

Manon Jendly

## Du risque clinique au risque algorithmique: perspectives et enjeux des pratiques expertales psychiatriques dans le système pénal

### Résumé

La mission d'aide à la décision judiciaire confiée aux psychiatres a connu de profonds bouleversements ces dernières années en Suisse. Alors que leur était traditionnellement dévolu la tâche d'évaluer la responsabilité pénale du sujet soumis à leur expertise, ils doivent depuis 2007 aussi évaluer les risques potentiels qu'il représente. Cet article rend compte de l'évolution progressive d'une évaluation du risque fondée sur les savoirs psychiatriques vers un risque actuariel, issu de recherches criminologiques empiriques. Il envisage aussi quelle pourrait être sa suite, déjà éprouvée à l'étranger, à savoir un risque «augmenté», fondé sur les mégadonnées et les instruments issus de l'intelligence artificielle destinés à les exploiter. Si cette perspective est réputée prometteuse pour certains, elle n'est pas exempte d'enjeux juridiques, sociaux, éthiques et politiques.

*Mots-clés:* risque, expertise, justice criminelle, big data, intelligence artificielle.

### Zusammenfassung

Die den Psychiatern anvertraute Funktion als Entscheidungsgehilfen der Strafjustiz hat in den letzten Jahren in der Schweiz tiefgreifende Veränderungen erfahren. Während sie traditionell für die Beurteilung der strafrechtlichen Verantwortlichkeit des Täters zuständig waren, müssen sie seit 2007 auch deren potenzielles Delikts- und Rückfallrisiko bewerten. Der vorliegende Artikel behandelt die schrittweise Entwicklung der Risikoevaluation. Diese erfolgte anfänglich gestützt auf die Grundlage psychiatrischer Erkenntnisse und wurde schliesslich zu einer aktuariellen Risikoanalyse auf Grundlage der empirischen, kriminologischen Forschung. Er zeigt zudem auch die im Ausland bereits getesteten Weiterentwicklungen auf: nämlich die Einschätzung eines «erhöhten» Risikos, die sich auf Big Data-Auswertungen und auf Instrumente künstlicher Intelligenz abstützen. Während diese Entwicklungsperspektiven für einige als vielversprechend angesehen werden, werfen diese doch nicht zu vernachlässigende rechtliche, soziale, ethische und politische Fragen auf.

*Schlüsselwörter:* Risiko, Gutachten, Strafjustiz, big data, künstliche Intelligenz.

### Summary

The mission entrusted to forensic psychiatrists as auxiliaries to the judicial decision-making process has undergone profound upheavals in recent years in Switzerland. While they were traditionally asked to assess the criminal responsibility of the defendant, since 2007 they are also called upon to assess the potential risks the defendant represents for public safety. This article discusses the gradual evolution of risk assessment practices, from a «clinical» risk based on psychiatric knowledge to an «actuarial» risk, based on empirical

criminological research. It also considers what could be the next step, already tested abroad, namely an «augmented» risk, fueled with Big data and Artificial intelligence. While some argue that this perspective is promising, it still raises substantial legal, social, ethical and political challenges.

*Keywords:* risk, expertise, criminal justice, big data, artificial intelligence.

### Introduction

Cet article discute des pratiques expertales psychiatriques en matière pénale en les situant dans le contexte plus large de sociétés occidentales obnubilées par les risques, quels qu'ils soient (Giddens 1990; Bauman 2007). Concept protéiforme, le risque façonne en effet une multitude d'interventions dans des champs aussi variés que la médecine, l'éducation, l'environnement ou encore la sécurité (Aven 2015). Pour certains, les politiques, pratiques et rhétoriques du risque révèlent l'état de nos situations socio-économiques, culturelles et politiques contemporaines. Ainsi est-il question d'une «société du risque» (Beck 1992) pour référer à la façon dont cette notion, dès le tournant des années 70, traduit les incertitudes inhérentes à notre condition humaine, nos activités et nos interactions. Mais le concept de risque peut aussi être appréhendé de façon plus restrictive, à l'aune des attributs techniques et probabilistes qui lui sont prêtés. Le risque et sa survenance sont alors présumés calculables, voire pronosticables (Prins/Reich 2018). Dans ce cadre, il permet d'opérationnaliser des connaissances à des fins prédéterminées, par exemple prévenir un dommage, et relaie alors en filigrane une idée de prédiction. Jamais ainsi n'avons-nous été aussi conscients des risques qui nous entourent, renseignés sur leur ampleur potentielle en termes chiffrés et le cas échéant, peu tolérants face à leur réalisation. Ces différents registres du risque ne sont donc pas mutuellement exclusifs: le risque dans sa définition la plus étroite s'ins-

crivant dans le contexte de son acception la plus large.

Pour le propos de cette contribution, le risque réfère à la probabilité qu'un événement se produise dans une certaine fenêtre temporelle. C'est cette perspective «technique» que retient le système de justice criminelle, qui a progressivement consacré la notion de risque au détriment de celle de «dangerosité», estimée trop nébuleuse et pseudo-scientifique (Ericson/Haggerty 1997; Garland 2003; Simon, 2007; O'Malley 2010). Depuis les années 90 outre-Atlantique et le début des années 2000 en Suisse, le risque criminel, son calcul et sa prédiction irriguent toute la chaîne pénale et influencent aussi les pratiques des experts psychiatres sollicités pour éclairer un magistrat dans ses prises de décision (Hannah-Moffat 2013). Plus récemment, des avancées techniques, technologiques et scientifiques fulgurantes lui donnent un nouveau souffle. A tel point que la justice pénale est susceptible de se distancer des expertises traditionnellement convoquées pour l'aider dans ses décisions, en faveur de modèles analytiques fondés sur la toute-puissance des données et des outils computationnels.

Partant, je formule ici l'hypothèse que si les pratiques expertales d'évaluation et de gestion des risques criminels oscillent encore aujourd'hui en Suisse entre appréciation clinique (semi) structurée et évaluation criminologique actuarielle et standardisée, leurs disputes seront bientôt réglées par l'arrivée sur le marché de nouveaux dispositifs. Je rejoins en cela ceux qui font le pari qu'elles seront demain concurrencées par des instruments automatisés d'extraction et d'analyse de données (partiellement) supervisés (Berk/Hyatt 2015). Ces dispositifs rendent compte de résultats en temps réel, présumés les mieux à même d'inspirer les décideurs et d'orienter les actions des intervenants. Il me sera rétorqué que nous sommes en Suisse et non en Amérique du Nord et que nos systèmes de justice diffèrent substantiellement. Je réponds que l'histoire est jalonnée d'initiatives de tous ordres, y compris en matière judiciaire, initiées outre-Atlantique et récupérées quelques années plus tard dans nos contrées, pour le meilleur ou pour le pire. La justice helvétique n'est pas hermétique aux tendances lourdes qui façonnent présentement ses homologues étrangers. Mon hypothèse par ailleurs ne peut que s'avérer plus prophétique (sans

mauvais jeu de mots) dans le contexte actuel de recueil massif de données sur nos personnes et nos activités dont sont tirées des analyses comportementales, ventilant les individus dans des catégories plus ou moins vulnérables en termes d'emploi, de crédit, de santé ou encore de déviance (O'Neil 2016; Eubanks 2018). Par conséquent, il y a fort à penser que les promesses allouées aux mégadonnées (*Big data*) et les instruments dédiés à les «faire parler» pour mitiger les risques criminels surinvestissent également les prétoires helvétiques, les établissements pénitentiaires et l'intervention en milieu ouvert très rapidement.

