Parliamentary Assembly
Assemblée parlementaire
http://assembly.coe.int

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

**Version provisoire**

## Commission des questions juridiques et des droits de l'homme

# Justice par algorithme – le rôle de l'intelligence artificielle dans les systèmes de police et de justice pénale

## Rapport*

Rapporteur : M. Boriss CILEVIČS, Lettonie, Groupe des socialistes, démocrates et verts

### A. Projet de résolution

1. On trouve désormais des applications de l'intelligence artificielle (IA) dans de nombreuses sphères de l'activité humaine, de la recherche pharmaceutique aux médias sociaux, de l'agriculture aux achats en ligne, du diagnostic médical à la finance et de la composition musicale à la justice pénale. Elles sont de plus en plus puissantes et influentes et les citoyens ignorent bien souvent quand, où et comment elles sont utilisées.

2. Le système de justice pénale représente l'un des principaux domaines de compétence de l'État : assurer l'ordre public, prévenir les violations de divers droits fondamentaux et déceler les infractions pénales, enquêter à leur sujet, poursuivre leurs auteurs et les sanctionner. Il confère aux autorités d'importants pouvoirs intrusifs et coercitifs, notamment la surveillance, l'arrestation, la perquisition et la saisie, la détention et le recours à la force physique et même à la force létale. Ce n'est pas un hasard si le droit international des droits de l'homme impose le contrôle juridictionnel de tous ces pouvoirs, c'est-à-dire un contrôle effectif, indépendant et impartial de l'exercice, par les autorités, de leurs compétences en droit pénal qui peuvent donner lieu à une profonde ingérence dans les droits de l'homme fondamentaux. L'introduction d'éléments non humains dans la prise de décision au sein du système de justice pénale peut donc présenter des risques particuliers.

3. Si l'on souhaite que les citoyens acceptent l'utilisation de l'IA et jouissent des avantages qu'elle pourrait présenter, ils doivent avoir confiance dans le fait que tout risque est géré de manière satisfaisante. Si l'IA doit être mise en place avec le consentement éclairé du public, comme on peut s'y attendre dans une démocratie, alors une réglementation efficace et proportionnée est nécessaire.

4. La réglementation de l'IA, qu'il s'agisse d'une autorégulation volontaire ou de dispositions légales obligatoires, doit reposer sur des principes éthiques fondamentaux universellement admis et applicables. L'Assemblée considère que ces principes peuvent être regroupés dans les grandes catégories suivantes :

   4.1. la transparence, y compris l'accessibilité et l'explicabilité ;

   4.2. la justice et l'équité, y compris la non-discrimination ;

   4.3. la prise de décision par une personne, qui en est responsable, et la mise à disposition de voies de recours ;

   4.4. la sûreté et la sécurité ;

   4.5. le respect de la vie privée et la protection des données.

5. L'Assemblée se félicite de la Recommandation Rec/CM(2020)1 du Comité des Ministres sur les impacts des systèmes algorithmiques sur les droits de l'homme, complétée par ses lignes directrices sur le traitement des impacts sur les droits de l'homme des systèmes algorithmiques, ainsi que de la recommandation de la

* Projet de résolution et projet de recommandation adoptés à l'unanimité par la commission le 9 septembre 2020.

Commissaire aux droits de l'homme du Conseil de l'Europe intitulée « Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme ». L'Assemblée approuve les propositions générales énoncées dans ces textes qui s'appliquent également dans le domaine des systèmes de police et de justice pénale.

6. L'Assemblée observe qu'un nombre important d'applications de l'IA à l'usage des systèmes de police et de justice pénale ont été développées à travers le monde. Certaines d'entre elles sont utilisées dans les États membres du Conseil de l'Europe, ou leur déploiement y est envisagé. Ces applications englobent notamment la reconnaissance faciale, la police prédictive, l'identification de victimes potentielles d'actes criminels, l'évaluation des risques en matière de détention provisoire, de peine prononcée et de libération conditionnelle, ou encore l'identification d'affaires non résolues qui pourraient l'être aujourd'hui grâce aux technologies modernes de criminalistique.

7. L'Assemblée considère que l'utilisation de l'IA dans les systèmes de police et de justice pénale risque d'être à de nombreux égards incompatible avec les principes éthiques fondamentaux mentionnés ci-dessus. Les situations suivantes sont particulièrement préoccupantes :

7.1. Les systèmes d'IA peuvent être fournis par des entreprises privées, qui peuvent invoquer leurs droits de propriété intellectuelle pour refuser l'accès au code source. Ces entreprises peuvent même acquérir la propriété des données traitées par le système, au détriment de l'institution publique qui fait appel à leurs services. Les utilisateurs et les sujets d'un système peuvent ne pas disposer des informations ou des explications nécessaires pour comprendre de manière élémentaire son fonctionnement. Il arrive que l'être humain ne soit pas en mesure de comprendre certains processus qui interviennent dans le fonctionnement d'un système d'IA. De telles considérations posent la question de la transparence (et, par conséquent, de la responsabilité/de l'obligation de rendre des comptes).

7.2. Les systèmes d'IA sont formés à partir d'énormes quantités de données, qui peuvent être entachées de préjugés de longue date, notamment par une corrélation indirecte entre certaines variables prédictives et des pratiques discriminatoires (comme l'utilisation du code postal pour identifier une communauté ethnique traditionnellement soumise à un traitement discriminatoire). Cette situation est particulièrement préoccupante dans les domaines de la police et de la justice pénale, en raison à la fois de la prévalence de la discrimination fondée sur divers motifs dans ce contexte et de l'importance des décisions qui peuvent être prises. L'apparente objectivité mécanique de l'IA peut masquer ces préjugés (« techwashing »), les renforcer, voire les perpétuer. Certaines techniques d'IA ne sont pas aisément contestables par les personnes concernées par leur application. De telles considérations soulèvent la question de la justice et de l'équité.

7.3. Les contraintes de ressources ou de temps, le manque de compréhension et la déférence ou la réticence à s'écarter des recommandations d'un système d'IA peuvent placer les fonctionnaires de police et les juges dans une situation d'extrême dépendance vis-à-vis de ces systèmes, qui les conduit à abdiquer leurs responsabilités professionnelles. De telles considérations soulèvent la question de la responsabilité de la prise de décision.

7.4. Elles ont également une incidence les unes sur les autres. Le manque de transparence d'une application d'IA réduit la capacité des utilisateurs humains à prendre des décisions en toute connaissance de cause. Ce manque de transparence et l'existence d'une responsabilité humaine incertaine compromettent la capacité des mécanismes de contrôle et de recours à garantir la justice et l'équité.

7.5. L'application de systèmes d'IA dans des contextes distincts mais liés, en particulier par des institutions différentes qui s'appuient successivement sur le travail des autres, peut avoir des effets cumulatifs inattendus, voire imprévisibles.

7.6. De même, l'ajout d'éléments basés sur l'IA aux technologies existantes peut avoir des conséquences d'une gravité imprévue ou non intentionnelle.

8. L'Assemblée conclut que, si l'utilisation de l'IA dans les systèmes de police et de justice pénale présente des avantages importants lorsqu'elle est correctement réglementée, elle peut avoir une incidence particulièrement grave sur les droits de l'homme dans le cas contraire.

9. Par conséquent, pour ce qui est des systèmes de police et de justice pénale, l'Assemblée appelle les États membres à :

9.1. adopter un cadre juridique national pour réglementer l'utilisation de l'IA, sur la base des principes éthiques fondamentaux mentionnés ci-dessus ;

9.2. tenir un registre de toutes les applications d'IA utilisées dans le secteur public et à s'y référer lors de l'examen de nouvelles applications, de manière à identifier et évaluer les effets cumulatifs éventuels ;

9.3. s'assurer que l'IA sert des objectifs politiques généraux et que ces objectifs ne se limitent pas aux domaines dans lesquels l'IA peut être appliquée ;

9.4. veiller à ce qu'il existe un fondement légal suffisant pour chaque application d'IA et pour le traitement des données pertinentes ;

9.5. garantir que toutes les institutions publiques qui mettent en œuvre des applications d'IA disposent d'une expertise interne qui leur permet d'évaluer la mise en place, le fonctionnement et l'incidence de tels systèmes et de dispenser des conseils en la matière ;

9.6. consulter les citoyens de manière significative, y compris les organisations de la société civile et les représentants des communautés, avant de mettre en place des applications d'IA ;

9.7. s'assurer que toute nouvelle application d'IA est justifiée, sa finalité précisée et son efficacité confirmée avant sa mise en service, en fonction du contexte opérationnel particulier ;

9.8. procéder à des évaluations d'impact initiales et périodiques des applications d'IA afin d'analyser, entre autres, les questions de respect de la vie privée et de protection des données, les risques de préjugés/discrimination et les conséquences pour les individus des décisions prises à l'aide de l'IA, en accordant une attention particulière à la situation des minorités et des groupes vulnérables et défavorisés ;

9.9. veiller à ce que les processus décisionnels essentiels des applications d'IA puissent être expliqués à leurs utilisateurs et aux personnes concernées par leur mise en service ;

9.10. mettre uniquement en œuvre que les applications d'IA qui peuvent être examinées et testées sur leur lieu d'exploitation ;

9.11. examiner avec soin les conséquences possibles de l'ajout d'éléments basés sur l'IA aux technologies existantes ;

9.12. mettre en place des mécanismes de contrôle éthique efficaces et indépendants pour la mise en place et l'exploitation des systèmes d'IA ;

9.13. veiller à ce que la mise en place, l'exploitation et l'utilisation des applications d'IA puissent faire l'objet d'un contrôle judiciaire efficace.

## B. Projet de recommandation

1. L'Assemblée renvoie à sa Résolution … (20…) intitulée « Justice par algorithme – le rôle de l'intelligence artificielle dans les systèmes de police et de justice pénale ». Elle observe que cette résolution a été adoptée alors que des travaux étaient en cours au sein du Conseil de l'Europe, menés par le Comité ad hoc sur l'intelligence artificielle (CAHAI).

2. L'Assemblée rappelle que tous les États membres du Conseil de l'Europe sont soumis aux mêmes normes fondamentales en matière de droits de l'homme et d'état de droit, notamment celles qui sont établies par la Convention européenne des droits de l'homme, selon l'interprétation retenue par la jurisprudence de la Cour européenne des droits de l'homme. Elle estime qu'un patchwork réglementaire – avec des normes différentes selon les pays – pourrait conduire les entreprises à rechercher les normes éthiques les plus avantageuses pour elles et à délocaliser le développement et l'utilisation de l'IA dans des régions soumises à des normes éthiques moins exigeantes.

3. L'Assemblée appelle par conséquent le Comité des Ministres à tenir compte de l'impact particulièrement grave que pourrait avoir sur les droits de l'homme le recours à l'intelligence artificielle dans les systèmes de police et de justice pénale lorsqu'il évaluera la nécessité et la faisabilité d'un cadre juridique européen applicable à l'intelligence artificielle.

