



Encadrement des risques techniques et juridiques des activités de police prédictive

Céline Castets-Renard, Philippe Besse, Jean-Michel Loubes, Laurent Perrussel

► To cite this version:

Céline Castets-Renard, Philippe Besse, Jean-Michel Loubes, Laurent Perrussel. Encadrement des risques techniques et juridiques des activités de police prédictive. [Rapport de recherche] Centre des Hautes Etudes du Ministère de l'Intérieur. 2019. hal-02190585

HAL Id: hal-02190585

<https://hal.science/hal-02190585>

Submitted on 22 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



MINISTÈRE DE L'INTÉRIEUR

Encadrement des risques techniques et juridiques des activités de police prédictive

Rapport final (Mars 2019)

Porteuse du projet

Professeur Céline Castets-Renard (IRDEIC, Université Toulouse Capitole)

Équipe du projet

Professeur Philippe Besse (IMT, INSA Toulouse)

Professeur Jean-Michel Loubès (IMT, Université Paul Sabatier)

Professeur Laurent Perrussel (IRIT, Université Toulouse Capitole)



SOMMAIRE

I. INTRODUCTION : ORIGINE ET CONTEXTE DE LA POLICE PREDICTIVE AUX ETATS-UNIS ET EN EUROPE	5
II. CARTOGRAPHIE ET ANALYSE DES OUTILS DE POLICE PREDICTIVE : COMPARAISON US ET EUROPE	13
A. PRESENTATION DES PRINCIPAUX OUTILS DE POLICE PREDICTIVE	13
1. <i>Analyse des principaux outils externes commerciaux déployés aux Etats-Unis : PredPol, Palantir, HunchLab, Beware</i>	<i>14</i>
a) PredPol	14
b) Palantir	20
c) HunchLab	23
d) Beware	27
2. <i>Analyse des principaux outils internes déployés aux Etats-Unis et en Europe.....</i>	<i>28</i>
a) La Strategic Subject List (SSL) à Chicago (ciblage de personnes)	28
b) PredVol et Paved en France (ciblage de lieux)	30
c) Les outils internes déployés ailleurs en Europe (ciblage de lieux)	32
B. CATEGORISATION DES OUTILS DE POLICE PREDICTIVE.....	33
1. <i>Outils internes des services de police et gendarmerie versus outils externes proposés par des entreprises privées.....</i>	<i>34</i>
2. <i>Ciblage des lieux versus ciblage des personnes</i>	<i>35</i>
III. ENJEUX JURIDIQUES ET ETHIQUES DES OUTILS DE POLICE PREDICTIVE.....	39
A. PRINCIPALES REGLEMENTATIONS APPLICABLES AUX OUTILS DE POLICE PREDICTIVE.....	39
1. <i>Respect des droits fondamentaux.....</i>	<i>39</i>
a) Respect des droits fondamentaux avant les actions de police prédictive.....	39
b) Respect des droits fondamentaux après les actions de police prédictive	40
2. <i>Egalité de traitement et discrimination</i>	<i>41</i>
a) Risques de biais et discrimination des outils de police prédictive.....	41
b) Discrimination et exercice de l'autorité publique	42
3. <i>Protection des données personnelles et transparence algorithmique.....</i>	<i>44</i>
a) Outils ciblant des personnes et protection des données personnelles	44
b) Outils ciblant des personnes et transparence algorithmique.....	47
4. <i>Police prédictive et missions de la police et gendarmerie : police administrative versus police judiciaire</i>	<i>49</i>
B. ENJEUX ETHIQUES DES ALGORITHMES PREDICTIFS : TRANSPARENCE, LOYALTE, EXPLICABILITE ET EFFICACITE.....	52
IV. CONCLUSION – RECOMMANDATIONS	59
A. CONCLUSION	59
B. RECOMMANDATIONS	61
1. <i>Recommandations au législateur pour l'encadrement des outils de police prédictive.....</i>	<i>61</i>
2. <i>Recommandations au Ministère de l'intérieur pour l'intégration des outils de police prédictive</i>	<i>62</i>
3. <i>Recommandations aux opérationnels pour l'utilisation des outils de police prédictive.....</i>	<i>64</i>
V. ANNEXES	67
A. ANNEXE 1 : GLOSSAIRE ET DEFINITION DE L'IA.....	67
1. <i>Glossaire de l'intelligence artificielle et police prédictive</i>	<i>67</i>
2. <i>Définition de l'IA</i>	<i>71</i>
B. ANNEXE 2 : LISTE DES PERSONNES INTERROGÉES PAR LES RAPPORTEURS	79
VI. BIBLIOGRAPHIE	81

I. Introduction : origine et contexte de la police prédictive aux Etats-Unis et en Europe

Origine de la police prédictive aux Etats-Unis

Les technologies du *big data* et les analyses prédictives ont fait évoluer les méthodes de maintien de la paix et sécurité publique, ainsi que du management des forces de police aux Etats-Unis, dès le début des années 2000¹. L'outil précurseur fut Compstat déployé par le département de la police de New York (NYPD) à partir de 1995², dans le but de rassembler les données de la criminalité et rendre compte de l'activité de la police. Cet outil a par la suite été déployé dans de multiples villes des Etats-Unis³. À partir du début des années 2000, sont apparus des outils proposés par des sociétés commerciales, comme PredPol, HunchLab ou encore Palantir.

Contexte social, criminel et légal aux US

Le fait que les activités de police prédictive se soient développées en premier lieu sur le territoire des Etats-Unis n'est pas l'effet du hasard. Même si la criminalité en France et en Europe n'est pas négligeable, en particulier en présence d'un risque terroriste élevé depuis plusieurs années, les Etats-Unis voient s'étendre une criminalité de grande ampleur et surtout une violence urbaine armée très significative. Les problèmes de criminalité se sont révélés particulièrement graves dans la plupart des grandes villes des Etats-Unis, comme en témoignent Chicago, New York, Los Angeles, Miami par exemple. Les règles de certains Etats autorisant la possession et le port d'armes, ainsi que la protection que confère le 2^e amendement de la Constitution, forment un contexte légal très différent de l'Europe. À titre d'exemple, Chicago, troisième mégapole aux Etats-Unis derrière New York et Los Angeles, est détentrice du record du taux de criminalité aux Etats-Unis avec 593 meurtres en 2017⁴. Chicago compte aussi une quantité impressionnante de gangs rivaux qui comportent plus de 100 000 membres, dont la plupart sont prêts à utiliser la violence armée. Plus largement, les Etats-Unis font face à des violences armées comme les *shootings*, notamment dans les écoles.

Aux Etats-Unis, aucune politique nationale ne régit la police. Les forces de police sont dirigées et rémunérées localement et chaque ville décide des directions à prendre par le biais

¹ A. G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, NYU Press, 2017.

² William Bratton, Commissaire au sein du Département de police de New York City dans les années 1990, avec l'aide de Jack Maple, ont mis en place une approche centrée sur les données pour gérer la police. En utilisant CompStat, les commandants des districts reportent les infractions toutes les semaines et les commandants de police évaluent la réduction des crimes et le taux d'arrestation.

³ William Bratton a été nommé chef de la police de Los Angeles et a apporté CompStat sur la côte ouest. Il a été également autorisé à mener la première expérience de police prédictive au sein de LAPD.

⁴ <https://www.chicagotribune.com/news/data/ct-shooting-victims-map-charts-htm1story.html>.

de débats politiques et d'élections locales. Dans la majorité des villes, une forte volonté politique d'éradiquer la criminalité urbaine a vu le jour spécialement dans les années 2000-2010. Dans un contexte de crise financière et économique à partir de 2007 et, conséquemment, de coupes budgétaires dans les départements de police, les outils de police prédictive sont apparus comme pouvant permettre d'atteindre cet objectif de « faire mieux » avec moins de moyens humains, lors même que ce contexte économique et social difficile faisait craindre une recrudescence de la criminalité. Les outils de police prédictive firent alors figure de recette miracle, de nature à compenser le manque de moyens humains et atteindre une meilleure efficacité dans la lutte contre la criminalité⁵. Parallèlement, si les départements de police de nombreuses villes ont vu leur budget diminuer, au niveau fédéral, le département de justice finança de nombreux projets pour encourager les villes à « essayer » ces nouvelles technologies⁶.

Enfin, l'émergence de ces outils de police prédictive s'est trouvée de fait encouragée par la crise de confiance des citoyens envers les forces de l'ordre de nombreuses villes à cette période. Les violences policières, en particulier envers la population des jeunes afro-américains, ont conduit à chercher des méthodes plus « objectives » pour améliorer le climat, ainsi que les conditions de maintien de l'ordre. Les risques de discrimination inhérents aux méthodes traditionnelles furent amplement dénoncés par la société civile et les mouvements sociaux comme « *Black lives matter* » mais également officiellement par les investigations du Département Américain de la Justice (*US Department of Justice* (DoJ)) portant sur le fonctionnement de la police de Ferguson après le décès de Michael Brown⁷. L'objectif était alors de trouver des solutions plus modernes et, si possible, non biaisées envers certaines catégories de population. L'inconstitutionnalité de méthodes, tel le *stop-and-frisk* (arrestations et fouilles) à New York ou, dans le reste du pays, le *Terry Stop*⁸ a convergé avec l'émergence de ces nouvelles solutions technologiques parées *a priori* de toutes les vertus. Ces politiques reposent sur la décision de la Cour Suprême rendue en 1968 dans l'arrêt *Terry v. Ohio* et ont contribué à réduire les exigences constitutionnelles du 4^e amendement. En principe, le 4^e amendement de la Constitution américaine interdit les recherches et saisies déraisonnables non fondées sur une « probable cause » mais la Cour suprême considère qu'il n'y a pas d'atteinte au 4^e amendement quand un policier arrête un suspect dans la rue et le maintien sans « cause probable », sur le fondement d'un simple

⁵ A. G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, op. cit. p. 21.

⁶ *Ibid.*

⁷ Voir les émeutes ayant entouré la mort de, un afro-américain âgé de 18 ans, abattu le 9 août 2014 par Darren Wilson, policier à Ferguson (Missouri). Une enquête a été menée par le FBI et le département américain de la justice ayant confirmé des pratiques racistes du département de la police de Ferguson : *US DEPARTMENT OF JUSTICE, INVESTIGATION OF THE FERGUSON POLICE DEPARTMENT* (2015).

⁸ Tiré du nom de l'arrêt *Terry v. Ohio* de la Cour Suprême (392 U.S. 1 (1968)).

“suspçon raisonnable” que cette personne a commis, commet ou va commettre une infraction et a une croyance raisonnable que cette personne est armée et est présentement dangereuse. Ces politiques de *stop-and-frisk* servaient aussi d’instrument de gestion de la police pour quantifier l’activité individuelle des policiers (quotas de fouilles). Cependant, une telle interprétation du 4^e amendement entraîna de nombreuses fouilles systématiques, spécialement dirigées envers les communautés afro-américaine et latino. Elle a finalement été jugée inconstitutionnelle par une juge fédérale en 2010⁹. Aujourd’hui, si ces programmes sont abandonnés, il en résulte un amoncellement de données massives sur la criminalité, proportionnellement plus nombreuses envers ces populations et de nature à engendrer des discriminations. Les données amassées depuis les années 1960 jusqu’à l’abandon de ces politiques en 2010 n’ont pas été effacées et sont venues alimenter les bases de données utilisés pour mettre en œuvre les outils de police prédictive, en dépit de leur défaut de précision et de l’impact disproportionné (*disparate impact*) envers les afro-américains. En conséquence, les prévisions des outils de police prédictive vont plus souvent pointer vers des populations afro-américaines et latinos ou encore vers les quartiers dans lesquels ils habitent majoritairement. Il en est de même des bases de données concernant les gangs. À New York, une telle base de données est alimentée à 99% par une population latino et afro-américaine, alors même qu’entrer dans cette base de données ne nécessite pas la preuve d’une criminalité ou suspicion d’un mauvais comportement¹⁰. Les bases de données historiques portent l’empreinte de ces discriminations. En utilisant les données historiques pour créer les algorithmes, les programmes ont appris en se fondant sur des données biaisées racistes, au risque de renforcer les biais¹¹.

Police prédictive, entreprises technologiques et déploiement militaire

La pression et le lobbying des entreprises technologiques expliquent aussi le choix fait par les conseils municipaux des villes de recourir à ces solutions pour combattre une criminalité grandissante avec des moyens faiblissant. Pour les entreprises américaines de technologie, il s’agissait là d’une opportunité de déployer sur le territoire intérieur, des technologies mises au point dans le cadre d’opérations militaires menées à l’international par les Etats-Unis, en particulier en Afghanistan ou en Irak. Le rôle joué par la société *Palantir* dans l’élaboration de ce type d’outils, initialement à usage militaire, illustre la place prise par l’industrie militaire dans les méthodes de contrôle de la criminalité sur le territoire national.

⁹ L’usage du contrôle à dose élevée est jugé non constitutionnel en 2010 par la Cour de justice de l’Etat de New York lors de l’affaire *Floyd vs City of New York* : 739 F. Supp. 2d 376 (S.D.N.Y. 2010).

¹⁰ <https://www.legalaidnyc.org/nypd-gang-database/>.

¹¹ *Ibid.*

Définition de la « police prédictive »

Aujourd'hui, les outils de police prédictive ne sont plus l'apanage des Etats-Unis et sont aussi introduits dans plusieurs Etats membres de l'Union européenne. Toutefois, les méthodes utilisées ne sont pas identiques, aussi faut-il s'entendre sur les notions.

De façon générale, la « police prédictive » consiste pour la police (en France, police ou gendarmerie) à utiliser des données et autres indicateurs pour calculer la probabilité du risque de survenance d'un crime dans le futur. Il peut également s'agir d'un calcul de probabilité qu'une personne, précédemment repérée en raison d'activités suspectées d'illégalité, puisse passer à l'action. De telles informations sont utiles, d'une part, pour prévenir la commission d'infractions et, d'autre part, pour mieux déployer les forces de police. Notons d'ailleurs que certaines entreprises de développement revendiquent l'utilité de leur outil essentiellement pour assurer cette seconde fonction.

De façon générale, les analyses prédictives ainsi réalisées se basent sur des données du passé, relatives à la localisation et l'heure de faits criminels, afin de faire des prévisions futures pour déterminer l'endroit et le moment où le risque de survenance de crimes est le plus élevé. L'objectif des analyses prédictives est d'anticiper ce qui peut se produire dans le futur par une utilisation plus efficace des informations passées. Grâce à cette connaissance, les forces de police et moyens peuvent être mieux déployés et utilisés pro-activement dans le but de prévenir les crimes.

Il existe différents types d'analyse en criminologie : la prédiction des auteurs d'infraction (focalisée sur la prévention de la récidive), la prédiction des victimes (concentrée sur des groupes cibles) et, enfin, la prédiction des lieux et temps du risque le plus élevé de survenance d'une infraction. Les analyses prédictives ne permettent pas d'expliquer le phénomène de la criminalité et constituent plutôt une approche *ad hoc* au sein de la lutte et prévention du crime. Une telle approche doit être complétée par d'autres stratégies de maintien de la sécurité publique. L'aspect innovant résiderait dans le fait que la police prédictive permettrait à la police d'anticiper le crime grâce à des données empiriques. Une telle approche suppose de tenir pour vrai les théories selon lesquelles les incidents criminels ne se produisent pas par hasard et présentent un pattern prévisible qu'il convient de trouver. Dans ce rapport, nous n'analyserons pas le bien-fondé de ce prérequis qui pourrait être discuté. Nous nous concentrerons plus précisément sur la définition de ce que sont l'intelligence artificielle et le *machine learning*, sur la description critique des principaux outils commercialisés, ainsi que sur le cadre légal à respecter en France, différent de celui applicable en particulier aux Etats-Unis.

Terminologie

La terminologie « police prédictive » est à utiliser avec précaution. L'« analyse

décisionnelle » est parfois préférée à « l'analyse prédictive »¹², dans la mesure où il s'agit d'aider la prise de décision, plutôt que de prédire la criminalité. En outre, parmi les différents outils existant, il convient aussi de se demander quelle place est laissée à l'homme. Les forces de police sont-elles libres de suivre ou non les préconisations faites par la machine ? Le chef opérationnel garde-t-il son libre arbitre ? Ces questions sont essentielles pour déterminer la place laissée à la technologie et l'imputation des responsabilités.

Notons cependant que pour les besoins de ce rapport, nous continuerons d'utiliser le terme « police prédictive » par simplicité.

Technologie et données

La police prédictive exploite des algorithmes d'apprentissage machine intégrant des modèles statistiques complexes de *machine learning*, afin par exemple de produire des cartes, fondées sur des unités spatiales placées en ville, montrant les lieux et temps où le crime a le plus de risque de survenir. La complexité des modèles utilisés et le recours à des algorithmes d'apprentissage automatique peuvent être discutés. Il apparaît en effet que si une masse importante de données de nature différentes peuvent être agrégées, les moyens utilisés ne sont pas nécessairement très complexes ni ne reposent sur des algorithmes sophistiqués d'apprentissage. Cela nous a conduit à apporter des précisions techniques sur ce qu'est l'IA, ainsi qu'à donner des définitions de certains mots clés¹³.

Quoi qu'il en soit, l'utilisation d'outils décisionnels suppose de disposer de données sur les actes criminels passés en nombre suffisant. Dès lors, une telle méthode ne peut être utilisée que pour prévenir des infractions ayant un fort impact et survenant suffisamment souvent pour produire des données en quantité et permettre la création de modèles. Cela suppose que les infractions en question doivent donner lieu à des plaintes et déclarations des victimes permettant de pointer relativement facilement des lieux ou périodes de survenance. Concrètement, il est plus efficace de considérer essentiellement les infractions, tels des cambriolages et vols (à l'arraché, vols de voiture ou vélos). Il est clair que les infractions pénales peu dénoncées, comme par exemple les violences domestiques, se prêtent mal à l'utilisation de ce type d'outils. De même, les crimes de sang surviennent en nombre insuffisant avec des motivations variées et sans qu'il y ait de pattern de répétition le plus souvent, si bien qu'il n'est pas possible de créer des modèles portant sur ce type de criminalité. Notons cependant, par exemple, que Chicago subit une criminalité armée tellement forte qu'un outil de prédiction des fusillades entre les gangs a été élaboré. La connaissance des gangs et leurs relations peut permettre de cerner des répétitions d'actes

¹² Voir l'article dans l'Essor de la gendarmerie nationale, à l'occasion du Forum International de la Cybersécurité, janv. 2018 : <https://essor.org/operationnel/gendarmerie-de-lanalyse-predictive-a-lanalyse-decisionnelle>.

¹³ Voir l'annexe 2 qui pose les définitions de l'IA et le glossaire.

criminels dans le passé pour espérer leur anticipation dans le futur.

Au centre des dispositifs déployés aux Etats-Unis, par exemple par le LAPD (*Los Angeles Police Department*), se trouvent les données « de police » sur la criminalité et éventuellement aussi des données « extra-police », telles les données environnementales et de météorologie¹⁴. Ces données massives ont émergé du fait de l'augmentation des capacités de collecte, stockage, de tri et d'analyse des indices numériques du crime. Des *data centers* de stockage des données ont également été mis en place, dans le but d'agréger et fusionner les données, de mieux partager l'information entre les agences locales, étatiques et fédérales compétentes notamment en matière de justice pénale¹⁵.

Police prédictive et collecte de données personnelles aux Etats-Unis

L'activité de police prédictive implique l'utilisation d'une grande quantité de données. La collecte des données est régie par des règles très différentes en France et dans l'Union européenne, en particulier s'agissant de la collecte et du stockage des données personnelles. Tout d'abord dans l'Union européenne et en France, la loi informatique et libertés du 6 janvier 1978, plusieurs fois révisée, et la directive 95/46/CE du 24 octobre 1995, abrogée par le règlement général de protection des données personnelles (RGPD) entré en application le 25 mai 2018 encadrent strictement le traitement des données personnelles, spécialement les données sensibles (par exemple les données raciales, de santé, d'opinion politique). Ainsi, le principe de finalité des traitements des données personnelles interdit le croisement de données. Aux Etats-Unis, la protection des données personnelles n'est pas globale mais sectorielle, aussi de nombreuses situations ne sont-elles pas régies par la loi et laissent alors la possibilité d'un usage sans limite des données. Il existe de nombreux cas dans lesquels les données peuvent être librement utilisées, agrégées, croisées et revendues. Dès lors, les risques d'atteinte à la vie privée sont *a priori* plus élevés aux Etats-Unis qu'en Europe, compte tenu des différences de contrainte légale. Ces différences de réglementation ont un impact sur l'utilisation des données, mais aussi sur la constitution de bases de données concernant la criminalité. Le contexte législatif aux Etats-Unis a ainsi permis un déploiement d'une grande ampleur des outils de police prédictive. Ce déploiement tient au contexte d'une criminalité grave et massive difficile à endiguer, aux règles souples permettant la collecte massive de données personnelles, ainsi qu'à la pression commerciale d'entreprises technologiques développant leurs outils de prédiction, dont la mise en œuvre a été amplement financée par le budget fédéral du département de la justice.

¹⁴ A. G. Ferguson, *THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT* (2017), p. 2.

¹⁵ https://it.ojp.gov/documents/fusion_center_guidelines_law_enforcement.pdf.

Police prédictive et contexte dans l'Union européenne

Dans l'Union européenne, les systèmes de police prédictive déployés paraissent plus modestes, en comparaison des Etats-Unis. La réglementation en Europe, en particulier la réglementation des données personnelles, a sans doute constitué et continue de constituer un obstacle au déploiement de ces outils, ou pour le moins de certaines catégories d'entre eux, portant un risque d'atteinte à la vie privée et aux droits fondamentaux des individus.

Cependant, tous les outils de police prédictive ne supposent pas la collecte de données personnelles. On constate en effet que sont utilisés dans un certain nombre d'Etats membres de l'Union européenne de outils ne reposant pas sur le recours à des données personnelles. Suivant la littérature¹⁶ et l'analyse des études disponibles en libre accès, on constate qu'au moins cinq Etats européens utilisent déjà de tels outils : le Royaume-Uni, les Pays-Bas, l'Allemagne, l'Autriche (projet pilote) et la Belgique (projet pilote) viennent s'ajouter à la France qui déploie l'outil Paved. L'Espagne, la Suède et le Danemark se déclarent intéressés par ces outils. Si on inclut l'usage d'outils prédictifs dans le domaine du renseignement, neuf Etats membres déclarent alors en utiliser.

Il convient de préciser le type d'outils dont il s'agit afin de les cartographier et catégoriser selon certains critères pertinents eu égard à la réglementation.

¹⁶ https://eucpn.org/sites/default/files/content/download/files/recommendation_paper_predictive_policing_update.pdf.

II. Cartographie et analyse des outils de police prédictive : comparaison US et Europe

Sans prétendre à l'exhaustivité¹⁷, seront présentés ici les principaux outils dits de « police prédictive », utilisés en Europe et aux Etats-Unis (A). Leur variété et hétérogénéité invitent à les catégoriser, dans la mesure où leur mode opératoire diffère et n'implique pas les mêmes conséquences sociales (B). En outre, les règles juridiques susceptibles de s'appliquer doivent conséquemment être différenciées, ce qui sera approfondi dans la 3^{ème} partie.

A. Présentation des principaux outils de police prédictive

La police prédictive est une méthode basée sur l'application de modèles statistiques et d'algorithmes d'apprentissage automatique, le plus souvent protégées par un brevet. Notons toutefois que les méthodes exploitées par les algorithmes et utilisées par les différentes entreprises ne sont pas nécessairement innovantes ou originales. Ce qui les différencie tient surtout aux données utilisées et aux règles de ciblage choisies en amont.

En pratique, il existe une multitude d'application. Certaines d'entre elles sont de puissantes applications commerciales bien connues, parmi lesquelles on trouve *PredPol*, *Palantir* ou encore *HunchLab*. Ces trois outils seront présentés ici, essentiellement à partir de la technologie décrite dans les brevets de *Palantir* et *PredPol*, alors que *HunchLab* a fait le choix de ne pas protéger ses technologies par brevet ou secrets commerciaux, arguant que leur plus-value tient aux données agrégées¹⁸ (1).

Parallèlement, il existe aussi des outils internes, développés directement par les services de police eux-mêmes, comme par exemple l'application CAS mise en œuvre aux Pays-Bas ou encore *PredVol* ou *Paved* en France. Seront évoqués les principaux outils connus et utilisés en Europe (2).

On peut d'ores et déjà remarquer que les services de police en Europe tendent à l'heure actuelle à développer leurs propres outils, plutôt que de se fier aux services proposés par les entreprises américaines. À l'inverse, comme précédemment relevé, les services de police des villes américaines ont dû faire face à une certaine urgence et ont été financièrement encouragés par l'Etat fédéral à utiliser ces services, afin de permettre le développement technologique des entreprises américaines.