La première partie de cette contribution présente un bref panorama de la consécration de l'approche dite «orientée risque» dans le système de justice pénale et les controverses l'entourant. La deuxième partie s'intéresse aux principales stratégies d'évaluation des risques criminels à des fins d'aide à la détermination de la sanction, et révèle un glissement progressif d'une «clinique du risque» à un «risque actuariel». Les outils d'aide à la décision qui découlent de cette évolution sont objets de débats, exposés dans la troisième partie. Dans la foulée, la quatrième partie discute brièvement la production exponentielle de données exploitées à des fins multiples et variées dans nos sociétés contemporaines, et à leur suite le déploiement de dispositifs d'analyse toujours plus perfectionnés destinés à les interroger, y compris au moyen d'algorithmes réputés prédictifs relevant de l'intelligence artificielle. La cinquième partie rend compte de l'appropriation progressive dans le système de justice pénale de ces techniques de *machine learning* sous la forme de «*automated data-driven risk assessment tools*». Ce faisant, elle confirme le passage d'une prise de décision inspirée de la clinique et de la criminologie à une prise de décision algorithmique auto-apprenante. La sixième et dernière partie identifie les principaux enjeux entourant ces instruments, avant de les discuter plus largement sous l'angle du projet politique qu'ils sont susceptibles de charrier.

## 1. L'approche orientée risque: un paradigme dominant mais controversé

Depuis plusieurs décennies une attention toujours plus importante est prêtée au concept de

risque dans les systèmes de justice pénale occidentaux. Cette tendance s'exprime notamment à travers des normes, pratiques et outils destinés non seulement à évaluer les risques criminels, à savoir estimer leur ampleur, mais aussi les réduire, à l'appui d'interventions dédiées (Viljoen *et al.* 2018). Ensemble, ces éléments forment ce qui est désormais courant de nommer «l'approche orientée risque». Initialement circonscrite à la prévention de la récidive, cette approche se déploie désormais tout au long de la chaîne pénale, de la police à la probation. S'agissant de la justice criminelle, elle imprègne toutes les décisions afférentes aux différentes étapes qui la jalonnent, particulièrement en matière de détention préventive, de détermination de la peine et de libération conditionnelle (Garrett/Monahan 2018).

A l'origine de cette approche figurent indubitablement les travaux de deux canadiens, J. Andrews et D. A. Bonta, ainsi que leurs collègues P. Gendreau et S. Wormith. On leur doit notamment la boîte à outils d'évaluation des risques la plus utilisée dans les milieux correctionnels occidentaux: les «*LS family tools*» (Dume/Roque 2017)<sup>1</sup>. Dans leur ouvrage-maître publié pour la première fois en 1994, *The Psychology of Criminal Conduct*, Andrews et Bonta présentent un cadre d'évaluation et de traitement des personnes placées sous main de justice articulé sous l'acronyme RNR, et dont le premier R, précisément, réfère à la notion de risque: *Risk-Need-Responsivity*. Cette approche représente désormais le paradigme dominant en matière d'évaluation et de gestion des risques au pénal. Erigée en «modèle», elle a ainsi inspiré maints systèmes d'exécution des sanctions dans le monde entier, y compris en Suisse sous l'acronyme ROS pour *Risikoorientierter Sanktionenvollzug*. Lancé en 2010 sous la forme d'un projet pilote au sens de l'art. 387 al. 4 CP, ROS a depuis lors été adopté dans les deux Concordats alémaniques en matière pénitentiaire et est en voie d'adaptation en Suisse latine (Brägger 2015).

Si l'approche orientée risque est consacrée dans la grande majorité des systèmes de justice pénale occidentaux, elle n'en demeure pas moins sujette à controverses. Elle fait ainsi

l'objet d'intenses débats dans les milieux académiques. Là où ses promoteurs évoquent à son appui une diminution de l'arbitraire inhérente à la justice «humaine», ainsi qu'une meilleure allocation des ressources, ses opposants insistent sur son caractère profondément totalisant et son inclinaison à exacerber des discriminations préexistantes (Starr 2014; Harcourt 2015). Les milieux professionnels intéressés la discutent également âprement. L'importance allouée à cette approche a notamment révélé une scission entre des décideurs qui l'imposent au motif d'une gestion plus «efficace» des populations judiciarisées, et des intervenants qui peinent à trouver du sens aux missions de «gestionnaires de risques» qui leur sont assignées, au détriment estiment-ils de leurs tâches d'aide à la réinsertion sociale (Jendly 2012; Dubourg/Gautron 2014). Enfin, cette approche ne fait pas non plus l'unanimité dans le monde politique, qui demeure divisé sur les finalités de la peine. Une approche focalisée sur les risques de délit privilégie en effet un but d'incapacitation (sélective) pour protéger la société, au détriment éventuellement d'autres objectifs, orientés sur les personnes et fondés sur la croyance en un Homme perfectible, bien que faillible (Monahan/Skeem 2016).

Ces débats placent les uns et les autres dans une opposition dichotomique que n'aident pas à réconcilier les tenants des études empiriques sur le risque. En effet, si l'on dispose maintenant d'une littérature robuste constituée de méta-analyses et revues systématiques des modèles et outils orientés risque, elle n'en demeure pas moins tout aussi sujette à controverses. S'agissant des publications portant sur l'évaluation des risques, cette littérature ne reflète pas de consensus quant à la validité et la fiabilité des outils qui en relèvent. La situation est similaire en matière de gestion des risques, puisque les résultats des évaluations des programmes destinés à les réduire sont tantôt mitigés, tantôt contradictoires (Prins/Reich 2018). En définitive, si le risque est omniprésent dans la justice criminelle actuelle, son appréhension est pétrie d'ambivalences: tout le monde en parle mais personne ne s'accorde ni sur ses délimitations conceptuelles, ni sur ses objectifs réels, ni sur ses potentiels. Dans ce contexte, trois corpus de savoirs disposent encore du monopole pour en discuter, à savoir la psychiatrie, la psychologie et la criminologie.

<sup>1</sup> *Level of Supervision Inventory* (LSI, 1982); *Level of Service Inventory-Revised* (LSI-R, 1995); *Level of Service/Case management Inventory* (LS/CMI, 2004).

## 2. D'une «clinique du risque» à un «risque actuariel»

Les pratiques expertales dont il est question ici couvrent des activités d'évaluation et de gestion des risques liés aux personnes. L'évaluation se rapporte à un exercice d'appréciation, en un temps et un lieu donnés, de la propension d'un individu à adopter un comportement contraire aux normes établies (tel est le cas par exemple de l'évaluation d'un risque de récidive spécifique), ou de se soustraire à une injonction qui lui est faite (par exemple, pour un prévenu en attente de jugement, de ne pas se présenter devant le juge). La gestion du risque se rapporte elle aux interventions susceptibles de réduire son occurrence ou à tout le moins sa gravité. Pour certains auteurs, l'évaluation des risques criminels devrait toujours comporter des indications en termes de gestion et prioritairement poursuivre un but d'intervention (Barabas *et al.* 2018). Elle devrait alors être considérée comme un exercice menant à un diagnostic, plutôt qu'un pronostic. Pour d'autres, bien au contraire, la notion de risque, par nature, en appelle toujours à un pronostic et traduit un idéal prédictif (Helmus/Balbschishin 2016). Si l'évaluation n'offre donc pas de façon systématique d'avenue en matière d'intervention, elle s'essaie en revanche toujours à prédire de façon quantifiable la propension d'un individu à développer des «problèmes de comportement», voire passer à l'acte.

