Strasbourg, 10 mars 2019 CDPC(2019)7

COMITÉ EUROPÉEN POUR LES PROBLÈMES CRIMINELS

(CDPC)

Groupe de travail d'experts sur l'intelligence artificielle et le droit pénal

DOCUMENT DE TRAVAIL II
1re réunion, Paris, 27 mars 2019

Document établi par Sabine Gless, Professeure, Rapporteure spéciale

A. Contexte : initiative du CDCP sur l'IA et la justice pénale

À la suite de la session thématique sur l'intelligence artificielle et le droit pénal organisée le 28 novembre 2018, les membres du CDPC ont décidé de constituer un groupe de travail chargé pour l'essentiel d'analyser l'incidence de l'IA sur la justice pénale et de présenter diverses pistes possibles pour les activités futures du CDPC. Le comité a convenu que ce (premier) projet devait s'articuler autour d'une vision commune des questions qui sont au cœur même de son travail, afin d'éviter les incidences indésirables dues à l'emploi de l'IA et d'empêcher que la robotique ne cause de graves préjudices.

Il est entendu que le Conseil de l'Europe peut jouer un rôle de premier plan en aidant ses États membres à élaborer des normes juridiques communes qui formeraient un système réglementaire adéquat, complet et simple, créant ainsi un équilibre viable qui reconnaît aux technologies leurs nombreux bénéfices tout en obligeant les auteurs d'abus et d'actes néfastes à rendre compte de leurs actes. Ce projet, tenant dûment compte du *caractère d'ultime recours* de la réglementation pénale dans un domaine aussi complexe, s'intéresse exclusivement aux situations dans lesquelles le degré de préjudice ou l'importance de l'obligation violée pourrait, ou devrait, engager une responsabilité pénale et/ou dans lesquelles l'utilisation de l'IA a une incidence directe sur les systèmes de justice pénale.

Pour les besoins de notre groupe de travail :

(1) nous partirons du postulat que les environnements intelligents ambiants, ou que les situations dans lesquelles l'informatique ubiquitaire répond à des besoins humains, débouchent sur une coopération accrue entre l'homme et la machine au quotidien. Celle-ci est évidente dans le domaine de la *conduite automatisée, lequel laisse déjà présager des effets considérables sur le droit pénal et servira d'exemple de base aux fins de ce groupe de travail*.
Cela étant, des questions fort similaires se posent dans chaque domaine d'application de l'IA, qu'il s'agisse de robots mis au service de personnes âgées ou de systèmes informatiques réalisant des évaluations de risques dans les établissements pénitentiaires.

(2) nous utiliserons les définitions suivantes :

- **Intelligence artificielle (IA)** : combinaison de disciplines – dont la logique mathématique, la statistique, le calcul de probabilités, la neurobiologie informatique et l'informatique – visant à permettre à la machine d'imiter, voire de surpasser, les facultés cognitives de l'être humain.

- **Robot** : dispositif doté d'une intelligence artificielle intégrée dans une structure matérielle, capable d'accomplir des actions ayant des incidences sur le monde réel ; désigne également un « **bot** », c'est-à-dire un logiciel autonome capable d'interagir avec d'autres programmes ou avec un utilisateur humain.

- **Preuve électronique** : donnée générée automatiquement lors d'une coopération homme-machine fondée sur l'IA, présentée comme preuve dans une enquête pénale.

- **Automatisation de la conduite** : remplacement progressif du conducteur humain par l'IA grâce à des assistants de conduite exécutant (provisoirement) les tâches du conducteur. L'industrie distingue actuellement cinq[1] niveaux de conduite automatisée.

B. IA et justice pénale

En novembre 2018, au cours de la session thématique, le CDPC a recensé les domaines d'intérêt dans lesquels l'IA a une incidence sur la justice pénale et qui pourraient avoir des effets sur les principes généraux du droit pénal, tels que la procédure régulière et l'équité, et sur les concepts fondamentaux de la coopération transfrontalière :

- Droit pénal matériel : en particulier, risque d'un vide de responsabilité ;
- Droit procédural : en particulier, question de la preuve électronique ;
- Entraide judiciaire (MLA) : en particulier, double incrimination et transfert transfrontalier des éléments de preuve ;
- Droit pénitentiaire et droit répressif : en particulier, évaluation des risques.

