



L'algorithme de dangerosité pénale aux États-Unis : vers une érosion des droits fondamentaux du procès

Clarisse Valmalette

► To cite this version:

Clarisse Valmalette. L'algorithme de dangerosité pénale aux États-Unis : vers une érosion des droits fondamentaux du procès. *Annuaire International de Justice Constitutionnelle*, 2020, XXXV. <hal-03169476>

HAL Id: hal-03169476

<https://hal.science/hal-03169476v1>

Submitted on 23 Mar 2021

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.



HAL Authorization

Institut Louis Favoreu
Groupe d'Études et de Recherches
sur la Justice Constitutionnelle
Équipe associée au CNRS (UMR 7318)
Aix-en-Provence

Annuaire International de Justice Constitutionnelle

XXXV

**2019
(extraits)**

ECONOMICA
49, rue Héricart
75015 Paris

**PRESSES UNIVERSITAIRES
D'AIX-MARSEILLE**
3, Avenue R. Schuman
13628 Aix-en-Provence cedex 01

2020

L'algorithme de dangerosité pénale aux États-Unis : vers une érosion des droits fondamentaux du procès

Clarisse VALMALETTE*

« Vous êtes identifié par l'outil d'évaluation COMPAS comme présentant un risque élevé de dangerosité pour la communauté. Considérant le poids des différents facteurs, je rejette votre demande de sursis avec mise à l'épreuve [...] car vos antécédents, votre comportement en liberté surveillée ainsi que l'outil d'analyse du risque suggèrent que votre risque de récidive est extrêmement élevé¹. » Le juge Horne de la Cour de circuit de La Crosse, au Wisconsin, condamne le 12 août 2013 Eric Loomis à huit ans et six mois de prison ferme en s'appuyant sur COMPAS², un outil d'évaluation algorithmique de la dangerosité pénale. La réalité semble avoir rejoint la fiction. Dans ce scénario tout droit sorti du film *Minority Report*, les agents de *Précrime* inventés par Steven Spielberg qui interceptaient les criminels sur le point de commettre une infraction, sont remplacés par de vrais juges guidés par des algorithmes. Ces dernières permettent au juge, au stade de la fixation de la peine, de prédire le taux de récidive d'un individu et d'indiquer le traitement le plus adapté à sa réinsertion sociale.

Promu comme équitable, neutre, efficace et précis par ses développeurs, l'algorithmique prédictif permet aux services juridictionnels de trancher, plus rapidement, à moindres frais et sous couvert d'impartialité, le sort de milliers de délinquants. Leur efficacité, leur transparence et leur rationalité s'accordent avec les objectifs contemporains de la justice, en partie conditionnés par la réduction des fonds publics alloués aux ministères de la Justice³ et la lutte contre la défiance des justiciables à l'égard du système judiciaire⁴.

Aux États-Unis, le recours à ces outils est en plein essor depuis les années 2000. Il se généralise également au Canada et au Royaume-Uni⁵. Il existe

* Doctorante contractuelle, Aix Marseille Univ, Université de Toulon, Univ Pau et Pays Adour, CNRS, DICE, ILF, Aix-en-Provence, France.

1 Cour de circuit de La Crosse, Wisconsin, 12 août 2013. Cité par la Cour suprême du Wisconsin, 13 juillet 2016, *State v. Loomis*.

2 *Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)*.

3 Commission européenne pour l'efficacité de la justice (CEPEJ), *Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement*, Strasbourg, 3-4 décembre 2018, p. 58 : « Leurs promoteurs font souvent valoir leur caractère neutre et le fait qu'ils se baseraient sur des moyens factuels et objectifs qui aideraient à rendre la justice plus précise et transparente. Un autre grand atout serait leur efficacité, parfois dépassant les capacités humaines, qui ne peut qu'être extrêmement précieuse dans un contexte général de réduction de fonds publics, voire de pénurie de ressources. »

4 Par exemple : SciencesPo-CEVIPOF, *Baromètre de la confiance politique-vague 10 : En qu(o)i les Français ont-ils confiance aujourd'hui ?*, janvier 2019, p. 34 : 44 % des personnes interrogées disent avoir confiance en la justice.