À l'heure actuelle en Suisse, les magistrats au pénal ne sollicitent pas d'évaluation du risque à proprement parler. C'est une expertise psychiatrique qu'ils sollicitent, soit parce que la loi la leur impose, soit parce qu'ils l'estiment nécessaire pour mieux se déterminer. Les acteurs érigés ici en experts doivent impérativement être des spécialistes de la psychiatrie, détenteurs d'un titre FMH et idéalement d'une formation *ad hoc* en psychiatrie légale. Sans entrer dans les détails de leur exercice, il est désormais bien établi que leur rôle en matière d'évaluation a progressivement glissé ces dernières années d'une évaluation de la responsabilité (art. 20 CP) à celle du risque de commission d'une (nouvelle) infraction (art. 56 al. 3 CP). En matière de détermination de la peine très spécifiquement, l'importance de condamner une personne eu égard à l'infraction commise dans le passé s'est déplacée vers l'attention accrue portée à son évaluation prospective des-

tinée à envisager quels seront ses agirs demain (Loup 2018). Les résultats de l'expertise, et en particulier le pronostic qu'elle relaie, viennent en soutien aux décisions des magistrats qui ne peuvent s'en distancer qu'en argumentant solidement leur prise de position.

Pour les experts psychiatres, l'obligation qui leur est faite de pronostiquer un risque de (nouveau) passage à l'acte a des conséquences pratiques, cliniques et éthiques largement discutées dans la littérature, y compris par les psychiatres eux-mêmes (Fonjallaz/Gasser 2017; Gravier 2009). Cette situation engendre des confusions profondes. Dans un système de sanctions dualiste comme le nôtre, elle est source notamment de fausses croyances selon lesquelles le crime serait une maladie et le criminel a fortiori un malade. Ainsi l'inflation des thérapies institutionnelles ordonnées en vertu de troubles mentaux (art. 59 CP), respectivement l'augmentation de la durée moyenne des séjours en exécution de mesure, ne sont-elles probablement pas si étrangères à cette «nouvelle» vocation allouée à l'expertise psychiatrique (Ferreira/Maugué 2017; Fink 2017). Or, comme les actes incriminés ne sont que rarement en lien avec des problèmes de santé mentale, il serait légitime que les psychiatres collaborent plus systématiquement avec des criminologues. Dans certains cantons, des premières tentatives de travail en équipe pluridisciplinaire, alliant criminologues, psychiatres et psychologues, reposent précisément sur ce constat. De plus, de gros efforts ont été fournis par des psychiatres pour une meilleure structuration des expertises légales (Dittman 1999; Moulin/Gasser 2012; Hachtel *et al.* 2018) et l'élaboration de cursus de formation dédiés (Delacrausaz 2017). Leurs pratiques ne reposent donc plus seulement sur leur orientation, leur expérience, leur intuition et leurs lectures. Elles font désormais l'objet d'une spécialisation et revendiquent aussi s'appuyer toujours plus sur des outils relevant du jugement professionnel structuré (JPS). Ces instruments systématisent l'évaluation autour de facteurs réputés nécessaires d'être investigués par le professionnel, lequel demeure toutefois libre d'apprécier le poids de chacun et leur rapport respectif (Gravier/Lustenberger 2005; Delacrausaz/Gasser 2012).

En parallèle à ces guides issus des sciences psychiatriques, une explosion de recherches empiriques en psychologie et en criminologie

sont à l'origine de l'élaboration de questionnaires standardisés, inspirés des sciences actuarielles. Accessibles sous format électronique, ces questionnaires s'articulent autour des principaux facteurs identifiés comme constitutifs de bons prédicteurs de problèmes de comportement. Les instruments ciblent des problématiques définies soit en termes généraux (par ex. violence, récidive), soit en termes plus spécifiques (par ex. violence domestique, délit sexuel). Leurs produits se présentent toujours sous un résultat chiffré (score) rendant compte d'un niveau (cote) de risque qualifié le plus souvent de faible, modéré ou élevé. Les outils relevant de la famille précitée des «LS» comptent parmi les dizaines d'instruments de ce type recensés en matière d'évaluation des risques criminels liés aux personnes, adultes et mineures, à usages pré-, inter- et post-sentenciels (Kehl *et al.* 2017).

A titre de synthèse et en l'état actuel des connaissances, il est d'usage de ventiler les différentes stratégies d'évaluation existantes en quatre générations (Guay *et al.* 2015; Barabas *et al.* 2018). La première réfère aux évaluations cliniques non structurées, n'incluant aucun mécanisme statistique susceptible de valider les décisions en ressortant. Elles sont largement critiquées pour leur caractère trop subjectif et leur capacité prédictive légèrement supérieure au hasard, multipliant au demeurant les faux positifs. Capitalisant sur les travaux de Burgess (1928), la deuxième génération s'appuie sur des modèles de régression statistiques réputés plus performants en matière d'analyse prédictive, permettant aux chercheurs d'identifier des variables liées empiriquement à la récidive, sans toutefois nécessairement avoir à comprendre pourquoi certaines sont plus significatives que d'autres. De type actuariel, les instruments qui en relèvent reposent sur des facteurs statiques, tels que l'âge ou les antécédents judiciaires, facilement codifiables. Moins ouverte à l'interprétation, cette deuxième génération donne lieu à des évaluations standardisées et qui ne sont pas susceptibles de mesurer le changement. Surtout, elle fait entrer la justice criminelle dans une ère que certains observateurs ont qualifié de «nouvelle pénologie», à savoir le passage d'un système axé sur la punition et/ou le traitement des individus à celui misant sur la surveillance et le contrôle de groupes, identifiés au moyen d'outils spécifiques de calculs de risques, de type probabi-

liste (Feeley, Simon 1992). Les facteurs statiques autour desquels s'articulent les stratégies de deuxième génération figent en effet les individus autour d'indicateurs qui ne peuvent être modifiés, donnant lieu à des approches managériales et incapacitantes. Le début des années 80 marque ainsi l'arrivée d'une troisième génération d'outils, destinée à assouplir le caractère implacable de la précédente en incorporant aux variables statiques des facteurs dynamiques, susceptibles d'évolution. La valeur prédictive des facteurs dynamiques est moins grande que celle des facteurs statiques, mais ils peuvent explorer les possibilités de changements et orienter les interventions en conséquence, palliant de la sorte les critiques émises à l'égard de ces prédécesseurs. Enfin, dans la foulée des efforts entrepris pour donner des indications de traitement, la quatrième génération d'instruments mise sur une évaluation intégrée des variables statiques et dynamiques, à laquelle sont ajoutés des facteurs de réceptivité, entendus comme des éléments susceptibles d'améliorer les résultats des interventions. Parmi ces derniers, une attention particulière est allouée au niveau d'intelligence, à l'estime de soi ou encore à des troubles de la santé mentale. Cette quatrième génération vise donc non seulement à poser un pronostic sur des risques spécifiques, mais poursuit aussi, désormais ouvertement, l'ambition de les réduire.

Là où les JPS mettent l'accent sur des dimensions en lien avec une souffrance psychique, la quatrième génération d'instruments s'articule autour des facteurs associés à huit principaux domaines dits «criminogènes» (les «*big eight*»), à savoir les antécédents criminels, la personnalité à tendance anti-sociale, les attitudes anti-sociales, les pairs anti-sociaux, l'éducation et/ou l'emploi, la famille, la consommation et les loisirs et activités récréatives (Andrews, Bonta 2010). Elaborés à partir d'études empiriques, le plus souvent des enquêtes longitudinales de cohortes de sortants de prison, ces outils s'inscrivent dans un mouvement de «scientification» de l'évaluation, essentiellement orientée sur le pronostic. L'idée dans cet article n'est toutefois pas tant de confronter ces différentes sources d'expertise, dès lors que leurs champs d'investigation respectifs plaideraient plutôt en faveur de leur maillage en la matière. En revanche, il convient d'insister sur ce qu'ils ont *en commun*, à savoir leur objectif

principal de guider les décideurs dans leur prise de décision, laquelle apparaît toujours plus orientée sur le futur.