¹⁷ Pour une présentation très complète et détaillée des différents outils existants, voir le rapport fait en 2015 pour la République Tchèque par le *Science and Research Institute* ACCENDO : Maps of the Future : A Modern Crime Analysis and Crime Prediction Based Tool to increase the Effectiveness and Quality of Public Administration performance in Crime Prevention : www.mvcr.cz/mvcren/file/maps-of-the-future-pdf.aspx.

¹⁸ D'après les propos de Jeremy Heffner, responsable produit et data scientist de l'entreprise Azavea qui commercialise HunchLab, reproduits par le journaliste Hubert Guillaud pour LeMonde : <http://internetactu.blog.lemonde.fr/2017/09/24/police-predictive-12-depasser-la-prediction-des-banalites>.

Enfin, il est important de relever aussi que ces outils, internes ou externes, se différencient par la méthode de ciblage choisie. Il est ainsi possible de distinguer : les outils ciblant les lieux et temps de survenance d'une ou plusieurs infractions déterminées par l'agrégation de données passées ; les outils ciblant les personnes susceptibles d'être impliquées dans une infraction, qu'elles en soient victimes ou auteurs.

Les outils les plus fréquemment utilisés sont des outils de cartographie des lieux de criminalité.

1. Analyse des principaux outils externes commerciaux déployés aux Etats-Unis : PredPol, Palantir, HunchLab, Beware

a) PredPol

PredPol est un logiciel commercial proposé par une société américaine, PredPol Inc., située à Los Angeles où il a été expérimenté en premier lieu par le LAPD¹⁹. L'outil a pour objectif de prédire, avec précision et en temps réel, le lieu et le moment où les crimes ont le plus de risques de survenir²⁰. Autrement dit, cet outil identifie les zones à risque (*hotspot*), suivant le modèle statistique utilisé en sismologie. Ce service a convaincu plusieurs dizaines de villes aux Etats-Unis. Outre Los Angeles, il a par exemple été utilisé dans le Comté du Kent au Royaume-Uni ou encore à Chicago. Les données entrantes sont : les archives de la police d'une ville ou d'un territoire spécifique (procès-verbaux, suivis d'arrestations, appels au secours), afin d'en déduire les endroits où les crimes sont les plus fréquents pour « prédire » les lieux à surveiller en priorité. La cible visée porte sur des lieux et non sur des personnes. Les types d'infractions concernées sont les cambriolages, vols de voitures et vols dans les lieux publics.

Analyse du brevet de PredPol

Le brevet US 8,949,164²¹ décrivant l'invention "*Event Forecasting System*" a été déposé le 6 septembre 2012 et obtenu le 3 février 2015 auprès de l'office américain des brevets et marques (USPTO)²². Le déposant et inventeur est Georges O. Mohler²³ qui a cédé son brevet à la société PredPol Inc. (cessionnaire). La demande US 9,805,311 B1 déposé le 31 octobre 2017 en est une continuation et revendique le bénéfice du brevet US 8,949,164, ainsi que du

¹⁹ <https://www.wired.com/story/los-angeles-police-department-predictive-policing>.

²⁰ PredPol affirme que son algorithme est jusqu'à deux fois plus précis que les analystes spécialisés : PredPol Inc. PredPol Predicts Gun Violence. http://www.predpol.com/wp-content/uploads/2013/06/predpol_gun-violence.pdf, 2013.

²¹ <https://patents.justia.com/patent/8949164>.

²² USPTO : United States Patent and Trademark Office.

²³ Georges O. Mohler est professeur d'informatique à Indiana University - Purdue University Indianapolis (IUPUI) et co-fondateur de la société PredPol.

procédé *Epidemic Type Aftershock Sequence (ETAS) Point Process Crime Forecasting*²⁴ qui constitue l'algorithme utilisé. Tous deux sont incorporés par référence dans leur intégralité et incluses aux Appendices A and B. Le brevet a été déposé dans la classification « représentation de la connaissance et raisonnement technique »²⁵ et l'invention a fait l'objet d'un soutien du Gouvernement²⁶.

L'invention ne décrit pas une suite logicielle mais une plateforme offrant un ensemble de services (SaaS : Software as a Service) disponible sur le centre de gestion de données de la société PredPol. Par des transferts sécurisés, cryptés, un service de police adresse des historiques de criminalité et reçoit en retour des prévisions de criminalité par type de délit et zone géographique sur une carte Google. Les patrouilles de police peuvent directement se connecter au serveur de PredPol afin de planifier leur circuit.

Les revendications du brevet permettent de déterminer le champ du brevet et donc l'appropriation par son titulaire. Ces revendications ne précisent pas la façon dont les données sont utilisées ni les calculs mis en œuvre. Elles portent essentiellement sur le procédé utilisé par le système de police prédictive, spécialement sur la méthode organisationnelle utilisée (les trois types de données (lieu, temps, infraction), la division géographique en cellules, le transfert d'informations par un système de télécommunications, le procédé de réception des données historiques, le recours aux données GPS, le lien avec des informations légales issues du code pénal...), plutôt que sur les aspects techniques. Il n'y a aucune description précise des méthodes algorithmiques utilisées.

Le brevet insiste plus particulièrement sur les interfaces graphiques et fonctionnalités proposées aux utilisateurs mais également critique sur le fond les cartes de *hotspot* (ou heatmap) représentant des lissages spatio-temporels des historiques de la criminalité. Si les lieux où la criminalité est la plus élevée sont connus des forces de police, les informations habituellement données par des logiciels comme CompStat ou CrimeStat ne sont pas suffisamment dynamiques et n'indiquent pas les lieux et le moment où la criminalité est la plus susceptible de se produire à un instant t , ce qui est prétendu pouvoir être fait avec le logiciel PredPol. En outre, la société PredPol indique pouvoir aider à une meilleure allocation des ressources dans le déploiement des patrouilles. Enfin, l'outil intègre aussi en temps réel la position de toutes les patrouilles et permet non seulement de savoir où elles se situent mais aussi de contrôler la position des policiers. Ce que prétend apporter PredPol par rapport à l'état de l'art antérieur est un système de police prédictive éprouvé qui fournit des prévisions

²⁴ ETAS est une application n° 61/573,541, déposée le 8 septembre 2011 auprès du USPTO.

²⁵ Current U.S. Class: Knowledge Representation And Reasoning Technique (706/46). International Classification: G06N 7/00 (20060101).

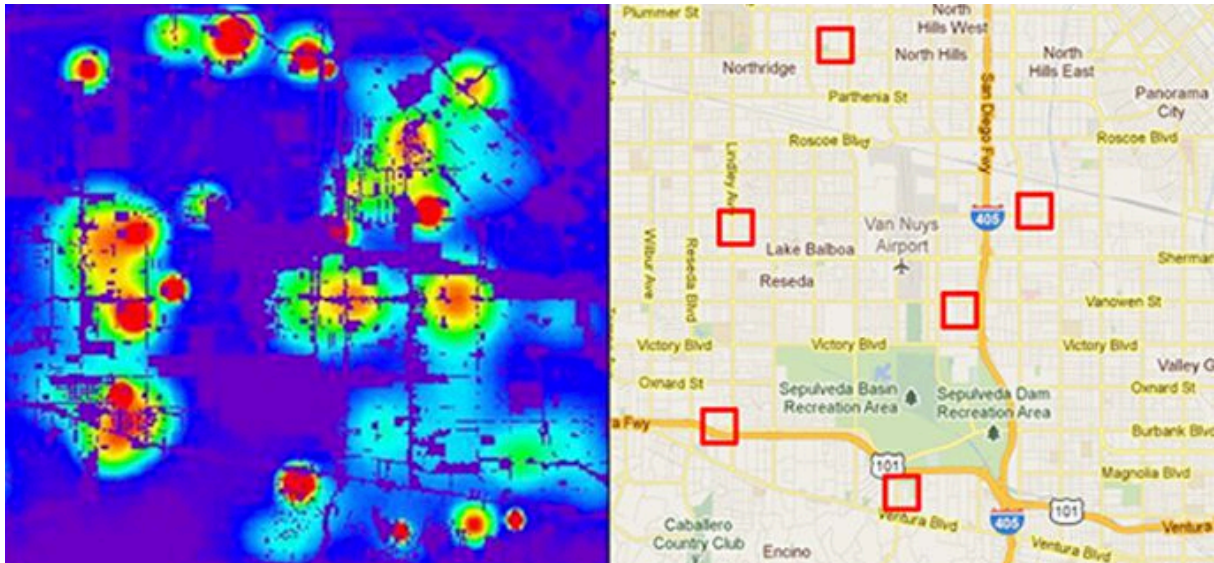
²⁶ Des parties de l'invention ont été réalisées avec le soutien du gouvernement sous le contrat DMS-0968309, ainsi que le contrat BCS-0527388, financé par la National Science Foundation (NSF), et enfin financé par le Multidisciplinary University Research Initiative (MURI) sous les contrats ARO MURI 50363-MA-MUR, 58344-MA and AFOSR FA9550-10-1-0569. Le gouvernement détient donc certains droits sur cette invention.

de criminalité ciblées et en temps réel en tenant compte des données historiques mais aussi des informations récentes spontanées qui soient présentées sous un format simple et pratique pour les forces de police et permette de renseigner les officiers afin de gérer au mieux les ressources et envoyer les patrouilles dans les lieux les plus pertinents. Le fait de mettre à disposition l'information sur des outils variés et mobiles comme des tablettes, smartphones, ordinateurs portables, en plus des ordinateurs de bureaux, a aussi constitué une rupture en comparaison des solutions précédemment utilisées.

Le modèle de prévision est basé sur l'hypothèse que la criminalité se diffuse selon un processus analogue à une contagion épidémique ou encore comme les secousses secondaires d'un tremblement de terre, ce qui serait particulièrement vrai pour les atteintes à la propriété et la criminalité des gangs. Les auteurs reconnaissent l'utilité des approches de type *hotspot* ou par estimation par noyau de la densité de la criminalité passée à condition d'en estimer correctement les paramètres, étape délicate à optimiser. Ils insistent sur les limitations de cette approche, incapable d'anticiper des événements spontanés ou leur réplication et ne faisant pas appel à une procédure rigoureuse d'estimation des paramètres par maximum de vraisemblance comme c'est d'usage en Statistique inférentielle classique. En plus du *hotspotting* historique, la prévision est donc basée aussi sur l'estimation des paramètres d'un modèle de type *Epidemic-Type Aftershock Sequence* (ETAS) qui est une application des processus ponctuels spatio-temporels auto-excitables (processus de Poisson non homogène) dont Reinhart²⁷ (2018) illustre l'utilisation pour les prévisions de la criminalité, des épidémies ou des tremblements de terre.

Les données d'apprentissage ou d'estimation des paramètres du processus ponctuel sont constituées par les historiques de trois types de données : les délits documentés par type, la localisation, la date. Les paramètres du modèle sont estimés par un algorithme EM (*expectation maximisation*). En temps réel, le serveur fournit des prévisions ou probabilités d'occurrence par type de délit et cellule géographique (150m x 150m), afin de proposer des recommandations tactiques pour optimiser les ressources de police. Une patrouille fournissant sa localisation GPS est informée en retour de sa présence dans une zone à risque.

²⁷ <https://arxiv.org/pdf/1708.02647.pdf>.



Source :

<https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/>

Remarques techniques :

- les données recueillies sont strictement limitées aux historiques de criminalité et non aux personnes impliquées ;
- le modèle de prévision est focalisé sur l'estimation des paramètres d'un processus ponctuel ETAS.

Remarques juridiques :

Le brevet détenu par PredPol permet l'appropriation d'une méthode dans sa globalité mais ne porte pas sur l'algorithme prédictif. Les aspects techniques ne font donc pas l'objet d'un droit de propriété mais vont être couverts par le secret commercial. En effet, la contrainte du brevet est d'obliger l'inventeur à révéler l'invention au public au travers des revendications qui constituent la délimitation légale de son droit de propriété. PredPol ne souhaitant pas dévoiler son algorithme a donc opté pour le secret commercial sur ces aspects. Le brevet en lui-même renseigne donc très peu sur les aspects techniques et si PredPol prétend à une certaine transparence de sa démarche, elle porte surtout sur la méthode mais non sur l'algorithme et les méthodes mathématiques utilisées. Il convient cependant de nuancer ce propos par la publication d'articles par les inventeurs sur le sujet²⁸.

Analyse critique de PredPol : inefficacité et biais algorithmique

Les études techniques de l'outil *PredPol* sont rares en raison d'une politique très limitée

²⁸ Voir les articles de G. Mohler dans la bibliographie.

d'ouverture des données utilisées et *a fortiori* du code développé par la start-up, protégé par le secret commercial. Quelques études techniques²⁹ ont néanmoins pu être menées, en se basant sur quelques jeux de données mis en accès libres par des villes comme Chicago et sur des modèles semblables à *PredPol*. Il est difficile d'évaluer la valeur ajoutée de cette prévision par rapport aux cartes historiques hotspots ou par estimation de la densité par noyau. En effet, les rares travaux publiés d'évaluation de cette approche ne concerne pas la qualité de prévision mais les statistiques de criminalité, statistiques qui sont sans doute plus sensibles à la stratégie de gestion des patrouilles qu'à une amélioration de la prévision par rapport à un historique classique. Contrairement à ce qui a pu être affirmé par *PredPol*³⁰, l'efficacité est finalement modeste, dépendant à la fois de la quantité de données disponibles sur une échelle temporelle et selon le type d'infraction commise. Les études précitées montrent le plus souvent que la prédiction des crimes a lieu majoritairement dans les zones historiquement les plus criminogènes de la ville, si bien que le logiciel n'apprend rien aux policiers expérimentés amenés à l'utiliser. Ce résultat n'est guère surprenant si on considère que la grande majorité des données utilisées portent sur les infractions passées.

Il est d'ailleurs à noter que si la police du Kent fut la première à introduire le "*predictive policing*" en Europe (Angleterre et Pays de Galles) en 2013, il est officiellement reconnu qu'il est difficile de démontrer si le système a véritablement permis de réduire la criminalité. Les officiers de police indiquent que le logiciel a été utile pour mener une politique proactive mais la preuve de son efficacité n'est pas véritablement apportée. Il apparaît même que la criminalité ait augmenté les cinq premières années de son utilisation³¹, ce qui fut alors justifié par des difficultés de maîtrise du logiciel ou par manque de temps de formation et d'utilisation du logiciel par les policiers. Dans le cas du Kent, les soi-disant avantages de *PredPol* en terme d'efficacité des résultats et de meilleure gestion des ressources policières, n'ont pas été clairement prouvés.

Au demeurant, le coût de l'outil lui-même n'est pas négligeable puisque, d'après ce que rapporte la BBC, il aurait coûté £130,000. Le coût élevé et le manque d'efficacité ont finalement conduit à l'abandonner en 2018, dès lors que la police du Kent était dans l'impossibilité de montrer la capacité à réduire la criminalité³². De façon générale, les coûts des outils proposés par les entreprises commerciales vont dépendre de la surface

²⁹ I. Benslimane, Étude critique d'un système d'analyse prédictive appliqué à la criminalité : *PredPol*®. *Cortex Journal*, 2014.

³⁰ G. Mohler. Marked point process hotspot maps for homicide and gun crime prediction in chicago. *International Journal of Forecasting*, 30(3) :491–497, 2014.

³¹ <https://www.bbc.com/news/uk-england-kent-32529731>.

³² <https://www.bbc.com/news/uk-england-kent-46345717>.

géographique couverte, de la taille de la population, de la durée du contrat. Les prix peuvent donc énormément varier. Il est à noter que, parallèlement, la police britannique s'est lancée dans un nouveau projet d'outil interne pour réduire les coûts et chercher une meilleure efficacité. Ce projet, intitulé NDAS (*National Data Analytics Solution*), a pour objectif de cibler des personnes risquant de mener des attaques à main armée ou des personnes susceptibles d'en être victimes, afin de les aider et offrir des thérapies. Il devrait entrer en application en mars 2019³³.

En résumé, l'expérience du comté de Kent tend à montrer que cette première tentative en Europe d'utiliser un logiciel prédictif élaboré aux Etats-Unis par une entreprise privée n'a pas fait la preuve de son efficacité. Les autorités britanniques suivent finalement l'exemple d'autres Etats européens de se doter de leurs propres outils. La tendance en Europe est de concevoir des outils internes, plutôt que de faire confiance à des outils externes élaborés par des sociétés privées, qui plus est américaines. Outre des raisons de coûts, les enjeux d'indépendance à l'égard d'acteurs privés justifient sans doute ses choix. Au demeurant, il est probable qu'un outil développé dans un certain contexte ne soit pas nécessairement pertinent dans un autre contexte criminogène, alors que les populations, configurations géographiques des villes et organisation des groupes criminels ne sont pas les mêmes.

Enfin, puisque le brevet ne décrit pas les aspects techniques de l'outil, l'algorithme étant protégé par secret commercial, il est difficile d'en vérifier l'efficacité mais aussi la loyauté. Son « auditabilité » est exclue et des critiques ont pu être formulées contre le caractère biaisé de l'outil qui a tendance à envoyer les patrouilles systématiquement dans les quartiers considérés comme les plus criminogènes où se situent principalement les populations afro-américaines et latinos³⁴. Les données historiques montrent certes des risques élevés dans ces quartiers, mais la plupart de ses données sont issues des politiques précédentes de *Terry Stop* et *Stop-and-frisk*, considérées comme biaisées, discriminantes et finalement inconstitutionnelles. Mais le système n'interroge pas ni ne remet en cause la loyauté et qualité de telles données. Au demeurant, le choix du type d'infractions, essentiellement liées aux atteintes à la propriété (cambriolages, vols de voitures...), constitue un type de criminalité susceptible d'être pratiquée par les populations les plus pauvres et les plus fragiles qui orientera vers le type de population précitée. Les résultats seraient naturellement différents si on considérait la criminalité financière par exemple, totalement exclue des outils de police prédictive aujourd'hui, eu égard aux difficultés de modélisation et en l'absence de données en suffisamment grand nombre. Le fait pour les services de police de vouloir prévenir

³³ <https://ici.radio-canada.ca/nouvelle/1138610/intelligence-artificielle-prediction-crimes-violents-royaume-uni-angleterre-grande-bretagne>.

³⁴ Sur les biais algorithmiques, voir l'étude réalisée sur le logiciel COMPAS de justice prédictive : Jeff Larson et al., *HOW WE ANALYZED THE COMPAS RECIDIVISM ALGORITHM* PROPUBLICA (2016), <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> (last visited Aug 12, 2018).

certaines infractions plutôt que d'autres au travers d'outils automatisés n'est pas neutre socialement.

Réponses de PredPol aux critiques formulées

Parmi les critiques formulées contre PredPol, le problème de biais algorithmique et biais des données est le plus souvent énoncé. Les fondateurs de PredPol et développeurs de la solution ont donné des réponses dans plusieurs articles publiés en 2017 et 2018 et mettent l'accent sur l'audit des données d'apprentissage³⁵. La qualité des données d'apprentissage est essentielle pour éviter ou réduire les biais. Mais si les données utilisées par PredPol sont biaisées, cela démontre que la société dans son ensemble est elle-même biaisée. PredPol ne fait que souligner ce constat sans pour autant créer la discrimination. Dès lors, les biais de l'outil ne sont pas plus importants que ceux générés précédemment par les données collectées par les policiers sur le terrain.

b) Palantir

Crime Risk Forecasting est le brevet déposé et détenu par la société **Palantir Technologies Inc.** (située à Palo Alto en Californie). Ce dispositif a été déployé par exemple à Los Angeles, New York ou à la Nouvelle-Orléans qui a souscrit un contrat en 2012 afin d'aider le département de police à cibler les chauffeurs violents. Ce contrat n'a cependant pas été renouvelé en raison des controverses suscitées par son objet ciblant la population. Les plaques d'immatriculation sont en effet des données personnelles. Mais le plus souvent, les contrats conclus avec les villes sont secrets, si bien qu'il est difficile de les contester³⁶. Les inventeurs de cette solution sont Duncan Robertson, Alexander Sparrow, Mike Lewin, Meline Von Brentano, Matthew Elkherj, Rafael Cosman³⁷.

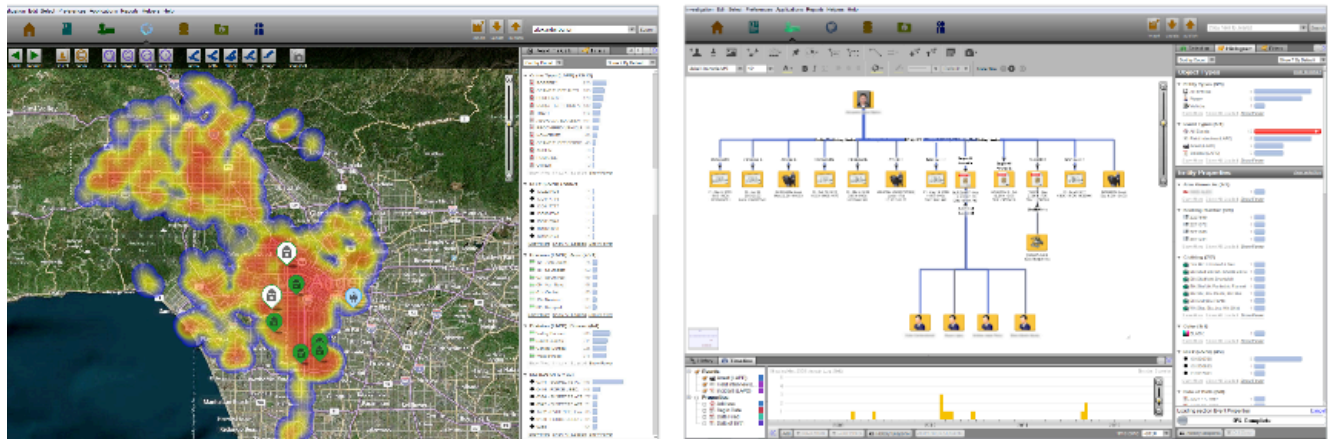
Crime Risk Forecasting est un ensemble de logiciels et matériels qui constitue une « invention » décrite dans le brevet US 9,129,219 déposé le 13 juin 2013 et obtenu le 8 septembre 2015. Il associe plusieurs composants fonctionnels dont un gestionnaire de base de données, des outils de visualisation, notamment de cartographie géographique interactive, et de prévision de la criminalité. Nous insisterons plus particulièrement sur les données archivées et utilisées ainsi que sur les méthodes et algorithmes de prévision, alors

³⁵ P. Jeffrey, Brantingham, The Logic of Data Bias and Its Impact on Place-Based Predictive Policing, *Ohio State Journal of Criminal Law* (2017). Brantingham, Valasik et Mohler, Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial, 2018. Notons que P. Jeffrey Brantingham est professeur d'anthropologie à l'Université de Californie à Los Angeles (UCLA), spécialiste du crime environnemental. Il est aussi le chef de la recherche-développement auprès de PredPol, Inc.

³⁶ Par exemple à la Nouvelle Orléans : <https://www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd>.

³⁷ <https://patentimages.storage.googleapis.com/60/94/95/5dbde28fe6eea2/US9129219.pdf>.

que le brevet est principalement orienté pour protéger les interfaces graphiques et fonctionnelles développées pour l'utilisation du produit.



L'intention affichée de Palantir est de mettre en œuvre une technologie pour prévenir les risques criminels, afin d'aider la police à savoir où et quand la criminalité va survenir dans le futur. La prévision du risque de criminalité est associée à une cellule géographique et temporelle de base, par exemple de 250 m carré, sur une période de 8h correspondant à la durée d'une patrouille de police. Ces valeurs, comme beaucoup d'autres du logiciel, peuvent être paramétrées. Des conventions graphiques : transparence, couleur... permettent de visualiser par cellule de la carte interactive l'intensité des risques, selon une échelle à définir, en fonction des types de criminalité choisis.

Données potentiellement utilisables ou utilisées.

- historique de la criminalité décrite par date, type, localisation ou différents cumuls de celle-ci ; date et heures peuvent être fournies avec précision ou par un intervalle de temps sur lequel le risque est uniformément réparti. De même la localisation peut être plus ou moins précise, par adresse, coordonnées ou zone géographique. La base peut inclure les modalités ou modes opératoires du crime selon une nomenclature ou en texte libre. Notons que les infractions envisagées sont larges car il peut s'agir de cambriolages, vols de véhicules ou à l'intérieur des véhicules, violences, etc.
- historiques d'informations peu ou pas directement liées à la criminalité : météo, présence de patrouilles dans la cellule ou à proximité ; répartition des personnels d'aide d'urgence,
- *custody data* ou données d'incarcération indiquent les personnes appréhendées ou en détention connues pour certains types de crimes. Ces données peuvent aider à faire décroître le risque d'une zone ou encore de renforcer celui où est opérée une libération.

Le prospective *hotspotting* est utilisé. Il s'agit du calcul d'une somme pondérée de tous les crimes ou de ceux d'un type particulier, associés à des voisinages temporels et géographiques déterminés par seuils paramétrables. La pondération peut apparaître sous la

forme d'un lissage exponentiel impliquant une décroissance dans le temps ou l'espace. L'option « histogramme » se limite à une cellule ou une zone déterminée.