**Exposé des motifs par M. Cilevičs, Rapporteur**

## 1. Introduction

1. La proposition de recommandation sur laquelle repose le présent rapport, que j'ai déposée le 26 septembre 2018, a été renvoyée devant la commission par le Bureau le 12 octobre 2018, à la suite de quoi j'ai été nommé rapporteur le 21 janvier 2019[1]. La commission a organisé une audition d'experts lors de sa réunion des 14-15 novembre 2019 à Berlin, Allemagne, à laquelle ont participé M. Michael Veale, chargé de cours en réglementation et droits numériques, University College London, Royaume-Uni et Mme Marion Oswald, chercheuse principale en droit du vice-président, Université de Northumbria, Royaume-Uni. La visite d'information prévue auprès de la police des West Midlands, Royaume-Uni, a été annulée en raison de la pandémie de COVID-19, mais elle a été remplacée par des vidéoconférences avec Tom McNeil, de la commission d'éthique du Commissariat aux questions de police et de criminalité des West Midlands, et avec Chris Todd, commissaire principal, et Nick Dale, inspecteur principal des West Midlands. Je souhaite remercier tous les intéressés de leurs contributions au présent rapport.

2. L'intelligence artificielle (IA) n'est plus un sujet de science-fiction, même si elle ne correspond pas encore à tout ce que prédisait la science-fiction. Nous ne disposons pas de machines sensibles capables d'égaler ou de surpasser l'être humain dans de multiples domaines (ce que l'on qualifie d'intelligence artificielle « générale » ou « forte »), mais il existe déjà des systèmes capables d'effectuer des tâches précises, comme la reconnaissance de modèles ou de catégories ou la prévision des comportements avec un certain degré « d'autonomie », si l'on peut dire (l'intelligence artificielle « restreinte » ou « faible »). Ces systèmes sont présents dans de très nombreuses sphères de l'activité humaine, de la recherche pharmaceutique aux médias sociaux, de l'agriculture aux achats en ligne, du diagnostic médical à la finance et de la composition musicale à la justice pénale. Ils sont de plus en plus puissants et influents et les citoyens ignorent bien souvent quand, où et comment ils sont utilisés. De fait, plus ils sont perfectionnés, moins il arrive qu'ils soient apparents : la société de recherche OpenAI a récemment annoncé qu'elle s'abstiendrait de rendre public un nouveau système de traitement par intelligence artificielle du langage naturel (c'est-à-dire du langage humain), car celui-ci était capable de produire des textes qu'il était impossible de distinguer des textes créés par l'homme et il était trop facile d'en faire une utilisation abusive.

3. Comme l'indique la proposition de recommandation, le système de justice pénale représente l'un des principaux domaines de compétence de l'État : assurer l'ordre public, prévenir les violations de divers droits fondamentaux et déceler les infractions pénales, enquêter à leur sujet, poursuivre leurs auteurs et les sanctionner. Il confère aux autorités d'importants pouvoirs intrusifs et coercitifs, notamment la surveillance, l'arrestation, la perquisition et la saisie, la détention et le recours à la force physique et même à la force létale. Ce n'est pas un hasard si le droit international des droits de l'homme impose le contrôle juridictionnel de tous ces pouvoirs, c'est-à-dire un contrôle effectif, indépendant et impartial de l'exercice, par les autorités, de leurs compétences en droit pénal qui peuvent donner lieu à une profonde ingérence dans les droits de l'homme fondamentaux. La mise en place d'éléments non humains de prise de décision au sein du système de justice pénale peut donc présenter des risques particuliers.

4. L'Assemblée a déjà abordé certaines des questions pertinentes pour le présent rapport dans sa Recommandation 2102 (2017) intitulée « La convergence technologique, l'intelligence artificielle et les droits de l'homme ». Elle y indiquait que « le législateur a de plus en plus de mal à s'adapter à l'évolution de la science et des technologies, et à élaborer les textes réglementaires et les normes qui s'imposent ». L'Assemblée concluait que « la préservation de la dignité humaine au XXI$^{e}$ siècle suppose le développement de nouvelles formes de gouvernance et de débat public ouvert, éclairé et contradictoire, de nouveaux mécanismes législatifs et surtout l'instauration d'une coopération internationale permettant de relever ces nouveaux défis de la manière la plus efficace ».

5. Bien que la commission des questions juridiques et des droits de l'homme n'examine pas pour la première fois l'intelligence artificielle, nous analyserons dans un premier temps certaines questions essentielles et générales, en nous appuyant sur les travaux menés par la commission de la culture, de la science, de l'éducation et des médias lors de l'élaboration de la Recommandation 2102, avant de nous pencher sur le cas précis de l'utilisation de l'intelligence artificielle et des algorithmes dans le système de justice pénale.

[1] Renvoi n$^{o}$ 4407.

### *1.1. Notions clés*

6. L'expression « intelligence artificielle » a été inventée pour la première fois en 1955 par, entre autres, John McCarthy du Dartmouth College au New Hampshire, dans une proposition de projet de recherche qui reposait sur l'hypothèse de travail que « tout aspect de l'apprentissage ou toute autre caractéristique de l'intelligence peut en principe être décrit avec une telle précision qu'il est possible de réaliser une machine capable de l'imiter ». Malheureusement, il n'existe toujours pas de définition universellement admise de l'intelligence artificielle et les désaccords sont nombreux sur ce que devrait en être la définition[2]. L'un des aspects centraux de ce problème est l'absence de définition commune de l'intelligence humaine : on suppose que l'intelligence artificielle doit s'entendre en gros comme une imitation de l'intégralité ou d'une partie des caractéristiques de l'intelligence humaine, ce qui par définition ne serait pas le cas de toute forme future d'intelligence artificielle générale ou forte « surintelligente »[3]. Pour une description générale de l'IA, veuillez consulter l'annexe au présent rapport.

7. On entend parfois dire que l'une des caractéristiques qui définit l'intelligence artificielle est son « autonomie ». Il convient toutefois d'être très prudent avec cette notion, car elle peut avoir de graves répercussions en matière d'obligation de rendre des comptes et de responsabilité, jusqu'à la question de savoir si l'intelligence artificielle doit être considérée ou non comme un agent moral, voire comme une personne morale. C'est ce qu'a très bien décrit le Groupe européen d'éthique des sciences et des nouvelles technologies. « Le terme "autonomie" vient de la philosophie et désigne la capacité des êtres humains à légiférer pour eux-mêmes, à formuler, penser et choisir les normes, les règles et les lois qu'ils suivront […]. Par conséquent, l'autonomie au sens éthique du terme peut uniquement s'appliquer aux êtres humains. C'est donc en quelque sorte un abus de langage que d'appliquer le terme "autonomie" à de simples objets, et même à des systèmes complexes très avancés qui sont adaptatifs, voire "intelligents" […]. Comme aucun objet ni système, aussi avancé et sophistiqué soit-il, ne peut en soi être qualifié "d'autonome" au sens éthique premier du terme, il ne peut se voir reconnaître le statut moral d'être humain ni hériter de la dignité humaine […]. Les êtres humains doivent avoir la capacité de déterminer les valeurs au service desquelles la technologie doit être mise, les éléments importants sur le plan moral, ainsi que les objectifs ultimes qu'il convient de poursuivre et la conception du bien qui mérite d'être défendue. Ces choix ne doivent pas être abandonnés à des machines, quelle que soit la puissance de leurs capacités »[4].

### *1.2. Considérations générales sur les possibilités et les risques de l'utilisation de l'intelligence artificielle*

8. On a également affirmé que l'intelligence artificielle pouvait « redéfinir le travail ou améliorer les conditions de travail des êtres humains et diminuer le besoin de participation, d'intervention et d'interférence humaine dans la production. Elle permet d'aider les êtres humains ou de les remplacer par une technologie intelligente dans les travaux difficiles, sales, ennuyeux ou dangereux, et même au-delà »[5]. De fait, comme nous l'avons indiqué au paragraphe 2, l'intelligence artificielle est déjà appliquée dans de nombreux domaines, le plus souvent avec des résultats positifs.

9. Le pouvoir que pourrait avoir l'intelligence artificielle présente également des risques. Sa rapidité, sa complexité et son évolutivité lui permettent de surpasser largement les êtres humains dans certaines tâches. L'impénétrabilité éventuelle des algorithmes autoproduits signifie qu'il peut être impossible de connaître la méthode et le raisonnement utilisés pour produire un résultat particulier, même pour le développeur de l'intelligence artificielle. Certains estiment que « les systèmes d'intelligence artificielle et les systèmes autonomes exécuteront des tâches beaucoup plus complexes et qui auront plus d'impact que les générations

[2] Une définition a par exemple été établie tout spécialement pour pouvoir apprécier la nécessité de réglementer l'intelligence artificielle : « l'intelligence artificielle est la capacité d'une entité non naturelle à faire des choix selon un processus d'évaluation » (*Robot Rules: Regulating Artificial Intelligence*, Jacob Turner, Palgrave Macmillan, 2018). On peut également dire que l'intelligence artificielle existe « dès lors qu'un comportement ne provient pas uniquement du programmeur, mais d'autres moyens, par exemple des bases de connaissances » (Arvind Narayanan, Université de Princeton). Malgré leurs différences, aucune de ces définitions n'est erronée.

[3] L'IA a également été définie comme une forme d'intelligence très différente de l'intelligence humaine : « l'intelligence artificielle est la continuation de l'intelligence par d'autres moyens [...]. C'est grâce à ce découplage que l'intelligence artificielle a la capacité d'entreprendre des tâches chaque fois qu'elle peut les réaliser sans faire preuve de compréhension, de conscience, de sensibilité, d'intuition, d'expérience ou même de sagesse. En somme, c'est précisément lorsqu'on cesse de chercher à reproduire l'intelligence humaine qu'on parvient véritablement à la remplacer ». (« *A Fallacy that Will Hinder Advances in Artificial Intelligence* », Professor Luciano Floridi, Financial Times, 1er juin 2017). Cette description met en lumière une caractéristique de l'intelligence artificielle particulièrement pertinente pour son éventuelle application dans le système de justice pénale.

[4] « Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems », Groupe européen d'éthique des sciences et des nouvelles technologies, Commission européenne, mars 2018.

[5] Groupe européen d'éthique des sciences et des nouvelles technologies, op. cit.

antérieures des technologies, en particulier avec les systèmes qui interagissent avec le monde matériel, ce qui accroîtra l'importance du préjudice que pourrait causer un tel système »[6]. Cet argument peut être poussé plus loin encore : « comme l'intelligence artificielle de pointe pourrait entraîner un profond bouleversement de l'histoire de la vie sur terre, il importe de la planifier et de la gérer avec toute l'attention et toutes les ressources qui conviennent »[7].

### *1.3. Considérations générales sur la réglementation de l'intelligence artificielle*

10. La question de savoir si, quand et comment il convient de réglementer l'intelligence artificielle a fait couler beaucoup d'encre. Certains commentateurs jugent cette réglementation peu souhaitable, car elle étoufferait l'innovation ou inciterait les entreprises d'intelligence artificielle à rechercher les normes éthiques les plus avantageuses pour elles, prématurée, puisque la technologie continue à évoluer, et même impossible, du fait de la nature intrinsèque de l'intelligence artificielle. Lorsqu'on examine ces questions, il convient tout d'abord de réfléchir aux expériences antérieures, en particulier à l'histoire de la réglementation d'internet.

