représentant 48 des 50 états américains[7]. Quand bien même Amazon aurait accepté de suspendre, pendant une année, la vente de ses outils de reconnaissance faciale, il ne s'agit là que d'un intermède d'une année, visant à permettre au Congrès américain d'adopter les dispositions topiques en la matière. A la différence de IBM et Microsoft, Amazon ne semble donc pas prêt à renoncer à ce marché fort lucratif. Ensemble, 44 associations de défense des libertés individuelles l'ont enjointe, au mois de mai 2021, de renoncer à la commercialisation de tels outils, à l'image des sociétés précitées[8]. On peut douter que cette simple missive face plier Amazon ; seule la fuite de talents ne souhaitant pas contribuer à une stratégie avec laquelle ils sont en désaccord et préférant quitter le navire, voire un éventuel détournement de ses clients, entraînant une baisse du chiffre d'affaires (peu probable), pourra, à mon avis, dissuader Amazon de poursuivre sur cette voie.

Comme l'ont souligné plusieurs défenseurs des libertés individuelles[9], ces développements sont inquiétants, ce à double titre :

Tout d'abord, quand bien même les Etats-Unis s'apparentent sans doute davantage à un Etat de droit que la Chine, au sens où nous l'entendons, que penser d'une société où les individus contribuent, pour ne pas dire, se substituent largement aux pouvoirs étatiques dans le cadre d'enquêtes policières ? Si la dystopie orwellienne apparaît au grand jour à l'échelle d'un pays en Chine, elle pointe le bout de son nez à l'échelle de certaines communautés, toujours plus importantes, aux Etats-Unis. Autre sujet d'inquiétude, la concentration des pouvoirs d'un groupe réduit d'acteurs qui, loin d'offrir de simples solutions, propose désormais un véritable écosystème, mêlant hardware, software et plateforme en la matière. Axon en est sans doute aujourd'hui l'exemple le plus criant[10]. Or, une telle concentration de pouvoirs par une entité privée sur des données privées, sensibles et collectées par des autorités sur la base de prérogatives publiques, n'est pas sans susciter quelques interrogations[11].

Ensuite, on ne saurait bien entendu ignorer le rôle joué, notamment dans le domaine informatique et numérique, par des acteurs privés dans les tâches dévolues au secteur public. Bien naïf celui qui pense qu'un gouvernement est à

7 ‹https://www.theverge.com/2021/1/31/22258856/amazon-ring-partners-police-fire-security-privacy-cameras› (consulté le 31.5.2021).

8 ‹https://www.forbes.com/sites/davidjeans/2021/05/11/amazon-facial-rekognition-ban-civil-rights-groups/?sh=47343bb8325e› (consulté le 8.6.2021).

9 Voir par exemple : ‹https://www.eff.org/deeplinks/2020/01/ring-doorbell-app-packed-third-party-trackers› (consulté le 21.5.2021).

10 ‹https://axon.com› (consulté le 31.5.2021).

11 Elizabeth Joh/ Thomas Joo, The Harms of Police Surveillance Technology Monopolies (27.4.2021), Denver Law Review Forum, disponible sous : ‹https://papers.ssrn.com/sol3/papers.cfm ?abstract_id=3834777› (consulté le 8.6.2021).

même de développer en parfaite autonomie les outils qui lui sont nécessaires. Les partenariats public-privé sont, dès lors, essentiels et monnaie courante. Force est toutefois d'admettre qu'un tel partenariat fait sourciller lorsqu'il est le fait d'un acteur dominant sur le marché, comme peut l'être Amazon, dont le client principal serait en passe de devenir le gouvernement américain – il est vrai, davantage en relation avec AWS qu'avec les caméras *Ring*[12]. Le contrôle croissant exercé par les *Big Tech* sur nos données, la masse d'informations relatives à nos habitudes comportementales qu'ils peuvent en extraire et l'influence aujourd'hui avérée qui peut en résulter ne sont pas sans susciter des craintes légitimes quant au bon fonctionnement de la démocratie. Ces craintes sont d'autant plus fortes qu'elles reposent sur une asymétrie d'informations patente entre leurs détenteurs et… le reste du monde ; une asymétrie source de bien des inquiétudes.

Un certain réveil, tardif il est vrai, semble toutefois se dessiner, avec des efforts entrepris au niveau européen pour réglementer les *Big Tech* au travers de différentes propositions intervenues en fin d'année 2020, comme le *Digital Services Act*, le *Digital Market Act*[13], ou encore la proposition de réglementation sur la gouvernance des données[14], sans compter la dernière en date, du 21 avril 2021, sur l'intelligence artificielle[15], sur laquelle j'aurai l'occasion de revenir. Les Etats-Unis ne sont pas en reste, puisque Joe Biden a récemment nommé deux des plus sévères et respectés critiques des *Big Tech* à de hautes fonctions[16], à savoir Lina Khan comme commissaire à la *Federal Trade Commission*[17], et Tim Wu au *National Economic Council*[18]. Reste à savoir si ces efforts et nominations suffiront à rétablir une situation qui, à ce jour, a largement échappé à la sphère d'influence des législateurs. L'avenir le dira.