### 3. Des instruments d'aide à la décision aussi adulés que critiqués

Peu importe la nature du système juridique en question, l'ensemble des instruments discutés jusqu'à présent ont pour objectif principal d'informer les décideurs et jamais de s'y substituer (Monahan/Skeem 2016). En matière de détermination de la sanction plus spécifiquement, ce sont aux juges auxquels revient le difficile exercice, *in fine*, de se prononcer sur le seuil de risque acceptable. Leurs décisions sont toutefois façonnées par un certain nombre de contraintes, liées aux moyens limités à leur disposition, notamment en termes d'information et de temps, mais aussi au contexte social, politique et médiatique les entourant. Elles peuvent aussi être viciées par les sensibilités, les motivations et les préjugés des magistrats. Dans ces conditions, les outils actuariels sont appréhendés comme de bons moyens pour légitimer les biais et interférences inhérents à la justice humaine, à tout le moins en partie. Dans les pays qui y recourent de façon systématique, tels qu'en Grande-Bretagne, au Canada et dans plusieurs états des États-Unis, ils sont ainsi plébiscités au motif qu'ils objectivent la décision judiciaire et améliorent la célérité du processus de prise de décision. Dès lors qu'ils offrent des prévisions basées sur des données historiques, investiguant les mêmes facteurs pour chaque personne, et notant/scorant tout le monde suivant le même procédé, ils sont réputés homogénéiser les éléments à la source de l'évaluation, neutraliser tout glissement arbitraire et décharger une justice saturée (Ciccolini/Conti 2018). Pour autant, ces outils sont aussi vertement critiqués, au motif notamment qu'ils reposent sur une logique nomothétique, et qu'ils se focalisent essentiellement sur les risques, lesquels ne sont au demeurant pas toujours bien qualifiés. Leur prétendue simplicité est également âprement débattue. Car s'ils sont réputés aisés à appliquer, interpréter leurs résultats requiert des connaissances dont les magistrats ne sont pas tous détenteurs, donnant lieu à des raccourcis, voire des conclusions erronées. Leur propension à déresponsabiliser les magis-

trats enfin est aussi pointée du doigt, l'outil pouvant servir de moyen bien commode pour se décharger de la motivation du jugement, en se limitant à invoquer ses résultats (Bennett Moses/Chan, 2014).

Il n'empêche, à l'heure actuelle, l'argument principal évoqué à l'appui du recours à ces instruments n'est pas tant celui de réduire l'arbitraire que celui de gérer les incertitudes. Car c'est en matière de pronostic que les attentes sont les plus grandes et les limites humaines les plus évidentes. Certes, la prédiction actuarielle surpasse la prédiction clinique (Dawes *et al.* 1989; Mossman 1994; Meehl 1996; Grove *et al.* 2000). Mais elle est largement perfectible. La méta-revue de Singh/Fazel (2010), par exemple, conclut à une grande hétérogénéité des résultats entre les 31 méta-analyses et neuf revues systématiques soumises à leur examen, s'agissant notamment de la validité et de la performance prédictives des outils les plus couramment utilisés. Au-delà, ces derniers se voient aujourd'hui concurrencés par de nouveaux dispositifs sur le marché de l'évaluation, auxquels précisément sont prêtés des potentiels prédictifs plus importants. Cette évolution s'inscrit dans le contexte plus général de l'avènement des *Big data* et de la numérisation croissante de nos sociétés, devenues «digitales».

### 4. Les promesses allouées aux «Big data»

Plus qu'un buzz word, le terme *Big data*, littéralement mégadonnées, est généralement défini en référence au *volume*, à la *variété* et à la *vitesse* des données susceptibles d'être collectées, stockées et analysées par une diversité de technologies pour générer des informations sur des comportements et/ou des situations (Smith *et al.* 2017)<sup>2</sup>. De façon plus située et suivant la définition qu'en donnent boyd/Crawford (2012)<sup>3</sup>, les *Big data* traduisent un phénomène culturel, technologique et scientifique conjuguant technologie, analyse et mythologie. La technologie tout d'abord, puisqu'à l'heure où une multitude de supports différents pro-

<sup>2</sup> Aux trois «V» sont désormais ajoutés sept nouvelles dimensions parmi lesquelles l'exhaustivité, l'indexation, la flexibilité et la nature relationnelle. V. Kitchin (2014).

<sup>3</sup> Le premier nom étant intentionnellement écrit sans majuscule (danah boyd), selon la graphie choisie par cette auteure.

duisent au quotidien une abondance de données sur nos personnes, des dispositifs toujours plus perfectionnés sont développés pour optimiser la puissance de calcul et la précision algorithmiques pour les réunir, les lier et les comparer afin d'en extraire de l'information. L'analyse ensuite, car ces vastes ensembles de données sont soumis à des méthodes d'investigation destinées à faire ressortir des valeurs (par exemple sous forme de classements ou de hiérarchies), scores (par exemple sous forme de probabilités) et schémas (par exemple sous forme d'échelles et/ou arbres décisionnels), à l'appui desquels sont formulées des recommandations, voire des revendications d'actions. Une pensée mythologique, enfin, car il demeure encore une croyance répandue selon laquelle les *Big data* offrent une forme supérieure de connaissances, gage de vérité, d'objectivité et de précision. Plus spécifiquement en l'espèce, les mégadonnées et les instruments voués à les exploiter revigorent l'idéal prédictif. Ce Graal n'est pas poursuivi seulement par Amazon, Google, Walmart, Alibaba ou Netflix au motif d'«améliorer l'expérience de leurs clientèles» avant même qu'elles n'expriment leurs désirs de consommation (et de la sorte augmenter leurs revenus). Il séduit aussi les agences gouvernementales, et en particulier celles préposées au contrôle et à la réaction sociale. De fait, les *Big data* sont de puissants vecteurs de renseignements sur nos habitudes et nos attitudes. Enjoint de répondre à une stricte logique de précaution dans un contexte intolérant à tout passage à l'acte, le système pénal y voit une magistrale opportunité pour parfaire ses activités d'évaluation des risques.

Car à la source de toute pratique d'évaluation, il est une constante: celle qui consiste à capitaliser sur de l'information pour tenter de mieux anticiper le futur. À l'ère des *Big data*, le risque se présente donc comme un «problème d'information», susceptible d'être jugulé pour autant de suivre une démarche structurée qui consiste à 1) recueillir massivement des données 2) procéder à leur analyse statistique pour en faire ressortir des modèles et relations et 3) activer les bons algorithmes qui interrogeront

cette somme d'informations pour influencer sur nos comportements et prises de décision. Ces données sont produites et analysées en temps réel à travers un ensemble de dispositifs technologiques, mais elles peuvent aussi être des données amoncelées dans le passé sous des formes diverses et qui sont aujourd'hui progressivement numérisées. Quand Ekbia *et al.* (2015) parlent d'obsession contemporaine pour le futur et la prédiction, ils précisent aussi combien le passé peut être intéressant dans la mesure où il nous apprend quelque chose sur l'avenir. La numérisation croissante de nos sociétés procède directement de cette idée. De nombreuses archives, transcriptions audio, images et documents sont actuellement numérisés pour nourrir demain une multitude de projets, de la planification d'une campagne de vaccination à la restauration d'une œuvre d'art ou le contrôle des flux d'un segment autoroutier. La numérisation des dossiers et décisions judiciaires s'inscrit aussi dans cette tendance, portée par l'idée que ces dossiers numériques pourront à l'avenir être interrogés, pour mieux se prémunir de potentiels risques criminels.