Le tour d'horizon proposé ci-après décrit succinctement les problématiques propres à chaque domaine, puis s'ensuit la question pratique de la rédaction d'un questionnaire. Cela étant, le *sujet principal* reste bien évidemment *les enjeux pour le droit pénal matériel* lorsque l'humain disparaît (lorsque l'on passe du carrosse à la voiture puis au véhicule sans conducteur, par exemple), mais que la nécessité d'imputer la responsabilité demeure en cas de préjudice causé par la coopération homme-robot. Le droit pénal matériel doit se pencher sur les questions du libre arbitre, de la négligence lors la collaboration menée par les différents acteurs à l'origine du préjudice et des risques socialement acceptés. En cas d'accident, les éléments de preuve pertinents proviendront très certainement de la machine, sous forme de données générées par le robot prenant part à la relation coopérative – ce qui soulève également des questions en matière d'entraide judiciaire. Les données doivent être récupérées et prises en compte à bon escient lors de l'établissement des faits lors de l'enquête pénale, ce qui peut influer sur les droits de la défense et soulever la question de la territorialité en fonction de l'endroit où sont stockées les données. Il s'agit de problèmes paradigmatiques dès lors que les humains et les machines coopèrent, parfois directement au service de la justice pénale. Ainsi, lorsqu'un système d'IA réalise une évaluation de risques en vue d'une libération anticipée, et que des humains acceptent ses décisions, les questions de la responsabilité, de la crédibilité et de la transférabilité en dehors des frontières peuvent également se poser.

I. Droit pénal matériel

1. Existe-t-il un vide de responsabilité ?

Un vide de responsabilité se crée-t-il (lorsque l'acteur humain disparaît) ?

Lorsque des acteurs humains coopèrent avec des robots et que les actes de l'homme cèdent la place à ceux du robot (*cas des systèmes de conduite automatisée remplaçant progressivement le conducteur, par ex.*), la question est de savoir qui est responsable des préjudices causés par le robot au cours de la conduite (*conducteur, fabricant, fournisseur, par ex.*) se pose.

[1] Norme SAE J3016_201401 <https://www.sae.org/standards/content/j3016_201401>.

L'objectif final est de s'assurer qu'un mécanisme de responsabilité adéquat s'applique aux actes du robot. C'est pourquoi il est important que les gouvernements légifèrent en matière de responsabilité. La portée et le contenu des dispositions restent à définir, mais elles pourraient viser des cas (spécifiques) de coopération homme-robot (comme la responsabilité des conducteurs en situation de conduite automatisée, que prévoit la loi allemande[2]) ou édicter des règles générales de responsabilité (comme le fait l'article 12 de la *Convention sur la cybercriminalité, STE n° 185*, qui place les États sous l'obligation d'engager la responsabilité des personnes morales). Des obligations de conformité pourraient également être imposées aux entreprises de l'industrie de l'IA, afin de garantir une représentation juridique et des structures de gouvernance interne adéquates.

2. Comment définir le risque socialement accepté ?

La vie moderne en général, et plus particulièrement la circulation routière, comportent un risque : celui de se blesser, parfois mortellement. La loi permet toutefois de conduire une voiture, car la société accepte certains risques liés à la circulation routière. Les constructeurs automobiles et d'autres experts estiment que la conduite automatisée (fondée sur une coopération homme-robot) présente plus de garanties de sécurité, dans nombre de situations, que la conduite humaine. Une question cruciale en matière de développement de systèmes de conduite automatisée concerne le type de risques que différentes sociétés sont disposées à accepter. Il importe donc de se demander si tous les États partagent la même notion du « risque socialement accepté ». Cette question peut aussi se poser dans le domaine de l'entraide judiciaire (voir ci-après, MLA, point xxx) et de la double incrimination. En arrêtant une définition commune, les États réduiraient les risques de conflits potentiels.

3. La responsabilité peut-elle être imputée (à l'un des acteurs de la chaîne d'approvisionnement de systèmes d'IA) ?

Les robots fonctionnent grâce à de multiples sources de données et services. En cas de préjudice, il peut être difficile d'identifier le responsable d'une donnée d'entrée ou d'un résultat de sortie. Ces problèmes sont déjà pris en compte au niveau technologique, grâce à des tests en boîte noire par exemple, mais il faut aussi qu'une réglementation les encadre. Cette dernière pourrait énoncer la responsabilité de la chaîne d'approvisionnement[3], définir des sphères de risques, etc.