12 Encore faut-il toutefois qu'Amazon l'emporte dans la bataille judiciaire qui l'oppose à Microsoft, laquelle s'est vue adjuger, au travers de son service Azure, le contrat JEDI d'une valeur de 10 milliards de dollars. L'affaire, pendante, semble cependant loin d'être gagnée par Amazon (‹https://www.crn.com/news/cloud/microsoft-azure-creates-top-secret-government-cloud-as-jedi-battle-rages-on›, consulté le 31.5.2021).

13 ‹https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2347› (consulté le 31.5.2021).

14 ‹https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-data-governance-data-governance-act› (consulté le 31.5.2021).

15 ‹https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence› (consulté le 31.5.2021).

16 ‹https://www.npr.org/2021/03/22/975220122/big-tech-showdown-looms-as-biden-taps-top-critics-lina-khan-tim-wu ?t=1622495538587› (consulté le 31.5.2021).

17 Lina Khan, Amazon's Antitrust Paradox, Yale Law Journal Vol. 126/2017, 710 (disponible sous : ‹https://www.yalelawjournal.org/note/amazons-antitrust-paradox› [consulté le 31.5.2021]).

18 Tim Wu, The Curse of Bigness: Antitrust in the New Gilded Age, 2018.

1.2 Le risque de dévier vers une société de surveillance

L'engouement suscité par le développement des systèmes d'intelligence artificielle et leurs potentialités dans le cadre de la justice pénale n'est, ainsi, pas sans danger. Preuve en est, par exemple, le peu de fiabilité dont témoignent encore les outils de reconnaissance faciale tels que *Rekognition*, proposé par Amazon, qui a faussement identifié 28 membres du Congrès comme des personnes ayant été arrêtées par le passé, en 2018, avec – qui plus est – une marge d'erreurs plus importante lors de l'identification de personnes de couleur[19], par ailleurs démontrée sur le plan scientifique[20]. Sans surprise, le taux de reconnaissance s'améliore toutefois constamment, une étude publiée au mois d'avril 2020 par le *National Institute of Standards and Technology (NIST)* ayant conclu à un taux d'erreur de 0.08% comparé à 4.1% en 2014[21].

Une étude publiée le 5 septembre 2019 par *Pewresearch* souligne cependant que 56% des Américains ont confiance en l'utilisation potentielle de tels outils par les autorités et que 59% d'entre eux comprennent l'intérêt présenté par leur emploi dans des espaces publics, même si ce taux baisse auprès du public plus jeune et, sans guère de surprise, auprès des minorités[22]. Une interdiction pure et simple de ces outils apparaît sans doute excessive en raison de l'avantage que procure leur exploitation aux forces de l'ordre, en leur proposant un panel de suspects potentiels, dont la pertinence sera ensuite appréciée en cours d'enquête. En revanche, leur utilisation se doit d'être soumise à certaines cautèles au vu de l'ingérence importante qu'ils représentent dans la vie privée des individus.

Conscients de l'utilité de ces systèmes, mais aussi de leurs risques, certains Etats, agissant le plus souvent par l'entremise de leurs autorités en matière de protection des données, se sont prononcés sur les conditions liées à l'exploitation de tels systèmes, par exemple dans les aéroports pour la France[23]. D'autres, plus restrictifs, se sont prononcés contre leur utilisation, comme l'Italie[24]. Le 1er mars 2021, la Commission australienne des droits de l'homme a

19 ‹https://www.theguardian.com/technology/2018/jul/26/amazon-facial-rekognition-congress-mugshots-aclu› (consulté le 1.6.2021).

20 Joy Buolamwini/Timnit Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, Proceedings of Machine Learning Research Vol. 81/2018, 1 ss.

21 William Crumpler, How Accurate are Facial Recognition Systems – and Why Does it Matter? (24.4.2020), disponible sous : ‹https://www.csis.org/blogs/technology-policy-blog/how-accurate-are-facial-recognition-systems-%E2%80%93-and-why-does-it-matter› (consulté le 1.6.2021).

22 ‹https://www.pewresearch.org/internet/2019/09/05/more-than-half-of-u-s-adults-trust-law-enforcement-to-use-facial-recognition-responsibly/› (consulté le 1.6.2021).

23 Voir la prise de position de la CNIL du 29.10.2020 : ‹https://www.csis.org/blogs/technology-policy-blog/how-accurate-are-facial-recognition-systems-%E2%80%93-and-why-does-it-matter› (consulté le 1.6.2021).

24 Prise de position de l'autorité italienne du 16.4.2021 : ‹https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9575842› (consulté le 1.6.2021).

enjoint les autorités à interdire temporairement le recours aux outils de reconnaissance faciale aussi longtemps qu'un cadre réglementaire n'aurait pas été adopté en la matière[25]. Au Royaume-Uni, la *High Court of Justice* a, pour sa part, jugé dans un arrêt du 4 septembre 2019 le recours à de tels outils par les forces de l'ordre comme admissible, à partir du moment où leur utilisation reposait sur une base légale, que cet usage était limité dans le temps et sur le plan géographique, qu'il poursuivait un but bien déterminé, et que toutes les données récoltées étaient détruites si aucune d'entre elles ne méritait de retenir l'attention[26].