11. En 1996, alors que l'utilisation d'Internet commençait à se répandre et que les géants d'internet d'aujourd'hui en étaient encore à leurs balbutiements ou n'avaient pas même été créés, John Perry Barlow a présenté une « Déclaration sur l'indépendance du cyberespace » au Forum économique mondial de Davos, en Suisse. « Gouvernements du monde industriel, géants fatigués de chair et d'acier, je viens du cyberespace, le nouveau foyer de l'esprit. Au nom de l'avenir, je demande au passé de nous laisser tranquilles. Vous n'êtes pas les bienvenus parmi nous. Votre souveraineté ne s'exerce pas sur l'espace où nous nous réunissons. [...] Je déclare que l'espace social mondial que nous construisons est naturellement indépendant des tyrannies que vous cherchez à nous imposer. Vous n'avez aucun droit moral à nous gouverner et vous ne possédez aucun moyen de coercition que nous avons de bonnes raisons de craindre ».

12. Cette Déclaration était au départ destinée aux individus, mais son esprit cosmopolite et libertaire est devenu depuis caractéristique de nombreuses entreprises d'internet. Jusqu'à ces derniers temps, le comportement de ces entreprises n'était pas véritablement remis en cause par les autorités nationales, qui n'avaient pas pris conscience de ses conséquences ou refusaient d'en prendre conscience. Comme l'a indiqué le professeur Paul Nemitz, « cette incapacité, aussi bien du législateur que des entreprises technologiques, à attribuer et à assumer des responsabilités à l'ère d'internet [...] a fait d'internet un fiasco à divers égards, puisqu'elle a permis la propagation d'une surveillance massive, du recrutement du terrorisme, de l'incitation à la haine raciale et religieuse et de multiples autres catastrophes pour la démocratie, dont dernièrement le scandale de Cambridge Analytica et la montée des populistes, qui ont souvent le plus tiré parti de l'aide que leur offraient Facebook, Youtube, Twitter et consorts, en combinant les techniques publicitaires et l'utilisation des réseaux pour élaborer une publicité ciblée à des fins de propagande politique »[8].

13. En dehors du fait que nous disposons aujourd'hui de cette expérience dont nous pouvons tirer des enseignements, la question de la réglementation de l'intelligence artificielle se distingue également par l'état d'avancement de cette technologie. Citons, une fois encore, le professeur Nemitz : « l'intelligence artificielle, contrairement à internet, n'est d'emblée pas une innovation naissante présentée principalement par des universitaires et des idéalistes, mais une technologie largement conçue et déployée sous le contrôle des plus puissantes entreprises de technologie d'internet ». En d'autres termes, nous en savons d'ores et déjà en pratique suffisamment sur l'intelligence artificielle pour réglementer son application et nous connaissons suffisamment les entreprises qui l'utilisent pour nous demander s'il peut être indispensable de privilégier une réglementation obligatoire et assortie de sanctions à une autorégulation volontaire fondée sur l'éthique.

### *1.4. Réglementation et confiance des citoyens*

14. Si l'on souhaite que les citoyens acceptent l'utilisation de l'intelligence artificielle et jouissent des avantages qu'elle pourrait présenter, ils doivent avoir confiance dans le fait que tout risque est géré de manière satisfaisante. Deux chercheurs de premier plan dans ce domaine ont fait remarquer que « nous savons qu'il n'existe aucune formule qui permette de gagner la confiance des citoyens, mais nous savons aussi par expérience qu'ils font en général confiance à la technologie dès lors que celle-ci présente des avantages, est sûre et bien réglementée »[9]. À moins que l'intelligence artificielle ne soit imposée au grand public contre sa volonté ou par la force, autrement dit si elle doit être mise en place avec le consentement du public, alors une réglementation efficace et proportionnée en devient une condition nécessaire, mais pas suffisante.

[6] « General Principles », IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems.
[7] « Asilomar AI Principles », Conférence d'Asilomar 2017.
[8] « Constitutional democracy and technology in the age of artificial intelligence », Phil. Trans. R. Soc. A 378:20180089.
[9] « Ethical governance is essential to building trust in robotics and artificial intelligence systems », Alan F.T. Winfield et Marina Jirotka, Phil. Trans. R. Soc A 376:20180085.

*1.5. Cohérence et harmonisation réglementaires*

15. Si une réglementation doit être mise en place, elle doit être cohérente et harmonisée à une échelle aussi étendue que possible. Le Groupe européen d'éthique des sciences et des nouvelles technologies, par exemple, a attiré l'attention sur « les risques inhérents à toute approche dépourvue de coordination et d'équilibre de la réglementation de l'intelligence artificielle et des technologies "autonomes". Un patchwork réglementaire pourrait conduire les entreprises à rechercher les normes éthiques les plus avantageuses pour elles et à déplacer le développement et l'utilisation de l'intelligence artificielle dans des régions soumises à des normes éthiques moins exigeantes. Le fait de laisser certaines régions, règles, données démographiques ou acteurs de l'industrie dominer le débat risque d'en exclure un ensemble plus large d'intérêts et de perspectives sociétales »[10].

*1.6. Gouvernance éthique*

16. Ceux qui jugent prématurée la mise en place d'une réglementation de l'intelligence artificielle admettent pourtant qu'un minimum de gouvernance éthique est indispensable. Cette dernière a été définie comme « un ensemble de processus, de procédures, de cultures et de valeurs conçu pour garantir les normes de comportement les plus exigeantes. La gouvernance éthique va donc au-delà de la simple bonne gouvernance, c'est-à-dire de la gouvernance efficace, en ce qu'elle inculque des comportements éthiques à la fois à chaque concepteur et aux organisations dans lesquelles ils travaillent »[11].

*1.7. Principes éthiques*

17. La définition de principes éthiques est indispensable, qu'il s'agisse d'en faire le simple fondement de codes de conduite éthique ou d'établir une réglementation qui en soit imprégnée. Les études montrent que la teneur essentielle des principes éthiques qui devraient être appliqués aux systèmes d'intelligence artificielle fait néanmoins l'objet d'un large consensus ; c'est notamment le cas des principes suivants[12] :

- *Transparence*. Le principe de transparence peut faire l'objet d'une interprétation élargie, de manière à englober l'accessibilité et l'explicabilité d'un système d'intelligence artificielle, en d'autres termes la possibilité donnée à un individu de comprendre le fonctionnement de ce système et le mode de production de ses résultats.
- *Justice et équité.* Ce principe comprend la non-discrimination, l'impartialité, la cohérence et le respect de la diversité et du pluralisme. Il suppose également que la personne à laquelle est appliqué un système d'intelligence artificielle puisse en contester les résultats, disposer d'une voie de recours et obtenir réparation.
- *Responsabilité*. Ce principe englobe le fait d'exiger qu'un être humain soit responsable de toute décision qui a des conséquences sur les droits et libertés individuels, et que l'obligation de rendre des comptes et la responsabilité juridique de ces décisions soient définies. Il est donc étroitement lié au principe de justice et d'équité.
- *Sûreté et sécurité.* Ce principe suppose que les systèmes d'intelligence artificielle fassent preuve de solidité, de sécurité contre toute ingérence extérieure et de sûreté contre la commission d'actes non intentionnels, conformément au principe de précaution.
- *Respect de la vie privée.* Si le respect des droits de l'homme en général peut être considéré comme inhérent aux principes de justice et d'équité, de sûreté et de sécurité, le droit au respect de la vie privée est particulièrement important chaque fois qu'un système d'intelligence artificielle procède au traitement de données à caractère personnel ou privé. Les systèmes d'intelligence artificielle doivent par conséquent respecter les normes contraignantes du règlement général sur la protection des données (RGPD) de l'UE et de la Convention 108 du Conseil de l'Europe sur la protection des données (et de sa version actualisée, la Convention 108+), le cas échéant.

[10] Groupe européen d'éthique des sciences et des nouvelles technologies, op. cit.
[11] Winfield et Jirotka, op. cit.
[12] Voir *Lignes directrices sur l'éthique en matière d'IA : situation en Europe et dans le monde*, Rapport provisoire commandé par le Comité ad hoc sur l'intelligence artificielle du Conseil de l'Europe (CAHAI), Ienca et Vayena, mars 2020. De même, Winfield et Jirotka (op. cit.) font remarquer qu'une « enquête informelle menée à la fin de 2017 a révélé qu'il avait été prévu de proposer au total 10 ensembles différents de principes éthiques [...] d'ici le mois de décembre 2017, dont sept ont vu le jour en 2017 [...]. Ces principes présentent de nombreux points communs ; ils préconisent notamment que les systèmes autonomes intelligents (i) ne soient pas préjudiciables, (ii) respectent les droits de l'homme et les libertés fondamentales, notamment la dignité et la vie privée, tout en favorisant le bien-être, et (iii) soient transparents et fiables, tout en garantissant que la responsabilité et l'obligation de rendre des comptes en la matière continuent à peser sur les individus qui les conçoivent ou les exploitent ». Ils expliquent également comment l'éthique inspire les normes sur lesquelles se fonde la réglementation.

18. En décembre 2018, la Commission européenne pour l'efficacité de la justice (CEPEJ) du Conseil de l'Europe a adopté une Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, qui montre comment les principaux éléments de la liste énumérée ci-dessus s'appliquent en la matière. La Charte éthique européenne de la CEPEJ énonce cinq principes : respect des droits fondamentaux ; non-discrimination ; qualité et sécurité ; transparence, neutralité et intégrité intellectuelle (notamment en rendant les méthodologies accessibles et compréhensibles – ce qui équivaut à les rendre explicables) ; et maîtrise par l'utilisateur.

*1.8. Préoccupations d'ordre éthique concernant l'utilisation de l'IA dans les systèmes de justice pénale*

19. Mme Oswald et d'autres personnes ont procédé à un examen détaillé du fonctionnement de HART (un système basé sur l'IA utilisé par la police de Durham – voir plus loin), en faisant un certain nombre d'observations et en tirant des conclusions de portée plus générale[13]. L'impact de l'IA doit être apprécié dans son contexte opérationnel, ce qui, dans le cas des services de police, comprend les opérations de routine, les objectifs et les processus décisionnels. L'action de la police, expression de la puissance publique, peut être soumise à un contrôle juridictionnel, mais l'IA ne conçoit pas un algorithme dans le but de le rendre compréhensible par l'homme. La police de Durham a délibérément choisi une variante de l'algorithme HART qui favorise les « erreurs prudentes », compromis entre (plus de) résultats positifs erronés qui prévoient un risque élevé et (moins de) résultats négatifs erronés (« erreurs dangereuses »). « La question de savoir si le bénéfice global pour la société d'un jugement de valeur particulier intégré dans un algorithme peut en justifier les conséquences négatives éventuelles pour les individus dépend dans une large mesure de la gravité de ces conséquences ». Par ailleurs, l'algorithme HART est alimenté par des données provenant uniquement de la police de Durham, et non d'autres administrations locales ou forces de police. Lorsque les décisions sont prises par des personnes, celles-ci ont accès à des sources d'information plus diversifiées : c'est pourquoi l'algorithme ne doit jamais avoir le dernier mot. Cela dit, Mme Oswald et ses collègues n'ont pas pu exclure l'éventualité que certains agents préfèrent « (consciemment ou non) renoncer à prendre la responsabilité d'une décision risquée et s'en décharger sur l'algorithme, ce qui entraîne une déresponsabilisation et une atrophie du jugement ». Ce type de situation pourrait aussi avoir des conséquences négatives à long terme. Le décideur humain peut s'adapter rapidement à l'évolution d'une situation, contrairement à un algorithme, qui aurait besoin de procéder à « un examen minutieux et constant des prédicteurs utilisés et à une mise à jour fréquente de l'algorithme grâce à des données plus récentes ».