5 Certains sont conçus par l'État, d'autres par des entreprises privées. Il existe aussi l'outil LSA (Level of Service Inventory) au Canada. En Europe, le programme HART (*Harm Assessment Risk Tool*) élaboré par l'Université de Cambridge à partir des archives communiquées par la police de Durham entre 2008 et 2012 est actuellement en expérimentation au Royaume-Uni. M. OSWALD, J. GRACE, S. URWIN, G. BARNES, « Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and 'Experimental' Proportionality », *Information et Communications*

aujourd'hui une soixantaine d'outils prédictifs en matières policière et judiciaire. Certains évaluent un taux global de dangerosité, d'autres sont spécialisés dans le risque d'addiction à l'alcool, aux drogues⁶, aux risques de commission d'infractions sexuelles ou de violences conjugales⁷. En 2008 déjà, vingt-huit États faisaient couramment usage de tables de prédiction pour fonder des décisions en matière criminelle⁸. COMPAS est le plus connu et le plus moderne de ces instruments⁹. La médiatisation de l'affaire « Loomis » citée ci-dessus a déclenché la controverse entourant l'usage de cet algorithme par les juridictions américaines suspectées d'avaliser des biais discriminatoires et des faux positifs et négatifs inhérents à ce type d'instruments¹⁰.

Ils calculent le risque de récidive d'un individu en confrontant les informations concernant le prévenu et celles d'un échantillon de population (données gemellaires). Concrètement, l'individu doit répondre à une centaine de questions parfois éloignées des faits reprochés¹¹, comme par exemple le code postal, la difficulté à payer des factures, la présence ou non d'un téléphone à la maison ou le contexte familial. L'algorithme pondère les différentes réponses, ainsi que les autres données recueillies par l'enquête judiciaire, à partir de facteurs prédéfinis. Le score produit est alors associé à un sous-groupe auquel est attaché un profil type de risques ainsi que les traitements recommandés pour traiter chacun des profils type¹². Ce fonctionnement rappelle dangereusement l'anthropométrie judiciaire ou « bertillonnage » par laquelle le criminologue Alphonse Bertillon proposait dans les années 1880 d'identifier les récidivistes à l'aide de leurs mensurations, de leurs photographies « face/profil » ainsi que de leurs stigmates physiques (tatouages, grains de beauté, cicatrice, etc.). C'est d'ailleurs peu de temps après que naît la pénologie actuarielle. Le fonctionnement de l'algorithme se distingue des mécanismes de l'intelligence artificielle puisque les suites de règles mathématiques (opérations logiques et instructions) ainsi que les données d'entrée sont déterminées à l'avance¹³. Bien que le processus de calcul demeure maîtrisé par les programmeurs,

Technology Law, août 2017, [Source en ligne] URL : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3029345 (consulté le 28 janvier 2020).

- 6 Par exemple, l'instrument SASSI (*Substance abuse Subtle Screening Inventory*) définit la probabilité d'une dépendance à des substances illicites.
- 7 Il existe les outils DV (*Domestic Violence*) utilisé dans le Colorado ou encore ODARA (*Ontario Domestic Assault Risk Assessment Tool*).
- 8 Qu'il s'agisse de condamnation ou d'aménagement de peines mais également d'opération de police ciblées, N. JEANNE et V. MAQUET, « Chapitre V, "Jacob, Megan, Adam..." : Les nouvelles figures juridiques de la dangerosité aux États-Unis », in G. GIUDICELLI-DELAGE et al., *La dangerosité saisie par le droit pénal*, Paris, Puf, coll. « Les voies du droit », 2011, p. 231.
- 9 L'instrument COMPAS a été créé en 1989 par Tim Brennan et Dave Wells, deux statisticiens de l'Université du Colorado à Boulder. Commercialisé par la société américaine Northpointe, c'est l'outil de prédiction le plus connu et le plus répandu aux États-Unis.
- 10 P. BERTAIL, D. BOUNIE, S. CLÉMENTON et P. WAELBROECK, *Algorithmes : biais, discrimination et équité*, Télécom ParisTech, février 2019, p. 10 sq. Il existe des solutions pour remédier aux biais techniques et cognitifs mais elles sont généralement coûteuses et délicates à mettre en œuvre.
- 11 Le questionnaire soumis au prévenu dans le cadre de l'analyse COMPAS est accessible à partir du lien suivant : <https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html> (consulté le 28 janvier 2020).
- 12 N. JEANNE et V. MAQUET, *op. cit.*, p. 231 ; B. HARCOURT, « Critique du champ pénal à l'âge actuariel », *Cahiers parisiens*, n° 3, 2007, p. 789.
- 13 S'agissant de la vérification de la qualité des données transmises à l'algorithme, les difficultés sont moindres car les juridictions doivent régulièrement réaliser des mises à jour des données transmises, afin que l'échantillon à partir duquel le profil du délinquant sera évalué soit représentatif de la population de l'État ou de la région concerné.