Au niveau européen, le dernier mot reviendra sans doute à la Commission européenne, qui, désireuse d'harmoniser la mise en œuvre des systèmes d'intelligence artificielle, a publié le 21 avril 2021 une ambitieuse proposition de réglementation sur l'intelligence artificielle[27]. Structurée autour du niveau de risques engendrés par les systèmes concernés, la Commission considère que le déploiement, par les autorités, de systèmes ayant pour objectif de permettre l'identification biométrique à distance dans les espaces publics, et donc les outils de reconnaissance faciale, sont interdits, à moins qu'ils ne poursuivent trois objectifs[28] :

- La recherche de victimes de crimes, ainsi que les enfants disparus ;
- La prévention d'un risque d'atteinte imminente à la vie ou à l'intégrité physique d'individus, dont les attaques terroristes font partie ;
- La découverte, localisation et identification de l'auteur d'un délit énoncé à l'art. 2.2 de la Décision-cadre 2002/584/JAI du Conseil du 13 juin 2002 relative au mandat d'arrêt européen et aux procédures de remise entre Etats membres – Déclarations de certains États membres sur l'adoption de la décision-cadre[29], pour autant que ce délit soit sanctionné d'une peine d'emprisonnement minimale de trois ans dans l'Etat en question.

---

25 ‹https://tech.humanrights.gov.au/sites/default/files/2021-05/AHRC_RightsTech_2021_Final_Report.pdf› (consulté le 8.6.2021) commenté le 28.5.2021 dans la revue en ligne ITNews : ‹https://www.itnews.com.au/news/human-rights-commission-calls-for-temporary-ban-on-high-risk-govt-facial-recognition-565173› (consulté le 8.6.2021).

26 [2019] EWHC 2341, disponible sous : ‹https://www.judiciary.uk/wp-content/uploads/2019/09/bridges-swp-judgment-Final03-09-19-1.pdf› (consulté le 1.6.2021).

27 Cf. *supra* n. 19.

28 Art. 5 de la proposition, c. 18 à 22 du préambule.

29 ‹https://eur-lex.europa.eu/resource.html ?uri=cellar :3b151647-772d-48b0-ad8c-0e4c78804c2e.0007.02/DOC_1&format=PDF› (consulté le 1.6.2021). Figurent sur cette liste les infractions suivantes : la participation à une organisation criminelle ; le terrorisme ; la traite des êtres humains ; l'exploitation sexuelle des enfants et la pédopornographie ; le trafic illicite de stupéfiants et de substances psychotropes ; le trafic illicite d'armes, de munitions et d'explosifs ; la corruption ; la fraude ; le blanchiment du produit du crime ; le faux monnayage ; la cybercriminalité ; les crimes contre l'environnement ; l'aide à l'entrée et au séjour irréguliers ; l'homicide volontaire, coups et blessures graves ; le trafic illicite d'organes et de tissus humains ; l'enlèvement, la séquestration

A supposer que ces conditions soient réalisées, le recours à de tels systèmes demeurerait en toute hypothèse subordonné au respect du principe de proportionnalité et exigerait ainsi que soit prise en considération la nécessité d'y recourir, en tenant compte, d'un côté, du préjudice qui pourrait exister en cas de non recours à un tel outil, et de l'autre, de l'atteinte aux droits fondamentaux résultant d'un tel usage. L'autorisation permettant une telle mise en œuvre devrait émaner d'une autorité judiciaire et résulter d'une requête dûment motivée (hormis urgence très particulière). Toujours en ligne avec le principe de proportionnalité, cette autorisation devrait toutefois être limitée au strict nécessaire, aussi bien sur le plan temporel que géographique. Enfin, la Commission considère qu'il appartient à chaque Etat membre de consacrer dans sa législation nationale la possibilité de mettre en œuvre des outils de reconnaissance faciale aux conditions fixées par la proposition ; autrement dit, si la proposition était adoptée, il appartiendrait à chaque Etat d'édicter une base légale permettant la mise en œuvre de la réglementation sur ce point.

L'approche de la Commission témoigne de la volonté de tirer parti des innovations que l'intelligence artificielle permet d'envisager, tout en entourant les risques qui peuvent en découler de cautèles propres à les minimiser. On ne peut, à l'évidence, que plébisciter une telle approche, toute la question étant de savoir si les cautèles envisagées sont les bonnes. L'avenir le dira.

2. Les outils de prévisibilité en matière criminelle

« Avec l'aide de vos mutants précogs, vous avez audacieusement et efficacement aboli le système punitif post-crime fondé sur l'emprisonnement et l'amende. Comme nous le savons tous, la perspective du châtiment n'a jamais été très dissuasive ; quant aux victimes, une fois mortes elles n'en retiraient guère de réconfort »[30].

Bien que nul ne prétende que le recours à l'intelligence artificielle soit un jour susceptible de renverser le système judiciaire en matière pénale au point où Philip K. Dick l'avait fait dans *Minority Report*, en anticipant purement et simplement des crimes à venir, le souci de prévenir le crime et ainsi le réduire a toujours été au cœur de la criminologie.