20. Au cours de son audition par notre commission, M. Veale a abordé l'utilisation de l'IA par la police dans un contexte social et politique plus large. Il a fait remarquer que les coupes opérées dans les services d'aide sociale, par exemple, pouvaient engendrer des problèmes qui nécessitaient l'intervention ultérieure d'autres administrations, notamment la police, dont les ressources propres sont également très sollicitées. L'IA ayant été présentée comme un moyen de réduire les coûts, la politique menée en matière de police pourrait privilégier les approches fondées sur l'IA. Les développeurs de l'intelligence artificielle peuvent certes chercher à éviter les résultats discriminatoires ou autres résultats indésirables, mais ils n'ont pas à l'esprit le contexte social dans son ensemble. Ainsi, les solutions qu'ils proposent définissent les paramètres du problème en fonction des réponses qu'il est techniquement possible d'apporter. Cela équivaut à privatiser l'élaboration des politiques, ce qui pourrait exacerber les divisions sociales et renforcer les injustices. Le secteur public, y compris la police, a donc besoin d'avoir sa propre expertise interne pour évaluer l'utilité réelle des solutions fondées sur l'IA et doit être prêt à les refuser.

21. Jamie Grace, l'un des collègues de Mme Oswald au sein de la commission d'éthique de la police des West Midlands (voir plus loin), observe qu'au Royaume-Uni, « comme David Lyon l'avait prédit en 2007, "l'État sécuritaire" a écarté l'État-providence du discours public et de la politique publique et que, soucieux de la protection civile au-delà de toute considération d'autonomie individuelle, "cet État sécuritaire s'appuie largement sur les données de surveillance" ». M. Grace est d'avis que l'IA « exacerbera et attisera les tensions existantes en matière de droits de l'homme dans le cadre de la justice pénale. Ces questions inhérentes aux droits de l'homme comprennent les préoccupations relatives au respect de la vie privée, la limitation de la liberté d'expression, les problèmes liés à la possibilité de discrimination raciale et le droit des victimes d'actes criminels à être traitées avec dignité »[14].

22. Depuis la présentation de ma note introductive en avril 2019, le recours à l'intelligence artificielle dans les systèmes de police et de justice pénale se heurte à une hostilité croissante généralisée. J'aborderai plusieurs situations spécifiques un peu plus loin, mais à ce stade, j'aimerais m'arrêter sur les principaux points d'une déclaration publiée par plusieurs chercheurs et universitaires américains de haut niveau en juillet 2019 :

[13] « Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality », Oswald, Grace, Urwin et Barnes, 2018.
[14] « Machine learning technologies and their inherent human rights issues in criminal justice contexts », novembre 2019.

> « Les évaluations actuarielles des risques avant procès souffrent de graves défauts techniques qui compromettent leur exactitude, leur validité et leur efficacité. Elles ne mesurent pas avec précision les risques que les juges sont tenus par la loi de prendre en compte. Lorsqu'ils prédisent la fuite ou un danger, de nombreux outils utilisent des définitions inexactes et excessivement larges de ces risques. En ce qui concerne la prédiction de la violence, aucun outil ne permet à l'heure actuelle de distinguer correctement le risque de violence d'une personne par rapport à une autre. Des qualifications de risques trompeuses cachent l'incertitude de ces prédictions aux enjeux élevés et peuvent amener les juges à surestimer les risques et la prévalence de la violence avant un procès. Pour générer des prédictions, les évaluations des risques s'appuient sur des données profondément erronées, comme l'historique des arrestations, des inculpations, des condamnations et des peines prononcées. Ces données ne constituent pas une mesure fiable et neutre de l'activité criminelle sous-jacente. Des décennies de recherche ont montré que, pour un même comportement, les Afro-américains et les Latino-américains étaient plus susceptibles d'être arrêtés, poursuivis, reconnus coupables et condamnés à des peines plus lourdes que leurs homologues blancs. Les évaluations des risques qui intègrent ces données faussées produisent des résultats faussés. Ces problèmes ne peuvent être résolus par des solutions techniques. Nous recommandons fortement d'envisager d'autres réformes »[15].

## 2. Les applications concrètes des algorithmes et de l'intelligence artificielle dans les systèmes de justice pénale

23. Cette partie sera principalement consacrée à trois exemples d'utilisation de l'intelligence artificielle dans les systèmes de justice pénale : PredPol, qui prévoit à quel endroit les infractions peuvent se produire et calcule à partir de ces éléments comment affecter au mieux les ressources policières ; HART (Outil d'évaluation des risques de préjudice), qui prévoit le risque de récidive au moment où est prise la décision d'engager ou non des poursuites ; et, enfin, COMPAS (Profilage de la gestion des délinquants passibles d'une peine d'emprisonnement à des fins de peine de substitution), qui prévoit lui aussi principalement le caractère récidiviste ou non d'un délinquant. Il s'agit, dans les trois cas, de « boîtes noires » – des systèmes propriétaires dont le fonctionnement interne n'est pas accessible au public.

### *2.1. Prévoir les infractions et affecter les ressources policières : PredPol*

24. PredPol, société californienne née d'un projet établi entre UCLA et la Police de Los Angeles, définit « le maintien de l'ordre prédictif [comme] une pratique consistant à déterminer les moments et les endroits où des infractions précises ont le plus de probabilité de se produire, puis à patrouiller dans ces secteurs pour empêcher la commission de ces infractions ». PredPol utilise l'historique des données d'un service de police client sur une période de deux à cinq ans pour former un algorithme d'apprentissage machine, qui est ensuite mis à jour quotidiennement. Seuls trois éléments de données sont utilisés : le type d'infraction, le lieu et la date et l'heure de sa commission. Selon PredPol, « aucune information démographique, ethnique ou socio-économique n'est utilisée. Cette méthode supprime le risque d'atteinte à la vie privée ou aux droits civils constaté avec d'autres modèles de maintien de l'ordre prédictifs ou fondés sur le renseignement ». Cette affirmation est cependant contestée.

25. Cette technologie n'est pas bon marché. Au Royaume-Uni, la police du Kent a utilisé PredPol de décembre 2012 à mars 2018, pour un coût de 100 000 GBP par an. Elle a indiqué qu'au cours d'un essai initial de quatre mois, l'utilisation de PredPol avait entraîné une diminution de 6 % de la criminalité sur la voie publique[16]. Mais la police du Kent a dernièrement déclaré que « si PredPol donne de bons résultats en matière de prévision des endroits où les infractions sont susceptibles d'avoir lieu, il est plus difficile de démontrer que ces informations nous ont permis de faire baisser la délinquance ». La police du Kent a néanmoins été suffisamment impressionnée par les capacités de cette technologie pour développer son propre système[17].

26. Bien que PredPol affirme qu'il n'existe aucun risque de violation des droits de l'homme, d'aucuns reprochent à ce système de perpétuer les préjugés classiques de la pratique policière, tout en dissimulant ce parti pris derrière une façade ou une présomption de neutralité mécanique, au moyen d'un discours destiné à parer la technologie de toutes les vertus. Même si les renseignements personnels ou socio-économiques ne figurent pas dans les données de formation, ils peuvent néanmoins être inhérents aux données, notamment en ce qui concerne le lieu des infractions. Si les forces de police avaient par le passé centré leur attention de façon disproportionnée sur un quartier donné, les infractions commises dans ce même quartier auraient eu plus de chance d'être décelées. Cette orientation fausserait alors artificiellement l'historique des données

[15] « Technical Flaws of Pretrial Risk Assessments Raise Grave Concerns », juillet 2019.
[16] « Predictive Policing statistics from Kent Police », Kent Online, 13 février 2018.
[17] « Kent Police stop using crime predicting software », The Telegraph, 27 novembre 2018.

utilisées pour créer l'algorithme PredPol, qui prévoirait par conséquent une plus forte probabilité que des infractions y soient commises à l'avenir. La police affecterait davantage de ressources au maintien de l'ordre dans cette zone, ce qui conduirait à perpétuer (et, en fait, à renforcer) ce parti pris ancien, mais cette fois sur une base prétendument « objective ». Si le quartier en question était habité majoritairement par des personnes d'une certaine ethnie ou religion, ce qui peut avoir été la raison même de l'attention particulière que la police lui avait au départ portée (« profilage ethnique »), les résultats de l'algorithme pourraient être discriminatoires, et ce pour des motifs pourtant interdits par le droit international des droits de l'homme, notamment l'article 14 de la Convention européenne des droits de l'homme.

27. D'autres services de police du Royaume-Uni expérimentent actuellement une autre forme de police prédictive. Neuf services, dirigés par la police des West Midlands et qui comportent également la police métropolitaine de Londres et la police du Grand Manchester, élaborent en ce moment un Outil national d'analyse des données (NDAS – voir plus loin). Il s'agira, par exemple, d'évaluer le risque qu'un individu commette ou soit victime d'une infraction commise à l'aide d'une arme à feu ou d'un couteau, ainsi que la probabilité qu'une personne soit victime de l'esclavage moderne, en combinant, d'une part, l'intelligence artificielle utilisée pour l'apprentissage de la machine et, d'autre part, les statistiques. Ce système repose sur des données relatives à environ cinq millions d'individus, à partir desquelles il définit près de 1400 indicateurs permettant de prévoir les infractions, dont 30 indicateurs particulièrement importants. La police des West Midlands collaborera avec l'autorité britannique chargée de la protection des données pour veiller à ce que le système soit conforme à la législation en matière de protection de la vie privée. L'agent responsable du projet a reconnu qu'il s'agissait en partie de répondre aux importantes réductions budgétaires subies par la police ces dernières années et à l'impératif conséquent d'accorder en priorité une attention aux personnes qui ont le plus besoin d'interventions urgentes[18].

28. Je vais maintenant examiner plus en détail le cadre réglementaire éthique qui a été élaboré autour des projets de la police des West Midlands relatifs à l'IA.