En plus du *hotspotting*, différents modèles, algorithmes ou combinaisons de ceux-ci peuvent être exécutés pour produire la prévision. Les principaux algorithmes d'apprentissage automatique sont cités dont la régression logistique, les machines à vecteur support, les réseaux de neurones... Régression logistique et réseaux de neurones sont plus particulièrement décrits. Ils prennent en compte la criminalité de la cellule concernée et celles d'un voisinage ou d'une échelle plus large l'englobant. Les variables prédictives sont donc la criminalité antérieure indicées par le temps et le lieu ou encore de nouvelles variables réalisant des cumuls. Bien d'autres variables peuvent être introduites comme la distance au plus proche poste de police, le temps depuis la fermeture des bars, des composantes saisonnières annuelles ou hebdomadaires, la météo... les données d'incarcération.

Des algorithmes plus complexes peuvent être exécutés par agrégation de méthodes associant *hotspotting*, histogramme, modèles issus de la criminologie, algorithmes d'apprentissage, le tout pouvant être pris en entrée d'un autre algorithme d'apprentissage, par exemple un réseau de neurones.

Remarques techniques :

* Les possibilités de combinaison, agrégation de modèles et algorithmes, associées à celles des très nombreuses variables explicatives qui peuvent être prises en compte, engendrent une très forte complexité et donc un nombre considérable de paramètres à estimer et d'hyper paramètres à optimiser. Le brevet ne dit rien sur la façon dont sont optimisés ces paramètres ni sur la qualité attendu des prévisions. Le brevet liste plutôt une immense combinatoire de combinaisons ou perspectives de combinaisons possibles sans rien dire sur la stratégie de choix à adopter.

- Il nous semble difficile qu'un service de police s'empare concrètement de cet outil sans une aide constante de spécialistes de la société Palantir.
- Tout algorithme ou combinaison produisant une prévision binaire ou plutôt une probabilité d'occurrence d'un crime ou délit peut être implanté sans rien changer aux implications juridiques de l'outil.

Remarques juridiques :

Le point sensible concerne les données archivées dont certaines d'entre elles constituent des données personnelles :

- Quels sont les risques de ré-identification possible des victimes à partir du fichier des historiques ? Quelles précautions sont prises pour les anonymiser et éviter la ré-identification ?
- *Quid des custody data* ou données d'incarcération qui sont non seulement des données personnelles mais qui, en principe, ne font l'objet de traitements que par les

services gouvernementaux de police ou justice.

Notons par ailleurs que Palantir propose d'autres services d'analyse de données de criminalité comme *Gotham* utilisé par le NYPD³⁸ mais dont il n'a à ce jour pas été possible d'en connaître les caractéristiques techniques.

c) HunchLab

L'outil HunchLab est développé par la société Azavea (Philadelphia). Si on en croit les informations commerciales données par cette société³⁹, le système détermine automatiquement les données concernant les crimes commis, les cycles temporels, la météo, la localisation des lieux qui sont typiquement des lieux de criminalité (bars, restaurants, commerces, stations de métro), des événements organisés dans la ville, qui sont ensuite agrégées et utilisées dans le but de générer une prédiction. Le système est indépendant et peut être contrôlé par les forces de police sans connaissance particulière des méthodes statistiques et analytiques.

³⁸ On en veut pour preuve la requête pour demander l'accès à cet outil, formulée par le Brennan Center for justice art New York University, School of Law envers le NYPD, portée devant le juge Supreme Court of the State of New York State dans l'affaire n° 160541/2016 rendu le 22 décembre 2017 sur la base d'une requête fondée sur le FOIL Act (Freedom of Information Law).

³⁹ <https://www.hunchlab.com/resources>.



Source : <https://www.officer.com/command-hq/technology/communications/automated-notification-systems/product/10224229/azavea-hunchlab>



Source : <https://www.hunchlab.com/features/>

Si HunchLab est un concurrent direct de PredPol, l'approche commerciale est un peu différente. Cet outil de police prédictive ne s'intéresse pas seulement à la prédiction du crime, mais aussi à aider la police à répondre à ces prédictions, afin d'apporter non seulement une mesure de la criminalité, mais surtout d'évaluer son impact. Le service ne se définit d'ailleurs pas comme un outil de prédiction du crime, mais comme « *un logiciel de gestion de patrouille de police proactif* »⁴⁰. Tout l'enjeu est de réussir à créer un système d'aide à la décision avec une rétroaction plus solide, permettant de mieux évaluer ce qui peut être prédit et ce qui ne peut l'être et surtout de prendre garde aux contre-effets.

Comme la plupart des systèmes s'attachant à des zones, HunchLab agrège des données hétérogènes et les traduit en carte, en points chauds. Le système produit des cibles ou missions selon un code couleur pour le type de crime. L'outil doit alors proposer aux patrouilles de se rendre dans ces points chauds, selon des modalités différentes. HunchLab enregistre enfin les retours des patrouilles : les policiers doivent répondre à une série de questions pour évaluer leur travail et l'efficacité des propositions faites par le système. Contrairement à PredPol qui favorise le sur-contrôle aux points chauds, HunchLab utilise un modèle probabiliste pour faire varier les patrouilles et les encourager à patrouiller sans vraiment surestimer une zone sur une autre, pour essayer de ne pas transformer la réaction de la police à l'information qu'ils reçoivent.

HunchLab produit un jeu de données pour entraîner le système, utilisant les premières

⁴⁰ <http://internetactu.blog.lemonde.fr/page/4>.

années pour faire des prévisions et les années suivantes pour les vérifier et les améliorer. Le système produit des cartes de prédiction du crime selon le type de crime (vol, agression, cambriolage...), mais sans donner aux policiers de niveaux de risques. L'objectif est en effet d'éviter le problème posé par ce type de logiciels qui ont tendance à envoyer les policiers uniquement là où le risque calculé est le plus grand, ce qui influence le comportement des policiers. En outre, HunchLab introduit de l'aléatoire, par exemple en n'envoyant pas les policiers dans un endroit pourtant évalué à risque, afin de les inciter à rester vigilants et ne pas faire une confiance absolue à l'outil. L'intérêt est de ne pas sur-interpréter et surreprésenter les zones visualisées en rouge. Cela permet de prendre du recul sur l'information livrée et ne pas vouloir interpellé n'importe quel individu se trouvant dans une zone où le risque de crime est représenté comme fort. Enfin, HunchLab présente des options aux policiers et tente également d'évaluer leurs effets. En effet, de l'aveu même des développeurs de HunchLab, ces systèmes de police prédictive ne cessent de faire des erreurs, c'est pourquoi il est important d'intégrer une rétroaction permanente, afin de tenter de les réduire. Ainsi, lorsque le système montre aux officiers une cellule avec une possibilité de crime, ces derniers ne sont pas informés du niveau de risque, c'est-à-dire qu'ils ignorent si c'est une zone avec un risque très élevé ou une zone sélectionnée aléatoirement, dans le but de ne pas influencer leur comportement *a priori*. HunchLab prétend aussi qu'il inclut des mesures pour réduire le potentiel préjudice causé par une présence policière excessive. Il utilise un modèle statistique pour faire varier les lieux où les policiers sont envoyés. En outre, s'agissant des infractions susceptibles de comporter des biais, sont utilisés les rapports publics des événements plutôt que les rapports des officiers, afin de ne pas utiliser de données susceptibles d'être biaisées.

Autre différence majeure avec PredPol, HunchLab est par ailleurs transparent sur les données qu'il utilise, ses théories, ses modèles, ses algorithmes et n'a déposé aucun brevet ni ne se prévaut de secrets commerciaux. Il est donc *a contrario* difficile de connaître les spécificités techniques des outils utilisés en se basant sur la description du brevet comme cela a pu être fait avec PredPol et Palantir. HunchLab prétend que les outils de police prédictive en général ne constituent pas de réelles inventions et que les algorithmes ou techniques de machine learning utilisés par les entreprises sont souvent proches et connus des spécialistes. La réelle plus-value constituerait en l'agrégation d'une multitude de données et dans le choix d'intégrer de l'aléa et ne pas donner une surreprésentation au *hotspotting*. HunchLab annonce vouloir viser le marché européen et proposer des modes opératoires compatibles avec les réglementations pénales des Etats membres.

Remarques juridiques : HunchLab semble présenter les mêmes caractéristiques que l'outil de Palantir, en raison de l'agrégation d'une très grande quantité de données dont certaines d'entre elles sont susceptibles d'être des données personnelles si aucune précaution

technique d'anonymisation des données n'est prise pour empêcher la ré-identification. En l'absence de description technique du brevet, il est difficile de rendre compte de l'effet d'un tel outil.

d) Beware

Beware, outil commercial de la société *Intrado*, est utilisé par certains départements de police aux Etats-Unis depuis 2012. Il s'agit non pas de cibler des lieux pour déterminer des hotspots mais de cibler des personnes⁴¹.

Selon les informations données par les médias, Il s'agirait d'une application mobile basée dans le cloud, qui fonctionne avec Motorola Solutions new Intelligent Data Portal (IDP), plateforme qui agrège en quelques secondes des informations contextuelles à partir des appels d'urgence (9-1-1). Sont agrégées à ces appels des données commerciales disponibles publiquement, ainsi que les informations pénales et les données issues des réseaux sociaux. L'algorithme de *Beware* détermine alors un score et un niveau de risque (vert, jaune ou rouge) des personnes qui sont automatiquement envoyés au policier ayant fait la requête. Le logiciel fut piloté pour la première fois dans le département de police de la ville de Thornton dans le Colorado et était utilisé par le département de la police de Fresno en Californie⁴², d'après ce que rapporte en janvier 2016 le *Washington Post*⁴³. Il est cependant à noter que certaines villes, comme Bellingham (Etat de Washington), ont décidé de ne pas se doter de ce logiciel, après que ses administrés aient manifesté leurs inquiétudes sur les coûts et implications sur la vie privée, bien que la ville ait reçu une aide fédérale pour couvrir une part du coût annuel de *Beware* de \$36,000⁴⁴. Cet outil est en effet controversé en raison de la multitude de données, y compris personnelles, agrégées par un opérateur privé, alors que le procédé et l'algorithme utilisés ne sont pas connus. Tout comme pour les autres outils commerciaux précédemment présentés, la preuve de son efficacité n'a pas été faite. Par ailleurs, l'utilisation de données massives dont la qualité et la source ne sont pas vérifiées voire même non connues, ne permet pas de garantir la transparence attendue. Enfin, le risque d'erreurs (faux positifs et faux négatifs) est encore plus élevé eu égard à l'utilisation massive de données dont la source et la nature sont très variées. Il semblerait que toutes ces faiblesses aient finalement conduit le conseil municipal de la ville de Fresno à ne plus utiliser cet outil.

⁴¹ https://www.youtube.com/watch?v=e_s8sPrv1PI.

⁴² Hoggard, C. (2015) Fresno police scanning social media to asses threat, *ABC30 Action News*, <http://abc30.com/news/fresno-police-scanning-social-media-to-assess-threat/525999/>.

⁴³ https://www.washingtonpost.com/local/public-safety/the-new-way-police-are-surveilling-you-calculating-your-threat-score/2016/01/10/e42bccac-8e15-11e5-baf4-bdf37355da0c_story.html?noredirect=on&utm_term=.d79edd09e433.

⁴⁴ Johnson, T. (2014) Intrado intrusion. City council backs away from social software, *Cascadia Weekly*, Wednesday July 9 2014 : <http://www.cascadiaweekly.com/cw/currents/17003> (consulté le 12 décembre 2018).

À défaut d'accéder à des informations techniques fiables et officielles issues d'un dépôt de brevet, aucune analyse technique n'a pu être faite.

Remarques juridiques : dans la mesure où cet outil cible des personnes et non des lieux, les risques d'atteinte au droit des données personnelles sont ici particulièrement élevés. On peut douter du fait qu'un tel outil puisse être conforme à la réglementation européenne de protection des données personnelles, *a fortiori* depuis l'entrée en application du règlement européen n° 2016/679/EU de protection des données (RGPD).

2. Analyse des principaux outils internes déployés aux Etats-Unis et en Europe

Le déploiement d'outils internes présente plusieurs avantages. L'un d'entre eux est sans doute de permettre une meilleure pertinence de l'outil et, partant, une meilleure acceptabilité s'il a été élaboré par les services destinés à l'utiliser. L'acceptabilité n'est pas négligeable car le rejet d'un outil voue le système à l'échec⁴⁵.

Si de nombreuses villes américaines ont eu massivement recours à des outils externes proposés par des entreprises commerciales, des outils internes ont également été développés, par exemple par la ville de New York⁴⁶ ou encore Chicago. Ces villes ont ainsi pu cumuler différents outils et services, internes et externes. Les outils internes sont souvent difficiles à connaître car ils ne font le plus souvent pas l'objet de brevets et tout dépendra de la communication qui pourra en être faite par les services de police. Seront présentés succinctement un outil utilisé à Chicago (a), les outils déployés en France (b) et ailleurs en Europe (c).

a) La Strategic Subject List (SSL) à Chicago (ciblage de personnes)

Certains départements de police des villes américaines ont développé leurs propres outils. L'un des plus connus aux Etats-Unis est l'outil prédictif utilisé à Chicago pour permettre d'identifier des auteurs potentiels d'infractions sur le fondement de données pénales passées et l'analyse de réseaux sociaux entre les individus. Cet outil, dénommé *Strategic Subject List* (SSL), est plus communément connu sous le nom de « Chicago Heat List ». Ce *Strategic*

⁴⁵ Voir par exemple les outils utilisés en matière de justice prédictive. Les travaux de la sociologue Angèle Christin présentés en 2017 lors du colloque organisé à l'INHESJ sur « Sécurité et justice : le défi des algorithmes » : <https://inhesj.fr/evenements/videotheque/securite-et-justice>.

⁴⁶ Par exemple le logiciel Patternizr élaboré par la police de New York depuis deux ans automatise le travail traditionnel de terrain afin de trouver des patterns et tendances des crimes sans considération de lieu, race ou genre. Ce logiciel regroupe des algorithmes de machine learning utilisant des données du passé sur dix années concernant la distance, l'heure, les méthodes et le type de forces utilisés pour découvrir des modèles : <https://nypost.com/2019/03/10/nypd-uses-new-tool-to-find-crime-patterns-officials>. Il faut noter que le New York Civil Liberties Union n'a pas réussi à avoir accès à Patternizr mais a demandé à NYPD d'être transparent sur son utilisation.

Subject Algorithm a été créé par l'Institut de Technologie de l'Illinois (Illinois Institute of Technology) et financé par le Département Fédéral de la Justice (DoJ - Bureau de l'assistance de Justice) à partir de données anonymisées de personnes arrêtées entre le 1^{er} août 2012 et le 31 juillet 2016. Il constitue un score d'étude de risque qui reflète la probabilité qu'un individu soit engagé dans une fusillade, comme victime ou auteur. Les scores sont calculés et placés sur une échelle de 0 (très faible risque) à 500 (risque extrêmement élevé). La ville de Chicago met ses données en libre accès sur le portail *open data* de la ville⁴⁷. Les individus présentant des antécédents pénaux sont classés selon huit attributs, à l'exclusion de la race et du sexe : le nombre de fois où la personne a été victime d'une fusillade ; l'âge lors de la dernière arrestation ; le nombre de fois où la personne a été victime d'une agression ; le nombre de précédentes arrestations pour violence ; l'affiliation à des gangs ; le nombre de précédentes arrestations liées aux narcotiques ; la tendance à mener des activités criminelles récentes ; le nombre de précédentes arrestations pour usage illégal d'armes. Ces attributs sont mis à jour sur le fondement d'un procédé permanent. Les jeux de données sont aussi révisés.

Selon les médias⁴⁸, cette liste basée sur le profil des individus serait la plus importante utilisée aux Etats-Unis. Il faut rappeler que la plupart des outils mis en œuvre par les polices des villes américaines, en particulier ceux proposés par les entreprises commerciales, sont principalement basées sur des lieux pour déterminer où et quand le crime est susceptible de survenir. À l'inverse, la liste de Chicago tend à anticiper les personnes susceptibles d'être engagées dans la criminalité. Cette liste est aussi controversée car aujourd'hui plus de 400000 citoyens de Chicago sont sur cette liste dont 1400 jeunes hommes considérés à haut risque. Une des interrogations porte sur les critères d'identification des suspects potentiels. Au demeurant, cet outil n'a pas non plus démontré son efficacité car la violence criminelle n'a fait que progresser depuis 2016⁴⁹. En particulier, il semble que ne soit pas garantie une stricte distinction entre la liste des personnes risquant d'être victimes de criminalité et celle portant sur les auteurs. En outre, la protection des victimes potentielles n'est pas aussi importante que la surveillance des auteurs⁵⁰.

L'objectif de la liste est aussi de prévenir les risques par l'organisation de visites des personnes ciblées, auteurs ou victimes potentielles d'infractions. Ces visites sont organisées par un représentant du département de la police de Chicago, avec un travailleur social et un représentant de la communauté. Le but est d'informer les personnes ciblées du fait qu'elles

⁴⁷ <https://data.cityofchicago.org/Public-Safety/Strategic-Subject-List/4aki-r3np>.

⁴⁸ <https://medium.com/equal-future/how-strategic-is-chicagos-strategic-subjects-list-upturn-investigates-9e5b4b235a7c>.

⁴⁹ Voir le rapport de RAND Corporation : WALTER L. PERRY ET AL., PREDICTIVE POLICING: THE ROLE OF CRIME FORECASTING IN LAW ENFORCEMENT OPERATIONS (2013), https://www.rand.org/pubs/research_reports/RR233.html (last visited Nov 29, 2018) (RAND Corporations).

⁵⁰ Andrew Guthrie Ferguson, p. 39.

sont surveillées et de les inciter à rester en dehors de toute activité criminelle.

Les analyses montrent que l'efficacité du procédé est limitée si cette démarche n'est pas accompagnée d'autres mesures sociales et économiques. En comparaison, les résultats obtenus par la ville de la Nouvelle Orléans (Louisiane) sont bien meilleurs grâce au programme global *NOLA for life*. En effet, d'autres dispositifs d'amélioration des services sociaux ont été mis en œuvre, ainsi que des accompagnements individuels (mentorat) ou encore de plus grandes aides financières aux écoles⁵¹. Vingt-neuf programmes ont accompagné les outils de ciblage de la population à risque et l'ensemble de ces mesures a permis de faire baisser la criminalité. Il semble toutefois que ce projet ait été élaboré avec la participation de Palantir et ne soit donc pas un outil interne propre à la police de Nouvelle-Orléans⁵².

b) PredVol et Paved en France (ciblage de lieux)

Quant à la France, un premier outil interne *Predvol* fut expérimenté dans l'Oise pour anticiper la délinquance commise sur les véhicules dans une zone géographique associée à un ilot (IRIS) au sens de l'INSEE. Le logiciel fournit une représentation cartographique du département et pointe des zones à risque selon une prévision à la semaine, à partir d'un historique qui remonte à 5 ans. La personne consultant le logiciel a toujours accès à l'évolution dans le temps. En outre, le logiciel intègre une évaluation entre ce qui a été prédit et les infractions constatées, permettant d'en mesurer l'efficacité. Les résultats montrent surtout que « l'outil reproduit ce que l'on connaît déjà »⁵³. Lancée opérationnellement en septembre 2016, la police et la gendarmerie ont fait un premier bilan de l'expérimentation fin février 2017. Le constat est que les périmètres sont parfois trop larges et surtout que les calculs ont tendance « à faire ressortir toujours les mêmes spots, les mêmes points chauds aux mêmes endroits ». Autrement dit, il n'apporte pas d'information supplémentaire aux opérationnels. En outre, l'insuffisance des données rend l'information peu réactive.

⁵¹ Cinq composants clés du projet sont : arrêter les fusillades, investir dans la prévention, promouvoir le travail et les opportunités, reconstruire les quartiers, améliorer le département de police de la Nouvelle Orléans (NOPD).

⁵² L'outil technologique semble avoir été élaboré par Palantir, sur le modèle du ciblage des membres d'Al-Qaeda en Syrie, sans que le conseil municipal de la ville ait été informé, du fait de la qualification philanthropique de la participation de Palantir dans le programme *Nola for life* : <https://www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd>. Cela montrerait que Palantir n'utilise pas seulement le Crime Forecasting System qui cible des lieux mais aussi des outils de ciblage des individus, comme l'outil Gotham à New York, en coopération avec les départements de police concernés mais sans que l'on sache officiellement qu'elle est la place prise par ces outils et le degré de participation des services de police. Les contrats ainsi conclus sont secrets et les procédés manquent de transparence, y compris à l'égard des membres des conseils municipaux des villes concernées.

⁵³ <http://www.internetactu.net/2017/07/26/ou-en-est-la-police-predictive>. Compte rendu des propos de Monsieur Yves Gallot, commissaire divisionnaire, chef de la Division des Systèmes d'Information Opérationnelle, Direction Centrale de la Sécurité Publique, tenus lors d'un colloque organisé à l'INHESJ sur « Sécurité et justice : le défi des algorithmes » : <https://inhesj.fr/evenements/videotheque/securite-et-justice>.

Compte tenu de ces faiblesses, l'outil fut finalement abandonné⁵⁴ mais a servi d'expérimentation permettant à ETALAB, en charge du programme, de tester et comparer les performances de très nombreux algorithmes d'apprentissage. Il est finalement ressorti qu'une simple prévision par moyenne mobile sur les neufs derniers mois proposait une prévision aussi bonne que celle obtenue par des algorithmes d'apprentissage (boosting, random forest....) plus sophistiqués.

Parallèlement, Paved est un autre outil algorithme d'aide à l'analyse décisionnelle dans la lutte contre la délinquance, développé par la gendarmerie nationale⁵⁵. En terme opérationnel, l'objectif est d'occuper le terrain pour dissuader toute commission d'infraction. Le choix a été fait de garder la main sur un logiciel créé en interne, avec des ingénieurs, programmeurs et *data scientists* gendarmes. Ce logiciel d'analyse prédictive de la délinquance (qui n'utilise aucune donnée personnelle) a été testé dans onze départements français avant d'être élargi à l'ensemble du territoire français en septembre 2018. Il permet d'accéder à une carte de France avec des zones de chaleur correspondant à des cambriolages ou atteintes aux véhicules⁵⁶. À l'heure actuelle⁵⁷, ces outils ne fonctionnent qu'avec des données de la police et de la gendarmerie, combinées avec des données socio-économiques (type INSEE). Les données externes conjoncturelles, comme la météo et l'agenda événementiel, ne sont pas intégrées. N'est pas non plus considérée l'organisation de l'activité des gendarmes (planning, maladie) ni le circuit des patrouilles.

Dans les dispositifs français, les données traitées proviennent essentiellement des logiciels utilisés dans le traitement des procédures à caractère judiciaire (Logiciel de Rédaction des Procédures Police Nationale -LRPPN- ou Gendarmerie Nationale -LRPGN). Ces données font l'objet avant traitement d'une « anonymisation », condition préalable à leur exploitation à des fins opérationnelles. L'anonymisation des données garantit la protection de la vie privée et fait sortir les données du champ d'application de la loi du 6 janvier 1978 sur la protection des données personnelles. Ce contrôle interne est indispensable et protecteur. Il limite, par ailleurs, les risques en cas d'intrusion des systèmes ou de piratage des algorithmes. En conséquence, les risques juridiques évoqués dans ce rapport relatifs à la protection des données personnelles ne concernent pas le système Paved. Néanmoins, les services de

⁵⁴ Voir le bilan de l'expérimentation fait par le commissaire divisionnaire Yves Gallot le 24 mai 2018.

⁵⁵ Des entretiens téléphoniques ont été menés avec le Colonel Laurent Collorig et le Colonel Patrick Perrot. Le Colonel Laurent Collorig a succédé à la tête de la division du renseignement au Colonel Patrick Perrot, Docteur en informatique, Commandant le groupement de gendarmerie départementale de la Haute-Marne (Chaumont). Il est à l'origine du projet PAVED, lorsqu'il était chef de la division du renseignement au SCRC (Service central du renseignement criminel), et auteur de nombreux articles scientifiques (voir la bibliographie). Le Colonel Collorig est l'actuel responsable hiérarchique de l'équipe en charge des algorithmes du programme PAVED.

⁵⁶ <https://www.franceinter.fr/emissions/grand-angle/grand-angle-02-octobre-2018>.

⁵⁷ Propos de Monsieur Yves Gallot, *op. cit.*

renseignement utilisent les outils de la société Palantir⁵⁸, susceptibles de cibler des données personnelles. Dès lors, la question de la conformité à la réglementation des données personnelles se pose dans ce cas.