---

et la prise d'otage ; le racisme et la xénophobie ; les vols organisés ou avec arme ; le trafic illicite de biens culturels ; l'escroquerie ; le racket et l'extorsion de fonds ; la contrefaçon et le piratage de produits ; la falsification de documents administratifs et trafic de faux ; la falsification de moyens de paiement ; le trafic illicite de substances hormonales et autres facteurs de croissance ; le trafic illicite de matières nucléaires et radioactives ; le trafic de véhicules volés ; le viol ; l'incendie volontaire ; les crimes relevant de la Cour pénale internationale ; le détournement d'avion/navire ; le sabotage.

30 Philip K.Dick, Minority Report, 1956.

Les premières recherches en la matière semblent avoir été menées en France en 1829 lorsque Adriano Balby et André Michel Guerry étudièrent les corrélations qui pouvaient exister entre le niveau éducatif et la perpétration d'infractions[31]. Les années ont passé, et la technologie a évolué. Le *Big Data*, couplé aux techniques d'apprentissages par la machine, permet aujourd'hui, on s'en doute, d'aller beaucoup plus loin. Plusieurs prestataires, là encore privés, se sont lancés sur ce marché, tels Palantir[32] ou Predpol[33]. L'analyse des données et leur traitement par des algorithmes permettent assurément aux systèmes d'intelligence artificielle de découvrir des tendances (*patterns*) plus facilement et infiniment plus rapidement qu'un être humain. Toutefois, au fil de l'entraînement des algorithmes, ces probabilités ont rapidement tendance à se transformer en vérités pour ces algorithmes, sans l'approche critique que l'être humain conserve, aussi entraînés soient-ils.

Quel que soit le système utilisé, l'objectif demeure le même : permettre d'orienter les patrouilles sur des zones identifiées comme « sensibles » et éviter le passage à l'acte du criminel, avec comme point de départ l'hypothèse selon laquelle le meilleur prédicteur des crimes à venir résulte des crimes passés. L'idée suivant laquelle la propension à la création de *clusters* dépendrait de la commission de crimes dans une certaine zone, par le passé, semble, il est vrai, confirmée par certaines études. Etre à même d'associer facteurs contextuels et événements délictuels permet non seulement de mieux anticiper la commission d'infractions, mais aussi d'allouer les ressources humaines et matérielles adéquates, ce qui, au final, permet également un meilleur contrôle des budgets.

Autrement dit, l'objectif n'est pas d'appréhender un individu déterminé avant qu'il ne commette un crime, mais bien plutôt de décourager la commission d'éventuelles infractions par le déploiement adéquat de forces de l'ordre en des zones jugées sensibles.

En 2019, l'Institut Paris Région (IPR) publiait un rapport dans lequel il soulignait l'intérêt de tels outils, tout en admettant que leur recours suscitait maintes questions sur le plan des libertés individuelles et le risque d'une surenchère de dispositifs de contrôle et un renforcement des logiques de surveillance massive[34]. Quelques années plus tôt, en 2013, RAND[35] avait de sur-

31 Pour un survol de cette histoire, voir : Joel Hunt, From Crime Mapping to Crime Forecasting: The Evolution in Place – Based Policing (10.7.2019), disponible sous : ‹https://nij.ojp.gov/topics/articles/crime-mapping-crime-forecasting-evolution-place-based-policing› (consulté le 1.6.2021).

32 ‹https://www.palantir.com› (consulté le 1.6.2021).

33 ‹https://www.predpol.com› (consulté le 1.6.2021).

34 ‹https://www.institutparisregion.fr/fileadmin/NewEtudes/Etude_1797/Etude_Police_Predictive_V5.pdf› (consulté le 2.6.2021).

35 RAND est un *think tank* américain, actif dans un grand nombre de domaines de la société et d'action de l'Etat.

croît attiré l'attention sur certaines limites de ces outils dans le cadre d'une étude[36], parmi lesquelles :

- Le fait que les prédictions reposent, par définition, sur des données passées et, par conséquent, impropres à anticiper d'éventuelles évolutions ne reposant pas sur elles (un reproche auquel on objectera que tel est généralement le cas de toute prédiction) ;
- Le fait que la qualité des outils dépend également de la qualité des données ingérées (*input*), avec un risque de biais importants. A titre d'exemple, le rapport mentionne le fait qu'une indication suivant laquelle des vols ont avant tout lieu en une zone donnée entre 7h et 8h le matin ne précise pas forcément si ces données ont trait à l'heure effective des vols ou de leur annonce à la police.

Si le risque de surenchère de dispositifs de contrôle est un souci important, évoqué également plus haut relativement aux outils de reconnaissance faciale, le risque d'atteinte aux libertés fondamentales paraît, à mon sens, beaucoup plus limité ici, aussi longtemps que ces outils ne servent qu'à permettre un déploiement adéquat de forces de police à vocation dissuasive. Dans ce cas, les données collectées ne revêtent pas de caractère personnel et relèvent plus de la statistique. Il en ira évidemment différemment si ces outils sont un jour utilisés pour appréhender un criminel en passe de commettre une infraction, comme l'imagine Philipp K. Dick, ce qui, à ce jour, n'est pas leur but.