Certes, la justice pénale suisse se montre encore frileuse à l'idée de recourir aux mégadonnées et à la technologie en support à ses activités. De même, la conversion au dossier judiciaire numérique vient tout juste de débiter en Suisse sous l'égide du projet *Justitia 4.0*, mais n'est prévue de s'étendre à l'ensemble des cantons que progressivement d'ici à 2026<sup>4</sup>. Sous d'autres latitudes cependant, ces pratiques ont déjà cours. Ainsi l'exploitation de grands ensembles de données, y compris des dossiers judiciaires numérisés, est non seulement largement admise dans de nombreux états des Etats-Unis à titre d'aide à la justice, mais aussi préconisée pour la réformer de l'intérieur. Le *Public Safety Assessment Tool (PSA)*, par exemple, a été élaboré dans la perspective de capitaliser sur les informations judiciaires passées pour aider les juges à se prononcer sur l'opportunité d'une détention préventive. Financé par la *Laura et John Arnold Foundation*, le PSA se fonde sur l'analyse de 1500 000 dossiers pré-sentenciels couvrant 300 juridictions américaines, dont il ressort neuf facteurs réputés les mieux à même de prédire une nouvelle arrestation, une nouvelle infraction violente ou un défaut de comparaître en cour<sup>5</sup>. L'argument initial à l'appui de son développement était de remplacer un système de caution (*bail system*)

<sup>4</sup> Sur ce projet et les traditionnels arguments évoqués à l'appui du passage au «tout numérique» V. [www.his-programm.ch/fr/projets/Justitia-40](http://www.his-programm.ch/fr/projets/Justitia-40) (dernière consultation: 26. 12. 2018).

<sup>5</sup> V. sous <https://www.psapretrial.org/> (dernière consultation: 26. 12. 2018).

particulièrement injuste puisque fondé sur les ressources financières des prévenus, et réduire l’incarcération avant jugement. Ses premiers tests empiriques se révèlent toutefois assez décevants: l’outil embrasse trop d’ambitions et reproduit les mêmes discriminations, raciales en particulier (Stevenson 2018; Eaglin 2018). 38 juridictions américaines, couvrant des villes importantes telles que Chicago et Houston, et trois Etats (Kentucky, Arizona et New Jersey) l’ont néanmoins adopté depuis son lancement en 2012. Ils l’estiment plus neutre, objectif et performant en matière d’évaluation, au motif précisément qu’il repose sur un vaste ensemble de données judiciaires, susceptibles d’éclairer le futur (Garrett/Monahan 2018).

A n’en pas douter, le recours aux *Big data* pour anticiper des problèmes sociaux de toutes natures constitue une tendance toujours plus lourde dans nos sociétés rivées sur le futur et la gestion des incertitudes qui lui sont inhérentes (Rose/Abi-Rached, 2013). En regard de cette évolution, il ne semble donc pas farfelu d’avancer que l’aide à certaines prises de décision judiciaires, celles en l’espèce initialement réservées à la personne d’un expert psychiatre, puisse prochainement faire l’objet d’un processus décisionnel de type uniquement algorithmique, fondé sur une masse de données extrêmement volumineuse (Hannah-Moffat 2018). Dans la foulée de ce mouvement, les sciences psychiatriques, la psychologie et la criminologie sur lesquelles lesdites évaluations étaient traditionnellement fondées, pourraient progressivement céder le pas aux sciences computationnelles (Selbst/Barocas 2018). En résumé, il y a fort à penser que l’approche orientée risque, actualisée dans sa forme la plus aboutie (bien qu’insatisfaisante) par des outils actuels, se voie prochainement suppléée par l’approche guidée par les mégadonnées (*smart data-driven approach*), auréolée de capacités prédictives renouvelées grâce à la mise en œuvre d’instruments relevant de l’intelligence artificielle.

## 5. Le recours à l’intelligence artificielle: du risque «augmenté»

L’amoncellement de données produites chaque jour dans le monde, aussi attractif soit-il, présente des défis importants en termes de stockage, mais aussi d’analyse et d’interprétation.

L’être humain ne peut à lui seul les traiter, les ordonner, les digérer et en extraire des informations. Dans les espaces scientifiques et industriels, les recherches se sont ainsi multipliées ces récentes années pour développer des machines, méthodes et procédés qui puissent les exploiter de façon aussi «*smart*» que possible, selon le terme désormais consacré. Leurs découvertes aujourd’hui dépassent toutes les prévisions, même celles jugées il y a peu comme relevant de la science-fiction, et vont en s’accéléralant. Elles sont réunies sous le chapeau de l’intelligence artificielle (IA), qui ne fait pas l’objet d’une définition figée, dès lors qu’elle recouvre depuis les années 50 une pluralité de réalités qui évoluent très rapidement<sup>6</sup>.

De façon très simplifiée pour les besoins de cet article, on pourrait dire de certains systèmes «intelligents» qu’ils sont le vecteur opérationnel des *Big data*: ils se nourrissent des données disponibles pour en extraire du renseignement qui pourra être utilisé à titre d’aide aux décisions et de soutien aux actions, peu importe leur domaine d’application. L’IA recouvre un large spectre de technologies, de méthodes computationnelles et d’applications mises en œuvre pour que des acteurs non humains puissent exécuter des tâches normalement associées à l’intelligence humaine (Russell/Norvig 1995; Raso *et al.* 2018). Ces dispositifs sont à l’heure actuelle ventilés en deux grandes catégories (Hessekiel *et al.* 2018). La première réfère à l’«IA faible» (*Narrow AI*) qui réunit des systèmes à même de simuler la pensée humaine pour effectuer une tâche bien spécifiquement délimitée. C’est le cas par exemple des «systèmes experts» qui fonctionnent suivant des règles de logique formelle pour engager un raisonnement sur une question particulière. La deuxième catégorie, l’«IA forte» (*General AI*), va encore plus loin en ce qu’elle promet le développement de machines susceptibles de réaliser seules une simulation complète du cerveau humain, douées de conscience de soi, voire même de sentiments («*thinking machines*»).

Une multitude de dispositifs relevant de l’IA faible alimentent nos vies quotidiennes, tel que l’assistant personnel Siri pour les détenteurs

<sup>6</sup> Pour un retour socio-historique, v. l’excellente synthèse de Elish/boyd 2018.

des produits Apple qui l'ont activé. L'IA forte en revanche n'en est qu'à ses débuts. Elle nécessite encore une meilleure compréhension du système cérébral et des techniques d'investigation plus perfectionnées en termes de puissance de calcul. Un exemple qui tend à s'en rapprocher est le programme développé par IBM, Watson, mis à l'épreuve dans plusieurs domaines, tels que le diagnostic médical, un jeu télévisé (*Jeopardy!*), l'analyse financière et jurisprudentielle, et dont l'ambition est de saisir les questions qui lui sont soumises en langage naturel et d'y répondre<sup>7</sup>. S'il a largement été revisité depuis son introduction dans les années 2000, il poursuit toujours la même idée: comprendre les requêtes d'un être humain et les problèmes qu'il a à résoudre, mais surtout parvenir à les solutionner et en énoncer les réponses de façon plus performante. Ainsi ses promoteurs préfèrent-ils parler d'*intelligence augmentée*<sup>8</sup>, tout en admettant (très récemment) ses écueils et impasses (Elish/boyd 2018).

A l'heure actuelle, une des avancées réputées les plus prometteuses en matière d'intelligence artificielle est le *Machine Learning*. Avec l'apprentissage automatique, l'idée est de donner au système la possibilité d'assimiler de façon *autonome* comment réaliser certaines tâches spécifiques, à partir des données mises à sa disposition. La machine est programmée de telle sorte à apprendre *par elle-même*, à classer et interpréter les données, mais aussi à tirer des enseignements de ses succès et échecs. Elle est dite «auto-apprenante». Beaucoup d'espoirs sont en particulier placés dans les différentes techniques d'apprentissage automatique pour tracer, identifier et anticiper des comportements et/ou situations. Leur concept opérationnel phare, les algorithmes prédictifs, viennent en appui de nombreux exercices de projection, notamment en analyse commerciale ou dans le cadre d'un recrutement (Cardon, 2015). Et ce,

de façon toujours plus autonome et moins supervisée.