L'outil Paved génère des enjeux techniques quant à la fusion des données. Beaucoup de données des bases sont séquentielles et il faudrait pouvoir, par exemple, suivre les verbalisations d'un véhicule sur toute la France pour mieux comprendre ses déplacements, comme c'est le cas de véhicules volés. Enfin, reste l'enjeu du circuit de décision : comment ces données sont-elles utilisées et intégrées dans les décisions opérationnelles ? Il n'y a pas encore d'interface pour faire se croiser l'expérience et le ressenti des équipes, leur perception du territoire et les prédictions d'un logiciel. En tout état de cause, le décideur opérationnel organise ses patrouilles et conserve sa liberté d'action.

c) Les outils internes déployés ailleurs en Europe (ciblage de lieux)

Le comté du Kent au Royaume-Uni utilise PredPol mais semble l'avoir abandonné par manque de temps et formation des services de police mais aussi à défaut d'efficacité avérée⁵⁹.

D'autres systèmes spécifiques et internes aux services de police ont été développés en Europe. Par exemple, le système CAS (*Crime Anticipation System*) est mis en œuvre depuis 2015 aux Pays-Bas par la police d'Amsterdam et sera déployé dans quatre autres unités de police (Nord des Pays Bas, Nord de la Hollande, Est des Pays Bas et la Hague). Le système fournit des prédictions, à l'aide d'un réseau de neurones, sur la survenance la plus probable d'infractions pour les deux prochaines semaines concernant les cambriolages et désormais plus largement les atteintes à la propriété (vols dans la rue, vols de bicyclettes) qui se déplacent après deux semaines de renforcement de la surveillance dans une zone. Les

⁵⁸ Le logiciel de la société Palantir est utilisé par la DGSI en matière de terrorisme. Le choix a été fait de recourir à une solution commerciale extérieure en l'absence d'outil interne permettant d'apporter le même service. d'après Patrick Calvar, à l'époque directeur général de la sécurité intérieure, lors d'une audition devant la Commission de la défense nationale et des forces armées à l'assemblée nationale le mardi 10 mai 2016 : <http://www.assemblee-nationale.fr/14/cr-cdef/15-16/c1516047.asp>. On ne manque pas de données mais de systèmes pour les analyser. En situation d'urgence, le choix a donc été fait d'utiliser un outil déjà prêt. Cette solution devait être provisoire mais semble toujours utilisée aujourd'hui.

Des initiatives se mettent toutefois en place pour proposer des alternatives dans le but de construire une ère post-palantir : <http://forcesoperations.com/comment-le-cluster-data-intelligence-ouvre-le-re-post-palantir>. Dans une volonté de soutenir une souveraineté numérique, un consortium nommé Artemis a été constitué par la Direction générale de l'armement (DGA), afin de lancer son propre programme d'analyse des données massives, en travaillant avec de grands groupes de service informatique, comme Atos, CapGemini, Thales et SopraSteria : <https://www.defense.gouv.fr/dga/actualite/big-data-et-ia-la-dga-presente-le-projet-artemis>. Egalement, le Data Intelligence Cluster GICAT s'est constitué dans le but de proposer une offre alternative souveraine à l'offre de Palantir mais toujours par des sociétés privées : <http://www.dataintelligencecluster.com>. Lancé par le Groupement des industries de défense et de sécurité terrestres et aéroterrestres (GICAT), un cluster dédié à la data intelligence réunit depuis octobre 2018 une vingtaine de sociétés, afin de répondre aux besoins et contraintes de l'administration et des entreprises françaises. Alternative souveraine à l'offre de Palantir, ce cluster a pour vocation de booster les industriels français à l'export.

⁵⁹ <https://www.bbc.com/news/uk-england-kent-46345717>.

zones urbaines sont découpées en cellules de 125m x 125m et les informations sont fournies toutes les 4 heures. De nombreuses données sont agrégées à ces cellules. Le choix des données est basé sur les infractions précédentes et inclut les données de criminalité en général, les données statistiques nationales et les données géolocalisées. Les cellules s'allument dans l'une des trois couleurs (rouge, orange et jaune), les cellules rouges étant les hotspots où la probabilité que le cambriolage ait lieu soit la plus élevée. Le résultat de cette prédiction aura pour conséquence d'augmenter la surveillance dans les zones rouges pour prévenir la survenance des cambriolages. Cependant, le principal objectif de CAS est de rendre la police plus efficace en optimisant les lieux où elle doit concentrer ses efforts.

Egalement, Precobs (*Pre-Crime Observation System*) est un outil développé par le German Institute for Pattern-based Prediction Technique (Institut für Musterbasierte Prognosetechnik (IfmPt)) (www.ifmpt.de) en 2011. Il est déployé en Allemagne⁶⁰ (Länder de Bavière et du Bade-Wurtemberg, Berlin et Munich), ainsi qu'à Zurich (Suisse). Les infractions en principe visées par *Precobs* sont les vols et cambriolages le plus souvent des habitations. Notons toutefois que les infractions visées à Zurich sont aussi les cambriolages commerciaux (bureaux, centres commerciaux et restaurants), à la demande des services de police de la ville. Ce programme identifie les zones de répétition des infractions, ce qui constitue la base automatique de la criminalité. Les matériels utilisés par la police incluent les données sur les lieux et dates pour planifier des contre-mesures. Selon la façon dont *Precobs* traite les dernières données, sont générées des prédictions que les forces de police peuvent utiliser pour opérer et prévenir la commission d'infraction.

À Milan, en Italie, le logiciel *Keycrime* a été développé avec l'aide de la police par une entreprise privée, afin de prédire la survenance des cambriolages⁶¹. Il intègre aujourd'hui également les vols des banques. Ce logiciel cible aussi des zones et non des personnes. La preuve scientifique de l'efficacité de cet outil a toutefois été contestée⁶².

B. Catégorisation des outils de police prédictive

Quelques remarques générales doivent être faites sur les conséquences qu'emportent cette classification et leur impact sur la réglementation qui sera vue dans la troisième partie, spécialement dans le choix d'un outil interne ou externe (1) ou encore dans le ciblage de lieux ou de personnes (2).

⁶⁰

<https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/predictive.policing.pdf>.

⁶¹ <https://www.emmeviemme.com>.

⁶² GIOVANNI MASTROBUONI, CRIME IS TERRIBLY REVEALING: INFORMATION TECHNOLOGY AND POLICE PRODUCTIVITY (2017), <https://papers.ssrn.com/abstract=2989914> (last visited Dec 12, 2018).

1. Outils internes des services de police et gendarmerie versus outils externes proposés par des entreprises privées

Le choix par des services de police d'utiliser une application commerciale ou de créer leur propre outil n'est pas indifférent. En utilisant un outil commercial, les polices sont dépendantes du marché. Elles achètent alors une sorte de 'black box' dont on ne sait pas exactement ce qui se passe avec les données collectées par les polices ni comment les algorithmes marchent et donc quels aspects sont pris en compte par les prédictions. Cette solution est rapide mais peu transparente et répond souvent à la pression de démarrer une méthode attendue, davantage pour des raisons politiques que de bonne pratique des activités de sécurité et lutte contre la criminalité. Quand les polices développent leurs propres outils, cela nécessite plus de temps mais les forces de police et analystes sont mieux en capacité de comprendre la méthodologie utilisée et les enseignements qu'ils pourront tirer de son utilisation, en fonction des résultats issus des données. À l'heure où la pression sociale pour plus de transparence algorithmique est forte, il est difficilement concevable que les outils de police prédictive soient obscurs et incompris des services qui les utilisent, *a fortiori* des citoyens. Plus encore, la loi impose à l'administration française des obligations pour les traitements automatisés de données personnelles⁶³.

Les enjeux dans le choix du type d'outils sont lourds de conséquences :

- enjeu de souveraineté nationale (beaucoup d'outils sont d'origine américaine ou israélienne) ;
- indépendance, autonomie de la mission de service public de la police en l'absence de contrôle sur les entreprises privées ;
- maintien dans le temps et pérennité des solutions élaborées en interne, alors que les acteurs privés peuvent changer à tout moment leur solution ou cesser de la fournir ou ne plus l'améliorer et la laisser périr ;
- coût élevé des solutions proposées par les entreprises privées, non seulement en coût d'acquisition mais aussi en coût annuel de maintenance, contraire au souci de bonne gestion des fonds publics, alors qu'une solution en interne peut s'avérer moins chère ;
- partage risqué des données de la criminalité en France avec des acteurs privés qui plus est étrangers ;
- réappropriation des données par les acteurs privés via les droits de propriété intellectuelle sur les outils (ex. brevet) et les secrets d'affaire, laissant flotter l'ambiguïté sur la « propriété » des données, ainsi que la maîtrise d'un accès matériel ;

⁶³ Voir la troisième partie sur les contraintes légales.

- appropriation de l'information dans le cadre d'une mission régalienne de service public du maintien de l'ordre ;
- appropriation de l'information au détriment de la transparence et du rendre compte (*accountability*) à la population sur le sujet sensible de la prévention de la criminalité et de la surveillance de zones voire personnes.

Titularité des données utilisées dans les dispositifs de police prédictive

Il convient par ailleurs de se demander aussi d'où proviennent les données utilisées dans un système de police prédictive. Il n'y a en principe aucune difficulté lorsque l'outil est élaboré par les forces de police elles-mêmes qui vont alors logiquement utiliser leurs propres données. Le risque est alors plutôt inversé, en ce que les données de la police ne doivent pas pouvoir être utilisées et récupérées par ailleurs par des entreprises commerciales, sauf lors de la mise à disposition explicite de ces données (données ouvertes par exemple).

Mais dans l'hypothèse où d'autres données seraient utilisées ou dans celle, plus probable, où l'outil utilisé est externe et élaboré par une entreprise commerciale, l'origine des données et leur titularité risquent d'être plus incertaines. Il convient alors de s'assurer de détenir l'autorisation, le plus souvent obtenue par contrat, d'utiliser de telles données.

2. Ciblage des lieux versus ciblage des personnes

La question de la cible visée par l'outil d'analyse prédictive est essentielle. Comme vu précédemment, la majorité des outils utilisés aux Etats-Unis et *a fortiori* en Europe vise à cartographier la criminalité en cherchant la prédiction des lieux et temporalité de survenance des infractions comme le font les systèmes *Paved* et *PredPol* par exemple. D'autres outils, comme *Beware* ou *Strategic Subject List* (SSL), peuvent viser des personnes, victimes ou auteurs d'infractions.

Selon la réponse donnée à la question du ciblage, les outils techniques utilisés seront différents et les données collectées seront également de nature différente. L'enjeu majeur sur ce point est de savoir si des données personnelles sont collectées, qu'elles soient directement ou indirectement identifiantes, puisque la définition des données personnelles en France et en Europe est particulièrement large⁶⁴. En effet, même si les données collectées ne sont pas des données personnelles *per se*, elles peuvent toutefois permettre une identification suite au croisement d'une grande quantité d'informations. Ainsi, par exemple, même dans l'hypothèse où la cible visée n'est pas une personne mais un lieu, si le maillage est très fin, on peut aisément parvenir à viser un quartier ou une rue avec peu d'habitants et réussir à les identifier individuellement. Dans ce cas, la réglementation sur la protection des

⁶⁴ Cf. réglementation partie 3 du rapport.

données personnelles aura vocation à jouer.

La plupart des dispositions de police prédictive mis en œuvre en Europe concerne la prédiction des lieux et temps de survenance des infractions, par un dispositif de *hotspotting*, portant le plus souvent sur des infractions de cambriolages et vols. Les outils de police prédictive visant les personnes, victimes ou auteurs d'infractions, sont plus difficiles à déployer en Europe qu'aux Etats-Unis, compte tenu de la différence de réglementation et d'un niveau de protection de la vie privée et des données personnelles plus élevé en Europe.

Champ d'application matériel et spatial de l'outil en cas de ciblage d'un lieu

Dans l'hypothèse où la cible visée est un lieu et non une personne, le champ d'application de l'outil de police prédictive est à déterminer, tant d'un point de vue matériel que spatial. Il faut ainsi préciser les infractions concernées par l'outil, l'étendue géographique totale de la zone à surveiller, ainsi que la finesse du maillage appliquée au découpage de la zone.

D'un point de vue technique, la question de l'étendue est déterminante pour s'assurer que la quantité de données collectées soit suffisante pour garantir une efficacité de l'outil. Par exemple, il semble que l'expérimentation PredVol qui portait sur les vols de voiture dans l'Oise collectées pendant 9 mois n'ait pas généré suffisamment de données pour permettre une prédiction satisfaisante⁶⁵. À l'inverse, un champ géographique ou un maillage trop large peuvent générer une trop grande quantité de données et l'outil peut alors ne pas être suffisamment précis pour permettre une surveillance opérationnelle.

D'un point de vue juridique, il est essentiel de savoir quelle infraction est visée puisque la qualification de l'infraction détermine le régime juridique et en particulier le cadre d'action des forces de police. Il convient en effet de se demander s'il s'agit simplement de prévenir une infraction ou bien de la découvrir, ce qui distingue les actions de police administrative et police judiciaire. Par ailleurs, de la qualification de l'infraction découle le régime juridique en droit pénal, ainsi que les actes autorisés ou non à accomplir par les autorités de police, tels que prévus par le code de procédure pénale.

Par ailleurs, d'un point de vue organisationnel, la détermination de l'infraction visée permet de savoir quels services de la police et gendarmerie sont susceptibles d'être compétents pour les doter de l'outil. L'étendue spatiale de l'outil a aussi une incidence sur la répartition territoriale des compétences entre la police et la gendarmerie. S'il s'agit de se concentrer sur une criminalité urbaine, tel le trafic de drogue, l'outil sera plus certainement utilisé par les services de police. Si la criminalité visée est large et non territorialisée, tels les cambriolages ou vols de voiture, visés par exemple par l'outil *Paved*, police et gendarmerie ont vocation à utiliser le même outil.

⁶⁵ <https://agd.data.gouv.fr/2018/01/12/predire-les-vols-de-voitures>.

Enfin, la question du maillage est délicate et n'est pas neutre pour des raisons socio-économiques. Les territoires urbains et ruraux sont souvent organisés en considération de facteurs sociaux, ethniques et économiques. La surveillance systématique de zones spécifiques peut conduire à cibler involontairement certaines catégories de population et conduire à des discriminations indirectes. Cela revient en effet à soupçonner plus souvent certaines populations de vouloir commettre ou planifier de commettre des infractions, par exemple de vols de voitures. À l'inverse, surveiller des zones que l'on va considérer par preuve statistique comme étant « à risque », par exemple de cambriolage, peut revenir à protéger davantage les quartiers privilégiés. Ces facteurs socio-économiques voire éthiques posent alors une question politique tenant à la répartition des moyens de la police dont il faut avoir conscience dans l'utilisation des outils.

La discrimination est d'autant plus délicate à percevoir qu'elle résulte parfois de l'outil même de police prédictive, entraîné à partir de données biaisées. Même si par la suite l'outil n'est pas alimenté par des données biaisées, les conditions d'apprentissage en amont peuvent s'avérer déterminantes pour créer une discrimination et conduire systématiquement à identifier toujours le même type de zones ou quartiers comme étant à risque, eu égard aux facteurs sociaux, économiques et ethniques de ses habitants et à la répétition passée des infractions intégrées dans les données historiques du système.

Champ d'application matériel et spatial de l'outil en cas de ciblage d'une personne

Outre le respect de la réglementation des données personnelles, il convient aussi de respecter les droits fondamentaux de la personne concernée, en particulier le principe d'égalité et la non-discrimination⁶⁶.

Le ciblage d'une personne pose des questions différentes pour déterminer en amont le champ d'application matériel et spatial de l'outil, compte tenu de l'imprévisibilité de son action et de sa mobilité. Ainsi, si l'outil de police prédictive a pour objectif de viser une personne, identifiée comme « à risque », il ne pourra probablement pas être déterminé en amont quelle infraction on cherche à prévenir ou poursuivre. Le danger est alors que l'on surveille une personne pour tous ses faits et gestes, y compris pour des actes illégaux pénalement qualifiables que l'on ne soupçonnait pas avant la mise en œuvre de la surveillance. Il faut en outre s'interroger sur la loyauté des modalités de collecte de la preuve. Par ailleurs, la surveillance d'une personne conduit indirectement à surveiller partiellement son entourage, ce qui peut permettre de mettre en lumière l'existence de complices ou d'auteurs d'autres infractions qui n'étaient pas initialement visés. Il convient donc d'être prudent dans les modalités de mise en œuvre de l'outil et les conséquences qui pourront en être tirées, pour

⁶⁶ Voir partie 3 sur la réglementation.

s'assurer de respecter le cadre d'action de la police administrative ou judiciaire. Ces hypothèses visent toutefois plus probablement des modes opératoires de police judiciaire qui supposent l'existence en amont de soupçons sur le fait que la personne ait commis ou aurait voulu commettre une infraction.

Des actions plus étendues des forces de police requièrent le plus souvent une autorisation judiciaire. Or, il peut parfois être difficile de respecter ces conditions procédurales, eu égard au fait que la frontière ne soit pas toujours clairement établie entre ce qui entre dans le ciblage déterminé *a priori* de l'outil de police prédictive et ce qui est découvert *a posteriori* et n'y entre pas. Cette difficulté peut se trouver renforcée en présence d'un ciblage de personne qui ne serait pas lié à un seul type d'infraction bien identifié plutôt que d'un ciblage de lieu.

Par ailleurs, si le lieu n'est pas ciblé et que la personne surveillée se déplace, pourra se poser la question de la compétence géographique des forces de l'ordre agissante, compte tenu de la spécialisation des services et autorités de police, non seulement à l'intérieur du territoire national, mais aussi éventuellement en dehors si la personne quitte le territoire français. Cette hypothèse empêchera probablement de fait la poursuite d'une surveillance par un outil de police prédictive mais peut donner lieu à un échange de données avec des polices étrangères dont certaines émanent de l'outil de police prédictive. Une telle situation est susceptible de tomber sous le coup des dispositions de la directive 2016/679/UE relative à l'échange de données personnelles à des fins de police et justice pour ce qui est du territoire de l'Union européenne mais aussi au-delà puisque ce texte prévoit aussi des flux de données transfrontaliers hors UE, y compris avec des organisations internationales. Ces dispositions ont été transposées en droit français au sein de loi informatique et libertés dans sa dernière révision par la loi n° 2018-493 du 20 juin 2018. Il convient donc d'être prudent sur la façon d'utiliser un outil visant un individu au ciblage géographique large.

Résumé des différentes catégories d'outils

Ciblage de zones		Ciblage de personnes	
Outils internes aux forces de l'ordre (outils publics) Ex. Paved	Outils externes aux forces de l'ordre (outils privés) Ex. PredPol	Outils internes aux forces de l'ordre (outils publics) Ex. SSL (Chicago)	Outils externes aux forces de l'ordre (outils privés) Ex. Beware
		Personne : victime potentielle d'infraction Ex. SSL (Chicago)	Personne : auteur potentiel d'infraction Ex. SSL (Chicago)

III. Enjeux juridiques et éthiques des outils de police prédictive

La loi encadre partiellement les activités de police prédictive et est en mesure de répondre à un certain nombre de risques. Il convient évidemment de s'y conformer (A). Cependant, l'ensemble des risques identifiés à ce jour n'a pas encore été pris en compte par le législateur, aussi est-il souhaitable de poser de surcroît des limites éthiques à l'usage d'algorithmes prédictifs, conformément aux préconisations faites dans le rapport *AI For Humanity* par le député et mathématicien Cédric Villani (B).

A. Principales réglementations applicables aux outils de police prédictive

Les outils de police prédictive soulèvent un certain nombre d'enjeux juridiques, obligeant à se conformer à différents types de législations, tenant principalement au respect des droits fondamentaux (1), à la lutte contre la discrimination (2), à la protection des données personnelles (3), ainsi qu'à l'encadrement des missions de police et des règles de responsabilité (4).

1. Respect des droits fondamentaux

a) Respect des droits fondamentaux avant les actions de police prédictive

Les activités de police prédictive peuvent devenir attentatoires aux droits fondamentaux des personnes si des précautions ne sont pas prises et si l'usage qui en fait est abusif. Ainsi, si les outils de police prédictive sont utiles à la prévention des infractions et la gestion des forces de police, ils ne doivent par exemple pas être utilisés comme moyen suffisant pour interpellier les individus, sous peine de porter atteinte à leurs droits fondamentaux. Plusieurs droits fondamentaux sont susceptibles d'être violés en cas d'usage abusif, disproportionné et non justifié d'outils de police prédictive : le droit à l'intégrité physique et mentale de la personne (charte des droits fondamentaux de l'UE, art. 3), le droit à la liberté et à la sûreté (CDFUE, art. 6), le droit au respect de la vie privée et familiale, de son domicile et de ses communications (CDFUE, art. 7), le droit à la protection des données personnelles (CDFUE, art. 8), le droit à la liberté de réunion et d'association (CDFUE, art. 12), le droit à l'égalité (CDFUE, art. 20) et à la non-discrimination (CDFUE, art. 21).

Les risques d'atteinte à ces droits sont plus élevés en cas d'usage d'outils de police prédictive visant des personnes et non des lieux. Le choix majoritairement fait par les services de police en Europe, mais aussi aux Etats-Unis, de considérer des lieux limite ces risques d'atteinte aux droits des personnes. Il n'en demeure pas moins que l'identification d'une zone à risque ne donne naturellement pas plus de droits à la police qui, en principe, reste dans le cadre

d'actes de police administrative de prévention et maintien de l'ordre. Aux Etats-Unis, le 4^e amendement de la Constitution américaine relatif à la protection du domicile contre les intrusions de l'Etat et, par extension, de la vie privée prohibe les recherches et saisies déraisonnables en l'absence de mandat judiciaire fondé sur une cause probable. Il s'agit là de la principale protection contre les intrusions injustifiées des forces de police fédérales. En outre, chaque Etat dispose de sa propre constitution avec des dispositions comparables, applicables à l'égard des forces de police étatique et municipales.

b) Respect des droits fondamentaux après les actions de police prédictive

À ces risques d'atteinte aux droits fondamentaux situés en amont, peuvent s'ajouter des risques en aval. *A posteriori*, les services de police peuvent se trouver dans l'obligation de répondre des conditions d'utilisation des outils de police prédictive et des conclusions qui ont pu être tirées sur le terrain à l'encontre des individus au cas par cas. Ainsi par exemple, aux Etats-Unis, le département de la police de New York a dû répondre en justice de l'utilisation de l'outil *Palantir Gotham* et de ses caractéristiques techniques⁶⁷. Le manque d'information sur l'existence et l'utilisation de l'outil prédictif, ainsi que sur la nature des données concernées et les conditions de mise en œuvre des résultats algorithmiques du traitement automatisé ont été contestés en justice sur la base d'un défaut de transparence et l'impossibilité de faire respecter les droits de la défense⁶⁸. La violation du principe du **due process** (amendements 5 et 14) a été invoquée par les parties demandereses.

En comparaison, l'Union européenne prévoit aussi l'obligation de respecter le droit à un recours effectif et à accéder à un tribunal impartial (CDFUE, art. 47), ainsi que le droit au respect de la présomption d'innocence et des droits de la défense (CDFUE, art. 48). Dans les affaires précitées, le manque de transparence et connaissance des outils de police prédictive est susceptible de porter atteinte aux droits de la défense, si la partie désignée par le système comme étant potentiellement dangereuse ne connaît ni les données utilisées ni le fonctionnement de l'outil, ce qui l'empêche de remettre en cause la pertinence des résultats obtenus et donc d'exercer pleinement ses droits de la défense. Le principe de transparence algorithmique est indispensable non seulement pour le traitement des données personnelles mais aussi pour l'exercice des droits de la défense en justice.

En droit français, le I de l'article préliminaire du code de procédure pénale (CPP) rappelle le principe du contradictoire selon lequel « la procédure pénale doit être équitable et contradictoire et préserver l'équilibre des droits des parties ». Un principe d'égalité découle

⁶⁷ Supreme Court of the State of New York, Brennan Center for Justice at New York University School of Law vs New York City Police Department (NYPD), 27 Dec. 2017. Le juge a fait droit à la demande d'accès au système de *Palantir Gotham* utilisé par la NYPD (*New York Police Department*).

⁶⁸ *State v. Loomis*, Supreme Court of Wisconsin, July 13th, 2016, 2016 WI 68.

de l'affirmation selon laquelle « les personnes se trouvant dans des conditions semblables et poursuivies pour les mêmes infractions doivent être jugées selon les mêmes règles. En outre, selon le II de l'article préliminaire du CPP, « l'autorité judiciaire veille à l'information et à la garantie des droits des victimes au cours de toute procédure pénale ». Au demeurant, « toute personne suspectée ou poursuivie est présumée innocente tant que sa culpabilité n'a pas été établie. Les atteintes à sa présomption d'innocence sont prévenues, réparées et réprimées dans les conditions prévues par la loi » (point III). Cette personne « a le droit d'être informée des charges retenues contre elle et d'être assistée d'un défenseur ». Ces garanties procédurales fondamentales correspondent au *due process* américain. Il sera cependant plus difficile de respecter ces exigences si les outils de police prédictive utilisés reposent sur des données et modes de raisonnement non connus par les services de police. Le recours à des outils internes plutôt qu'externes limite ce risque.