Apparaît en revanche beaucoup plus préoccupant le risque de voir un algorithme travailler sur des données de mauvaise qualité avec, pour conséquence, l'obtention de résultats biaisés et donc trompeurs. Richardson, Schultz et Crawford s'en sont fait l'écho dans une contribution publiée en 2019, se consacrant à trois cas d'étude concernant le recours à de tels systèmes par les forces de police de Nouvelle-Orléans, de Chicago et du comté de Maricopa[37]. Déplorant l'absence de standards en relation avec le traitement des données jugées pertinentes, les auteurs relevaient que l'outil utilisé à Chicago reposait sur le nombre d'arrestations, et non de condamnations, celui employé à la Nouvelle-Orléans sur des pratiques policières discriminant manifestement les personnes de couleur, et celui du comté de Maricopa sur des pratiques discriminant la communauté hispanique. Etant donné le caractère propriétaire des algorithmes des sociétés privées mettant à disposition de tels outils, on imagine difficilement que ces biais puissent être écartés avec, comme conséquence, des résultats finalement peu probants. Face aux réticences rencontrées, la police de Los Angeles a, pour sa part, décidé dix ans après l'avoir adopté de ne plus

36 ‹https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf› (consulté le 2.6.2021).

37 Rashida Richardson/Jason Schultz/Kate Crawford, Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice, 94 N.Y.U. L. Rev. 192, 2019.

recourir à l'outil Predpol, qui n'aurait d'ailleurs pas démontré son efficacité dans la lutte contre la criminalité[38]. Ce constat est partagé par John Hollywood, analyste au sein de RAND, pour qui l'amélioration des prédictions obtenue grâce à ces outils n'est que de l'ordre de 10 à 25 %[39].

L'efficacité de ces outils resterait donc à démontrer. Une étude récente portant sur l'utilisation faite au Mexique d'outils prédictifs recourant aux réseaux neuronaux aurait toutefois constaté une nette amélioration de la précision, avec un taux de prévisibilité atteignant 81 %[40]. Leur utilisation semble en toute hypothèse être considérée par la Commission européenne comme un système présentant un degré de risque élevé, dont l'exploitation est soumise à plusieurs obligations. En effet, le chiffre 6 lettre g de l'Annexe III considère comme tels les systèmes d'intelligence artificielle utilisés dans le domaine de la criminalistique pour permettre aux autorités de mettre au jour des tendances (*patterns*) inconnues ou des relations cachées entre des données diverses recoupées au travers de différentes bases de données. Quand bien même le préambule ne détaille guère ce qu'il faut entendre en la matière[41], la disposition ne semble pas exiger que les données soient nécessairement qualifiées de « personnelles », et les outils de pronostics devraient donc. à mon sens, être couverts par cette disposition, dans la mesure où leur but consiste à déterminer, par un recoupement de données, certaines tendances en matière de criminalité.

A partir de là, toute entité qui entend développer ou faire développer un tel système en vue de sa mise sur le marché devra satisfaire à de nombreuses obligations décrites sommairement plus loin[42]. Il suffit ici de préciser que la proposition de règlement tient compte du risque majeur que constitue la mauvaise qualité des données potentiellement ingérées dans le système, puisqu'elle exige de la part du développeur qu'il mette sur pied une gouvernance en matière de données, qui porte notamment sur une appréciation du risque des biais potentiellement existants.

38 ‹https://www.latimes.com/california/story/2020-04-21/lapd-ends-predictive-policing-program-Los Angeles Times› (consulté le 2.6.2021).

39 ‹https://www.bbc.com/news/business-46017239› (consulté le 2.6.2021).

40 Ana Laura Lira Cortes/Carlos Fuentes Silva, Artificial Intelligence Models for Crime Prediction in Urban Spaces, Machine Learning and Applications: An International Journal (MLAIJ) Vol. 8, No 1, March 2021, disponible sous : ‹https://papers.ssrn.com/sol3/papers.cfm ?abstract_id=3822346› (consulté le 8.6.2021).

41 Voir c. 38 du préambule.

42 Cf. *infra* 3.2.

3. Le recours à des systèmes d'intelligence artificielle dans l'enceinte des tribunaux

Si nombreuses sont les applications possibles des systèmes d'intelligence artificielle dans le cadre de la prévention et de la détection des infractions, de tels outils sont également susceptibles d'être utilisés par les magistrats lorsqu'il s'agit de fixer la peine de l'accusé reconnu coupable.

Je me pencherai ici plus particulièrement sur deux aspects :

- L'affaire *Loomis* (3.1) ;
- La proposition de règlement sur l'intelligence artificielle, dans la mesure où elle est susceptible de s'appliquer à ces différents outils (3.2).

3.1 L'affaire Loomis

« *Ultimately, we conclude that, if used properly as set forth herein, a circuit court's consideration of a COMPAS risk assessment at sentencing does not violate a defendant's right to due process and that the circuit court did not erroneously exercise its discretion here* »[43].

A la lecture de ces mots, Eric Loomis a compris que sa condamnation à six ans d'emprisonnement, reposant sur le recours par l'autorité inférieure à un algorithme ayant apprécié le risque de récidive qu'il faisait courir à la société, venait d'être confirmée par la Cour Suprême de l'Etat du Wisconsin. Comment la Cour en est-elle arrivée à ce résultat ?