II. Procédure pénale

1. Défis liés aux preuves générées par la machine

Lorsque la coopération homme-robot cause un préjudice et que le conducteur humain est accusé de négligence, il y a de fortes chances pour que l'élément de preuve pertinent présenté à charge soit une « preuve machine » ou soit constitué de données générées par le robot au

[2] § 1a Strassenverkehrsgesetz / Loi fédérale sur le transport routier adoptée en 2017 (www.gesetze-im-internet.de/stvg) : *« Conduire un véhicule automatisé est légal sous réserve que les systèmes de conduite automatisée soient utilisés conformément à l'autorisation…»*

[3] Dans la mesure où les entités coopérant à fournir des systèmes d'IA savent probablement mieux que quiconque comment mettre en place des garanties de responsabilité et de transparence au sein de la chaîne d'approvisionnement (origines et utilisation des données d'entraînement, données de test, modèles, interfaces de programmation applicative (API), etc.)

cours de l'activité en question. La question se pose alors de savoir comment récupérer et exploiter ces données (voir aussi au point xxx, MLA, ci-après) et, si elles sont versées au dossier comme éléments de preuve, comment évaluer leur crédibilité. Il est surprenant de voir à quel point les procédures d'enquête, et surtout les évaluations de la fiabilité de la preuve, sont axées sur l'être humain. Des travaux théoriques suggèrent que la procédure accusatoire se prête bien, par certains de ses aspects, à l'élaboration de méthodes d'évaluation efficaces pour vérifier la fiabilité d'une preuve machine. Ce n'est pas le cas pour l'instant du système inquisitoire.

2. Droits de la défense (article 6, paragraphe 3 de la CEDH, notamment le droit « d'interroger un témoin ») ?

Le recours à une preuve machine pourrait affaiblir les droits de la défense. Imaginons le cas d'un accident mortel se produisant lors d'un trajet en voiture effectué pour partie par un conducteur humain et pour partie en conduite automatisée, et à la suite duquel seul le conducteur est poursuivi. La question se pose de savoir si ce dernier disposera de moyens adéquats pour préparer sa défense en connaissance de cause par rapport aux données générées par un robot qui seront présentées à charge. Le droit d'une personne accusée d'une infraction pénale à interroger des témoins elle-même, conféré par l'article 6, paragraphe 3, alinéa d de la CEDH, peut être un point de départ pour obliger à divulguer le code source, le paramétrage d'un système d'apprentissage machine, des données d'entraînement, etc.

3. Contrôle des codes et industrie (source ouverte/secret des affaires/lanceurs d'alerte)

La possibilité d'assurer sa défense en connaissance de cause (après examen des preuves machine) nécessite des systèmes d'IA transparents et contrôlés par des organes indépendants dès le départ. Ce n'est qu'en permettant à des tiers experts de réaliser des contrôles et de publier des informations sur des systèmes clés indépendamment d'un procès pénal et en comprenant les infrastructures d'IA depuis leur phase de configuration jusqu'au déploiement en passant par leur entraînement que l'on peut élaborer des stratégies de défense. Se pose alors la question des sources ouvertes, du privilège dit des secrets d'affaires et de la protection des lanceurs d'alerte, comme l'on appelle ces employés que le sens du devoir pousse à dévoiler devant un tribunal pénal les informations sensibles auxquelles ils ont accès.

4. Investigations intrusives et droits de l'homme

Les agents de services répressifs, les magistrats et les représentants d'autres autorités risquent, lorsqu'ils utilisent des systèmes d'IA (équipant leur véhicule par exemple), d'altérer la validité des enquêtes pénales. Les progrès considérables accomplis par l'IA dans le domaine de la reconnaissance faciale peuvent aider à identifier des suspects ou des prisonniers en cavale. Cela étant, le recours à ces systèmes pourrait être considéré comme une violation de l'article 8 de la CEDH et donc nécessiter une réglementation suffisamment stricte pour garantir le respect de la vie privée. L'IA est utilisée à des fins de profilage, notamment dans des programmes de police prédictive, ce qui soulève la question de la discrimination raciale ou de la validité de la présomption d'innocence.