En résumé, l'utilisation d'outils de police prédictive désignant des individus ou zones à risques doit pouvoir s'accompagner d'explications sur l'outil et les données utilisées pour permettre aux individus de comprendre ces outils mais également les conséquences négatives qui pourraient être tirées contre eux, afin de pouvoir contester de tels résultats, le cas échéant. À défaut, non seulement il y a un risque d'atteinte à la réglementation des données personnelles mais surtout, plus généralement, à la protection des droits fondamentaux et garanties procédurales du droit pénal. L'exemple des Etats-Unis montre ainsi que les personnes concernées par les résultats d'outils de police prédictive ont tenté de remettre en cause la constitutionnalité de leur utilisation. Notons que les risques d'atteinte aux droits fondamentaux situés en aval ne sont pas moindres que ceux portant sur l'utilisation d'outils qui se contentent de cibler des lieux. En effet, les conséquences négatives pour les personnes peuvent être fortes lorsque leur domicile est considéré comme se trouvant dans une zone à risque. Dès lors, le respect des droits de la défense, en particulier le principe du contradictoire et le droit de contester la décision, suppose d'assurer une transparence et « explicabilité » des modalités d'utilisation de l'outil ayant guidé la décision, quel qu'il soit.

2. Egalité de traitement et discrimination

a) Risques de biais et discrimination des outils de police prédictive

Aux Etats-Unis, les médias⁶⁹, académiques⁷⁰ et organisations de protection et défense des droits civils⁷¹ se sont fait l'écho de problèmes de biais et discriminations des outils de police ou justice prédictive. La violation du 14^e amendement qui pose le principe d'une protection

⁶⁹ Par ex. : <https://bdtechtalks.com/2018/03/26/racist-sexist-ai-deep-learning-algorithms>.

⁷⁰ Voir notamment les travaux menés par AINow Institut : <https://ainowinstitute.org>.

⁷¹ Voir par exemple les actions d'ACLU : <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/secret-algorithms-are-deciding-criminal-trials-and>.

égale de tous les citoyens par la loi a pu être invoquée.

En droit français, l'article 95 alinéa 3 de la loi n° 78-17 du 6 janvier 1978 modifiée par la loi du 20 juin 2018 sur la protection des données personnelles interdit tout profilage entraînant une discrimination à l'égard des personnes physiques sur la base de données sensibles. De plus, l'article 225-1 du code pénal pose plusieurs critères de discrimination. L'alinéa 1^{er} précise ainsi que : « Constitue une discrimination toute distinction opérée entre les personnes physiques sur le fondement de leur origine, de leur sexe, de leur situation de famille, de leur grossesse, de leur apparence physique, de la particulière vulnérabilité résultant de leur situation économique, apparente ou connue de son auteur, de leur patronyme, de leur lieu de résidence, de leur état de santé, de leur perte d'autonomie, de leur handicap, de leurs caractéristiques génétiques, de leurs mœurs, de leur orientation sexuelle, de leur identité de genre, de leur âge, de leurs opinions politiques, de leurs activités syndicales, de leur capacité à s'exprimer dans une langue autre que le français, de leur appartenance ou de leur non-appartenance, vraie ou supposée, à une ethnie, une Nation, une prétendue race ou une religion déterminée ». L'alinéa 2 vise la discrimination opérée entre les personnes morales. L'article 225-3 du code pénal exclut toutefois certaines discriminations du champ d'application des discriminations en matière pénale. Enfin, la discrimination définie aux articles 225-1 à 225-1-2, commise à l'égard d'une personne physique ou morale, est punie de trois ans d'emprisonnement et de 45 000 euros d'amende (C. pénal, art. 225-2).

Par ailleurs, l'article 226-19 du Code pénal protège particulièrement les données personnelles sensibles, en raison des risques de discrimination qui y sont attachées. Il ajoute que : « le fait, hors les cas prévus par la loi, de mettre ou de conserver en mémoire informatisée, sans le consentement exprès de l'intéressé, des données à caractère personnel qui, directement ou indirectement, font apparaître les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses, ou les appartenances syndicales des personnes, ou qui sont relatives à la santé ou à l'orientation sexuelle ou à l'identité de genre de celles-ci, est puni de cinq ans d'emprisonnement et de 300 000 euros d'amende ».

b) Discrimination et exercice de l'autorité publique

L'article 432-7 du Code pénal dispose que « La discrimination définie aux articles 225-1 et 225-1-1, commise à l'égard d'une personne physique ou morale par une personne dépositaire de l'autorité publique ou chargée d'une mission de service public, dans l'exercice ou à l'occasion de l'exercice de ses fonctions ou de sa mission, est punie de cinq ans d'emprisonnement et de 75 000 euros d'amende lorsqu'elle consiste :

1° A refuser le bénéfice d'un droit accordé par la loi ;

2° A entraver l'exercice normal d'une activité économique quelconque.

Dès lors, les officiers de police et gendarmerie doivent être particulièrement vigilants dans

l'utilisation des outils de police prédictive pour ne pas être conduits à refuser le bénéfice d'un droit, comme le droit à l'égalité ou le droit à la présomption d'innocence, en raison soit d'une mauvaise utilisation de ces outils, soit d'une mauvaise qualité des données, soit des biais introduits dans les systèmes, conduisant à discriminer les personnes ciblées.

Discrimination directe, discrimination indirecte, discrimination de groupe

Par ailleurs, l'article 1^{er} de la loi n° 2008-496 du 27 mai 2008 portant diverses dispositions d'adaptation au droit communautaire dans le domaine de la lutte contre les discriminations prévoit que : « Constitue **une discrimination directe** la situation dans laquelle, sur le fondement de son origine, de son sexe, de sa situation de famille, de sa grossesse, de son apparence physique, de la particulière vulnérabilité résultant de sa situation économique, apparente ou connue de son auteur, de son patronyme, de son lieu de résidence ou de sa domiciliation bancaire, de son état de santé, de sa perte d'autonomie, de son handicap, de ses caractéristiques génétiques, de ses mœurs, de son orientation sexuelle, de son identité de genre, de son âge, de ses opinions politiques, de ses activités syndicales, de sa capacité à s'exprimer dans une langue autre que le français, de son appartenance ou de sa non-appartenance, vraie ou supposée, à une ethnie, une nation, une prétendue race ou une religion déterminée, une personne est traitée de manière moins favorable qu'une autre ne l'est, ne l'a été ou ne l'aura été dans une situation comparable ».

L'alinéa 2 ajoute que : « Constitue une **discrimination indirecte** une disposition, un critère ou une pratique neutre en apparence, mais susceptible d'entraîner, pour l'un des motifs mentionnés au premier alinéa, un désavantage particulier pour des personnes par rapport à d'autres personnes, à moins que cette disposition, ce critère ou cette pratique ne soit objectivement justifié par un but légitime et que les moyens pour réaliser ce but ne soient nécessaires et appropriés ».

Discrimination systémique et intentionnalité en matière pénale

Une discrimination systémique est un processus qui met en jeu un système d'acteurs dans lequel personne ne manifeste pas directement d'intention discriminatoire, mais dont le résultat sera de produire une situation de discrimination⁷². Le concept de discrimination systémique découle de la reconnaissance de l'existence de déséquilibres socioéconomiques ou d'inégalités sociales historiquement constitués. Les discriminations systémiques sont donc le résultat de processus qui produisent et reproduisent les places sociales inégalitaires en fonction de l'appartenance à une « classe », une « race » ou un « sexe », cette appartenance pouvant être réelle ou supposée. Elles ne sont pas intentionnelles et

⁷² Dictionnaire des dominations, Collectif Manouchian, 2012.

proviennent de la somme de plusieurs représentations qui, cumulées, forment un contexte discriminant.

Intentionnalité et preuve de la discrimination : testing

La lutte contre la discrimination est confrontée à deux principales difficultés. La première tient au fait qu'elle doit s'accompagner d'une intentionnalité pour être sanctionnée. La seconde est qu'il faut parvenir à la prouver.

En principe, l'élément intentionnel, tout comme l'élément matériel, est nécessaire en droit pénal à la constitution d'une infraction. Il existe toutefois des exceptions pour certaines infractions comme la contrefaçon par exemple. En l'absence d'élément intentionnel, la discrimination ne peut être sanctionnée juridiquement. La discrimination indirecte ne suppose pas d'intention mais sera particulièrement difficile à prouver.

La deuxième difficulté tenant à la preuve fait l'objet de mesures légales spéciales destinées à la faciliter. En ce sens, l'article 225-3-1 code pénal autorise le *testing*. Il dispose ainsi que : « Les délits prévus par la présente section sont constitués même s'ils sont commis à l'encontre d'une ou plusieurs personnes ayant sollicité l'un des biens, actes, services ou contrats mentionnés à l'article 225-2 dans le but de démontrer l'existence du comportement discriminatoire, dès lors que la preuve de ce comportement est établie ». Si cette mesure est utile dans certaines situations, elle n'est toutefois pas pertinente en matière de discrimination algorithmique.

En conséquence, si la discrimination algorithmique, comme toute discrimination, est interdite, elle sera difficile à prouver et donc à sanctionner.

3. Protection des données personnelles et transparence algorithmique

a) Outils ciblant des personnes et protection des données personnelles

Les outils de police prédictive visant des individus supposent un traitement massif de données, intégrant des données personnelles. En principe, les outils visant une zone n'intègrent pas de données personnelles, alors même que la majorité des outils utilisés en Europe et même aux Etats-Unis relèvent de cette seconde hypothèse. Les développements qui suivent n'ont donc en principe pas vocation à s'appliquer à ces outils. Il convient toutefois de rester vigilant dans l'usage qui est fait de ces outils, ainsi que des risques de ré-identification. Dans la première hypothèse concernant les outils visant des individus, le respect de la réglementation sur la protection des données personnelles est un enjeu majeur. Aux Etats-Unis⁷³, la protection des données personnelles par la *privacy law* est sectorielle, signifiant l'absence de réglementation générale de nature à couvrir tous les secteurs

⁷³ Voir D. Solove et P. Schwartz, *Information Privacy Law*, Wolters Kluwer, 6^e éd., 2018.

d'activité. Au demeurant, la protection de la vie privée relève des deux niveaux de compétence, fédérale et étatique. Les activités de police prédictive supposent l'utilisation de données personnelles par des acteurs publics. Le *Privacy Act* (1974) couvre l'utilisation par les seules agences fédérales d'enregistrement de données portant sur des personnes identifiables et incluses dans une base de données, ce qui exclut les traitements par les acteurs privés, ainsi que par les agences étatiques et municipales. Or, la police est gérée à plusieurs niveaux : fédéral, étatique, local (ou *county*, soit la partie administrative d'un Etat) et surtout municipal s'agissant des outils de police prédictive, mis en œuvre essentiellement par les villes. Dès lors, il convient de rechercher les textes applicables à la protection des données personnelles des Etats ou villes concernées, ce qui suppose de faire face à un millefeuille législatif. Ainsi par exemple, l'Etat de New York a adopté le *Personal Privacy Protection Law (Public Officers Law, Article 6-A, sections 91-99)* en 1984 qui vise les données personnelles traitées par les agences de sécurité étatiques. Par ailleurs, les dispositions de la constitution doivent être respectées, en particulier le 4^e amendement précité qui interdit les recherches et saisies déraisonnables, en l'absence d'un mandat judiciaire fondé sur une cause probable et protège la vie privée selon l'interprétation faite par la Cour Suprême.

La situation est cependant plus simple et plus protectrice en Europe, où le législateur, d'abord des Etats comme la France, puis de l'Union européenne, ont adopté des législations générales qui s'appliquent à tous les traitements de données personnelles, sans considération du secteur d'activité, à l'exception d'activités sensibles tenant par exemple à la sécurité nationale. En France, la loi n° 78-17 du 6 janvier 1978 dite informatique et libertés (LIL) a été récemment révisée par la loi n° 2018-493 du 20 juin 2018 pour intégrer les exigences du RGPD (règlement n° 2016/679/UE sur la protection des données personnelles en matière civile et commerciale), entré en application le 25 mai 2018. En outre, a aussi été transposée dans la loi du 20 juin 2018, la directive 2016/680/EU *relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales et à la libre circulation de ces données*. Ces textes ont été complétés par le décret n° 2018-687 du 1^{er} août 2018 et l'ordonnance n° 2018-1125 du 12 décembre 2018. En outre, le Conseil Constitutionnel a été saisi et a rendu une décision n° 2018-765 DC le 12 juin 2018 sur la loi du 20 juin 2018. Alors que le constat fait en France est que les fichiers de sécurité utilisant des données personnelles respectent la réglementation afférente, l'enjeu principal aujourd'hui est de les mettre en conformité avec les nouvelles règles européennes⁷⁴. Une

⁷⁴ Voir le rapport d'information *sur les fichiers mis à la disposition des forces de sécurité* présenté par les députés Didier Paris et Pierre Morel-À-L'huissier à l'Assemblée Nationale le 17 octobre 2018, p. 9.

des plus grandes difficultés est la protection des données personnelles dès la conception d'un projet de traitement (*privacy by design*).

Si le traitement des données massives par un outil de police prédictive porte notamment sur des données personnelles, il est susceptible de présenter un risque élevé pour les droits et libertés des personnes physiques. Auquel cas, le responsable de traitement devra effectuer une analyse d'impact relative à la protection des données à caractère personnel (LIL, art. 90). Si le traitement est mis en œuvre pour le compte de l'Etat, cette analyse d'impact doit être adressée à la Commission nationale de l'informatique et des libertés (CNIL). Notons que la récente loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles a maintenu le régime d'autorisation par arrêté ministériel ou par décret en Conseil d'État pris après avis motivé et publié de la CNIL pour ce type de traitements, afin de préserver les garanties de protection des libertés individuelles. Ce régime concerne les traitements mis en œuvre pour le compte de l'État qui « intéressent la sûreté de l'État, la défense ou la sécurité publique » ou « ont pour objet la prévention, la recherche, la constatation ou la poursuite des infractions pénales ou l'exécution des condamnations pénales ou des mesures de sûreté ». On peut probablement considérer que les traitements de données personnelles pour la mise en œuvre d'outils de police prédictive tomberaient dans cette catégorie et seraient donc soumis au régime de l'autorisation.

Transparence algorithmique et protection des données personnelles

En principe, l'article 10 de la loi Informatique et Liberté du 6 janvier 1978 et l'article 22 du RGPD interdisent toute décision individuelle fondée exclusivement sur un algorithme. Ces dispositions n'auront en principe pas vocation à s'appliquer aux outils de police prédictive qui utilisent une cartographie prévisionnelle du phénomène criminel. En effet, les mesures d'organisation du service, telles que l'orientation des patrouilles, ne sont pas des décisions individuelles produisant des effets juridiques. Il n'existe donc, en théorie, pas d'obstacle légal à une automatisation de l'orientation des patrouilles. Au demeurant, comme vu précédemment, de tels outils ciblant des lieux ne traitent en principe pas des données personnelles mais seulement des données historiques sur les infractions elles-mêmes.

Fondements légaux de la transparence algorithmique

La transparence algorithmique a été encouragée par la loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal dite loi CADA. Cette loi permet d'obtenir l'accès aux documents administratifs auprès de la Commission d'accès aux documents administratifs (CADA), alors même que le code est légalement considéré comme un

« document »⁷⁵ depuis la loi n° 2016-1321 pour une République numérique. Cette dernière loi prévoit des mesures de transparence des traitements algorithmiques faits par l'administration, codifiée aux articles L. 311-3-1 et s. du code des relations entre le public et l'administration. Dans le même sens, notons que la ville de New York a voté des dispositions en faveur de plus de transparence des traitements automatisés utilisés par les agences de la ville⁷⁶, dans le but notamment de diffuser une meilleure information à la population. Les associations de défense des droits civils aspirent à obtenir plus de transparence sur les outils de police prédictive utilisés par NYPD⁷⁷.

b) Outils ciblant des personnes et transparence algorithmique

Dans l'hypothèse où l'outil de police prédictive traiterait de données à caractère personnel, si cet outil a pour objectif de viser des personnes, les dispositions sur la transparence algorithmique auront vocation à s'appliquer. Le RGPD et la directive en matière pénale posent toutefois un principe selon lequel la personne concernée a le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire (article 22§1 du RGPD, art. 11 Directive 2016/680/UE).

Cependant, des exceptions sont prévues en cas de conclusion d'un contrat ou encore de consentement explicite de la personne concernée (art. 22§2). Si l'une de ces exceptions s'applique, des garanties sont prévues (art. 22§3) : le responsable du traitement doit mettre en œuvre des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée, au moins le droit de la personne concernée d'obtenir une intervention humaine de la part du responsable du traitement, d'exprimer son point de vue et de contester la décision. La loi française ajoute d'autres exigences, selon lesquelles les règles définissant le traitement, ainsi que les principales caractéristiques de sa mise en œuvre doivent être communiquées par le responsable de traitement à l'intéressé s'il en fait la demande, à l'exception des secrets protégés par la loi (conformément à l'article L. 311-3-1 du code des relations entre le public et l'administration).

Au demeurant, l'article 22§2 du RGPD donne la possibilité aux Etats de prévoir leurs propres exceptions (marge de manœuvre laissée aux Etats)⁷⁸. L'article 47 al. 2 de la loi n° 78-17 du 6 janvier 1978 modifiée par la loi du 20 juin 2018 dispose désormais que, par exception, des décisions administratives individuelles fondées exclusivement sur un traitement automatisé

⁷⁵ Voir les débats sur la diffusion en 2016 du code source du programme « Admission Post Bac » (APB) sur la répartition des étudiants en études supérieures après l'obtention du baccalauréat.

⁷⁶ Int. N° 1696-A.

⁷⁷ <https://www.aclu.org/blog/criminal-law-reform/reforming-police-practices/police-reform-coming-new-york-city-will-nypd>.

⁷⁸ Judith Rochfeld, *L'encadrement des décisions prises par algorithme*, DALLOZ IP/IT 474 (2018).

peuvent être prises, à condition que le traitement ne porte pas sur des données sensibles. De telles dispositions s'appliquent dans le respect de l'article L. 311-3-1 du code des relations entre le public et l'administration (telles que prévues par la loi n° 2016-1321 pour une république numérique). Ces décisions administratives individuelles comportent, à peine de nullité, la mention explicite prévue à l'article L. 311-3-1 du code des relations entre le public et l'administration, afin d'en informer l'intéressé. Les règles définissant ce traitement, ainsi que les principales caractéristiques de sa mise en œuvre, sont communiquées par l'administration à l'intéressé s'il en fait la demande. Des mesures d'information en ligne ont aussi été prévues⁷⁹. Le décret d'application n° 2017-330 du 14 mars 2017, codifié aux articles R. 311-3-1-1 et R. 311-3-1-2 du code des relations entre le public et l'administration, est venu préciser ces modalités de communication. À la demande de la personne concernée, l'administration doit transmettre sous une forme intelligible et sous réserve de ne pas porter atteinte à des secrets protégés par la loi : 1° Le degré et le mode de contribution du traitement algorithmique à la prise de décision ; 2° Les données traitées et leurs sources ; 3° Les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé ; 4° Les opérations effectuées par le traitement.

En outre, l'article 47 al. 2 de la loi du 20 juin 2018 prévoit désormais que, « pour ces décisions, le responsable de traitement s'assure de la maîtrise du traitement algorithmique et de ses évolutions afin de pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard ». Cependant, par exception à l'article 47 al. 2, aucune décision par laquelle l'administration se prononce sur un recours administratif ne peut être prise sur le seul fondement d'un traitement automatisé de données à caractère personnel.

Ajoutons que, dans sa décision n° 2018-765 DC du 12 juin 2018, le conseil constitutionnel a estimé dans son point 71 que pour satisfaire ces dispositions, « ne peuvent être utilisés, comme fondement exclusif d'une décision administrative individuelle, des algorithmes susceptibles de réviser eux-mêmes les règles qu'ils appliquent, sans le contrôle et la validation du responsable du traitement ». Vont donc être exclus certains outils d'intelligence artificielle dont les règles de modification ne sont pas guidées par leurs concepteurs⁸⁰.

Droit à l'information des personnes concernées

Les articles 13 à 15 du RGPD reconnaissent un droit à l'information des personnes concernées sur l'existence d'une prise de décision automatisée, y compris un profilage.

⁷⁹ Le décret n° 2016-1922 du 28 décembre 2016 prévoit les cas de publication en ligne des documents administratifs qui vise les administrations de 50 agents ou salariés au moins exprimé en équivalents temps plein.

⁸⁰ Voir les définitions de l'IA en annexe 1.

Doivent leur être indiquées les informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour la personne concernée.

Exclusion des décisions individuelles prises sur le seul fondement de traitements automatiques en matière pénale

Enfin, en matière pénale et pour les traitements intéressant la sûreté de l'Etat et la défense, les articles 95 al. 2 et 120 al. 2 de la loi n° 78-17 du 6 janvier 1978 modifiée par la loi n° 2018-493 du 20 juin 2018 confirment que « aucune autre décision (hors décision de justice) produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative ne peut être prise sur le seul fondement d'un traitement automatisé de données destiné à prévoir ou à évaluer certains aspects personnels relatifs à la personne concernée ». Aucune exception n'est prévue dans ces matières, contrairement à la matière administrative. Autrement dit, il est totalement exclu dans ces matières de mettre en place un dispositif de décision automatique reposant exclusivement sur un procédé technique, de nature à prendre des décisions individuelles ou à évaluer les individus. En revanche, rien n'empêche d'utiliser de tels dispositifs comme procédé d'aide à la décision.

4. Police prédictive et missions de la police et gendarmerie : police administrative versus police judiciaire

Police administrative versus police judiciaire

Traditionnellement, sont distinguées les missions de prévention (maintien de l'ordre public), des missions de répression, renvoyant à des compétences différentes et au principe constitutionnel de séparation des pouvoirs entre le juge administratif et le juge judiciaire.

Rappelons que la police judiciaire est exercée, sous la direction du procureur de la République, par les officiers, fonctionnaires et agents (C. de Proc. Pén., art. 12). Elle est chargée de constater les infractions à la loi pénale, d'en rassembler les preuves et d'en rechercher les auteurs, tant qu'une information n'est pas ouverte. Lorsqu'une information est ouverte, elle exécute les délégations des juridictions d'instruction et défère à leurs réquisitions (C. de Proc. Pén., art. 14). L'article 17 ajoute que les officiers de police judiciaire reçoivent les plaintes et dénonciations et procèdent à des enquêtes préliminaires dans les conditions prévues par les articles 75 à 78. En cas de crimes et délits flagrants, ils exercent les pouvoirs qui leur sont conférés par les articles 53 à 67. Ils ont le droit de requérir directement le concours de la force publique pour l'exécution de leur mission. Rappelons que le crime ou délit flagrant est « le crime ou le délit qui se commet actuellement, ou qui vient de se commettre. Il y a aussi crime ou délit flagrant lorsque, dans un temps très voisin de l'action, la personne soupçonnée est poursuivie par la clameur publique, ou est trouvée en possession d'objets, ou présente des traces ou indices, laissant penser qu'elle a participé au

crime ou au délit » (C. Proc. Pén., art. 53).

Police prédictive et ciblage de lieux

En matière de police prédictive, la question sera de savoir comment le recours à ces outils s'insérera dans les missions et pouvoirs des officiers de police. Les outils de ciblage des lieux visent les endroits où des infractions sont susceptibles de se commettre, sans pour autant que lesdites infractions n'aient encore été signalées, ni même que leur survenance ne soit sûre. Dès lors, ils relèveront en principe des activités de police administrative et de maintien de l'ordre. En ce sens, dans le but de respecter les exigences du code de procédure pénale, le système PAVED n'a pas pour objectif de permettre des interventions en situation de flagrance mais de maintien de l'ordre public en prévenant la commission des infractions, en particulier grâce à la présence des forces de l'ordre. L'objectif est de patrouiller dans les zones identifiées par l'outil afin de dissuader la commission d'une infraction. En conséquence, il convient de ne pas surévaluer les zones à risque signalées par les points rouges, en étant tenté d'outrepasser les pouvoirs de police administrative.

La flagrance suppose la commission d'une infraction ou sa tentative et non pas un simple risque élevé de survenance, aussi la police intervient-elle en principe dans le cadre des pouvoirs de police judiciaire. Il faut ainsi éviter que la seule révélation de zones rouges à risque conduise à soupçonner toutes les personnes se trouvant dans les zones en question et à faire basculer d'une mission de police administrative à une police judiciaire. Par exemple, les contrôles d'identité ou les fouilles ne pourront être justifiés par le simple fait que le lieu soit signalé à risque par le système, en l'absence de tout soupçon à l'égard des personnes visées. En effet, les officiers de police judiciaire peuvent inviter à justifier, par tout moyen, de son identité toute personne à l'égard de laquelle existe notamment une ou plusieurs raisons plausibles de soupçonner :

- qu'elle a commis ou tenté de commettre une infraction ;
- ou qu'elle se prépare à commettre un crime ou un délit ;
- ou qu'elle est susceptible de fournir des renseignements utiles à l'enquête en cas de crime ou de délit (C. de Proc. Pén., art. 78-2).