Dans le système pénal aux Etats-Unis, l'utilisation d'instruments d'évaluation des risques relevant des techniques de *machine learning*, les «*automated data-driven risk assessment tools*», est toujours plus encouragée. Les décisions en matière de détention avant jugement sont considérées comme un domaine d'application «particulièrement prometteur» pour l'apprentissage automatique, au motif qu'il s'agit d'une «*concrete prediction task for which there is a large volume of data available*» (Kleinberg *et al.* 2017, 2). Des dispositifs similaires sont aussi employés par les agences préposées au suivi de délinquants à leur sortie de prison, pour évaluer leur potentiel de récidive en temps réels. Par exemple, la deuxième version du *Minnesota Screening Tool Assessing Recidivism Risk* (MnSTARR 2.0), lancée en novembre 2016, s'appuie sur un processus de notation complètement automatisé, ne nécessitant plus d'intervention humaine (Duwe/Rocque 2017). Dans les polices de Richmond, Santa Cruz ou Chicago, les forêts d'arbres décisionnels (*random forest algorithm*) sont utilisées quotidiennement pour prédire la probabilité d'une infraction en un lieu et une fenêtre temporelle donnés (Uchida 2013; Ferguson 2017). Plus récemment, ces techniques rendent compte également d'indications sur la personne des infracteurs potentiels (Brayne 2017). Ainsi par exemple, le programme Gotham utilisé par la police de Los Angeles se fonde sur un ensemble d'informations issues notamment de procès-verbaux de police et de rapports de suivis de probationnaires pour en extraire des listes de personnes érigées en «délinquants chroniques». Ces listes sont ensuite distribuées aux patrouilles avec l'ordre de les surveiller à distance et de les contrôler ponctuellement. Si les analyses demeurent certes en les mains d'acteurs humains, c'est un algorithme dit «prédictif» qui diligente le processus, dont seule l'entreprise à son origine, Palantir, dispose du code source. Bien qu'à leurs balbutiements encore dans les juridictions européennes, ces instruments gagnent progressivement en intérêt et font l'objet de financements tant publics que privés pour les perfectionner. Les techniques de *machine learning* semblent donc gentiment donner naissance à une 5<sup>e</sup> génération d'instruments voués à orienter les décisions et actions de la justice criminelle.

<sup>7</sup> V. <https://www.ibm.com/watson/>.

<sup>8</sup> Ainsi que l'exprime par exemple Rob High, *Chief Technology Officer pour le programme IBM Watson*: «*AI's role is to augment. It's to assist us and amplify us (...) in many ways. AI is not about replicating the human mind. Frankly, we've got plenty of human minds out there already, and from an economic standpoint, replicating the human mind is probably either not useful, but it's certainly nowhere near as plausible in terms of the current technology. So rather than focusing on that, what we ought to be thinking about is what can AIs do to augment us? I like to call it Augmented Intelligence, not Artificial Intelligence*». V. l'entrevue sous: <https://www.zdnet.com/article/ai-augments-and-amplifies-human-cognition/> (dernière consultation: 26.12.2018).

## 6. Les principaux enjeux de l'approche fondée sur les mégadonnées

L'exploitation automatisée de mégadonnées pour renforcer la performance prédictive des pratiques évaluatives contemporaines ravive l'idée d'un oracle qui n'est pas sans poser de vifs questionnements épistémologiques, méthodologiques, techniques, juridiques et éthiques. En termes *épistémologiques*, il n'est envisageable de parler de «prédiction» que dans les domaines où existe la certitude que des événements passés se répètent de façon parfaitement similaire. Or, rarissimes sont ces domaines, raison pour laquelle parler de justice ou police «prédictive» relève de l'abus de langage ou de la pensée magique (Aradau/Blanke 2017).

En termes *méthodologiques*, l'approche orientée risque s'appuie sur un protocole de recherche empirique de type hypothético-déductif. Des hypothèses fondées sur des théories et recherches passées sont formulées, puis testées à l'aide de modélisations statistiques classiques, dont il ressort, cas échéant, des résultats significatifs en termes de prédicteurs. Le procédé est de type causal (Chandler 2015). L'apprentissage automatique privilégié dans l'approche guidée par les mégadonnées obéit à une autre logique. Au lieu de tester des hypothèses fondées sur des théories ou recherches antérieures, il analyse au moyen d'un algorithme un jeu de données qui lui est fourni (*training data*), par exemple des extraits de décisions judiciaires, et identifie l'hypothèse qui lui paraît la plus pertinente reliant les données d'entrée (*input data*) aux résultats (*output data*). Il évolue ensuite au gré des données qui le nourrissent et à partir desquelles il «apprend» des relations établies pour orienter les décisions en temps réel. Son procédé est de type corrélationnel (Bennett Moses/Chan, 2014). Dans ce cadre, la source, la nature et la qualité des données sont d'une importance capitale. Comme l'est le degré de supervision de la machine et le modèle algorithmique retenu.

Si des contraintes *techniques* en termes de puissance computationnelle existent, ce sont surtout des questions de fiabilité, d'exactitude et de précision qui interpellent (Eckhouse 2018). On peut craindre en effet des produits très disparates si les machines sont laissées totalement libres de colliger et interpréter des

données d'ordres et de rangs différents, sans effort de contextualisation, ni garantie de véricité. La critique classique émise à l'égard des processus de prise de décision algorithmiques est qu'ils sont tout aussi sujets à des biais que l'est la décision humaine, susceptibles alors de mener à ce que Koepke et Robinson (2018) nomment des «*zombie predictions*». Bien des débats tournent donc aujourd'hui autour de ce qui est communément appelé «la boîte noire» de ces dispositifs, en référence aux variables retenues, aux techniques utilisées pour les corrélées, et aux formats privilégiés pour en restituer les résultats.

Les enjeux sont également *juridiques*, puisque ces dispositifs modifient les règles de propriété intellectuelle et de la sphère privée, comme des droits procéduraux élémentaires, tels que l'accès à un procès équitable (Kehl *et al.* 2017). Qu'en est-il lorsque le juge, présumé indépendant et impartial, fait reposer sa décision sur un score inattaquable car opaque? Comment au juste les renseignements produits par la machine sont-ils convertis dans ses décisions? Le magistrat dispose-t-il des connaissances suffisantes pour les interpréter? A ce jour l'usage de ces dispositifs est très peu encadré, ni soutenu par des guides facilitant une exploitation compréhensive de leurs produits (Lightbourne 2017).

L'exploitation automatisée des mégadonnées pose enfin de sérieuses questions *éthiques*, dès lors qu'elle peut porter de sérieux préjudices aux personnes qui en sont l'objet. Typiquement, il lui est reproché de perpétuer la fracture raciale et socio-économique béante aux Etats-Unis, sous couvert de vernis techno-scientifique (Angwin 2016; Selbst 2017). Des notions aussi essentielles que la confiance, la transparence, l'égalité de traitement et la probité peuvent ici être ébranlées, d'où la multiplication d'initiatives récentes préconisant une meilleure «gouvernance algorithmique» (Danaher *et al.* 2017).