En 2013, Eric Loomis est inculpé pour avoir pris part à un brigandage. Dans le cadre de l'instruction, un rapport d'appréciation du risque de récidive est établi au moyen de l'outil COMPAS, distribué par *equivant*[44], qui repose sur des auditions de l'inculpé et son casier judiciaire[45]. Compte tenu du caractère propriétaire de l'algorithme, la méthodologie utilisée et son fonctionnement ne sont pas divulgués, au titre du secret d'affaires[46]. L'inculpé n'a connaissance que du résultat de l'appréciation (*output*), retenu par le Tribunal pour le condamner à six ans d'emprisonnement[47]. En appel, le condamné fait valoir que le recours à COMPAS viole son droit d'être entendu (*due process*), à tout le moins à trois égards : tout d'abord, l'algorithme fonctionnant sur la base de données de groupes, il contreviendrait, selon lui, au principe de l'individualisation de la peine ; ensuite, le fait d'ignorer la méthodologie à la base de son fonctionne-

43 *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), disponible sous : ‹https://caselaw.findlaw.com/wi-supreme-court/1742124.html› (consulté le 2.6.2021).

44 ‹https://www.equivant.com/northpointe-risk-need-assessments/› (consulté le 2.6.2021).

45 *State v. Loomis*, 754.

46 *Idem*, 761.

47 *Idem*, 755-756.

ment irait à l'encontre de l'obligation de fonder une condamnation sur la base d'informations précises ; enfin, les critères selon lesquels fonctionne COMPAS revêtiraient un caractère discriminatoire à l'encontre des accusés[48].

La Cour Suprême rejette un à un ces moyens de défense. Le fait de tenir compte du genre de l'accusé est un critère important, qui repose sur des données objectives[49]. Dans la mesure où l'appréciation se base sur des données connues de l'accusé et résulte de son passé criminel ou de ses auditions, il les connaissait toutes[50]. La Cour considère par ailleurs que le recours à COMPAS en soi n'implique pas que la peine n'est pas individualisée. Quand bien même COMPAS repose, il est vrai, sur des données de groupe[51], il ne s'agit que d'un outil, qui n'empêche nullement l'autorité de se forger sa propre opinion et de fixer la peine en conséquence[52].

Si le recours est rejeté, la Cour n'en rappelle pas moins que le recours à des outils tels que COMPAS présente des risques exigeant de l'autorité qu'elle motive clairement le résultat auquel elle aboutit, y compris quant aux facteurs pris en considération dans le cadre de la fixation de la peine, en dehors du seul résultat auquel a conduit l'exécution de l'algorithme[53]. Manifestement désireuse de témoigner de ses réticences à voir se propager le recours à de tels outils, la Cour attire par ailleurs l'attention des tribunaux de l'Etat du Wisconsin sur cinq points relatifs à COMPAS[54] : premièrement, son caractère propriétaire fait qu'il est impossible de comprendre le fonctionnement de l'algorithme sous-jacent ; deuxièmement, COMPAS ne permet pas d'identifier des individus au profil particulièrement dangereux, dès lors qu'il repose sur des données de groupe ; troisièmement, les données sur lesquelles repose l'algorithme sont des données nationales, qui ne tiennent pas forcément compte des particularités de l'Etat du Wisconsin ; quatrièmement, le risque de biais – notamment s'agissant des minorités – ne peut être exclu ; enfin, COMPAS a été développé, à la base, pour apprécier le risque présenté par un individu après sa condamnation, et non pour fixer la peine.

Cet arrêt a suscité de nombreuses critiques, notamment de la part de *ProPublica*, un journal d'investigation diffusé en ligne, qui, sous les plumes de Julia Angwin, Jeff Larson, Surya Mattu et Laurent Kirchner, a mis au jour les nombreux biais affectant COMPAS. Cet outil discriminerait ainsi systématiquement les minorités ; un afro-américain encourerait 77% de risques supplémentaires, par rapport à un caucasien, d'être considéré comme susceptible

48 *Idem*, 756-757.
49 *Idem*, 766-767.
50 *Idem*, 761-762.
51 *Idem*, 764.
52 *Idem*, 764-765.
53 *Idem*, 769.
54 *Idem*, 769-770.

de récidiver en commettant un crime violent, et 45 % de risque en plus s'agissant d'infractions de manière générale[55]. Pire encore, sur un examen de 7000 personnes, COMPAS n'aurait finalement correctement identifié un risque de récidive que dans 20 % des cas[56]. Une recension de cet arrêt, publiée dans la *Harvard Law Review*, évoquait le danger à recourir à des outils dont le fonctionnement, couvert par le secret d'affaires, les rend d'autant plus incompréhensibles pour les magistrats que ces derniers ne disposent pas des compétences nécessaires pour les comprendre ; favorisant la quantité d'informations au détriment de leur qualité, COMPAS fait courir le risque que les magistrats ne s'en remettent à des données empiriques, que l'être humain a tendance à suivre, plutôt qu'à leur propre raisonnement[57].

S'en sont suivies différentes propositions, visant à fournir au recours à de tels outils un cadre suffisant[58]. Dans son rapport, le consortium *Partnership on AI* considère que le développement et la mise en œuvre de tels outils doivent se faire autour de trois axes, subdivisés en dix recommandations :

- Précision, validité et biais :
  1. Les données d'entraînement doivent servir à mesurer les variables souhaitées ;
  2. Les biais pouvant en résulter doivent être mesurés et minimisés ;
  3. Les outils ne doivent pas être utilisés pour prédire des objectifs en réalité distincts ;

- Interface utilisateur :
  4. Les prédictions et la manière dont l'outil y parvient doivent être aisément compréhensibles ;
  5. Les outils doivent fournir des prédictions suffisamment précises ;
  6. Les utilisateurs de ces outils doivent être correctement formés sur le fonctionnement, l'utilisation et les limites de l'outil ;

55 ‹https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing› (consulté le 7.6.2021).