Même si le contrôle d'identité est possible avant la commission de l'infraction, encore faut-il qu'il y ait des raisons de soupçonner la personne concernée par le contrôle et le fait qu'elle se trouve dans une zone à risque élevé ne suffit pas en soi à constituer un tel soupçon. Le signalement d'une zone à risque n'est pas suffisant *per se* et ne remet donc pas en cause l'application de ces règles classiques du contrôle d'identité.

Police prédictive et ciblage de personnes

Il sera très certainement plus difficile de respecter ces principes, s'agissant des outils de

ciblage de personnes, spécialement des auteurs potentiels d'infractions. Le fait que l'outil de police prédictive les signale comme dangereux peut inciter à vouloir les fouiller voire les appréhender, ce qui ne peut se produire avant toute commission d'infraction. À titre d'exemple, la méthode utilisée à Chicago consistant à visiter les personnes soupçonnées d'être engagées dans une activité criminelle pour les inciter à abandonner toute activité criminelle ne conduit pas à outrepasser les pouvoirs de la police mais a pour effet de vouloir empêcher la commission d'infraction. En cas de passage à l'acte, dans la mesure où la personne a été prévenue des risques, la peine maximale sera appliquée.

Reste cependant à se demander si, en visant un individu, volontairement ou non, il peut être porté atteinte à la présomption d'innocence en procédant à une surveillance par des outils de police prédictive. Cette question est évidemment essentielle du point de vue de la protection des droits fondamentaux. Cependant, elle ne se pose réellement que dans l'hypothèse où les personnes ciblées sont des auteurs potentiels d'infractions et non pas les victimes potentielles que l'on cherche à protéger. Dans ce deuxième cas, il s'agira de prévenir la commission de l'infraction et l'action de la police s'inscrira le plus souvent dans le cadre de la police administrative, soumise aux règles de droit administratif et à la compétence du juge administratif. À l'inverse, dans le premier cas, il faut considérer l'action d'éventuels auteurs d'infractions. À l'égard de ces personnes susceptibles de commettre une infraction, tout dépendra si le but de l'action de police est simplement de dissuader la commission d'une infraction par une présence physique des forces de l'ordre (maintien de l'ordre de la police administrative) ou si on cherche plus directement à appréhender une personne qui commet ou s'apprête à commettre une infraction, dans le cadre d'un flagrant délit. L'enquête de flagrance a pour objet de découvrir des infractions et relève de la police judiciaire, pour en rechercher les auteurs, afin de les confier au tribunal de l'ordre judiciaire. Elle suppose alors de respecter les exigences de preuve et règles de procédure pénale afférentes.

Ces questions classiques de qualification des missions de police administrative et judiciaire, liées au contrôle étroit des mesures de police et de répartition des compétences des tribunaux de l'ordre administratif et judiciaire, peuvent être rendues plus difficiles encore dans le cadre de l'utilisation d'outils de police prédictive, si les missions ne sont pas clairement posées et encadrées au préalable.

Liberté et responsabilité du décideur opérationnel

Un autre risque juridique tient au fait que l'officier de police ou gendarmerie puisse être contraint de suivre les recommandations de l'algorithme prédictif et ne soit notamment plus libre de choisir son lieu de patrouille. Or, des règles déontologiques communes à la police et gendarmerie sont reprises à l'article R.434-4 du Code de la sécurité intérieure, selon lequel « l'autorité investie du pouvoir hiérarchique prend des décisions, donne des ordres et les fait

appliquer ».

Dès lors, peu importe que le décideur opérationnel s'adjoigne l'expertise technique d'une personne physique ou morale ou d'un outil de police prédictive, il demeure seul décideur. Il est dès lors important que ce principe soit rappelé aux décideurs opérationnels à qui est confié un outil d'aide à la décision, *a fortiori* si celui-ci est artificiellement autonome⁸¹.

Naturellement, la liberté va avec la responsabilité et s'il est crucial que les décideurs restent libres, il l'est tout autant qu'ils restent responsables des décisions prises. L'article R. 434-4 du Code de la sécurité intérieure précise que « *l'autorité hiérarchique assume la responsabilité des ordres donnés* ». Ainsi, le commissaire de police ou l'officier de gendarmerie est le seul décideur et également seul responsable des ordres donnés.

B. Enjeux éthiques des algorithmes prédictifs : transparence, loyauté, explicabilité et efficacité

Les règles légales, spéciales ou générales, précédemment rappelées, encadrent l'usage des algorithmes en Europe et en France. Ces mesures ne suffisent pas à régler toutes les difficultés, aussi faut-il les compléter par une approche éthique, s'agissant des questions restées sans réponse légale. Ainsi, un certain nombre de difficultés demeurent. Tout d'abord, on peut noter le problème posé par la survalorisation des résultats algorithmiques. Ainsi, la croyance en la neutralité, objectivité, supériorité du résultat algorithmique doit en réalité être contredite par l'existence de possibles biais et discrimination. Au demeurant, l'opacité des méthodes utilisées ne permet pas de remettre en cause les résultats. De façon générale, si des biais humains peuvent naturellement remettre en cause l'égalité des décisions prises à l'égard des personnes, de tels biais sont, de façon générale, mieux acceptés socialement, alors que la machine est supposée être infaillible.

Principe de loyauté

Dans son étude sur « l'Éthique des algorithmes »⁸², la CNIL a posé un certain nombre de principes et recommandations utiles. En présence de décisions prises par un algorithme, elle invite à respecter un principe de loyauté et un principe de vigilance, impliquant des questionnements, vérifications, audits... Le principe de loyauté dégagé par le Conseil d'État

⁸¹ Cyril Piotrowicz, La liberté des décideurs de la sécurité publique à l'épreuve des algorithmes prédictifs, in ALGORITHMES ET ESPACES NORMATIFS, Revue de la Gendarmerie Nationale, n° 261, p. 69.

⁸² CNIL, COMMENT PERMETTRE À L'HOMME DE GARDER LA MAIN ? RAPPORT SUR LES ENJEUX ÉTHIQUES DES ALGORITHMES ET DE L'INTELLIGENCE ARTIFICIELLE | CNIL, <https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes-et-de> (last visited Nov 29, 2018).

en 2014 dans son étude sur le numérique et les libertés fondamentales⁸³ suppose que les personnes concernées doivent être informées de manière loyale sur le traitement de leurs données à caractère personnel et de leur conférer notamment un droit d'accès. Selon le Conseil d'État, « la loyauté consiste à assurer de bonne foi le service de classement ou de référencement sans chercher à l'altérer ou à le détourner à des fins étrangères à l'intérêt des utilisateurs ». Le principe de vigilance viserait à surveiller l'aspect instable que peuvent potentiellement présenter les algorithmes, en particulier en ce qui concerne le *machine-learning* ou apprentissage-machine par laquelle la machine dispose d'une autonomie d'adaptation, sans assistance humaine.

Exigence de qualité des données

Notons par ailleurs que le principe de transparence algorithmique n'est pas suffisant pour permettre une explication de la décision prise et *a fortiori* pour garantir l'absence de biais et discrimination. La qualité des données constitue également une condition essentielle à une meilleure explication, laquelle suppose de connaître les données, les conditions de collecte de celles-ci et conditions d'entraînement des algorithmes, pour permettre une évaluation des risques et les méthodes utilisées pour remédier aux biais potentiels ou constatés.

Egalement, les données d'entrées sont essentielles à la qualité et fiabilité du résultat en sortie.

Exigence de quantité des données

La question de la quantité des données est essentielle pour l'efficacité des modèles. L'exemple de PredVol semble indiquer qu'en l'absence d'une quantité de données suffisantes (vols de voiture dans l'Oise), le modèle ne fonctionne pas ou n'est pas suffisamment réactif.

Le choix fait par la majorité des services de polices en Europe mais aussi par les entreprises produisant des produits commerciaux de se concentrer sur les vols, en particulier cambriolages, témoigne de cette exigence d'une quantité suffisante de données pour pouvoir établir des modèles. Ce type d'infractions, considérées à une échelle géographique suffisante, semble générer suffisamment d'informations pour permettre le bon fonctionnement de l'outil.

Au-delà de la quantité, il convient de considérer également la diversité des données, leurs sources, ainsi que le respect de la finalité des traitements concernées, en cas d'utilisation de données personnelles. Contrairement à ce que l'on pourrait croire, le nombre important de

⁸³ La Documentation française, ETUDE ANNUELLE 2014 DU CONSEIL D'ETAT - LE NUMERIQUE ET LES DROITS FONDAMENTAUX, <http://www.ladocumentationfrancaise.fr/rapports-publics/144000541/index.shtml> (last visited, Nov 21, 2018).

sources de données n'est pas nécessairement une garantie de meilleure performance. Il convient de faire attention au piège qui consisterait à penser que beaucoup de variables et de données vont conduire à de meilleurs résultats. Il ne suffit pas de le croire, il faut le montrer. Le piège serait d'introduire, par de mauvaises justifications, des données personnelles ou insuffisamment anonymisée ou encore des sources de biais. Les données socio-économiques des iris de l'INSEE peuvent avoir un effet de stigmatisation des quartiers populaires.

En outre, les modèles peuvent être faussés par une trop grande quantité de données disparates, de nature à rendre plus difficile la compréhension des phénomènes que l'analyste ne peut alors plus appréhender. C'est toute la difficulté des outils comme *HunchLab* ou *Crime Forecasting* de *Palantir*. À l'inverse, les modèles intégrant peu de sources et catégories de données sont plus simples, plus compréhensibles et, souvent, plus efficaces. Ceci est d'autant plus vrai que l'agrégation d'un nombre important de sources suppose d'évaluer les données et de mettre en place des méthodes de standardisation et de vérification de la valeur probatoire, de façon à savoir si des données aux caractéristiques et attributs différents peuvent être effectivement combinées. Ainsi par exemple, *Paved* ou *PredPol* ont l'avantage d'utiliser peu de catégories de données : l'heure, la date, la localisation et la nature de l'infraction de nature à être facilement évaluées et encodées.

L'utilisation des algorithmes et la transparence dans la modélisation de la prédiction du crime constituent des enjeux de confiance entre le public et les forces de l'ordre. Le fait d'utiliser peu de données permet de mieux expliquer les résultats et de mieux maîtriser aussi d'éventuels dysfonctionnements du modèle.

S'agissant des outils développés en interne par les forces de police et gendarmerie, l'hypothèse d'une utilisation des données personnelles et le croisement des fichiers de sécurité se heurtent à des obstacles légaux, mais aussi à la faisabilité technique et à la pertinence de l'information susceptible d'être générée. Même si des recommandations pourraient être faites pour faciliter les échanges de données issues des très nombreux fichiers de sécurité recensés⁸⁴, il n'est pas certain que la pertinence des modèles prédictifs serait améliorée, compte tenu des difficultés générées par la qualité des données et la variété de leurs sources.

Evaluation de l'efficacité des outils de police prédictive

Lorsqu'un algorithme de profilage d'un risque criminel produit trop de faux-positifs ou de faux-

⁸⁴ Voir les recommandations faites dans le rapport d'information *sur les fichiers mis à la disposition des forces de sécurité* présenté par les députés Didier Paris et Pierre Morel-À-L'huissier à l'Assemblée Nationale le 17 octobre 2018.

négatifs, se pose un problème de performance et donc de qualité de l'algorithme. Ce point sensible donne lieu à un lobbying actif aux États-Unis⁸⁵. Or, La justification de l'utilisation des outils de police prédictive, surtout si ceux-ci sont commerciaux donc très coûteux pour la collectivité, doit passer par leur évaluation.

Cette évaluation intervient à plusieurs niveaux correspondant à plusieurs couches du ou des logiciels mis en œuvre :

- pertinence, efficacité, simplicité d'utilisation, des outils de gestion et visualisation des données ;
- qualité des modèles et algorithmes de prévision des délits ;
- efficacité sur les statistiques de la criminalité.

Certains de ces points ont déjà été abordés lors de la description des principaux outils disponibles mais il est utile d'en résumer une approche synthétique.

Ergonomie

La simplicité et la transparence de la couche de gestion et visualisation des données par "hotspotting" (carte de chaleur) sont fondamentales pour que les services de police et de gendarmerie puissent s'approprier ces outils. Elle ne peut être que plus efficace, car évidemment plus flexible et interactive, en comparaison des punaises sur une carte. Cette évaluation est de la compétence des utilisateurs finaux.

Prévision

La couche "prédictive" nécessite une évaluation rigoureuse pour montrer, ou non, son efficacité. Le problème est que, comme le signalent Bennett Moses et Chan⁸⁶, très peu d'évaluations existent ou ont été publiées par un organisme ou une entité indépendante de la société qui commercialise l'outil. Les principaux travaux publiés à ce jour concernent en fait une évaluation de l'impact sur les statistiques de criminalité mais pas sur la précision effective de la prévision des délits sur une zone géographique déterminée. La mesure de cette précision nécessite la mise en place d'un protocole rigoureux d'estimation de l'erreur de prévision sur des échantillons tests indépendants de l'estimation des modèles ou de l'entraînement des algorithmes. Elle est indispensable, voire rédhibitoire, pour pouvoir justifier de l'utilisation d'algorithmes ou modèles sophistiqués donc opaques aux utilisateurs finaux. Elle est également indispensable pour justifier l'utilisation de variables, autres que celles des historiques des délits et qui peuvent s'avérer sensibles : risque de ré-identification des personnes concernées, risque de biais discriminatoires introduits par des informations socio-économiques comme les IRIS de l'INSEE. Rappelons que l'évaluation réalisée par les

⁸⁵ Elisabeth Grosdhomme, *Ethique des algorithmes, attention au trompe l'œil*, in *ALGORITHMES ET ESPACES NORMATIFS*, REVUE N° 261, p. 113.

⁸⁶ Lyria Bennett Moses & Janet Chan (2018). *Algorithmic prediction in policing: assumptions, evaluation, and accountability*, *Policing and Society*, 28:7, 806-822.

data scientists d'ETALAB⁸⁷ pour l'expérimentation de PredVol dans l'Oise a conclu à l'inutilité de mettre en œuvre des algorithmes sophistiquées sur des données complexes, comparativement à un simple calcul de "moyenne mobile" sur 9 mois.

Efficacité

Quelques études ont été publiées pour monter l'"efficacité" des outils commerciaux de police prédictive sur les statistiques de la criminalité mais celles-ci soulèvent des problèmes. Les principaux travaux ont été conduits par les créateurs de la société PredPol (e.g. G. O. Mohler) donc sans souci d'indépendance. En outre, ils ne mesurent pas la qualité de prévision des délits mais finalement la pertinence de la gestion des patrouilles de police sur le terrain. C'est finalement sur cet argument que sont principalement promus ces outils. Plus de présence physique, plus longtemps, peut amener une effective baisse, géographiquement et temporellement localisée, de la criminalité, ou son déplacement dans l'espace ou leur report dans le temps (effet "plumeau"). Rappelons que le Kent a abandonné l'utilisation de PredPol jugé trop couteux au regard des résultats sur les statistiques de la criminalité.

Trois méthodes d'évaluation peuvent être considérées⁸⁸ ou plus exactement trois critères différents peuvent être pris en compte : prédictibilité, efficacité, efficience.

La prédictibilité est une mesure concernant la précision des prédictions. La précision du système à prédire un délit est mesurée en donnant la fraction de délits correctement prédits quotidiennement.

L'efficacité est une mesure de la criminalité constatée. La seconde évaluation concerne le taux de criminalité en analysant les fluctuations du nombre de délits quotidiens avant et après la mise en place du système d'analyse prédictive. Cette dernière évaluation, sûrement la plus attractive, est cependant très délicate. En effet de nombreux biais peuvent amener à penser à tort que la mise en place d'un tel système a effectivement diminué ou augmenté la criminalité, comme de nombreuses entreprises commerciales peuvent le prétendre. Nous pouvons dès à présent évoquer quelques-uns de ces biais comme :

- un tri des données ;
- des confusions de corrélation/causalité ;
- des problèmes de dénombrement du nombre de délits.

L'efficience est enfin une évaluation des moyens nécessaires. Il s'agit là de considérer l'évaluation des possibilités techniques de mise en œuvre d'un tel système (ressources humaines, matérielles, informatiques, etc...).

Le principal critère utilisé pour mesurer l'efficacité de ces outils de police prédictive est le

⁸⁷ <https://agd.data.gouv.fr/2018/01/12/predire-les-vols-de-voitures>

⁸⁸ Voir Ismael Bensliman, *op. cit.*

changement constaté en matière de statistiques de la criminalité (2^{ème} critère). Mais un nombre important d'études sur la criminalité a montré que les fluctuations des manifestations du crime peuvent être le résultat d'un nombre important de variables que l'on ne peut contrôler sur un court projet piloté implanté sur une seule ville. L'échelle doit donc être la plus large possible pour que l'on puisse en tirer des conséquences significatives.

IV. Conclusion – Recommandations

A. Conclusion

Il existe différents types d'outils de police prédictive et une pluralité de façon de les utiliser, aussi la prudence est-elle de mise s'agissant de faire des recommandations générales. Quoi qu'il en soit, Il serait trompeur de les considérer comme une panacée pour éradiquer la criminalité. Ces outils ne peuvent pas remplacer le travail des analystes expérimentées et des policiers de terrain.

Cependant, de tels outils peuvent être complémentaires et rendre le travail des enquêteurs plus efficace et leur permettre d'économiser du temps. Les projets d'intelligence artificielle semblent attendus au sein des services de la police nationale et gendarmerie nationale, dans le but de libérer le personnel de certaines tâches à faible valeur ajoutée. En ce sens, semble prometteur le traitement des flux croissants d'informations, pour permettre de comparer les informations émanant de bases différentes, d'enrichir les données, de créer d'autres modèles prédictifs que ceux actuellement mis en œuvre. Des projets semblent déjà identifiés en matière de gestion et d'analyse d'images en masse (vidéo), de gestion des appels d'urgence (modernisation des centres d'information et de commandement (*M-CIC 2*), mise en œuvre des plates-formes d'appels d'urgence unifiées), de gestion de crise, de sécurité routière ou encore de lutte contre la fraude⁸⁹. Les techniques de reconnaissance faciale sont particulièrement prometteuses.

Si les potentialités de l'intelligence artificielle et de la police prédictive sont attractives, quelques précautions sont à prendre, traduites en recommandations dans ce rapport. Les risques et défauts de certains outils de police prédictive, notamment de ciblage des personnes ou encore les outils commerciaux quels qu'ils soient, ne sont pas de nature à donner confiance au public. La confiance est essentielle et le choix d'outils doit privilégier l'acceptabilité sociale, ce qui suppose de ne pas recourir à des outils utilisant une masse de données, spécialement des données personnelles, dont les procédés ne seraient ni transparents ni maîtrisés par les services de police conduits à les utiliser. Le contrôle, la connaissance de l'outil devront être les principaux objectifs, afin de garantir la maîtrise sur le terrain. En ce sens, le choix actuellement fait en Europe et en France de promouvoir des outils internes plutôt qu'externes paraît satisfaisant.

⁸⁹ Voir le rapport d'information *sur les fichiers mis à la disposition des forces de sécurité* présenté par les députés Didier Paris et Pierre Morel-À-L'huissier à l'Assemblée Nationale le 17 octobre 2018, p. 64.

Par ailleurs, il convient naturellement de respecter la loi et il est douteux que la collecte massive de données personnelles, telle que réalisée par les outils commerciaux utilisés aux Etats-Unis, puisse être conforme aux réglementations sur les données personnelles mises en œuvre dans l'Union européenne et ses Etats membres. Dès lors, le choix fait en Europe par les polices européennes, à l'exception des services de renseignement, de cibler des lieux plutôt que des personnes va dans le bon sens.

Egalement, les outils de police prédictive doivent rester des moyens mis à la disposition des services de police parmi d'autres, sans nécessairement prédominer et surtout sans devoir supplanter l'action humaine. En tout état de cause, il est essentiel que les opérationnels de la police et gendarmerie conservent sur le terrain leur libre arbitre et ne soient pas contraints de suivre les recommandations de la machine. Il convient de promouvoir clairement la supériorité de la décision humaine sur celle de la machine. Cette liberté doit être préservée et doit s'accompagner d'une responsabilité, laquelle ne peut être transférée à la machine (préservation du principe de responsabilité à la charge du décideur humain).

Si l'Europe semble avoir pris du retard sur les Etats-Unis dans l'utilisation d'outils de police prédictive, le moindre avancement technologique pourrait s'avérer être un avantage pour plusieurs raisons.

D'abord, l'Europe peut bénéficier du retour d'expérience sur les outils commerciaux déployés depuis plusieurs années et en mesurer l'efficacité, sous réserve qu'une telle information soit accessible et non partisane.

Ensuite, l'implantation de ces technologies dans les villes américaines s'est faite au détriment du contrôle des résultats et de la maîtrise par la police des actions à mener. Il y a là un risque de perte de compétence qui pourrait s'avérer catastrophique à l'avenir si les opérationnels ne se fient plus à leur expérience et connaissance de terrain et uniquement à une technologie obscure dont l'efficacité n'est pas clairement démontrée. Il est douteux que ce soit le sens vers lequel il faille aller aujourd'hui.

Enfin, le recours à ces outils s'est souvent fait aux Etats-Unis au détriment du respect des droits fondamentaux et de la vie privée des individus par l'utilisation de données dont l'origine et la fiabilité ne sont pas maîtrisées. Ces données ont été recueillies à l'occasion de pratiques policières partisans envers certaines populations et ont généré des biais racistes. De telles données ayant été utilisées pour entraîner les modèles d'apprentissage, les biais en sortent renforcés aussi peut-on douter de la bonne qualité de ces outils. Mieux vaut mettre en place une bonne politique de gestion de la donnée pour entraîner et utiliser correctement de tels outils.

B. Recommandations

1. Recommandations au législateur pour l'encadrement des outils de police prédictive

Renforcer les exigences légales de transparence algorithmique et « explicabilité »

Si la loi n° 2018-493 du 20 juin 2018 renforce les exigences de transparence algorithmique et impose que le responsable de traitement s'assure de la maîtrise du traitement algorithmique et de ses évolutions afin de pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard, de telles dispositions supposent l'utilisation de données personnelles. Or, la mise en œuvre d'un traitement algorithmique utilisant des données non personnelles peut avoir un effet discriminant, aussi convient-il d'étendre de telles exigences à tout type de traitement, au-delà de la seule hypothèse d'utilisation de données personnelles.

Légiférer sur la discrimination indirecte et systémique sans intentionnalité

Il convient de légiférer sur les discriminations indirecte et systémique qui constituent souvent des discriminations de groupe. À l'heure actuelle, ce type de discrimination est exclu du champ d'application de la réglementation sur la discrimination, en l'absence d'intentionnalité. Or, en situation d'utilisation des systèmes d'apprentissage machine, ce genre de discrimination va se multiplier, sans intentionnalité et sans même qu'il soit aisé de la détecter et *a fortiori* de la prouver. Ce type de discrimination devrait être juridiquement constitué et sanctionné même en l'absence de preuve d'intentionnalité, à condition que la discrimination algorithmique en tant que telle soit prouvée.

Maintenir les règles de liberté et d'imputabilité de la responsabilité à l'homme et non à la machine

Doit être préservée la liberté de décision de l'autorité investie dans l'utilisation des outils de police prédictive qui doivent être de simples aides à la décision. Les décideurs opérationnels doivent rester libres de leurs décisions de patrouiller et de déterminer les lieux, sans considération des conseils donnés par les outils de police prédictive, qu'ils soient internes ou externes, et sans avoir à en répondre.

Les règles classiques de responsabilité doivent rester applicables dans la mise en œuvre des outils de police prédictive quel que soit le type d'outil utiliser et quelle que soit le degré d'intervention de l'outil dans la prise de décision.

La responsabilité doit rester humaine et ne pas être transférée à la machine.

2. Recommandations au Ministère de l'intérieur pour l'intégration des outils de police prédictive

Définir les activités et usages de l'outil dans un contexte global

Il faut que les outils d'aide à la décision (IA ou pas) s'intègrent dans des activités.

Le ministère de l'intérieur devra réfléchir au processus qui intègre ces outils. Il convient de ne pas se limiter à voir l'outil isolément mais l'ensemble du dispositif dans sa globalité, intégrant les hommes, les procédures et les outils.

Egalement, la finalité (ou le périmètre) de l'outil doit être clairement définie.

Donner la préférence aux outils internes de police prédictive

Il convient de recommander aux services de police et gendarmerie de créer leurs propres outils ou d'ajuster des outils non-commerciaux, plutôt que d'utiliser des logiciels commerciaux, ce qui suppose alors de laisser la maîtrise à des entreprises privées, le plus souvent étrangères. La création en interne d'outils de police prédictive est de nature à permettre une meilleure maîtrise des outils et explicabilité par ses concepteurs et utilisateurs, par opposition aux outils commerciaux, plus opaques.