le fonctionnement de l'algorithme est couvert par les brevets industriels donc inaccessibles aux juges.

L'utilisation de ce type d'algorithme dans une matière aussi sensible que le procès pénal nourrit un certain nombre d'inquiétudes. Au-delà des suspicions portant sur leur fiabilité prédictive et sur l'usage ultérieur des données confiées à l'algorithme, la nature de ces dernières ainsi que le but assigné à l'instrument interrogent sa conformité aux droits fondamentaux. À la différence de la justice prédictive, les données utilisées pour alimenter l'algorithme ne sont pas des décisions de justice mais des informations à caractère personnel recueillies dans le cadre du processus pénal¹⁴. En outre, ce type d'algorithme n'a pas pour objectif de prédire l'issue d'un litige, mais de déterminer le risque de dangerosité d'un individu et ainsi influencer l'intensité de la peine effectivement prononcée¹⁵. S'il est admis que des barèmes de standardisation des sanctions sont transmis aux juges pour harmoniser certains contentieux répétitifs¹⁶, les algorithmes dont il est ici question interviennent dans un domaine qui ne peut se priver de l'interprétation casuistique d'un juge. C'est la difficulté centrale que soulève le recours à ces algorithmes en matière pénale.

Les méthodes actuarielles désignent d'une manière générale les calculs de probabilités appliqués aux questions d'assurance, de finances et aux prévisions sociales¹⁷. La pénologie actuarielle prend racine au début du XX^e siècle à l'initiative d'Ernest Burgess¹⁸ dont la volonté était d'éradiquer les discriminations raciales pratiquées dans le système judiciaire américain. Au-delà de la quête d'un outil de prédiction objectif des comportements humains, la méthode actuarielle a trouvé, dans la montée des exigences sécuritaires de la société des années 1970, un terrain privilégié de développement. Le besoin accru de sécurité s'est concrétisé dans le désir de maîtriser toujours davantage les risques liés à la dangerosité des individus¹⁹. L'absence de définition précise de la notion de dangerosité²⁰ contrastant avec son utilisation désormais déterminante impliquait que les outils algorithmiques prêtent

14 J.-P. JEAN, « À l'ère du numérique, ce que le criminel pourrait apprendre au civil en l'état », *Recueil Dalloz*, n° 17, 2019, p. 951.

15 Bien qu'incontournable dans le procès pénal américain, l'institution du jury ne sera pas traitée dans le cadre de cette étude puisque sa fonction est de déterminer la culpabilité du prévenu et non de décider du montant de la peine ce qui relève de la compétence du juge.

16 Sur la standardisation en matière pénale, v. V. GAUTRON, « La "barémisation" et la standardisation des réponses pénales saisies au travers d'une étude quantitative et qualitative de l'administration de la justice pénale », in I. Sayn, *Le droit mis en barèmes*, 2014, Dalloz, p. 85-97.