### *2.2. Prévoir la récidive et prévenir la récidive : HART*

29. Le système HART a été mis au point par la police de Durham, en collaboration avec des chercheurs de l'Université de Cambridge, à partir des données de formation de 104 000 personnes arrêtées sur une période de cinq ans. Il utilise des « valeurs prédictives », dont la plupart prennent en compte les antécédents criminels du suspect, ainsi que son âge, son sexe et sa zone géographique, pour le catégoriser comme un délinquant présentant un risque faible, moyen ou élevé de commettre de nouvelles infractions graves dans les deux ans à venir. Les personnes qui présentent un risque moyen, c'est-à-dire qui sont « susceptibles de commettre une infraction non grave », peuvent alors participer au programme « Checkpoint » de la police, qui « offre aux délinquants éligibles un contrat de 4 mois au cours duquel ils prendront part à une solution de substitution aux poursuites. Le contrat propose des interventions qui visent à remédier aux causes profondes qui les ont amenés à commettre une infraction, afin de prévenir leur récidive ». HART a été conçu pour réduire le nombre de personnes incarcérées alors qu'elles pourraient faire l'objet d'autres formes d'intervention qui seraient tout aussi efficaces, voire davantage, pour réduire le risque qu'elles récidivent.

30. Le policier à la tête du projet HART a publié, en compagnie d'universitaires et d'un policier australien, un article détaillé qui analyse le système HART « en prenant comme point de départ les principes de nécessité, de proportionnalité et de prévisibilité énoncés par le droit européen des droits de l'homme »[19]. Conscients que l'intelligence artificielle représente « une autre forme de prise de décision et non un cerveau humain amélioré », les auteurs posent une série de questions auxquelles il est difficile de répondre. « Le contrôle juridictionnel et les principes des droits de l'homme résisteront-ils à l'épreuve du temps ? Quel degré d'opacité sommes-nous prêts à accepter ? Quelle part d'erreur ? Quel degré d'incertitude à l'égard des avantages futurs de ce système ? Ce sont là de grandes questions, dont la pertinence vaut pour toute application de l'intelligence artificielle dans le domaine de la justice pénale.

31. Les auteurs de l'article reconnaissent eux-mêmes que ce système présente un risque important, déjà mentionné plus haut à propos de PredPol, qui justifie un regard très critique. « Certains facteurs de prévision utilisés par le modèle [...] (comme le code postal[20]) pourraient être considérés comme liés indirectement à des mesures de privation collective. […] [On] pourrait soutenir que [l'utilisation du code postal sous forme de] variable risque de créer une sorte de cercle vicieux qui pourrait perpétuer ou amplifier la schématisation actuelle de la criminalité. Si, à la suite de ces prévisions, la police réagit en accentuant son action sur les

[18] « Exclusive: UK police wants AI to stop violent crime before it happens », New Scientist, 26 novembre 2018.
[19] « Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality », Oswald, Grace, Urwin et Barnes, 3 avril 2018.
[20] C'est-à-dire un indicateur codifié d'un quartier donné, utilisé pour faciliter la distribution de courrier.

zones de codes postaux à plus haut risque, les habitants de ces zones attireront davantage l'attention de la police et seront arrêtées en plus grand nombre que les habitants des quartiers à faible risque, qui n'auront pas été la cible d'une surveillance policière. Les arrestations effectuées constituent alors des résultats dont l'utilisation entraîne une répétition ultérieure du même modèle, ce qui renforce de plus en plus l'attention portée par les forces de police à ces zones ». Le chef de la police de Durham, Michael Barton, a déclaré qu'il y aurait lieu de s'inquiéter si le système HART était utilisé par les tribunaux, mais qu'il servait uniquement à atténuer le risque de récidive et non à déterminer la peine à infliger[21]. Cette analyse omet cependant le fait qu'une personne qui serait identifiée à tort comme présentant un risque élevé, et ce en raison du caractère biaisé des données de formation, se verrait refuser l'accès au programme « Checkpoint » et pourrait au bout du compte être placé en détention provisoire.

32. En 2017, HART a été « rafraîchi » par l'apport de nouvelles données, afin de réduire le recours au code postal comme facteur de prévision. Cet ensemble de données, baptisé « Mosaic », a été développé par la société Experian essentiellement comme un produit commercial utilisé à des fins de marketing. Mosaic repose sur les profils des 50 millions d'adultes résidant au Royaume-Uni, constitués à partir de données recueillies auprès de sources publiques, notamment internet. Mosaic définit de manière explicite des catégories de personnes en fonction, par exemple, de leur groupe d'âge ou de leur appartenance ethnique (« jeunes déconnectés », « gris dépendants » ou « héritage asiatique »). L'ONG Big Brother Watch a fait valoir qu'il était « effrayant qu'une société de vérification des capacités de crédit recueille des millions d'informations sur les citoyens et vende des profils au plus offrant. Mais il est véritablement dystopique que la police alimente ces profils grossiers et offensants au moyen de l'intelligence artificielle pour prendre des décisions qui concernent la liberté et la justice au Royaume-Uni »[22]. La police de Durham a indiqué qu'elle avait « collaboré avec Experian pour mieux comprendre les communautés locales »[23] ; en 2018, elle a cessé d'utiliser Mosaic, mais cette décision aurait plutôt été motivée par des considérations financières que par l'éthique[24].

33. Bien que la police de Durham ait souligné que le système HART était uniquement utilisé à des fins de conseil et que la prise de chaque décision incombait à des fonctionnaires de police formés, l'application concrète de cette déclaration de principe suscite un certain scepticisme. Comme cela a été le cas avec l'utilisation de PredPol par la police du Kent, le chef de la police de Durham a indiqué que le recours aux nouvelles technologies avait été motivé par les réductions budgétaires successives subies par ses services[25]. Ces mêmes restrictions budgétaires peuvent avoir des conséquences sur le temps et l'attention dont disposent les fonctionnaires de police, qui sont autant de facteurs déterminants pour garantir que les décisions prises à l'aide du système HART seront effectivement prises par des êtres humains responsables. Andrew Wooff, de l'Université d'Édimbourg, a fait remarquer que dans l'univers du maintien de l'ordre, « pressé par le temps et coûteux en ressources », « on peut imaginer qu'un fonctionnaire de police puisse préférer s'en remettre au système que prendre lui-même une décision »[26]. Big Brother Watch a par ailleurs ajouté que « comme cet algorithme a été conçu pour détecter les cas qui pourraient échapper à l'attention de la police ou dont elle hésite à penser qu'ils présentent un risque élevé, peut-on vraiment s'attendre à ce que les policiers portent un jugement systématiquement contraire aux résultats donnés par l'intelligence artificielle ? Pour que cet algorithme fonctionne comme un outil d'aide à la prise de décision, il faut qu'il alerte la police sur les délinquants potentiels qu'elle n'a peut-être pas pris en compte. On peut par conséquent se demander si la police ignorerait tranquillement les propositions faites par cet algorithme »[27].

34. Big Brother Watch a également attiré l'attention sur les questions de l'explicabilité de la méthodologie et de l'obligation de rendre des comptes, en faisant remarquer que « les décisions prises par l'intelligence artificielle ne peuvent être contestées, puisqu'il n'est parfois pas même possible d'expliquer les conclusions auxquelles elle est parvenue ». La police de Durham a rétorqué à ce sujet qu'elle « était prête à révéler à une autorité de régulation compétente l'algorithme HART, ainsi que les données à caractère personnel et les données relatives à la détention d'une personne associées à cet algorithme »[28]. Le point de vue de la police, qui n'est pas déraisonnable, représente par ailleurs un solide argument en faveur de la création de telles autorités.

[21] « UK police test if computer can predict criminal behaviour », Financial Times, 6 février 2019.
[22] « Police use Experian Marketing Data for AI Custody Decisions », Big Brother Watch, 6 avril 2018.
[23] « Durham police criticised over 'crude' profiling », BBC News, 9 avril 2018.
[24] Op. cit., Financial Times, 6 février 2019.
[25] Op. cit., Financial Times, 6 février 2019.
[26] Citation tirée de « UK police are using AI to inform custodial decisions – but it could be discriminating against the poor », Wired, 1er mars 2018.
[27] « A Closer Look at Experian Data and Artificial Intelligence in Durham Police », Big Brother Watch, 6 avril 2018.
[28] Wired, op. cit.

*2.3. Prévoir la récidive et statuer sur la détention provisoire, le choix de la peine et la libération conditionnelle : COMPAS*

35. Le système COMPAS, désormais propriété de la société Equivant, est utilisé par plusieurs États aux États-Unis pour évaluer le risque de récidive d'un individu. Il a été décrit par Northpointe, filiale d'Equivant, comme un « instrument d'évaluation des risques et des besoins de quatrième génération ». Il s'agit d'un outil en ligne conçu pour évaluer les besoins criminogènes et le risque de récidive des délinquants, en utilisant trois « échelles » : le « risque de mise en liberté avant le procès » (c'est-à-dire le risque de non-comparution et de nouvelle arrestation pour infraction) ; la « récidive générale » (commission d'un nouveau délit ou d'un nouveau crime dans les deux ans) ; et la « récidive avec violence » (commission d'infractions assorties de violences)[29]. L'État de New York, par exemple, a commencé à utiliser COMPAS pour évaluer les personnes qui purgent une peine probatoire ; il y est à présent également utilisé par les juges pour le choix de la peine infligée. En Floride, il sert à décider si un accusé doit être placé en détention provisoire ou en liberté sous caution. Dans le Wisconsin, il est utilisé à chaque étape du système carcéral, de la détermination de la peine à la libération conditionnelle.

36. En 2014, le procureur général des États-Unis de l'époque, Eric Holder, a mis en garde contre le fait que, « bien que les systèmes d'évaluation des risques du type COMPAS aient été établis avec les meilleures intentions du monde, je crains qu'ils ne sapent sans le vouloir notre ambition de garantir une justice personnalisée et égale. [...] Ils risquent d'aggraver des disparités infondées et injustes, qui sont déjà bien trop fréquentes dans notre système de justice pénale et dans notre société ». Au vu de ces préoccupations, le site internet d'enquête ProPublica a analysé l'utilisation de COMPAS et est parvenu à deux conclusions particulièrement critiques à son égard. S'agissant de son efficacité, ProPublica constate que, « lorsqu'on prend en compte tout un éventail d'infractions, [...] l'algorithme est un peu plus précis qu'à pile ou face », et qu'il est « remarquablement peu fiable pour prévoir les infractions violentes ». Pour ce qui est de sa neutralité, ProPublica observe que, si COMPAS commet à peu près le même taux d'erreur pour les individus blancs ou noirs, il est bien plus susceptible de produire des résultats positifs erronés (c'est-à-dire de prévoir à tort un « risque élevé ») pour les populations noires et a plus de probabilité de produire des résultats négatifs erronés (c'est-à-dire de prévoir à tort un « risque faible ») pour les populations blanches[30].

37. Il convient de noter que les conclusions de ProPublica ont été critiquées à la fois pour des raisons techniques liées à la validité de l'analyse statistique et pour la manière de présenter ces conclusions[31]. Pourtant, l'auteur de l'une de ces réfutations critiques a lui-même indiqué que « ce n'est pas tant l'outil qui semble biaisé que le système »[32]. Cela nous ramène à l'idée de discours destiné à parer la technologie de toutes les vertus, alors que les conséquences discriminatoires préjudiciables de cette technologie pour les individus restent les mêmes, quelle que soit la manière de présenter le problème.