Une telle solution est à privilégier, de manière à pouvoir répondre à ces nouvelles exigences légales de transparence et explication des décisions individuelles prises sur le fondement des algorithmes, *a fortiori* lorsque ces outils sont exclusivement utilisés pour la prise de décision.

Poser des exigences méthodologiques et d'audit

Il convient de poser des exigences méthodologiques pour la conception des outils de police prédictive qu'ils soient internes ou externes, allant au-delà des exigences légales. Devraient être publiées les données, et leurs caractéristiques, avec lesquelles l'algorithme a été entraîné, ainsi que l'évaluation des risques et méthodes utilisées pour atténuer les biais potentiels. Egalement, il convient de publier la liste des données d'entrées utilisées par l'algorithme pour prendre une décision.

Les circonstances d'obtention de ces données doivent aussi être connues. Si les données émanent de la police, il n'y a pas de difficulté de légitimité dans leur utilisation (sous réserve de respecter les réglementations citées) mais, dans le cas contraire, des contrats doivent être conclus avec les titulaires des données en question. Si les services de police prédictive utilisent des produits proposés par des sociétés commerciales, il convient de s'assurer que lesdits fournisseurs n'acquiescent pas la « propriété » des données de la police ni que ces données pourront être cédées ou réutilisées pour quelque usage que ce soit. De telles

données touchent à la souveraineté et au pouvoir régalién de l'Etat.

Devrait également être mis en place un système d'homologation technique par audit, afin de vérifier la fiabilité des algorithmes avant l'autorisation de leur déploiement.

Former et acculturer les personnels aux outils de police prédictive

Mettre en place des plans de formation et acculturation aux outils de représentation cartographique et de *reporting* pour améliorer la connaissance et les pratiques professionnelles.

Coopérer avec les universités, écoles et laboratoires de recherche

Encourager la coopération avec les équipes scientifiques et laboratoires de recherche des universités, écoles ou CNRS qui comportent des ressources scientifiques du meilleur niveau, tant pour la conception que la mise en œuvre des outils et méthodes. L'expérience montre que les universitaires sont en mesure de donner des conseils et même une assistance, spécialement pendant la période d'implantation, et que les meilleurs outils ont été conçus à l'occasion d'une telle coopération.

Il faut noter que les outils développés aux Etats-Unis l'ont tous été dans un contexte académique. L'accès aux données par les universitaires dans un cadre de confiance peut permettre d'améliorer la connaissance commune et la pertinence des outils déployés sur le terrain. Elle est en outre de nature à favoriser l'acceptabilité sociale.

La coopération pourrait également concerner la formation et l'acculturation aux outils.

3. Recommandations aux opérationnels pour l'utilisation des outils de police prédictive

Bien déterminer la cible et les modalités d'action de la police administrative

L'outil de police prédictive doit clairement établir la cible visée, en précisant s'il s'agit de surveiller des lieux ou des personnes. Le ciblage des individus engendre l'applicabilité de la réglementation des données personnelles et implique le respect des droits fondamentaux.

Au demeurant, les modes opératoires mis en œuvre doivent rester dans le cadre de la mission de la police administrative, en ce que ces outils doivent en principe être utilisés pour prévenir l'infraction.

Ne pas « sur-policer » et surévaluer les quartiers signalés au risque de discriminer

Compte tenu du risque de discrimination indirecte et de groupe, voire systémique, il convient d'être vigilant dans l'utilisation des résultats algorithmiques et d'éviter de sur-policer certaines zones ou quartiers qui apparaîtraient systématiquement à risque du fait de données passées. Les données passées peuvent introduire un biais algorithmique en surreprésentant certaines catégories de population ou certains groupes ethniques.

Il convient également de ne pas surévaluer l'information donnée par l'outil de police prédictive révélant une zone à risque, laquelle n'a pas pour effet en soi de modifier les missions de la police qui reste dans le cadre de la police administrative de prévention de la commission de l'infraction et maintien de l'ordre.

Evaluer les outils en cours d'utilisation et en continu

Parallèlement à la mise en place d'outils de police prédictive, il convient de conduire systématiquement des évaluations de la qualité statistique de prévision sur des échantillons indépendants. Cela afin :

- de contrôler l'optimisation de certains paramètres des modèles mêmes élémentaires: coefficients de lissage temporels et géographiques des "moyennes mobiles". Ceux-ci peuvent en effet être liés aux territoires ainsi qu'aux types de délits.
- d'éviter une confiance aveugle en des algorithmes sophistiqués, donc opaques et donc difficilement acceptables par les utilisateurs finaux.
- d'éviter de prendre en compte inutilement dans les modèles des informations personnelles ou socio-économiques au risque de soulever des problèmes tant juridique que de biais des prévisions.

Il convient également de conduire des études globales, en intégrant ces autres variables mais sous un contrôle strict de l'anonymisation avec pour objectif de cerner les causes de la délinquance dans une perspective de prévention plus que de répression.

Cette évaluation ne doit pas nécessairement être lourde mais suppose de :

- varier la localisation ;
- se focaliser sur la qualité de prévision et non sur les statistiques de criminalité ;
- optimiser au mieux les hyper paramètres du modèle retenu.

Ceci peut être complété par une veille technologique pour tester de nouveaux modèles.

Faire des études d'impact sur les risques de discrimination

En ce sens, le Rapport Villani⁹⁰ indique que « Les lignes directrices adoptées par le G29⁹¹ demandent qu'un PIA⁹² soit effectué lorsqu'un risque de discrimination ou d'exclusion émerge avec un traitement de données. Cet axe central dans l'acceptabilité sociétale de l'IA doit faire l'objet d'une analyse à part entière. Il s'agirait d'accompagner le PIA d'un dispositif analogue pour les discriminations, un *discrimination impact assessment*, ou *DIA*, pour amener les concepteurs d'IA à s'interroger sur les conséquences sociales des algorithmes qu'ils produisent ».

⁹⁰ CÉDRIC VILLANI, AI FOR HUMANITY 235 (2018) (p. 147), https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf.

⁹¹ Groupe de l'article 29 de la directive 95/46/CE sur la protection des données personnelles et qui regroupe en particulier les autorités nationales de protection des données personnelles dont la CNIL.

⁹² *Privacy Impact Assesment* ou étude d'impact sur la vie privée.

V. Annexes

A. Annexe 1 : Glossaire et définition de l'IA

1. Glossaire de l'intelligence artificielle et police prédictive

Algorithme

Suite finie et non ambiguë d'opérations permettant d'atteindre un résultat. Par extension un ensemble de codes informatiques ou instructions réalisant les opérations.

Apprentissage automatique (machine learning)

Ensemble d'algorithmes, sous-ensemble de l'intelligence artificielle (IA), mimant les capacités humaines d'apprentissage et d'entraînement.

Apprentissage statistique (statistical learning)

Ensemble d'algorithmes qui apprennent à partir d'un jeu de données dit d'apprentissage. Les données sont constituées d'un ensemble de n individus ou instances décrits par p variables X_j ou caractéristiques (features) qui peuvent être quantitatives (e.g. âge) ou qualitatives (e.g. genre). Une autre variable Y dite à expliquer ou modéliser ou prédire est également observée sur ces mêmes individus. Un algorithme est entraîné en minimisant une erreur de prévision de Y , définie comme un risque statistique ou espérance (moyenne) d'une fonction perte. . D'où l'appellation : apprentissage statistique. Cette fonction perte ou erreur peut être quadratique ou en valeur absolue si Y est quantitative (problème de régression), un taux d'erreur ou une entropie si Y est qualitative (objectif de classification supervisée). Les principaux algorithmes d'apprentissage statistique sont cités ci-dessous.

Apprentissage

L'apprentissage est la procédure qui consiste à exécuter un algorithme conduisant à son entraînement ou encore à l'estimation des paramètres ou poids du modèle sous-jacent sur les données d'apprentissage. Cet apprentissage se décompose en deux étapes : une d'estimation qui ajuste au mieux les données en minimisant l'erreur apparente ou de prévision de l'échantillon d'apprentissage. Cette phase dépend d'un ou de plusieurs hyper-paramètres contrôlant la complexité ou flexibilité du modèle : e.g. le nombre de variables, neurones, feuilles, vecteurs supports... La deuxième étape concerne l'optimisation de la ou des valeurs de ces hyper-paramètres afin d'éviter le sur-ajustement du modèle ou sur-apprentissage de l'algorithme et atteindre la meilleure prévision du modèle, ou meilleure qualité de généralisation de l'algorithme.

Régression, régression logistique

Algorithmes basés sur l'estimation des paramètres d'un modèle statistique classique, fonction linéaire des variables d'entrée ou explicatives X_j du modèle. Les capacités explicatives (tests statistiques) du modèle sont négligées au profit du seul objectif de

prévision. On parle de *régression logistique* lorsque la variable à modéliser, prédire, Y est qualitative (e.g. type de délit) au lieu de quantitative (classification supervisée). C'est alors la probabilité d'occurrence d'une modalité de Y qui est modélisée. Le sur-ajustement est contrôlé par le nombre de variables ou par régularisation. Le modèle est interprétable si le nombre final de variables du modèle est réduit.

Arbres binaires de décision

Algorithme conduisant à la construction d'un modèle schématisé par une arborescence de décisions hiérarchisées. Chaque nœud de l'arbre est décrit par une règle binaire associant une variable explicative X_j et une valeur seuil ou une partition dyadique des modalités de X_j . La séquence de nœuds est construite en minimisant un critère : entropie si Y est qualitative, variance au sein d'une feuille si Y est quantitative. La prévision est fournie par la moyenne des valeurs des individus de l'apprentissage au sein d'une feuille (régression) ou par la modalité la plus fréquente (classification). Le sur-ajustement est contrôlé par le nombre de feuilles. Le modèle est interprétable à condition que sa complexité : le nombre de feuilles, soit réduit.

Réseaux de neurones

Algorithme conduisant à la construction d'un modèle schématisé par un graphe de neurones formels. Un neurone formel opère une transformation (fonction de transfert) d'une combinaison linéaire des variables en entrée. Les réseaux les plus courants sont issus du *perceptron* dont le graphe est organisé en couches successives. Chaque sortie des neurones d'une couche est connectée à l'entrée de la couche suivante. La première couche accepte les variables en entrée. La dernière couche produit en sortie la prévision ; valeur de Y quantitative ou probabilités des classes de Y qualitative. L'apprentissage des poids ou paramètres des combinaisons linéaires de chaque neurone est obtenu l'exécution d'un algorithme de *rétropropagation du gradient*. L'erreur observée, sur des séquences (*epochs*) aléatoires des données d'apprentissage, contribue à la correction itérative de chacun des poids jusqu'à l'obtention d'une convergence nécessairement locale. Le sur-apprentissage est contrôlé par les nombres de couches, neurones, des termes de régularisation ou encore la le nombre d'*epochs*, de l'apprentissage. Le modèle qui en découle est opaque, il n'est pas possible d'interpréter l'influence d'une variable X_j sur la sortie Y .

Random forest, boosting

Ces algorithmes produisent des ensembles de modèles, très généralement des arbres binaires de décision. Chaque modèle est estimé ou appris à partir d'un sous échantillon aléatoire (*bootstrap*) de l'échantillon d'apprentissage initial dans le cas des forêts aléatoires. Il est obtenu de façon adaptative dans le cas du *boosting* dans le but d'améliorer pas-à-pas la solution à chaque étape. La prévision finale est une moyenne des prévisions de chaque modèle dans le cas de Y quantitative (régression), la modalité la plus fréquente dans le cas

de Y qualitative (classification supervisée). Une forêt aléatoire n'est pas sensible au sur-apprentissage contrairement aux dernières versions de *boosting* (*gradient boosting machine*, *extrem gradient boosting*) dont il faut contrôler les valeurs d'un ou plusieurs hyper-paramètres. Ces algorithmes sont opaques même s'ils fournissent des indications sur l'importance des variables explicatives.

Machine à vecteurs supports

Dans le cas linéaire, cet algorithme de classification supervisée binaire recherche un meilleur hyper plan séparateur entre deux classes. Cet hyper-plan est solution d'un problème d'optimisation quadratique sous contraintes de bonnes classifications des éléments de la base d'apprentissage. L'hyper plan est défini par la seule connaissance de la définition du produit scalaire de l'espace des solutions. Le cas non linéaire est obtenu en définissant un produit scalaire spécifique et adapté au problème posé, au type des données, par une fonction noyau ; par exemple un polynôme ou la densité de la loi gaussienne. Le sur-apprentissage est contrôlé par un paramètre de pénalisation.

Autres algorithmes et combinaisons

Bien d'autres modèles et algorithmes sont proposés dans les librairies et suites logicielles ; régression PLS, analyse discriminante décisionnelle, réseaux bayésiens, *multiple additive regression splines*, *kernel regression least square*... de même que toute combinaison de tous ces algorithmes comme c'est le cas dans les solutions des concours de prévision de type Kaggle.

Cellule unitaire ou target geographical area x Time window

Les historiques de criminalité par type ou groupe de délit, leurs prévisions sont calculées et représentées, cartographiées par unité spatio-temporelle dont la taille est paramétrable. Par exemple, dans le cas des produits développés par Palantir la cellule par défaut est un cube de 250m x 250m x 8h. Ce type de cellule s'adapte bien à une ville américaine construite par bloc, mais moins à des topographies plus complexes ou la notion de rue ou de portion de rue est plus adaptée. Le projet abandonné PredVol test dans le département de l'Oise utilisait le découpage IRIS de l'INSEE⁹³. La taille des cellules doit être liée à la quantité de délits observés. Une petite cellule est géographiquement et temporellement plus précise à condition que suffisamment de délits aient pu y être observés. Dans le cas contraire, la précision est illusoire et il est préférable de considérer des unités plus grandes dans lesquels des délits ont pu être observés.

Crime risk forecasting

Nom de l'application logicielle développée, brevetée et diffusée par la société Palantir pour la représentation spatiaux-temporelle et la prévision de la criminalité.

⁹³ <https://agd.data.gouv.fr/2018/01/12/predire-les-vols-de-voitures>.

Hotspotting, heatmap

Modèle de représentation ou plutôt de visualisation géographique de la criminalité par type de délit ou groupes de délits et par unité cellulaire ou groupe d'unités cellulaires. Un code graphique (couleur) représente l'intensité de la criminalité définie par une formule faisant intervenir les nombres d'occurrences des délits et crimes passés. De très nombreuses options sont possibles pour faire intervenir des zones d'influence des cellules proches ou des historiques plus ou moins anciens. De même ces options peuvent considérer des seuillages strictes de ces influence ou faire intervenir des pondérations (lissage exponentiel, estimation de densité par la méthode du noyau) pour oublier plus ou moins vite le passé ou doser l'influence géographique. Le réglage de ces paramètres est fondamental d'autant que ce « hotspotting » du passé produit finalement une prévision raisonnable par moyenne mobile en révélant les zones et périodes de forte criminalité dont la reproduction est très attendue. Dans le cas de l'expérimentation PredVol, une moyenne mobile de 9 mois conduisait à une aussi bonne prévision que celles fournies par des algorithmes très sophistiqués.

Système d'information géographique (SIG)

Base de données répertoriant tous les crimes et délits passés en les répertoriant pas leur localisation géographique (adresse, latitude, longitude, GPS de la patrouille de police...) et temporelle plus ou moins précise ou à l'intérieur d'une zone géographique ou d'un intervalle de temps.

Processus Stochastique

Un processus stochastique est un ensemble de variables aléatoires qui sont liées les unes aux autres par opposition aux variables indépendantes. Le lien entre ces variables est défini par le type de processus considéré. On distingue en particulier les **Processus Gaussiens** dont la loi est donnée par la fonction moyenne et la covariance. Ce type de processus permet donc de modéliser des phénomènes spatiaux temporels dans lesquels les événements présentent des corrélations à la fois en temps (des événements survenant à peu de temps d'intervalle sont plus dépendants) mais également en espace (des événements observés dans des emplacements proches sont plus corrélés entre eux). Un autre cas de figure est donné par les **Processus de Poisson**. Ce sont des processus pour compter des événements et ainsi comprendre l'apparition de phénomènes ponctuels. Ils sont gouvernés par un paramètre décrivant l'intensité d'apparition des événements.

Les modèles Gaussiens sont souvent employés en raison de la facilité de leur mise en oeuvre en utilisant la méthode de **Krigeage**. L'estimation se fait en deux temps. Dans un premier temps, la covariance est estimée par maximum de vraisemblance. Puis les nouvelles valeurs sont prédites par combinaison linéaire des valeurs du processus observé dans des emplacements proches du lieu où l'on souhaite prédire. Cette formule ne dépend que des valeurs observées et de la covariance prédite et fournir le meilleur estimateur linéaire de la

prédiction.

Disparate impact

Le Disparate Impact est un indicateur permettant de mesurer la disparité d'une décision par rapport à deux groupes dans une population divisée en un groupe majoritaire et un groupe minoritaire. Le Disparate Impact d'une règle est le rapport entre la probabilité de prendre une décision positive pour le groupe minoritaire et celle pour le groupe majoritaire. Ce rapport est égal à 1 lorsque la décision ne dépend pas de l'appartenance au groupe. Plus ce rapport est faible, plus le groupe minoritaire est désavantagé. Ainsi un Disparate Impact faible correspond à une décision fortement déséquilibrée et qui désavantage le groupe minoritaire.

2. Définition de l'IA

Quelle Intelligence artificielle ?

Les organisations (privées ou publiques) ayant appris à stocker, gérer massivement leurs données depuis 10 ans, la phase suivante concerne leur analyse pour leur valorisation et l'aide à la décision, voire de la décision automatique. Après s'être appelée "big data analytics" puis "data science" cette phase fait maintenant référence à une pratique d'intelligence artificielle (IA), appellation largement médiatisée, notamment depuis les succès remarquables en reconnaissance d'images depuis 2012, en traduction automatique, d'AlphaGo en 2016 ou autour des expérimentations de véhicules autonomes.

L'IA n'est pas une invention récente car cette discipline ou plutôt cet ensemble de théories et techniques est apparue conjointement avec le développement des tous premiers ordinateurs (ENIAC en 1943), eux-mêmes conséquences des efforts, durant la deuxième guerre mondiale, pour produire rapidement des abaques de balistique, puis réaliser les calculs de faisabilité de la première bombe atomique. L'objectif initial était la simulation des comportements du cerveau humain. C'est aussi en 1943 que Mc Culloch (neurophysiologiste) et Pitts (logicien) ont proposé les premières notions de neurone formel. Notons le début de la théorisation de l'IA avec les travaux pionniers d'Alan Turing en 1950 et la première apparition de la notion de perceptron, le premier réseau de neurones formels, par Rosenblatt en 1957. Manque de moyens de calcul et d'algorithmes pertinents, l'approche connexionniste de l'IA est mise en veilleuse durant les années 70 au profit de la logique formelle (e.g. calcul des prédicats du premier ordre) comme outil de simulation du raisonnement. Les systèmes experts associant base de connaissance (règles logiques), base de faits et moteur d'inférence ont connu un certain succès, notamment avec le développement du langage Prolog, mais ont buté sur la complexité algorithmique explosive des problèmes NP complets. Ce fut alors, au début des années 80, le retour massif de l'approche connexionniste avec le développement de l'algorithme de rétropropagation du gradient qui a ouvert la possibilité, en

lien avec des moyens de calculs suffisamment performants, de l'apprentissage de réseaux de neurones complexes. Le développement de l'IA s'est ensuite focalisé dans les années 90 sur des objectifs d'apprentissage (*machine learning*), qui devint plus précisément l'apprentissage statistique (*statistical learning*) en conséquence de la publication du livre éponyme de Vapnik (1995). C'est toute une farandole d'algorithmes : séparateurs à vaste marge (SVM), *bagging*, *boosting*, *random forest*... qui provoqua à nouveau la mise en retrait des approches connexionnistes considérées, au même titre que bien d'autres algorithmes souvent plus performants. Mais certains chercheurs dont Yan le Cun, Yoshua Benjio, Geoffrey Hinton, ont continué à développer des structures connexionnistes spécifiques dont les fameux réseaux intégrant des produits de convolution (*convolutional neural networks*) sur des images. L'accumulation complexe de ces couches fut nommée apprentissage profond (*deep learning*) avec un réel succès marketing.

Actuellement, la présentation médiatique de l'IA diverge rapidement vers des questions philosophiques (transhumanisme), comme celle de singularité technologique lorsque les machines deviendront plus "intelligentes" que l'homme... Le développement de l'IA soulève des questions également anxiogènes de destruction de nombreux emplois qualifiés⁹⁴ et pas seulement des métiers manuels absorbés par la robotisation des entreprises. D'autres craintes sont liées aux menaces concernant la vie privée ainsi qu'aux questions éthiques abordées par ailleurs⁹⁵. Nous oublierons ces aspects pour nous focaliser sur les algorithmes d'IA en exploitation, ceux qui impactent nos quotidiens professionnels ou personnels, conséquences de la datafication de nos environnements. Ces algorithmes, capables de s'entraîner, en surfant sur la vague ou plutôt le tsunami des données, afin de construire des décisions automatiques, constituent le sous-ensemble historique de l'IA appelé apprentissage automatique ou *machine learning*. Plus précisément, nous laisserons de côté les algorithmes de renforcement ou de décision séquentielle (e.g. bandit manchot) qui sont des algorithmes d'optimisation stochastique trouvant leurs applications dans la gestion des sites de vente en ligne. Il reste alors le principal sous-ensemble des algorithmes d'apprentissage statistique au sens de Vapnik (1995), incluant également l'apprentissage profond ou *deep learning*. Ceux-ci construisent des règles de décision ou des prévisions par minimisation d'un risque <http://wikistat.fr/pdf/st-m-app-risque.pdf>, généralement une erreur moyenne de prévision. Leur succès et la généralisation de leur utilisation sont des conséquences directes de la datafication (*big data*) du quotidien.

Marketing et *data mining*, finance et trading algorithmique, traduction automatique et

⁹⁴ B. Stiegler B. et A. Kyrou A. L'Emploi est Mort, Vive le Travail ! Fayard, 2015.

⁹⁵ Philippe Besse, Céline Castets-Renard & Aurélien Garivier, LOYAUTÉ DES DECISIONS ALGORITHMIQUES (2017), <https://hal.archives-ouvertes.fr/hal-01544701> (last visited Nov 21, 2018).

traitement du langage naturel (*sentiment analysis*), reconnaissance faciale et analyse d'images en lien par exemple avec les véhicules autonomes, aide au diagnostic, détection d'anomalies, prévision de défaillance et maintenance préventive dans l'industrie... sont autant de domaines d'application des algorithmes d'apprentissage statistique, sous-ensemble de l'Intelligence Artificielle bénéficiant et valorisant les masses de données en croissance exponentielle.

Cartographie de l'intelligence artificielle

Dresser une cartographie consiste en la définition des différents domaines de l'intelligence artificielle. Cela oblige d'abord à s'interroger sur ce qui doit être entendu par domaine :

- Domaine défini selon les technologies et modèles employés (exemple : réseau de neurones)
- Domaine défini selon le champ d'application (exemple : traitement du langage)

Il n'est pas si facile que cela d'en privilégier un. En effet, les deux perspectives sont indissociables. Ainsi si on prend un champ d'application, on parlera des technologies généralement utilisées dans celui-ci. Ces technologies sont elles-mêmes inspirées des spécificités de ce champ d'application. Considérons par exemple le domaine du « web » : les interconnexions entre pages donnent un sens particulier à la notion de fouille d'opinions, les connexions jouant un rôle de première importance. Les technologies n'ont généralement de sens que pour un champ d'application. Nous allons toutefois adopter une présentation sous l'angle technologique dans un premier temps en considérant celle-ci de la manière la plus neutre possible et en la revisitant en intégrant les champs d'application.

Principaux modèles et technologies

Historiquement, l'intelligence artificielle a d'abord cherché à reproduire le **raisonnement** humain. L'homme est vu comme un être rationnel doué de capacités d'abstraction : il est capable d'acquérir puis raisonner avec des connaissances. Les progrès de l'intelligence artificielle ont conduit à spécialiser les connaissances : les connaissances sur les actions à donner naissance au domaine de la **planification**. L'acquisition de connaissances à partir de documents écrits ou sonores a fait naître le domaine du **traitement du langage**. La description de raisonnement peut être vue comme la recherche de solutions parfaites ou approchées à l'aide d'heuristiques. Ce domaine concerne plus généralement la définition d'**algorithmes**.

Simultanément, les progrès de l'informatique dans les réseaux et communication ont donné naissance à l'informatique distribuée. L'intelligence artificielle a intégré cette dimension : les applications « intelligentes » sont vues comme des agents et l'intelligence artificielle a cherché à reproduire les activités humaines de coopération et collaboration. L'objectif est de concevoir

des **systèmes multi-agents**. Différents types de systèmes sont alors considérés par l'intelligence artificielle et le besoin de définir des **architectures** conduit à l'émergence de ce domaine.