17 Définition issue du *Dictionnaire de l'Académie française*, 9^e éd., {Source en ligne} URL : <https://www.dictionnaire-academie.fr/article/A9A0482>

18 Il figure parmi les plus célèbres concepteurs de tables de prédiction à côté des époux Glueck. Ernest Burgess, professeur de sociologie à l'Université de Chicago, proposa en 1928 des critères tels le groupe d'appartenance sociale (*bobo*, gangster, drogué, *farm boy*...) ou le type de personnalité (egocentrique, émotionnellement instable...), afin de connaître les chances de réussite d'une mise en liberté conditionnelle, N. JEANNE et V. MAQUET, *op. cit.*, p. 230-231. De sensibilité démocrate, Ernest Burgess était motivé par « les meilleures intentions du monde » explique le président Jean-Paul Jean, « il voulait éliminer les discriminations de fait portant essentiellement sur la couleur de peau », J.-P. JEAN, *op. cit.* p. 949.

19 D. GARLAND, *The Culture of Control, Crime and Social Order in Contemporary Society*, Oxford University Press, 2001, p. 12.

20 La notion de dangerosité est particulièrement controversée en droit pénal. Au mieux peut-on dire qu'une personne dangereuse est un individu « ayant agi avec conscience et dont le comportement infractionnel est révélateur d'une certaine dangerosité », N. JEANNE et V. MAQUET, *op. cit.*, p. 225.

main-forte à la rationalisation de l'appréhension de la dangerosité par le juge répressif, amenuisant peu à peu son rôle dans la détermination de la peine²¹.

De surcroît, la crise de confiance des Américains dans leur système judiciaire²² doublée de l'augmentation brutale de la population carcérale dans les années 1980²³ ont achevé la volte-face de la politique pénale américaine : la justice réhabilitative encouragée par l'État providence et reposant sur les peines indéterminées, que les juges individualisaient de manière discrétionnaire²⁴ a peu à peu cédé le pas à une justice rétributive favorisant l'application stricte de peines égales²⁵. La fonction de la peine s'est ainsi progressivement réduite, au nom de l'ordre public, à la protection de la population contre les individus dangereux. Les algorithmes prédictifs ne pouvaient fournir une aide plus précieuse à la rationalisation de l'incertitude dans le domaine pénal²⁶.

Si l'individualisation des peines n'était à l'origine pas la priorité de cette justice rétributive, elle l'est devenue à partir des années 2000, du moins en apparence, grâce aux outils de prédiction algorithmique. Traditionnellement, le juge s'appuyait sur les éléments du rapport présentenciel appelé *presentence investigative report* (PSI) pour individualiser la peine du délinquant en fonction de son passé pénal et personnel. Ce document constituait sa source principale d'information. Cette partie du rapport a petit à petit été délaissée au profit de *guidelines* limitant strictement la marge de manœuvre du juge dans la fixation de la peine²⁷. Avec le développement des outils actuariels, l'officier de probation introduisit plus récemment l'analyse algorithmique de dangerosité dans le rapport présentenciel. Ce dernier se présente comme un document privilégié pour l'individualisation des sanctions puisqu'il détermine le risque de récidive du délinquant et son potentiel de réhabilitation²⁸. Certes la responsabilité pénale est en principe indissociable de la réalisation matérielle d'un acte criminel aux États-Unis, mais aujourd'hui la dangerosité influence les différentes étapes du procès pénal, y compris le montant de la peine²⁹. Ainsi de l'audience préliminaire (*preliminary hearing*) à l'issue de laquelle le

21 *Ibid.*, p. 230.

22 C. CASTETS, « L'IA en pratique : la police prédictive aux États-Unis », *Dalloz IP/IT*, n° 5, 2019, p. 314.

23 A. GARAPON et I. PAPADOPOULOS, *Juger en Amérique et en France*, Paris, Odile Jacob, 2003, p. 266.

24 Ces peines indéterminées ont fait l'objet de vives critiques dès 1973 par Marvin Frankel, juge fédéral, suite notamment aux soulèvements du pénitencier d'Attica en 1971, *ibid.*, p. 267.

25 Antoine Garapon rappelle à ce titre que « pour l'idéal puritain de l'homme autonome qui pourvoit seul à ses propres besoins, aucun magistrat n'est plus compétent pour le juger que sa propre conscience individuelle. Partant, personne ne devrait être jugé par le pouvoir discrétionnaire des hommes ; les sanctions pénales devraient être prévues par des règles de droit inflexibles » (*ibid.*, p. 256).