En définitive, l'ensemble de ces enjeux ont une portée *politique*. Les techniques, modèles et rhétoriques mystiques affiliés aux *Big data* et à l'IA génèrent de nouvelles formes de connaissances, d'expertises et de pratiques, dans un contexte bien particulier, avide de *hype* et de certitudes (Smith *et al.* 2017; Elish/boyd 2018). Les promesses qui leur sont allouées impactent nos modes d'existence et de production des connaissances, nos rapports aux institutions et

au marché, nos idéaux de justice et de vivre ensemble (Selbst/Barocas 2018). Dans le système pénal, l'approche fondée sur les mégadonnées bouscule déjà les savoirs et les professionnels jusqu'à présent privilégiés pour maintenir l'ordre, faire justice, ou encore accompagner des sortants de prison. Cette approche revisite les finalités de la peine et les principes qui diligent sa détermination, par exemple celui de l'individualisation (Deskus 2018). Elle crée de nouvelles classes de populations, plus ou moins vulnérables au contrôle social et à l'exclusion. Elle motive des investissements considérables injectés dans des modalités d'exécution de la sanction, toujours plus complexes au plan technologique<sup>9</sup>. Et comme ses outils le plus souvent sont développés et subtilement commercialisés par des entreprises commerciales, elle floute les frontières entre entités publiques et privées, entre logiques économiques et sociales (Carlson 2018).

## Conclusion

En 2014, les experts psychiatres français Daniel Zagury et Jean-Louis Senon dénonçaient dans leur éditorial d'un numéro de *L'information psychiatrique* une « désagrégation de la pratique médicale légale », alimentée notamment par une inflation « jusqu'à l'absurde » de requêtes d'expertises sur « la dangerosité » allant jusqu'à encourager des pratiques de « serials experts ». Ils regrettaient la « toute-puissance trompeusement accordée » à leur discipline, ou encore le peu de scrupule de certains psychiatres à poser de « lointains pronostics ». Ils en appelaient alors à une « refonte globale pour la défense de l'expertise ». Cet appel résonne encore et est pour partie transposable à la situation helvétique. A le considérer aujourd'hui, la question se pose de savoir s'il n'y sera pas répondu par une justice prétendument augmentée, précipitant les pratiques expertales psychiatriques sur la voie d'une obsolescence programmée. Sachant que ces pratiques ne seront pas seules à subir de profondes transformations, il semble approprié que l'ensemble des protagonistes concernés –

psychiatres, magistrats, avocats, psychologues, criminologues, forensiciens, spécialistes des données, ingénieurs et justiciables – en discutent plus étroitement ensemble. Voilà d'ailleurs peut-être ici la plus belle des exigences de notre temps: celle de penser ses défis au-delà de nos (in)différences.

## Bibliographie

- Andrews D.A., Bonta J., *The Psychology of Criminal Conduct*, 5<sup>e</sup> éd., Routledge, New York 2010
- Angwin J., Larson J., Mattu S., Kirchner L., Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks, *ProPublica*, 2016. Disponible en ligne: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Aradau C., Blanke T., Politics of Prediction: Security and the Time/Space of Governmentality in the Age of Big Data, *European Journal of Social Theory*, 2017, 20(3), 373–39
- Aven T., Risk Assessment and Risk Management: Review of Recent Advances on their Foundation, *European Journal of Operational Research*, 2016, 253(1), 1–13
- Barabas C., Dinakar K., Ito J., Virza M., Zittrain J., Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment, in: Friedler S.A., C. Wilson C. (Eds), *Proceedings of Machine Learning Research*, 2018, 81, 1–15
- Bauman Z., *Liquid Times: Living in an Age of Uncertainty*, Polity Press, Cambridge 2007
- Beck U., *Risk Society: Towards a New Modernity*, Sage, London 1992
- Bennett Moses L., Chan J., Using Big Data for Legal and Law Enforcement Decisions: Testing the New Tools, *University of New South Wales Law Journal*, 2014, 37(2), 643–678
- Berk R., Hyatt J., Machine Learning Forecasts of Risk to Inform Sentencing Decisions, *Federal Sentencing Reporter*, 2015, 27(4), 222–228
- boyd d., Crawford K., Critical Questions for Big Data, *Information, Communication & Society*, 2012, 15(5), 662–679
- Brägger B., Risikoorientierter Sanktionenvollzug (ROS): Eine Analyse mit Ausblick, *Jusletter*, 9 Mars 2015
- Brayne S., Big Data Surveillance: The Case of Policing, *American Sociological Review*, 2017, 82(5), 977–1008
- Cardon D., *A quoi rêvent les algorithmes. Nos vies à l'heure des big data*, Seuil, Paris 2015
- Carlson A., The Need for Transparency in the Age of Predictive Sentencing Algorithms, *Iowa Law Review*, 2017, 103(1), 303–329
- Chandler D., A World without Causation: Big Data and the Coming of Age of Posthumanism, *Millennium: Journal of International Studies*, 2015, 43(3), 833–851
- Danaher J. et al., Algorithmic governance: Developing a Research Agenda Through the Power of Collective Intelligence, *Big Data & Society*, 2017, 4(2), 1–21
- Dawes R. M., Faust D., Meehl P. E., Clinical versus Actuarial Judgment, *Science*, 1989, 243, 1668–1674

<sup>9</sup> Tel qu'en attestent notamment les dernières versions de bracelet électronique, visant un recueil massif de données utilisées à des fins de surveillance, de détection et de prédiction.