Comme précédemment indiqué, l'intelligence artificielle a cherché dans un premier temps à reproduire l'intelligence de l'homme. Une approche différente consiste à ne pas essayer de reproduire celle-ci mais à la simuler. Cette approche est principalement basée sur les réseaux de neurones et les algorithmes génétiques. Son domaine de prédilection est **l'apprentissage** automatique : un système initialement « inintelligent » acquiert des connaissances en établissant des connexions (apprentissage) entre des données initialement non connectées. Le système devient alors intelligent ou plus précisément simule l'intelligence : c'est en effet l'homme qui donne du sens aux connexions.

Comme on le voit, en intelligence artificielle, deux approches existent : *bottom-up* et *top-down*. Ces deux approches sont complémentaires et non opposées : la figure 1 détaille de manière non exhaustive ces différents modèles. La figure souligne que séparer modèles et applications n'est pas évident : le traitement du langage et la planification en sont deux illustrations évidentes.

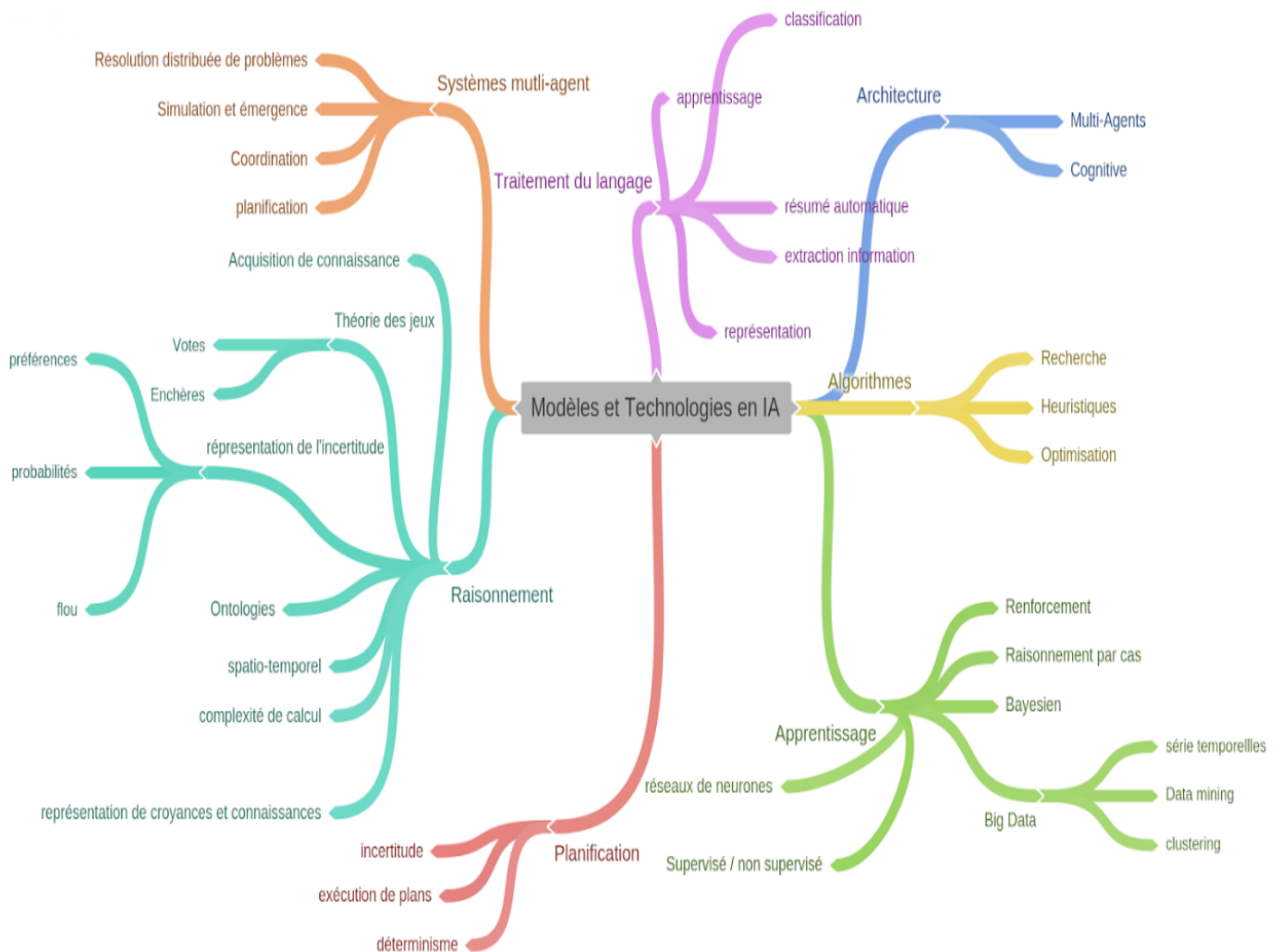


Figure 1 : Cartographie partielle des technologies en IA

Pour des raisons de clarté chaque modèle ou technologie est considéré de manière distincte. Or les interconnexions sont nombreuses : le traitement du langage utilise des techniques d'apprentissage et de raisonnement. Les architectures sont liées au raisonnement et au système multi-agent. De même pour la planification et les algorithmes. Potentiellement, chaque technologie peut se lier à une ou plusieurs autres.

- **Algorithmes** ce domaine consiste à élaborer des algorithmes de *recherches* de solutions pour un problème ayant une grande combinatoire de paramètres (ex : automatisation d'emploi du temps). Des *heuristiques* sont élaborées afin d'approcher la meilleure solution, le calcul de la solution *optimale* étant parfois inatteignables.
- **Architecture** ce domaine consiste à définir des architectures logicielles. Ces architectures vont mettre en application des concepts issus des *systèmes multi-agents* ou des techniques de raisonnement : architecture dite *cognitive* permettant de représenter des connaissances, buts, intentions et autres états mentaux.

- **Apprentissage** ce domaine consiste à définir des modèles permettant d'établir des connexions entre des ensembles de données. De fait, la proximité avec le *big data* et les techniques de fouille est importante (calcul de règles d'association). Le *raisonnement par cas* suit une approche similaire : face à une base de cas (exemple : des pannes), le système recherche un cas similaire ou proche et propose un diagnostic. Les techniques bayésiennes permettent de construire des graphes de dépendances entre données tout en permettant de gérer l'incertitude à l'aide de probabilités.

Les techniques les plus populaires aujourd'hui consistent à établir ces liens à l'aide de *réseaux de neurones*. Chaque neurone peut être vue comme un système actif si les données en entrée ont une certaine valeur. Les réseaux de neurones permettent de construire des relations de dépendance entre données extrêmement complexes.

La construction de ces relations, c'est-à-dire la phase d'apprentissage, peut être supervisée ou non : l'homme indique à la machine si les relations calculées sont correctes ou pas.

- **Planification** ce domaine consiste à *établir et exécuter une séquence d'actions* dont les effets sont connus afin de satisfaire un but (être dans un certain état). Par exemple, un robot « explorateur » doit déterminer les actions à exécuter afin de recueillir des échantillons. Les actions peuvent avoir des effets *incertains* ; de même l'évaluation de l'état courant peut être bruitée.
- **Raisonnement** ce domaine consiste à définir des modèles de représentation de la connaissance et des techniques de raisonnement pour les exploiter. Ces modèles peuvent être spécialisés afin de représenter des connaissances *spatio-temporels* (être avant, après, à côté...), des *connaissances sur les connaissances* (un agent 1 sait qu'un agent 2 connaît l'information A). Les connaissances peuvent être *imparfaites ou floues* et le raisonnement doit en tenir compte. Plus généralement, le raisonnement est *complexe* et l'évaluation de cette complexité est de première importance. La *théorie des jeux* issue des sciences économiques permet de représenter le comportement rationnel de multiples agents (enchères, votes, ...)
- **Système multi-agent** ce domaine revisite plusieurs domaines sous l'angle de la distribution : que signifie la *représentation de problèmes distribués* ? Quels sont les concepts spécifiques à la dimension multiple : *coordination, délégation*, connaissances à propos des autres agents et de leur expertise... Cette nouvelle dimension conduit à ne plus pouvoir considérer comme possible les algorithmes classiques ou les techniques de raisonnement. L'alternative est alors de se tourner vers les techniques de *simulation* (exemple : comportement de foules) ou des agents

avec des comportements et des comportements basiques font *émerger* des phénomènes.

- **Traitement du langage** ce domaine a pour objectif d'établir des modèles permettant de donner du sens à des documents sonores ou écrits. La première étape de reconnaissance des données d'entrée s'approche du traitement du signal. Les termes reconnus doivent être *représentés* pour ensuite être *classés*. Cette approche analytique se rapproche des techniques de raisonnement. L'apprentissage est de plus en plus populaire : le système apprend le sens du signal en entrée. Le système est alors capable après une étape d'apprentissage de prédire le sens d'une phrase. C'est la technique utilisée dans les assistants vocaux.

Application et modèle

Le mariage de ces applications offre une perspective sans fin à la définition d'applications embarquant de l'intelligence artificielle. La figure 2 présente une vue partielle des applications : ces applications peuvent elles-mêmes être placées dans différents domaines d'activité : ainsi la vision s'applique aussi bien en médecine que dans le domaine des transports. Le traitement du langage avec son impact sur la construction d'assistants vocaux se retrouve dans de multiples domaines comme par exemple les tuteurs intelligents ou le commerce en ligne.



Figure 2 : Applications et modèles

B. Annexe 2 : Liste des personnes interrogées par les rapporteurs

Colonel Laurent COLLORIG

Chef de la division du renseignement
Service Central du Renseignement Criminel (SCRC)
Directeur du programme PAVED

Commissaire Divisionnaire Yves GALLOT

Direction Centrale de la Sécurité publique
Sous-direction des missions de sécurité – Division des systèmes d'information opérationnelle
Auteur du bilan de l'expérimentation PredVol

Chef d'escadron Jérôme LAGASSE

Chargé de projets / chef du département études du Centre de Recherche de L'Ecole des Officiers de la Gendarmerie Nationale (CREOGN)
Doctorant

Colonel Patrick PERROT

Ingénieur Télécom – Docteur en Télécoms en IA
Commandant le groupement de gendarmerie départementale de la Haute-Marne (Chaumont)
À l'origine de l'emploi de méthodes mathématiques dans l'appréhension de la délinquance et de l'activité gendarmique et de l'outil PAVED lorsqu'il était chef de la division du renseignement au SCRC.
Auteur de nombreux articles sur l'analyse décisionnelle publiés dans des revues scientifiques

Cyril PIOTROWICZ

Doctorant en droit pénal, procédure pénale, sociologie criminelle
Université Jean-Moulin Lyon 3
Université de Technologie de Troyes
Ecole Nationale Supérieure de la Police
Contributeur au rapport pour le CHEMI "La police prédictive et la résilience organisationnelle des acteurs de la sécurité intérieure" sous la direction scientifique du Dr. Anne-Sophie Chavent-Leclere, Maître de conférences à la Faculté de Droit (Univ Lyon III).

VI. Bibliographie

Documentation produite en collaboration avec le Centre des Hautes Etudes du Ministère de l'Intérieur

Dir. sc. Anne-Sophie Chavent-Leclerc, *Rapport sur la police prédictive et la résilience organisationnelle des acteurs de la sécurité intérieure*, Projet porté par l'Equipe de recherche Louis Josserand de l'Université Jean-Moulin Lyon 3 en collaboration avec le CHEMI (2018).

Ministère de l'intérieur, *Le préfet et l'intelligence artificielle*, 37 CHEMI, JOURNEES D'ETUDES & DE REFLEXION (2018).

EUROPEAN CRIME PREVENTION NETWORK, PREDICTIVE POLICING RECOMMENDATIONS PAPER 13 (2016), https://eucpn.org/sites/default/files/content/download/files/recommendation_paper_predictive_policing_update.pdf.

WALTER L. PERRY ET AL., PREDICTIVE POLICING: THE ROLE OF CRIME FORECASTING IN LAW ENFORCEMENT OPERATIONS (2013), https://www.rand.org/pubs/research_reports/RR233.html (last visited Nov 29, 2018) (RAND Corporations).

Rapports en France et en Europe

CÉDRIC VILLANI, AI FOR HUMANITY 235 (2018), https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf.

CNIL, COMMENT PERMETTRE A L'HOMME DE GARDER LA MAIN ? RAPPORT SUR LES ENJEUX ETHIQUES DES ALGORITHMES ET DE L'INTELLIGENCE ARTIFICIELLE | CNIL, <https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes-et-de> (last visited Nov 29, 2018).

Communication from the commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, on Artificial Intelligence for Europe, Com(2018) 237 final.

The European Commission's High Level Expert Group on Artificial Intelligence, DRAFT ETHICS GUIDELINES FOR TRUSTWORTHY AI (2018).

N° 1335 - Rapport d'information de MM. Didier Paris et Pierre Morel-À-L'Huissier, Mission d'information sur les fichiers mis à la disposition des forces de sécurité, <http://www.assemblee-nationale.fr/15/rap-info/i1335.asp> (last visited, Nov 25, 2018).

La Documentation française, ETUDE ANNUELLE 2014 DU CONSEIL D'ETAT - LE NUMERIQUE ET LES DROITS FONDAMENTAUX, <http://www.ladocumentationfrancaise.fr/rapports-publics/144000541/index.shtml> (last visited, Nov 21, 2018).

ARTICLE 29 DATA PROTECTION WORKING PARTY, GUIDELINES ON AUTOMATED INDIVIDUAL DECISION-MAKING AND PROFILING FOR THE PURPOSES OF REGULATION 2016/679 (2017).

DON CASEY, PHILIP BURRELL & NICK SUMNER, DECISION SUPPORT SYSTEMS IN POLICING, <https://bulletin.cepol.europa.eu/index.php/bulletin/article/view/345/298> (last visited Dec 14, 2018).

Maps of the Future : A Modern Crime Analysis and Crime Prediction Based Tool to increase the Effectiveness and Quality of Public Administration performance in Crime Prevention, *Science and Research Institute* ACCENDO (2015).

Rapports aux Etats-Unis

US DEPARTMENT OF JUSTICE, INVESTIGATION OF THE FERGUSON POLICE DEPARTMENT (2015).
THE WHITE HOUSE, BIG DATA: A REPORT ON ALGORITHMIC SYSTEMS, OPPORTUNITY, AND CIVIL RIGHTS 29 (2016).

Executive Office of the President of the United States, National Science and Technology Council Committee on Technology, Preparing for the Future of Artificial Intelligence, October 2016.

Colloques

« Sécurité et justice : le défi des algorithmes », Colloque organisé à l'INHESJ, en juin 2017 : <https://inhesj.fr/evenements/videotheque/securite-et-justice>.

« Police prédictive », Journée d'étude professionnelle organisée par l'équipe de recherche Louis Josserand et l'IEJ de Lyon avec le soutien du CHEMI, Janv. 2018.

Revues et articles

Gendarmerie nationale, ALGORITHMES ET ESPACES NORMATIFS, REVUE N° 261 GENDARMERIE, /crngn/Publications/Revue-de-la-gendarmerie-nationale/Revue-N-261 (last visited Nov 24, 2018).

Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671–732 (2016).

Bilel Benbouzid & Dominique Cardon, MACHINES À PRÉDIRE | CAIRN.INFO, <https://www.cairn.info/revue-reseaux-2018-5-page-9.htm?contenu=resume> (last visited Nov 27, 2018).

Bilel Benbouzid, QUAND PRÉDIRE, C'EST GÉRER | CAIRN.INFO, <https://www.cairn.info/revue-reseaux-2018-5-page-221.htm> (last visited Nov 27, 2018).

Bilel Benbouzid, *Des crimes et des séismes*, n° 206 RESEAUX 95–123 (2017).

Ismael Benslimane, *Étude critique d'un système d'analyse prédictive appliqué à la criminalité : PredPol®*, CORTEX JOURNAL 4 (2014).

Philippe Besse, Céline Castets-Renard & Aurélien Garivier, LOYAUTE DES DECISIONS ALGORITHMIQUES (2017), <https://hal.archives-ouvertes.fr/hal-01544701> (last visited Nov 21, 2018).

Philippe Besse, Céline Castets-Renard, Aurélien Garivier & Jean-Michel Loubès, *L'IA du Quotidien peut-elle être Éthique ?*, [HTTPS://ARXIV.ORG/PDF/1810.01729 27](https://arxiv.org/pdf/1810.01729v2.pdf).

Lyria Bennett Moses & Janet Chan (2018). Algorithmic prediction in policing: assumptions, evaluation, and accountability, *Policing and Society*, 28:7, 806-822.

Rosamunde van Brakel, *Pre-emptive Big Data Surveillance and Its (Dis)Empowering Consequences: the Case of Predictive Policing*, in EXPLORING THE BOUNDARIES OF BIG DATA, VAN DER SLOOT ET AL (EDS.) (2016), http://www.academia.edu/24974497/Exploring_the_Boundaries_of_Big_Data_editions (last visited Dec 12, 2018).

P Jeffrey Brantingham, *The Logic of Data Bias and Its Impact on Place- Based Predictive Policing*, 15 14.

P. Jeffrey Brantingham, Matthew Valasik & George Mohler, DOES PREDICTIVE POLICING LEAD TO BIASED ARRESTS? RESULTS FROM A RANDOMIZED CONTROLLED TRIAL RESEARCHGATE, https://www.researchgate.net/publication/323027247_Does_Predictive_Policing_Lead_to_Biased_Arrests_Results_from_a_Randomized_Controlled_Trial (last visited Nov 29, 2018).

Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023–1045 (2017).

Alexandra Chouldechova, *Fair prediction with disparate impact: A study of bias in recidivism prediction instruments*, ARXIV:1610.07524 [CS, STAT] (2016), <http://arxiv.org/abs/1610.07524> (last visited Aug 12, 2018).

Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014).

Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. (2008).

Lucie Cluzel-Métayer, *La loi pour une République numérique : l'écosystème de la donnée saisi par le droit*, AJDA 340 (2017).

Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C.L. REV. 93 (2014).

Jean-Marc Deltorn, LA PROTECTION DES DONNEES PERSONNELLES FACE AUX ALGORITHMES PREDICTIFS | REVUE DES DROITS ET LIBERTES FONDAMENTAUX, <http://www.revuedlf.com/droit-ue/la-protection-des-donnees-personnelles-face-aux-algorithmes-predictifs/> (last visited Nov 24, 2018).

Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a Right to Explanation is Probably Not the Remedy You are Looking for*, SSRN ELECTRONIC JOURNAL (2017), <https://www.ssrn.com/abstract=2972855> (last visited Aug 12, 2018).

Lilian Edwards & Michael Veale, *Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions'?* (2018), <https://papers.ssrn.com/abstract=3052831> (last visited Dec 5, 2018).

Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion* (2012), <https://papers.ssrn.com/abstract=2050001> (last visited Dec 5, 2018).

ANDREW GUTHRIE FERGUSON, *THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT* (2017).

Batya Friedman & Helen Nissenbaum, *Bias in Computer Systems*, 14 ACM TRANS. INF. SYST. 330–347 (1996).

Antoine Garapon, *Les enjeux de la justice prédictive*, 1–2 SEMAINE JURIDIQUE, ED. GÉNÉRALE 4 (2017).

Bryce Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"*, 38 AI MAGAZINE 50 (2017).

Bryce W Goodman, *A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection*, 9.

Elizabeth E. Joh, *Policing by Numbers: Big Data and the Fourth Amendment* (2014), <https://papers.ssrn.com/abstract=2403028> (last visited Dec 5, 2018).

Elizabeth E. Joh, *Feeding the Machine: Policing, Crime Data, & Algorithms* (2017), <https://papers.ssrn.com/abstract=3020259> (last visited Dec 5, 2018).

Elizabeth E. Joh, *The Undue Influence of Surveillance Technology Companies on Policing* (2017), <https://papers.ssrn.com/abstract=2924620> (last visited Dec 5, 2018).

Dimitra Kamarinou, Christopher Millard & Jatinder Singh, *Machine Learning with Personal Data: Profiling, Decisions, and the EU General Data Protection Regulation*, 7.

Margot E. kaminsky, *THE RIGHT TO EXPLANATION, EXPLAINED* (2018).

Danielle Kehl, Priscilla Guo & Samuel Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY,

HARVARD LAW SCHOOL (2017), <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041> (last visited Aug 12, 2018).

Joshua Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633 (2017).

Jeff Larson et al., *HOW WE ANALYZED THE COMPAS RECIDIVISM ALGORITHM* PROPUBLICA (2016), <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> (last visited Aug 12, 2018).

Kristian Lum & William Isaac, *To predict and serve?*, 13 SIGNIFICANCE 14–19 (2016).

VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, *BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK* (2013).

GIOVANNI MASTROBUONI, *CRIME IS TERRIBLY REVEALING: INFORMATION TECHNOLOGY AND POLICE PRODUCTIVITY* (2017), <https://papers.ssrn.com/abstract=2989914> (last visited Dec 12, 2018).

Brent Daniel Mittelstadt et al., *The ethics of algorithms: Mapping the debate*, 3 BIG DATA & SOCIETY (2016).

George Mohler, *Marked point process hotspot maps for homicide and gun crime prediction in Chicago*, 30 INTERNATIONAL JOURNAL OF FORECASTING 491–497 (2014).

Helen Nissenbaum, *Accountability in a computerized society*, 2 SCI ENG ETHICS 25–42 (1996).

CATHY O'NEIL, *WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY* (First edition ed. 2016).

FRANK PASQUALE, *THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION* (2015).

Jean-Marc Pastor, *Accès aux traitements algorithmiques utilisés par l'administration*, AJDA 604 (2017).

Patrick Perrot, *What about AI in criminal intelligence? From predictive policing to AI perspectives*, 1 65–75 (2017).

Patrick Perrot, *L'analyse du risque criminel: l'émergence d'une nouvelle approche*, RESEARCHGATE, https://www.researchgate.net/publication/274071556_L'analyse_du_risque_criminel_l'emergence_d'une_nouvelle_approche (last visited Nov 29, 2018).

Jackson Polansky & Henry Fradella, *Does "Precrime" Mesh with the Ideals of U.S. Justice? Implications for the Future of Predictive Policing* (2016), <https://papers.ssrn.com/abstract=2832365> (last visited Dec 5, 2018).

David Pontille & Didier Torny, *La manufacture de l'évaluation scientifique. Algorithmes, jeux de données et outils bibliométriques*, 177 RESEAUX 23–61 (2013).

ENHANCING JUSTICE: REDUCING BIAS, (Sarah E. Redfield & American Bar Association eds., 2017).

Judith Rochfeld, *L'encadrement des décisions prises par algorithme*, DALLOZ IP/IT 474 (2018).

Judith Rochfeld (dir.), avec C. Castets-Renard, N. Martial-Braz, C. Zolynski, *Décryptage de la loi sur la protection des données personnelles*, Dalloz, 2019 (à paraître).

Antoinette Rouvroy, *La "digitalisation de la vie même": enjeux épistémologiques et politiques de la mémoire digitale*, 47 DOCUMENTALISTE - SCIENCES DE L'INFORMATION 63–64 (2010).

Salvatore Ruggieri, *Using t-closeness anonymity to control for non-discrimination*, 31 (2014).

ANDREW D. SELBST, DISPARATE IMPACT IN BIG DATA POLICING (2017), <https://papers.ssrn.com/abstract=2819182> (last visited Dec 5, 2018).

DANIEL SOLOVE & PAUL SCHWARTZ, INFORMATION PRIVACY LAW (6e ed. 2018), <https://www.informationprivacylaw.com/> (last visited Dec 13, 2018).

ALAIN SUPLOT, LA GOUVERNANCE PAR LES NOMBRES, COURS AU COLLEGE DE FRANCE (2012-2014) (Fayard ed. 2015).

LINNET TAYLOR, LUCIANO FLORIDI & BART VAN DER SLOOT, GROUP PRIVACY - NEW CHALLENGES OF DATA TECHNOLOGIES (2017), <https://www.springer.com/gb/book/9783319466064> (last visited Dec 4, 2018).

Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System* (2017), <https://papers.ssrn.com/abstract=2920883> (last visited Dec 5, 2018).

Rashida Richardson, Jason Schultz & Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, NEW YORK UNIVERSITY LAW REVIEW ONLINE, FORTHCOMING (2019).

Indre Zliobaite, *A survey on measuring indirect discrimination in machine learning*, ARXIV:1511.00148 [CS, STAT] (2015), <http://arxiv.org/abs/1511.00148> (last visited Nov 21, 2018).

Brevets

Palantir, Crime Risk Forecasting, Patent N°: US9,129,219B1

PredPol, Event Forecasting System, Georges O. Mohler (déposant et inventeur) et PredPol, Inc. (cessionnaire), Patent N°: US8,949,164