26 B. HARCOURT, « Une généalogie de la rationalité actuarielle aux États-Unis aux XIX^e et XX^e siècles », *RSC*, n° 1, 2010, p. 31.

27 Les *sentencing guidelines* sont des règles non contraignantes uniformisant la politique pénale américaine à l'échelle fédérale mais aussi au sein de chaque État. Développées depuis 1987, ces lignes directrices calibrent les sanctions de manière extrêmement précise à partir de facteurs déterminés. Davantage de précisions sur ces directives pénales seront données ci-dessous.

28 N. JEANNE et V. MAQUET, *op. cit.*, p. 229.

29 K. DANIELLE, P. GUO et S. KESSLER, « Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing », Berkman Klein Center for Internet & Society, 2017, [Source en ligne] URL : <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041>. En particulier, en Arizona, dans le Colorado, au Delaware, dans le Kentucky, en Louisiane, en Oklahoma, en Virginie, à Washington, ou encore au Wisconsin, les autorités judiciaires utilisent ces outils pour la détermination de la peine, S. M. SMYTH, « Can We Trust Artificial Intelligence in Criminal Law Enforcement », *Canadian Journal of Law and Technology*, vol. 17, n° 99, 2019, p. 5.

juge décide de l'opportunité d'une détention provisoire³⁰, au séjour carcéral³¹ en passant par le stade du jugement, l'évaluation algorithmique de la dangerosité est prise en considération. Pourtant, aucun encadrement juridique n'a été prévu aux États-Unis. La Cour suprême a pour le moment refusé de se saisir de la question³². Seuls quelques juges suprêmes d'États ont tenté de poser quelques consignes d'utilisation, à l'instar de la Cour suprême du Wisconsin dans l'affaire « Loomis » en 2016³³. Laurent Dartigues comme plusieurs auteurs s'indignent à ce titre de la passivité des communautés judiciaire, politique et universitaire face au changement qui est à l'œuvre dans nos États de droit³⁴.

Jusqu'à présent, le juge pénal s'appuyait sur des expertises psychiatriques réalisées par des cliniciens. Ces dernières n'étaient cependant pas exemptes de critiques notamment en raison de l'empreinte subjective du clinicien sur ses conclusions. L'analyse traditionnelle de la dangerosité pénale laissait au juge une grande marge d'appréciation : les résultats produits par les expertises cliniques étant réputés discutables, il revenait au juge de confronter les avis et arguments divergents en vue de trancher selon son intime conviction.

Pour autant que l'analyse algorithmique soit comparable à une expertise³⁵, l'objet est ici de savoir si les algorithmes prédictifs transforment l'office du juge

30 Les considérations de dangerosité peuvent selon la Cour suprême américaine constituer le fondement essentiel d'une détention provisoire. Cour suprême des États-Unis, 26 mai 1987, *United States v. Salerno*.

31 La présidence Trump a signé le 21 décembre 2018 le *First Step Act of 2018* (n° 115-391) étendant l'usage de ces outils algorithmiques à tous les établissements pénitentiaires américains. Le texte est accessible à partir du lien suivant : <https://www.govinfo.gov/link/plaw/115/public/391?link-type=pdf> (consulté le 28 janvier 2020).

32 Cours suprême des États-Unis, 18 septembre 2017, *Loomis Eric L. v. Wisconsin* (rejet du *certiorari* n° 16-6387). La Cour n'a pas justifié l'irrecevabilité de l'affaire « Loomis ». Sans doute cela s'explique-t-il par la nouveauté de cette pratique, sur laquelle très peu de cours d'État ont eu jusqu'à présent l'occasion de se prononcer.

33 Cour suprême du Wisconsin, 13 juillet 2016, *State v. Loomis* ; Cour suprême d'Indiana, 9 juin 2019, *Malenchik v. State* (la Cour indique dans une formule ambiguë que l'instrument algorithmique ne doit pas être utilisé en tant que circonstance aggravante, il doit simplement servir à informer ou suppléer le juge dans sa recherche d'informations et dans la détermination de la peine).