38. La controverse suscitée par COMPAS a également amené les commentateurs à souligner la nécessité d'un débat public, de la transparence et de l'obligation de rendre des comptes. « Les sociétés démocratiques devraient s'employer dès maintenant à déterminer le degré de transparence qu'elles attendent des systèmes de prise de décision automatisée. Avons-nous besoin d'une nouvelle réglementation du logiciel pour nous assurer qu'il peut faire l'objet d'une inspection satisfaisante ? Il importe que le législateur, les juges et le public aient leur mot à dire sur le choix des mesures d'équité auxquelles les algorithmes donnent la priorité. Mais si les algorithmes ne reflètent pas réellement ces jugements de valeur, qui en sera tenu responsable ? »[33]. D'aucuns soutiennent que cette transparence devrait englober l'accès à l'algorithme : « les questions épineuses soulevées par des logiciels comme [COMPAS] sont autant de raisons impérieuses d'en rendre les formules publiques ou du moins de les soumettre à un examen rigoureux et indépendant »[34].

*2.4. Identifier des anciennes affaires susceptibles d'être résolues – le « Calendrier des affaires non élucidées »*

39. Aux Pays-Bas, la police a mis au point un système d'apprentissage automatique basé sur l'IA pour l'aider à identifier des affaires anciennes, graves et non résolues (*cold cases*) qui pourraient désormais avoir de bonnes chances d'être élucidées. Ce système repose sur l'idée que plus de la moitié des affaires non

[29] « Practitioners Guide to COMPAS », Northpointe, 17 août 2012.
[30] « Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks », ProPublica, 23 mai 2016.
[31] « False Positives, False Negatives, and False Analyses: A Rejoinder to 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks' », Flores, Lowenkamp et Bechtel.
[32] Anthony Flores, cité dans « The machines that could rid courtrooms of racism », Washington Post, 18 août 2016.
[33] « Inspecting Algorithms for Bias », Matthias Spielkamp, MIT Technology Review, 12 juin 2017.
[34] Op. cit., Washington Post.

élucidées qui sont rouvertes sont résolues grâce à une nouvelle technologie qui n'existait pas au moment de la première enquête. (Dans près de la moitié des cas, l'affaire est résolue grâce à de nouveaux témoignages ; le terme « Calendrier des affaires non élucidées » désignait à l'origine une méthode consistant à interroger des détenus sur des affaires non résolues.) Une fois que les dossiers de ces affaires non résolues sont numérisés, ils sont introduits dans le système d'IA, qui identifie ceux qui comportent des éléments de preuve prometteurs susceptibles d'être réexaminés à la lumière de nouvelles techniques de criminalistique. Ce même travail effectué manuellement par des agents de police risquerait de prendre des semaines pour chaque affaire, avec d'infimes chances de succès. Les agents responsables de ce projet espèrent qu'il pourra être étendu pour identifier des affaires non élucidées qui pourraient être résolues en utilisant des données non-criminalistiques, telles que les sciences sociales, les réseaux sociaux et les témoignages. Ce système pourrait même permettre d'améliorer la capacité de la police à résoudre des enquêtes en cours[35].

40. L'un des problèmes posés par cette approche est que la durée légale de conservation des données de police qui n'ont pas été utilisées lors de l'enquête initiale est trop courte pour les enquêtes sur des affaires non résolues. Dès lors, la destruction de ces données peut entraîner la suppression d'éléments susceptibles de permettre la résolution d'une affaire grâce à de nouvelles techniques de criminalistique. Le chef de la police a donc décidé de ne pas supprimer ces anciennes données et de restreindre leur droit d'accès à un nombre limité de « gardiens ». Le gouvernement a accepté cette décision, considérant « qu'il était préférable d'accepter ce défaut dans le respect de la loi et de se contenter des mesures prises par le chef de la police pour limiter l'accès aux données au strict minimum. Le respect de la lettre de la loi ne pourrait être assuré qu'au moyen d'une méthode de sélection grossière, qui détruirait également des données pouvant permettre de déceler des éléments dans des affaires non élucidées. Cela compromettrait gravement la résolution de ces affaires. Avec cette décision, nous pouvons désormais éviter cela[36] ». Toutefois, il n'est pas certain que cette approche soit conforme au principe général de la législation sur la protection des données, selon lequel les données à caractère personnel ne doivent être traitées (et conservées) qu'en vertu d'un fondement légal ou sur la base du consentement de la personne concernée[37].

### *2.5. Autres exemples*

41. Tous les systèmes algorithmiques destinés à faciliter la prise de décision ne font pas appel à l'intelligence artificielle. Un projet de recherche américain a constaté que les tribunaux de 46 des 52 États américains, ainsi que du District de Columbia, utilisaient une forme d'outil d'évaluation des risques au cours du processus décisionnel de la détention provisoire[38]. Dans la plupart des cas, il s'agit d'un outil algorithmique sans apprentissage[39]. Les outils les plus courants sont l'évaluation de la sécurité publique (Public Safety Assessment – PSA), utilisée dans au moins cinq États et 59 comtés (les divisions administratives des États) couvrant 56,3 millions de personnes ; l'instrument d'évaluation des risques avant procès de Virginie (Virginia Pretrial Risk Assessment Instrument – VPRAI), utilisé dans au moins 43 comtés couvrant 19,9 millions de personnes, et sa version révisée (VRPAI-R, censée corriger les préjugés fondés sur la race et le genre), utilisée dans au moins un État et 16 comtés couvrant 14,3 millions de personnes ; l'outil d'évaluation des risques avant procès de l'Ohio (Ohio Risk Assessment System Pretrial Assessment Tool – ORAS-PAT), utilisé dans au moins cinq États et 48 comtés ; et COMPAS (voir plus haut), utilisé dans au moins 11 comtés couvrant 4,3 millions de personnes.

42. Le Pretrial Justice Institute (PJI), une organisation américaine à but non lucratif, a, par le passé, encouragé l'utilisation d'outils algorithmiques d'évaluation des risques pour réduire le recours à la libération sous caution en espèces (c'est-à-dire la mise en liberté conditionnelle contre le versement d'une somme d'argent), qui se traduisait souvent par le maintien en détention provisoire des prévenus les moins fortunés. Cependant, en juillet 2020, le PJI a revu sa position après avoir constaté que, si les taux de détention provisoire diminuaient, le profil ethnique des détenus restait aux alentours de 50 % de prévenus noirs et de 30 % de prévenus blancs. C'est ce qu'il a expliqué lors de l'annonce de sa nouvelle politique :

> « Nous constatons aujourd'hui que les outils d'évaluation des risques avant procès, conçus pour prédire la comparution d'un individu devant un tribunal sans nouvelle arrestation, ne peuvent plus faire partie de notre solution visant à construire des systèmes de justice équitables avant le procès. Quels que soient leur science, leur marque ou leur âge, ces outils s'appuient sur des données reflétant le racisme

[35] « How the Dutch police are using AI to unravel cold cases », The Next Web, 23 mai 2018.
[36] « Volunteers and artificial intelligence used in cold cases », Gouvernement des Pays-Bas, 4 février 2019.
[37] Voir, par ex., l'article 5 de la Convention 108+ du Conseil de l'Europe sur la protection des données.
[38] Mapping Pretrial Justice, a collaboration between the Movement Alliance Project and MediaJustice – voir pretrialrisk.com.
[39] « Artificial Intelligence in Adjudication and Administration. A Status Report on Government Use of Technology in the United States », Coglianese et Ben Dor, 2019.

structurel et les inégalités institutionnelles qui affectent nos politiques et nos pratiques en matière de justice et de maintien de l'ordre. L'utilisation de ces données ne fait qu'aggraver les inégalités »[40].

43. Les critiques sont également venues de l'administration publique. Un rapport destiné au tribunal général de l'État du Massachusetts a conclu que « bien qu'ils possèdent d'une bonne intention, les outils d'évaluation des risques peuvent avoir leurs propres limites, en raison de leur dépendance à l'égard de données dont la corrélation avec la prévisibilité est contestable et de la rigidité de l'application qui contraint la décision du juge. […] Les outils d'évaluation des risques s'appuient sur un historique de données pour calculer et déterminer un résultat probable lorsqu'ils sont appliqués à un justiciable donné. La qualité des prédictions dépend largement de la qualité des données insérées dans l'outil. De nombreux outils d'évaluation des risques présentent un défaut majeur : l'algorithme s'appuie sur les données d'arrestation, point de départ de l'analyse de prédiction d'un événement futur ; or ces données peuvent comporter un préjugé implicite. […] En outre, de nombreux outils d'évaluation des risques disponibles sur le marché ne révèlent pas leurs algorithmes ou leurs méthodologies, ce qui renforce la méfiance à l'égard du système, rend difficile la contestation de leurs résultats par les avocats de la défense et empêche un tribunal de traiter rapidement et précisément les résultats incohérents. Le fait d'utiliser les données d'arrestation comme facteur de prédiction a donné lieu à des résultats biaisés, ce qui va à l'encontre de l'objectif premier de l'utilisation de cet outil. […] Les inconvénients de la mise en œuvre d'un outil d'évaluation des risques actuellement disponible seraient probablement plus importants qu'une amélioration progressive des décisions de mise en liberté conditionnelle »[41].

44. Les juges de la Cour suprême de l'Ohio ont décidé de ne pas recommander l'utilisation d'outils d'évaluation des risques avant procès dans le cadre d'une réforme du système de mise en liberté sous caution menée par l'État, malgré les conclusions antérieures d'un groupe de travail spécial. Il semble que les juges aient été particulièrement sensibles aux arguments de l'American Civil Liberties Union, qui a surtout souligné le risque de préjugés raciaux[42].

## 3. Contrôle et réglementation

### *3.1. La commission d'éthique de la police des West Midlands au Royaume-Uni*

45. Cette partie portera sur la situation de la police des West Midlands, qui est particulièrement intéressante et pour laquelle nous disposons d'un grand nombre d'informations (notamment grâce à mes propres contacts avec les personnes concernées). La police des West Midlands réalise un projet sur les informations basées sur des données (« Data-Driven Insights »), destiné à améliorer l'utilisation des grandes quantités de données déjà intégrées dans ses différents ensembles de données, qui constituent des « silos d'informations » distincts, dont les contenus sont difficiles à combiner ou à recouper. Les mêmes outils de traitement des données étant facilement transposables aux 43 autres services de police du Royaume-Uni, ce projet a été étendu pour devenir la Solution nationale d'analyse de données (NDAS). La NDAS est mise en place en partenariat avec huit autres institutions ou services de police et avec le soutien technique d'une entreprise privée, Accenture. Le projet comprend trois éléments principaux : Insight Search, un moteur de recherche couvrant neuf bases de données, consultable par les policiers sur des appareils mobiles ; Business Insights, qui concerne les questions internes (comme le bien-être des policiers) ; et Insights Lab, qui développe des outils innovants d'analyse de données spécifiquement adaptés aux services de police. Le projet Data-Driven Insights était doté d'un budget de 17 millions GPB, sachant que la valeur des bénéfices générés par Insight Search a été estimée à 23 millions GPB rien que sur les trois premières années d'exploitation. Jusqu'à présent, ses systèmes basés sur l'IA ont été utilisés principalement pour identifier des individus susceptibles de commettre des infractions à l'aide d'une arme blanche ou d'une arme à feu (« forme de violence la plus grave ») et des personnes susceptibles d'être victimes de la traite des êtres humains (« esclavage moderne »).