56 *Idem*.

57 *State v. Loomis – Wisconsin Supreme Court Requires Warnings Before Use of Algorithmic Risk Assessments in Sentencing*, 130 ; Harvard Law Review 1530 (2017).

58 Voir, parmi d'autres : Berkman Klein Center (9.11.2017), ‹https://cyber.harvard.edu/publications/2017/11/openletter› (consulté le 7.6.2021) ; Berkman Klein Center (10.12.2018), ‹https://cyber.harvard.edu/publication/2018/assessing-assessments› (consulté le 7.6.2021) ; Partnership on AI, ‹https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/› (consulté le 7.6.2021) ; The Law Society, Algorithms in the Criminal Justice System (juin 2019), disponible sous : ‹https://www.lawsociety.org.uk/topics/research/algorithm-use-in-the-criminal-justice-system-report› (consulté le 7.6.2021).

- Gouvernance, transparence et responsabilité :
  7. Les législateurs doivent s'assurer qu'ils tiennent compte des objectifs de politique législative recherchés dans la manière dont ces outils sont conçus ;
  8. La manière dont l'outil a été conçu, son architecture et la façon dont il a été entraîné doivent lui permettre d'être l'objet de recherches et d'une approche critique ;
  9. Les outils doivent permettre la conservation des données de manière à pouvoir être testés, contrôlés et potentiellement contestés ;
  10. Les différents Etats doivent endosser la responsabilité quant à l'utilisation, l'appréciation, la veille et l'audit de ces outils.

Si diverses approches ont été retenues, toutes soulignent les impératifs de : (i) qualité des données ingérées pour maîtriser les biais et s'assurer qu'elles sont spécifiques au but recherché (risque de récidive, risque de fuite, risque de faire défaut, etc.), (ii) formation et compréhension des utilisateurs pour mieux appréhender ces outils, et (iii) gouvernance et transparence qui doivent les entourer.

Bien que l'approche européenne ne soit pas spécifique aux outils susceptibles d'être utilisés dans le cadre de la justice pénale, la proposition de règlement européen sur l'intelligence artificielle constitue, à ce jour, le seul exemple de tentative de réglementation de ces systèmes à l'échelle mondiale. Loin de simples recommandations, la proposition offre ainsi, pour la première fois, une esquisse de cadre juridique. Qu'en est-il ?

3.2 La proposition de règlement sur l'intelligence artificielle

A l'image des outils de prévisibilité criminelle, les systèmes d'intelligence artificielle susceptibles d'être utilisés dans l'enceinte des tribunaux sont considérés, dans le cadre de la proposition de règlement, comme présentant un degré de risques élevé. En font ainsi partie, au regard de l'annexe III, les systèmes permettant d'évaluer les risques de perpétration d'infractions ou de récidives comme COMPAS[59], ceux ayant pour objectif d'interpréter l'état émotionnel d'une personne[60] ou, plus généralement, tout outil de justice prédictive visant à interpréter les faits, les conséquences juridiques qui en découlent[61] et la pertinence des preuves avancées[62].

La mise en œuvre des systèmes d'intelligence artificielle considérés comme présentant un niveau de risques élevé sera sujette à un nombre important

59 Ch. 6 lit. a Annexe III.
60 Ch. 6 lit. b Annexe III.
61 Ch. 8 Annexe III.
62 Ch. 6 lit. d Annexe III.

d'obligations, qui n'est pas sans rappeler, à différents égards, l'approche prise par la Commission dans le cadre du Règlement 2016/679 (RGPD) ou encore du Règlement 2017/745 sur les appareils médicaux. A supposer qu'elle soit adoptée, la réglementation aurait une portée extraterritoriale, puisqu'elle s'appliquerait à tout système commercialisé ou utilisé ou encore dont le résultat serait utilisé au sein de l'Union Européenne[63] ; seuls les outils utilisés par des organismes publics pour assurer la mise en œuvre de traités internationaux dans le domaine de l'entraide judiciaire font figure d'exception dans le cadre qui nous intéresse[64].

Sans entrer dans les détails de cette proposition de règlement, qui dépasse allègrement les 100 pages, on peut résumer les obligations mises à la charge du développeur d'un tel système comme suit :

- Mise en place d'un système de gestion des risques, qui passe notamment par leur identification et les mesures prises pour les minimiser, avec une appréciation du risque résiduel[65] ;
- Mise en place d'une gouvernance en matière de données, dont l'objectif consiste, en particulier, à examiner la possibilité de biais et éventuelles lacunes, ainsi que la manière d'y remédier[66] ;
- Mise en place d'une documentation technique répondant aux exigences posées par l'Annexe IV, dont l'objectif consiste, en particulier, à assurer une compréhension quant au fonctionnement de l'algorithme (notamment la logique, les choix architecturaux et de classification, les méthodes et techniques d'entraînement, le mécanisme de contrôle par l'humain, la description de toute modification, etc.)[67] ;
- Mise en place d'un système permettant la conservation automatique des logs pour assurer un traçage de tous les événements liés à l'utilisation de l'outil[68] ;
- Informations des utilisateurs portant, en particulier, sur le contenu de la documentation technique[69] ;
- Assurance qu'un contrôle par l'humain du système est possible, avec notamment pour objectif de comprendre tout biais éventuel qui pourrait résulter