III. MLA et préoccupations relatives aux infrastructures

1. Double incrimination

L'exigence de double incrimination est un principe traditionnel de la MLA, en vertu duquel un État peut refuser de poursuivre une personne s'il considère que l'acte commis ne relève pas du droit pénal. Son champ d'application est plus restreint dans l'UE, mais il s'applique de façon générale dans le contexte du Conseil de l'Europe. Si certains États autorisent la conduite automatisée et d'autres non, et qu'une voiture franchit une frontière, son conducteur peut se retrouver dans une situation problématique s'il utilise une technologie qui n'est autorisée que dans un pays donné. Ce cas de figure ne présente en principe aucun problème, car les codes de la route s'appliquent au niveau national, avec ou sans conduite automatisée. Mais dans la pratique, des problèmes peuvent surgir dans le cas des véhicules homologués pour une conduite automatisée qui doivent être conduits manuellement dans certains pays.

2. Accès transfrontalier aux données/éléments de preuve

L'accès aux données impliquant des systèmes d'IA, tels que les systèmes de conduite automatisée, nécessitent souvent des opérations transfrontalières, si les serveurs sont situés sur une juridiction étrangère par exemple, ou si les données sont hébergées sur le cloud d'un fournisseur privé étranger. La Convention sur la cybercriminalité (STE n° 185) prévoit un accès transfrontalier aux données en fixant des conditions minimales pour les mesures d'enquête, parmi lesquelles des injonctions de produire (article 18) et des obligations de conservation de données (articles 16 et 17). Mais on ne sait pas très bien si les moyens prévus au titre de cette convention sont suffisants ou s'il est nécessaire de les actualiser.

3. MLA et acteurs privés

Aujourd'hui, les données sont souvent hébergées par des fournisseurs de services dans le Cloud (CPS), éventuellement dans un pays tiers. Dans certaines enquêtes pénales, les règles de territorialité ne semblent plus remplir leur rôle de garantes de l'accès à l'information – un rôle désormais dévolu aux CSP. Il faut garder ce point à l'esprit dans le domaine de l'entraide judiciaire.

IV. Droit pénitentiaire, maintien de l'ordre / Évaluation des risques

L'utilisation de l'IA peut aussi avoir une incidence sur le droit pénitentiaire et le maintien de l'ordre. Dans différents domaines, des programmes sont utilisés pour dresser le profil d'individus dans le cadre d'évaluations de risques – par exemple pour évaluer les risques de récidive d'un détenu avant d'accorder une libération anticipée. Des outils de reconnaissance efficaces pourraient être utilisés dans les prisons pour prendre des décisions concernant les conditions de détention de certains détenus.

On pourrait également envisager d'équiper des véhicules de tels systèmes, par exemple pour détecter les conduites dangereuses et les tendances à l'excès de vitesse.

C. Perspectives d'action

1. Étapes de travail

Le groupe de travail devrait examiner, sans arrêter de décision finale, l'objectif général :

(a) en matière de fond (par ex., combien de domaines couvrir parmi ceux décrits sous les intitulés B.I à IV ?) ;

(b) en matière de forme (quelle forme doit prendre l'activité du CDPC dans le domaine : (i.) actualisation d'une convention, par ex. la Convention sur la cybercriminalité, par l'ajout d'un protocole ; (ii.) rédaction et adoption d'une nouvelle convention intitulée « Convention sur l'IA et le droit pénal » ; (iii.) élaboration d'une recommandation ou d'un autre instrument non contraignant ?).

2. Questionnaire

Le groupe de travail est chargé d'établir un questionnaire.

3. Plan de travail

27 mars 2019	1ère réunion du groupe de travail (Paris) pour préparer le questionnaire en vue d'avoir une vue d'ensemble sur la situation concernant le droit pénal matériel et procédural dans les Etats membres
Mai 2019	Envoyer le questionnaire à toutes les délégations du CDPC (Etats membres)
Septembre 2019	Délai pour les Etats membres pour envoyer leurs réponses au questionnaire
Septembre/Octobre 2019	Préparation d'une analyse des réponses au questionnaire reçues par les Etats membres
Octobre 2019	2ème réunion du groupe de travail (Paris) pour discuter des résultats atteints et qui seront présentés à la réunion plénière du CDPC du mois de Décembre
3-6 Decembre 2019	Présentation des résultats atteints à la réunion plénière du CDPC (Strasbourg)
Janvier/Février 2020	3ème réunion du groupe de travail (Paris). Prise en compte des conclusions du CDPC et début du travail concernant un instrument du Conseil de l'Europe/ CDPC
Juin 2020	Conférence international sur les normes communes de droit pénal sur les préjudices causés par les véhicules autonomes (ou d'autres applications de l'IA)
Septembre/Octobre 2020	4ème réunion du groupe de travail (Paris) pour compléter le projet d'instrument en vue de sa présentation au CDPC