46. Le Commissaire aux questions de police et de criminalité des West Midlands (PCC) est un fonctionnaire élu indépendant, chargé d'assurer l'efficacité et l'efficience de la police des West Midlands, ainsi que de faire respecter l'obligation de rendre des comptes du chef de la police. Le PCC des West Midlands a mis en place une commission d'éthique pour conseiller le PCC et le chef de la police sur les projets de données scientifiques proposés par le laboratoire d'analyse des données de la police des West Midlands. L'objectif déclaré de cette commission est de « garantir que l'éthique et les droits individuels figurent au cœur du travail du laboratoire ». Son mandat l'oblige à prendre en compte un vaste éventail de principes éthiques ; malgré les réticences initiales de la police des West Midlands, il fait expressément référence aux droits de l'homme, de manière

[40] « Updated Position on Pretrial Risk Assessment Tools », 2 juillet 2020.
[41] « Final Report of the Special Commission to Evaluate Policies and Procedures Related to the Current Bail System », Tribunal général du Massachusetts, 31 décembre 2019.
[42] « Ohio Supreme Court proposes bail reforms that don't include risk assessments », cleveland.com, 24 janvier 2020.

générale et plus spécialement à propos de la non-discrimination et du respect de la vie privée[43]. La commission d'éthique respecte résolument le principe de diversité : la parité hommes-femmes est assurée et le recrutement est confié à une agence spécialisée dans la diversité ethnique, qui porte les postes vacants à l'attention des « minorités ethniques, des parents isolés, des personnes handicapées, des groupes religieux, des personnes de toutes orientations sexuelles et identités de genre ». La plupart des membres ont une certaine forme de formation technique spécialisée[44], mais certains d'entre eux ne sont pas des spécialistes. La commission comprend un haut représentant d'un autre service de police, ce qui, d'après M. McNeil, permet « d'ancrer la commission dans la réalité ». Ses activités sont extrêmement transparentes et passent par des actions de sensibilisation et de consultation de la collectivité, malgré un budget limité. Il a été décidé que, bien que la NDAS soit une collaboration entre plusieurs forces de police, la commission d'éthique du PCC des West Midlands donnerait également son avis sur les propositions de la NDAS, puisqu'il n'existe aucune autre structure nationale en matière d'éthique. La commission d'éthique peut toutefois présenter une faiblesse structurelle : comme son existence dépend du PCC des West Midlands, rien ne garantit qu'un prochain PCC la maintienne en place.

47. Lorsqu'il fait des propositions à la commission d'éthique, le laboratoire d'analyse des données utilise la matrice « ALGO-CARE ». Celle-ci a été développée par Mme Oswald et d'autres personnes après examen du programme HART de la police de Durham (voir ci-dessus)[45] et a été acceptée par le Conseil national des chefs de la police, qui la considère comme un modèle de bonne pratique en matière d'autorégulation. Elle vise à inciter les développeurs d'outils algorithmiques à l'usage de la police à prendre en compte une série de préoccupations éthiques et pratiques essentielles. L'acronyme ALGO-CARE est utilisé pour regrouper un ensemble de questions relatives à ces préoccupations, notamment sur les plans :

- Consultatif (*Advisory* – les résultats algorithmiques sont-ils utilisés à titre consultatif, la décision étant prise par l'homme ?)
- Juridique (*Lawful* – l'utilisation de l'algorithme poursuit-elle un objectif légal, est-elle proportionnée et les données utilisées ont-elles été obtenues et traitées dans le respect de la loi ?)
- de la Granularité (les propositions des algorithmes sont-elles suffisamment détaillées, les données saisies sont-elles fiables et leurs biais éventuels ont-ils été corrigés ?)
- de la propriété (*Ownership* – qui est propriétaire de l'algorithme et des données, la police est-elle titulaire de tous les droits nécessaires pour les utiliser et le système sera-t-il maintenu, mis à jour et sécurisé selon les besoins ?)
- de la Contestation (quels sont les mécanismes de contrôle et de vérification ? Les personnes concernées reçoivent-elles notification et sont-elles informées de l'utilisation de l'outil ?)
- de la précision (*Accuracy* – le niveau de précision spécifié correspond-il à l'objectif politique et peut-il être vérifié régulièrement ? Le taux de résultats positifs/négatifs erronés peut-il se justifier, quelles sont les conséquences des résultats erronés et le risque qui en découle est-il acceptable ? Les utilisateurs de l'outil disposent-ils de l'expertise nécessaire ?)
- de la Responsabilité (l'utilisation de l'outil est-elle objectivement équitable, transparente et responsable ? Peut-on considérer qu'elle est conforme à l'éthique et à l'intérêt général ?)
- de l'Explicabilité (existe-t-il des informations appropriées sur les dispositions applicables et la pondération des différents facteurs ? La police aurait-elle la possibilité de demander à un expert en sciences des données d'expliquer et de justifier l'outil ?)

48. La commission d'éthique ne rend pas de décisions contraignantes, mais donne plutôt des conseils au PCC et au chef de la police. Ces conseils ne portent pas sur la loi ou le respect des normes juridiques, mais sur l'éthique au sens large. En principe, le/la chef de la police peut ignorer ces conseils, mais il/elle reste responsable devant le/la PCC qui peut, en théorie, le/la démettre de ses fonctions. Dans la pratique, les conseils de la commission ont toujours été suivis. Par exemple, une proposition d'application de « gestion intégrée des délinquants » a subi trois révisions avant d'être approuvée par la commission d'éthique. M. McNeil m'a expliqué qu'un certain nombre de propositions avaient été rejetées, dont un modèle prédictif, suite aux réactions de diverses instances communautaires locales.

49. Ma conversation avec les deux fonctionnaires de police de grade supérieur, M. Todd, commissaire principal, et M. Dale, inspecteur principal, a clairement montré qu'ils comprenaient bien les questions éthiques liées à la NDAS. Nous avons discuté de la protection des données, de la confiance du public, de la transparence, de l'explicabilité, de l'obligation de rendre des comptes, de la responsabilité et du contrôle, ainsi que de la nécessité de démontrer l'efficacité des solutions technologiques, qui doit être évaluée afin de

[43] Pour en savoir plus, consultez www.westmidlands-pcc.gov.uk/wp-content/uploads/2019/07/Ethics-Committee-Terms-of-Reference-as-at-1-April-2019.pdf?x83908.
[44] Marion Oswald, qui a participé à notre audition en commission à Berlin en novembre 2019, en est membre.
[45] Oswald et autres, 2018, op. cit.

déterminer la proportionnalité des atteintes portées aux droits protégés. Les fonctionnaires de police ont insisté sur le fait que la décision finale serait toujours prise par une personne (qui en assumera la responsabilité) : les résultats algorithmiques ne font qu'éclairer la décision, sans jamais la déterminer. Ils ont également indiqué comment les éventuels problèmes de préjugés anciens dans les ensembles de données avaient été soulevés et résolus à l'occasion des échanges entre la police et la commission d'éthique.

50. Le commissaire principal Todd et l'inspecteur principal Dale ont bien conscience, tout comme M. McNeil (et Mme Oswald, présente lors de notre audition en commission), que cette commission d'éthique ne constitue pas une solution parfaite. Ils ont par ailleurs souligné l'absence de législation particulière sur l'utilisation de l'IA par la police au Royaume-Uni. L'école de police est en train de mettre au point une pratique professionnelle autorisée – une norme nationale commune – sur l'utilisation de l'analyse des données, mais elle ne sera pas non plus obligatoire. Le non-respect de cette norme pourrait toutefois être un élément pertinent à prendre en compte dans le cadre d'une procédure judiciaire.

### *3.2. Éléments d'une possible réglementation de l'IA utilisée dans les systèmes de police et de justice pénale*

51. M. Veale et Mme Oswald ont également formulé des propositions sur la manière dont l'IA pourrait être réglementée. L'utilisation de systèmes algorithmiques dans les processus décisionnels humains devrait se justifier au cas par cas, en tenant compte du contexte opérationnel. Cette démarche devrait inclure la prise en compte de la nature des interventions qui en résultent et de leur impact, surtout sur les communautés marginalisées. Un registre national des systèmes d'IA déployés dans le secteur public devrait être mis en place. Une attention particulière devrait être portée à l'effet cumulatif des systèmes algorithmiques : un système isolé peut très bien n'avoir que des effets négatifs proportionnés, mais plusieurs systèmes utilisés à différents stades du traitement des données peuvent avoir des effets cumulatifs bien plus importants et difficiles à prévoir. Les systèmes devraient être explicables à leurs utilisateurs immédiats et aux personnes concernées par le processus décisionnel. Il devrait être possible d'examiner et de tester le système d'IA sur son lieu d'exploitation. Il faudrait exclure les systèmes propriétaires qui ne le permettent pas. Les conséquences de l'introduction d'éléments d'IA dans les technologies existantes devraient être étudiées avec attention : par exemple, la reconnaissance faciale appliquée aux réseaux de caméras de surveillance a déjà créé une forme de surveillance de masse extrêmement puissante, qui va bien au-delà des attentes qui existaient lors de la première mise en service de cette technologie, et qui pourrait aussi permettre une lecture labiale automatique.

### *3.3. Le CAHAI et le cadre juridique possible de l'intelligence artificielle*

52. En septembre 2019, le Comité des Ministres a créé le Comité ad hoc sur l'intelligence artificielle (CAHAI), un organe intergouvernemental. Le CAHAI a été chargé d'examiner la faisabilité et les éléments possibles d'un cadre juridique applicable au développement, à la conception et à l'application de l'intelligence artificielle. Ses travaux se fondent sur les normes du Conseil de l'Europe en matière de démocratie, de droits de l'homme et d'état de droit, ainsi que sur d'autres instruments juridiques internationaux pertinents et sur les travaux en cours au sein d'autres organisations internationales et régionales. Outre les participants habituels, qui représentent les États membres et observateurs du Conseil de l'Europe et d'autres organes du Conseil de l'Europe (y compris l'Assemblée), le CAHAI bénéficie d'un niveau de participation exceptionnellement élevé de représentants d'organismes du secteur privé, de la société civile et d'établissements de recherche et universitaires.

53. Le CAHAI a tenu sa première réunion du 18 au 20 novembre 2019. Il a notamment décidé que figurerait, parmi les éléments essentiels de la future étude de faisabilité, une « cartographie des risques et des opportunités découlant du développement, de la conception et de l'application de l'intelligence artificielle, y compris l'impact de cette dernière sur les droits de l'homme, l'État de droit et la démocratie ». Le CAHAI prévoit actuellement d'adopter l'étude de faisabilité lors de sa troisième réunion, prévue en décembre 2020.