63 Art. 2 (1).
64 Art. 2 (4).
65 Art. 9.
66 Art. 10.
67 Art. 11, 16 (c) et 18.
68 Art. 12, 16 (d) et 20.
69 Art. 13.

de l'utilisation du système et de pouvoir apprécier la pertinence du résultat proposé[70] ;

- Garantie de la sécurité et la stabilité du système[71] ;
- Apposition du marquage CE de conformité[72] ;
- Etre à même de démontrer la conformité du système mis sur le marché avec les exigences précitées[73] ;
- Mise en place d'un plan de gestion de qualité, dont l'objectif consiste notamment à assurer une documentation de tout ce qui précède, ainsi que tout suivi de l'utilisation du système une fois sur le marché (*post monitoring activity*), des éventuelles annonces aux autorités nationales compétentes et des actions correctives entreprises en tant que de besoin[74].

Relevons d'emblée que si l'essentiel des obligations repose, sans surprise, sur les épaules du développeur, les utilisateurs potentiels, tels les autorités et les tribunaux, ne sont pas pour autant libres de toute obligation. Ainsi devront-ils, d'une part, s'assurer que les données qu'ils sont susceptibles d'enregistrer dans le système utilisé sont de qualité (soit dépourvues de biais et propres à atteindre le but recherché) et que les logs sont conservés et, d'autre part, être à même de mener des analyses d'impact en matière de données, en application de l'art. 35 RGPD, sur la base des informations que le développeur leur aura communiquées, conformément à ses propres obligations[75].

Au final, l'approche fondée sur les risques privilégiée par la Commission européenne convainc. Certes, les obligations sont lourdes, au point que l'on peut se demander si un tel cadre réglementaire ne reviendra pas à privilégier les *Big Tech* et autres multinationales, seules à même de supporter les coûts nécessaires à une telle mise en conformité, au détriment de plus petites sociétés pourtant bien souvent porteuses d'innovations. Quoi qu'il en soit, l'effort de la Commission européenne pour proposer un cadre juridique complet ne peut être que salué. Comme elle l'avait été pour le RGPD, l'Union européenne apparaît ainsi une nouvelle fois comme précurseuse dans un domaine particulièrement sensible pour l'avenir de nos sociétés.

70 Art. 14.
71 Art. 15.
72 Art. 16 (i).
73 Art. 16 (j).
74 Art. 16 (b), 17 et 21.
75 Art. 29. Autant dire que ces obligations ne s'appliqueront pas aux autorités et tribunaux suisses qui, en tant qu'utilisateurs situés en dehors du territoire de l'Union Européenne, ne seront pas soumis à la réglementation ; exception doit cependant être faite, à mon sens, de l'hypothèse où le résultat de l'utilisation du système le serait vis-à-vis d'un ressortissant de l'Union européenne (art. 1 [c]).

4. Conclusion

Arrivé au terme de ce survol, que faut-il en retenir ?

Tout d'abord, le fait que les systèmes d'intelligence artificielle sont appelés à jouer un rôle croissant dans l'administration de la justice pénale, à tout le moins en matière de prévisibilité et dans le cadre de l'instruction. Il est, en revanche, plus difficile d'apprécier le rôle que ces outils pourraient jouer dans l'enceinte des tribunaux – du moins à court ou moyen terme – où leur utilisation – du moins sous nos latitudes – me laisse plus circonspect, faute d'apporter la preuve d'une valeur ajoutée incontestable.

Ensuite, les interrogations que suscitent le rôle majeur joué, à ce jour, par des acteurs privés dans ces développements et les partenariats avec les gouvernements qui s'ensuivent, notamment compte tenu de l'exemple qui nous est offert par les Etats-Unis : (i) le risque de dépendance des autorités à l'égard d'entités privées ; (ii) le pouvoir de contrôle important joué par des entités privées sur des informations dont le traitement résulte de prérogatives publiques et (iii) l'importance de sociétés américaines dans ce secteur.

Enfin, le fait que le recours à de tels outils dans un domaine aussi sensible ne se conçoit que moyennant la mise en place de cautèles visant à éviter que ne se crée un état de surveillance redouté de la part des citoyens ; craintes, qui plus est, renforcées en cette période de pandémie. Ces cautèles doivent s'inscrire dans un cadre législatif au sens formel largement plébiscité et pour lequel la proposition de règlement sur l'intelligence artificielle publiée le 21 avril 2021 par la Commission européenne apparaît comme un fer de lance en la matière.

Porteuse de promesses, l'intelligence artificielle l'est aussi de dangers. Les Etats en sont conscients et des efforts se font jour pour fournir les garanties nécessaires à un développement positif de ces technologies dans l'intérêt de tous. Il ne reste qu'à espérer que ces efforts porteront leurs fruits, et que seules les promesses se concrétiseront.