- Delacrausaz P., Gasser J., La place des instruments d'évaluation du risque de récidive dans la pratique de l'expertise psychiatrique pénale: l'exemple lausannois, *L'information psychiatrique*, 2012, 88(6), 439-443
- Delacrausaz P., L'exécution des peines et des mesures orientée vers les risques en Suisse romande, point de vue d'un psychiatre forensique, *Revue suisse de criminologie*, 2017, 1(17), 50-57
- Deskus C., Fifth Amendment Limitations on Criminal Algorithmic Decision-Making, *New York University Journal of Legislation & Public Policy*, 2018, 21(1), 237-286
- Dittmann V., Was kann die Kriminalprognose heute leisten?, in: Bauhofer S., Bolle P.H., Dittmann V. (Eds), *Gemeingefährliche Straftäter*, Rüegger, Zurich 2000, 67-95
- Dubourg E., Gautron V., La rationalisation des méthodes d'évaluation des risques de récidive, *Champ pénal/ Penal field*, 2014, Vol. XI. Disponible en ligne: <http://journals.openedition.org/champpenal/8947>
- Duwe G., Rocque M., Effects of Automating Recidivism Risk Assessment on Reliability, Predictive Validity, and Return on Investment (ROI), *Criminology & Public Policy*, 2017, 16(1), 235-269
- Eaglin J., Constructing Recidivism Risk, *Emory Law Journal*, 2017, 67, 59-122
- Eckhouse L., Lum K., Conti-Cook C., Ciccolini J., Layers of Bias. A Unified Approach for Understanding Problems With Risk Assessment. *Criminal Justice and Behavior*, 2018. Disponible en ligne: <https://doi.org/10.1177/0093854818811379>
- Ekbia H., Mattioli M., Inna Kouper I., Arave G., Ghazinejad A., Bowman T., Ratandeeep Suri V., Tsou A., Weingart S., Sugimoto C., Big Data, Bigger Dilemmas: A Critical Review, *Journal of the Association for Information Science & Technology*, 2015, 66(8), 1523-1545
- Elish M. C., boyd d., Situating Methods in the Magic of Big Data and AI, *Communication Monographs*, 2018, 85(1), 57-80
- Ericson R. V., Haggerty K., *Policing the Risk Society*, University of Toronto Press, Toronto 1997
- Eubanks V., *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, New York 2018
- Feeley M. M., Simon J., The New Penology: Notes on the Emerging Strategy of Corrections and its Applications, *Criminology*, 1992, 30(4), 449-474
- Ferguson A. G., (2017). *The Rise of Big Data Policing: Surveillance, Race and The Future of Law Enforcement*, New York University Press, New York 2017
- Ferreira C., Maugué L., Prévenir le risque de récidive par l'obligation de soins: les apories de l'article 59 du Code pénal suisse, *Champ pénal/ Penal field*, 2017, Vol. XIV. Disponible en ligne: <http://journals.openedition.org/champpenal/9473>
- Fink D., *La prison en Suisse. Un état des lieux*, Presses polytechniques et universitaires romandes, Lausanne 2017
- Fonjallaz J., Gasser J., Le juge et le psychiatre; Une tension nécessaire, *Médecine et Hygiène/Stämpfli*, Chêne-Bourg/Berne 2017
- Garland D., The Rise of Risk, in: R. V. Ericson & A. Doyle (Eds), *Risk and Morality*, University of Toronto Press, Toronto 2003, 48-86
- Garrett B. L., Monahan J., *Judging Risk*, Virginia Public Law and Legal Theory Research Paper N° 2018-44. Disponible en ligne (SSRN): <https://ssrn.com/abstract=3190403>
- Giddens A., *Consequences of Modernity*, Polity Press, Cambridge 1990
- Gravier B., Psychiatrie et violence: quel est le risque acceptable? *Psychiatrie et violence*, 2009, 9(1). Disponible en ligne: <http://dx.doi.org/10.7202/038863ar>
- Gravier B., Lustenberger Y., L'évaluation du risque de comportements violents: le point sur la question, *Annales Médico-psychologiques, revue psychiatrique*, 2005, 163(8), 668-680
- Grove W. M., Zald D. H., Lebow B. S., Snitz B. E., Chad N., Clinical versus Mechanical Prediction: A Meta-Analysis, *Psychological Assessment*, 2000, 12(1), 19-30
- Guay J. P., Benbouriche M., Parent, G., L'évaluation structurée du risque de récidive des personnes placées sous main de justice: méthodes et enjeux, *Pratiques Psychologiques*, 2015, 21(3), 235-257
- Hachtel H., Vogel T., Graf M., Überarbeitung des Basler Kriterienkatalogs zur Beurteilung der Legalprognose («Dittmann-Liste»), *Forensische Psychiatrie, Psychologie, Kriminologie*, 2018, 1-8. Disponible en ligne: <https://doi.org/10.1007/s11757-018-0477-0>
- Hannah-Moffat K., Algorithmic Risk Governance: Big Data Analytics, Race and Information Activism in Criminal Justice Debates, *Theoretical Criminology*, 2018. Disponible en ligne: <https://doi.org/10.1177/1362480618763582>
- Hannah-Moffat K., Actuarial Sentencing: An «Unsettled» Proposition, *Justice Quarterly*, 2013, 30(2), 270-296
- Harcourt B., Risk as a Proxy for Race: The Dangers of Risk Assessment. *Federal Sentencing Reporter*, 2015, 27(4), 237-243
- Helmus L. M., Babchishin K. M., Primer on Risk Assessment and the Statistics Used to Evaluate Its Accuracy, *Criminal Justice and Behavior*, 2017, 44(1), 8-25
- Hessekiel K., Kim E., Tierney J., Yang J., Bavitz C. T., *AGTech Forum Briefing Book: State Attorneys General and Artificial Intelligence*, Harvard Law School, Berkman Klein Center for Internet & Society, Boston 2018
- Jendly M., Performance, transparence et accountability: une équation (dé) responsabilisante des professionnels exerçant en prison? *Déviance et Société*, 2012, 36(3), 243-262
- Kehl D., Guo P., Kessler S., *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, Harvard Law School, Berkman Klein Center for Internet & Society, Boston 2017
- Kitchin R., Big Data, New Epistemologies and Paradigm Shifts, *Big Data & Society*, 2014, 1(1), 1-12
- Kleinberg J., Lakkaraju H., Leskovec J., Ludwig J., Mullainathan S., *Human Decisions and Machine Predictions*, 2017, NBER Working Paper N° 23180
- Koepke J. L., Robinson D. G., Danger Ahead: Risk Assessment and the Future of Bail Reform. *Washington Law Review*, 2018, 93(4), 1725-1807

- Lightbourne J., Damned Lies & Criminal Sentencing Using Evidence-Based Tools, *Duke Law & Technology Review*, 2017, 15(1), 327-343
- Loup S., *L'expertise psychiatrique: LA carte à jouer dans la décision judiciaire pénale? Regards croisés sur les (en)jeux du dispositif expertal psychiatrique au prisme du jeu de poker*. Thèse de doctorat, Lausanne 2018, Ecole des sciences criminelles
- Meehl P. E., *Clinical versus Statistical Prediction. A Theoretical Analysis and a Review of the Evidence*, 2<sup>e</sup> éd., University of Minnesota, Minnesota 1996
- Monahan J., Skeem J., Risk Assessment in Criminal Sentencing, *Annual Review of Clinical Psychology*, 2016, 12(1), 489-513
- Moulin V., Gasser J., Intérêt et limites de l'évaluation du risque de récidive d'actes illégaux dans les expertises psychiatriques, *Revue médicale suisse*, 2012, 354(8), 1775-1780
- Mossman D., Assessing Predictions of Violence: Being Accurate About Accuracy, *Journal of Consulting and Clinical Psychology*, 1994, 62(4), 783-92
- O'Malley P., *Crime and Risk*, Sage, London 2010
- O'Neil, *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*, Broadway Books, New York 2016
- Raso P., Hilligoss H., Krishnamurthy V., Bavitz C., Levin K., *Artificial Intelligence & Human Rights: Opportunities & Risks*, Harvard Law School, Berkman Klein Center for Internet & Society, Boston 2018
- Rose N., Abi-Rached J. M., *Neuro: The New Brain Sciences and the Management of the Mind*, NJ: Princeton University Press, Princeton 2013
- Russell S. J., Norvig P., *Artificial Intelligence: A Modern Approach*. Prentice Hall Series in Artificial Intelligence, N. J.: Prentice Hall, Englewood Cliffs 1995
- Selbst A. D., Disparate Impact in Big Data Policing, *Georgia Law Review*, 2017, 52(1), 109-195
- Selbst A. D., Barocas S., The Intuitive Appeal of Explainable Machines, *Fordham Law Review*, 2018, 87(3), 1085-1139
- Simon J., *Governing Through Crime. How the War on Crime Transformed American Democracy and Created a Culture of Fear*, Oxford University Press, New York 2007
- Singh J. P., Fazel S., Forensic Risk Assessment: A Meta-review, *Criminal Justice and Behavior*, 2010, 37(9), 965-988
- Smith G. J. D., Bennett Moses L., Chan J., The Challenges of Doing Criminology in the Big Data Era: Towards a Digital and Data-driven Approach, *British Journal of Criminology*, 2017, 57(2), 259-274
- Starr, S. B., Evidence-Based Sentencing and the Scientific Rationalization of Discrimination, *Stanford Law Review*, 2014, 66(4), 803-872
- Stevenson M. T., Assessing Risk Assessment in Action, *Minnesota Law Review*, 2018, 103(1), 303-384
- Uchida C., Predictive Policing, in: G. Bruinsma, D. Weisburd (Eds), *Encyclopedia of Criminology and Criminal Justice*, Springer, New York 2013, 3871-3880
- Viljoen J. L., Cochrane D. M., Jonnson M. R., Do Risk Assessment Tools Help Manage and Reduce Risk of Violence and Reoffending? A Systematic Review, *Law and Human Behavior*, 2018, 42(3), 181-214
- Zagury D., Senon J.-L., L'expertise psychiatrique pénale en France, un système à la dérive, *L'information psychiatrique*, 2014, 90(8), 627-629

**Manon Jendly**

Professeure associée

Université de Lausanne

Ecole des sciences criminelles

Batochime

1015 Lausanne-Dorigny

Visiting scholar

University of California, Los Angeles

School of Law

Manon.Jendly@unil.ch