54. C'est dans ce contexte institutionnel que l'Assemblée débattra du présent rapport et de divers autres rapports relatifs à l'IA actuellement en préparation dans diverses commissions. L'Assemblée a choisi d'aborder le sujet sous un angle contextuel, en examinant les effets de l'IA dans différents domaines. La commission des questions juridiques et des droits de l'homme, par exemple, a également consacré des rapports aux technologies de l'interface cerveau-machine, aux véhicules autonomes et aux systèmes d'armes létales autonomes (ce dernier se trouve aux premiers stades de son élaboration). Les recommandations que l'Assemblée pourrait adopter sur la base de ces rapports fourniront donc au CAHAI des éléments d'orientation importants pour la cartographie des risques et des opportunités de l'intelligence artificielle et de son impact

sur les droits de l'homme, l'état de droit et la démocratie, ainsi que pour déterminer ensuite la nécessité d'un cadre juridique international contraignant.

55. L'utilisation d'outils d'intelligence artificielle dans les systèmes de police et de justice pénale présente des risques particuliers pour les droits de l'homme. D'une part, cela s'explique par l'importance des décisions qui peuvent être prises sur la base de résultats algorithmiques – décisions en matière de surveillance, de perquisition et saisie, d'arrestation, de détention, de condamnation, de libération conditionnelle avec ou sans caution, notamment, qui ont toutes une incidence particulièrement forte sur les droits de l'homme. D'autre part, cela tient au fait que les données policières classiques utilisées par l'IA pour former les algorithmes sont souvent entachées de préjugés de longue date et que l'introduction de l'IA dans un processus décisionnel aussi délicat peut porter atteinte à la responsabilité humaine et rendre les décisions non transparentes et difficiles à contester. Mes échanges avec la police des West Midlands au Royaume-Uni ont montré que la police elle-même est consciente de la nécessité d'encadrer son utilisation de l'IA par une législation ; en l'absence de cette dernière, elle prend déjà des mesures de son côté pour assurer un contrôle éthique et une normalisation des bonnes pratiques. Les exigences générales strictes en matière de légalité et de régularité énoncées aux articles 5 et 6 de la Convention – le droit à la liberté et à la sûreté et le droit à un procès équitable, qui sont tout spécialement conçus pour s'appliquer aux systèmes de police et de justice pénale – confirment la nécessité d'une réglementation juridique.

56. Au niveau des normes internationales, il existe déjà des normes applicables, notamment dans la Convention européenne des droits de l'homme et dans la jurisprudence de la Cour, ainsi que des instruments spécialisés tels que la Convention 108+. Toutefois, en raison de la nature *sui generis* de l'intelligence artificielle et de ses nouvelles applications possibles, il serait préférable de définir plus précisément les normes pertinentes dans un instrument spécialisé et juridiquement contraignant.

## 4. Conclusions et recommandations

57. La recherche sur l'intelligence artificielle numérique informatisée remonte au milieu du siècle dernier, mais son développement et son application ont fait des progrès spectaculaires ces dernières années, en grande partie grâce aux progrès des techniques d'apprentissage automatique rendus possibles par la puissance accrue du traitement informatique et par la disponibilité d'énormes quantités de données d'entraînement. L'intelligence artificielle est désormais utilisée dans un large éventail de situations toujours plus nombreuses, dont certaines peuvent avoir de profondes répercussions sur la démocratie, les droits de l'homme et l'état de droit, et notamment sur le système de justice pénale.

58. Ce rapport a examiné certaines des questions et préoccupations d'ordre général qui ont trait à l'IA, ainsi que quatre applications de l'IA dans divers aspects du système de justice pénale. Il s'est également intéressé à plusieurs préoccupations spécifiques soulevées par certaines applications, ainsi qu'à un système de contrôle éthique de l'utilisation de l'IA par la police. Mon analyse montre que les préoccupations générales relatives à l'IA valent également pour ses applications spécifiques dans les systèmes de police et de justice pénale que j'ai étudiés. Elle conduit également à penser que les arguments d'ordre général avancés en faveur d'une réglementation plus stricte de l'intelligence artificielle peuvent être particulièrement pertinents pour son application dans les systèmes de justice pénale. Mes propositions détaillées pour faire face à cette situation figurent dans l'avant-projet de résolution et de recommandation ci-joint.

**Annexe**

**Intelligence artificielle – Description et principes éthiques**

*On a tenté à plusieurs reprises de définir le terme « intelligence artificielle » depuis sa première utilisation en 1955. Ces initiatives s'intensifient aujourd'hui, car les organes normatifs, notamment le Conseil de l'Europe, réagissent aux capacités et à l'omniprésence croissantes de l'intelligence artificielle en œuvrant en faveur de son encadrement juridique. Il n'existe cependant toujours pas de définition « technique » ou « juridique » unique qui soit universellement admise[46]. Aux fins du présent rapport, il est toutefois indispensable de décrire cette notion.*

À l'heure actuelle, le terme « intelligence artificielle » désigne en général les systèmes informatiques capables de percevoir et d'extraire des données de leur environnement, puis d'utiliser des algorithmes statistiques pour traiter ces données, afin d'obtenir des résultats qui correspondent à des objectifs prédéterminés. Les algorithmes se composent de règles définies par l'homme ou par l'ordinateur lui-même, qui « forme » l'algorithme en analysant des ensembles de données considérables et continue à affiner ces règles à mesure qu'il reçoit de nouvelles données. Cette méthode, connue sous le nom « d'apprentissage automatique » ou « d'apprentissage statistique », est actuellement la technique la plus utilisée pour les applications complexes ; elle est uniquement devenue possible ces dernières années grâce à l'augmentation de la puissance de traitement des ordinateurs et à la disponibilité de données suffisantes. « L'apprentissage en profondeur » représente une forme particulièrement avancée d'apprentissage automatique, qui utilise plusieurs couches de « réseaux neuronaux artificiels » pour traiter les données. L'analyse ou la compréhension intégrale par l'homme des algorithmes développés par ces systèmes n'est pas toujours possible ; aussi sont-ils parfois qualifiés de « boîtes noires » (il arrive que ce terme désigne également, mais pour une raison différente, les systèmes d'intelligence artificielle propriétaires protégés par les droits de propriété intellectuelle).

Les formes actuelles d'intelligence artificielle sont toutes « restreintes », c'est-à-dire affectées à une tâche unique définie. L'intelligence artificielle « restreinte » est également qualifiée parfois de « faible », même si les systèmes modernes de reconnaissance faciale, de traitement du langage naturel, de conduite autonome et de diagnostic médical, par exemple, sont incroyablement sophistiqués et effectuent certaines tâches complexes avec une rapidité et une précision étonnantes. « L'intelligence artificielle générale », parfois qualifiée d'intelligence artificielle « forte », qui est capable d'exécuter toutes les fonctions du cerveau humain, est encore à réaliser. La « super-intelligence artificielle » désigne un système dont les capacités dépassent celles du cerveau humain.

---

*Comme le nombre de domaines dans lesquels les systèmes d'intelligence artificielle sont appliqués est en augmentation, puisqu'ils se propagent dans des domaines qui peuvent avoir un impact important sur les droits individuels et les libertés, ainsi que sur les systèmes démocratiques et l'État de droit, la dimension éthique de ce phénomène a fait l'objet d'une attention croissante et de plus en plus urgente.*

Un large éventail d'acteurs a formulé de nombreuses propositions d'ensemble de principes éthiques qui devraient être appliqués aux systèmes d'intelligence artificielle. Ces propositions sont rarement identiques et diffèrent à la fois dans les principes qu'elles énoncent et par la manière dont elles définissent ces principes. Les études montrent que la teneur essentielle des principes éthiques qui devraient être appliqués aux systèmes d'intelligence artificielle fait néanmoins l'objet d'un large consensus ; c'est notamment le cas des principes suivants[47] :

- *Transparence.* Le principe de transparence peut faire l'objet d'une interprétation élargie, de manière à englober l'accessibilité et l'explicabilité d'un système d'intelligence artificielle, en d'autres termes la possibilité donnée à un individu de comprendre le fonctionnement de ce système et le mode de production de ses résultats.
- *Justice et équité.* Ce principe englobe la non-discrimination, l'impartialité, la cohérence et le respect de la diversité et du pluralisme. Il suppose également que la personne à laquelle est appliqué un système d'intelligence artificielle puisse en contester les résultats, disposer d'une voie de recours et obtenir réparation.
- *Responsabilité.* Ce principe englobe le fait d'exiger qu'un être humain soit responsable de toute décision qui a des conséquences sur les droits et libertés individuels, et que l'obligation de rendre des

[46] Pour une vue d'ensemble élargie des tentatives de définition de « l'intelligence artificielle », voir *AI Watch: Defining Artificial Intelligence – Towards an operational definition and taxonomy of artificial intelligence,* Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F. et Delipetrev, B., European Commission Joint Research Centre, 2020.
[47] Voir *Lignes directrices sur l'éthique en matière d'IA : situation en Europe et dans le monde,* rapport provisoire commandé par le Comité ad hoc sur l'intelligence artificielle (CAHAI) du Conseil de l'Europe, M. Ienca et E. Vayena, mars 2020.

comptes et la responsabilité juridique de ces décisions soient définies. Il est donc étroitement lié au principe de justice et d'équité.

- *Sûreté et sécurité.* Ce principe suppose que les systèmes d'intelligence artificielle fassent preuve de solidité, de sécurité contre toute ingérence extérieure et de sûreté contre la commission d'actes involontaires, conformément au principe de précaution.
- *Respect de la vie privée.* Si le respect des droits de l'homme en général peut être considéré comme inhérent aux principes de justice et d'équité, de sûreté et de sécurité, le droit au respect de la vie privée est particulièrement important chaque fois qu'un système d'intelligence artificielle procède au traitement de données à caractère personnel ou privé. Les systèmes d'intelligence artificielle doivent par conséquent respecter les normes contraignantes du règlement général sur la protection des données (RGPD) de l'UE et de la Convention 108 du Conseil de l'Europe sur la protection des données (et de sa version actualisée, la Convention 108+), le cas échéant.

La mise en œuvre effective des principes éthiques applicables aux systèmes d'intelligence artificielle exige une approche intégrée de l'éthique, notamment une évaluation de leur impact sur les droits de l'homme, de manière à garantir le respect des normes établies. Il ne suffit pas que ces systèmes soient conçus uniquement sur la base de normes techniques et que des éléments soient ajoutés à un stade ultérieur pour tenter de faire respecter les principes éthiques.

Dans quelle mesure le respect de ces principes doit-il être intégré dans des systèmes particuliers d'intelligence artificielle ? Cela dépend des utilisations prévues et prévisibles de ces systèmes : plus leur impact sur l'intérêt général et les droits et libertés individuels est important, plus les garanties doivent être strictes. La réglementation éthique peut donc être mise en œuvre de différentes manières, depuis les chartes internes volontaires pour les domaines les moins sensibles jusqu'aux normes juridiques contraignantes pour les plus délicats. Dans tous les cas, il importe qu'elle prévoie des mécanismes de contrôle indépendants selon le niveau de réglementation.

Ces principes essentiels portent sur les systèmes d'intelligence artificielle et leur environnement immédiat. Ils n'ont pas vocation à être exhaustifs ni à exclure des préoccupations éthiques plus générales, telles que la démocratie (participation pluraliste des citoyens à l'élaboration de normes éthiques et réglementaires), la solidarité (qui admet les différences de point de vue des divers groupes) ou la durabilité (préservation de l'environnement de la planète).