34 L. DARTIGUES, « Compte rendu du livre de Bernard E. Harcourt *Against Prediction. Profiling, Policing, and Punishing in an Actuarial Age*, Chicago, The University of Chicago Press, 2007 », *Champ pénal/Penal field*, 27 janvier 2010, {Source en ligne} URL : <http://journals.openedition.org/champpenal/7707> (consulté le 17 janvier 2020) : « Si tel est le cas, il me semble que le terme de "transformation" employé par B. Harcourt est presque trop faible pour rendre compte de l'effet de la méthode actuarielle. Si celle-ci soutient une formule de dissuasion et neutralisation des délinquants présumés dangereux en lieu et place de l'ancien précepte de punition et réhabilitation qui est au cœur de la philosophie de la pénalité néo-classique, il s'agit d'un véritable changement de paradigme aux conséquences vraisemblablement majeures pour nos États de droit. On imagine mal dès lors qu'un tel processus n'entraînerait point de résistances, dans le monde judiciaire, dans la sphère politique, dans le champ intellectuel. »

35 Pour comprendre l'impact de l'analyse algorithmique de dangerosité sur l'office du juge, il est utile de la rapprocher de l'expertise clinique car toutes deux visent à fournir au juge des informations sur la personnalité du prévenu et ses besoins. La notion d'expertise concerne « des actes qui revêtent un aspect "scientifique" manifeste, caractérisés par une activité d'interprétation des faits ou des résultats par le "sachant" », D. BOURCIER et M. DE BONIS, *Paradoxes de l'expertise : Savoir ou juger*, Le Plessis-Robinson, Institut Synthélabo pour le progrès de la connaissance, 1999, p. 79. Si l'algorithme ne pourra jamais être qualifié de « sachant » assermenté, sa fonction s'approche de celle d'un expert en ce qu'elle consiste à interpréter les faits de manière à proposer au juge une meilleure compréhension du cas concret. Tous deux fournissent des éléments de connaissance objectifs. L'expertise comme le résultat de l'algorithme sont en principe soumis au débat contradictoire. Si

pénal. Il s'agit dès lors de montrer que l'effacement du libre arbitre du juge met en péril certains droits fondamentaux du justiciable, en particulier, les principes de légalité et d'individualité des peines. Non seulement le juge se voit contraint dans la détermination de la peine de mobiliser une donnée extérieure à la légalité pénale qu'il est fortement incité à respecter (I), mais en plus, sa tâche d'individualisation des peines lui est retirée au bénéfice de l'algorithme (II). L'équilibre entre les principes de légalité et d'individualisation des peines se trouve ainsi mis à mal. Transféré pour l'essentiel à une intelligence artificielle, le principe d'individualisation des peines n'est plus en mesure de contrebalancer le principe de légalité, dont il tempérait traditionnellement la dureté. Le premier assouplissait la marge de manœuvre du juge dans la détermination de la peine contrainte par le second. D'une manière générale, il semble que l'algorithme prédictif ne soit pas un simple outil supplémentaire à la disposition du juge et nourrisse la crainte d'une dégradation de la qualité de la justice aux États-Unis.

I.- UNE DONNÉE EXTÉRIEURE À LA LÉGALITÉ PÉNALE IMPOSÉE AU JUGE

Aux États-Unis, si le juge n'est pas formellement contraint de suivre le résultat de l'algorithme, il est parfois obligé d'en solliciter l'analyse, ce qui manifeste la défiance déjà ancienne du pouvoir politique à l'égard du pouvoir d'appréciation des juridictions (A). Si l'usage de l'algorithme est imposé d'en haut par le législateur, le résultat de l'algorithme de par son fonctionnement peut lui-même être perçu comme prescriptif pour le juge (B). Ces deux éléments suffisent à concurrencer de manière illégitime le principe de légalité des peines puisque celui-ci impose que la loi exclusivement de toutes autres sources normatives détermine les infractions et les peines³⁶.

A.- Le recours à l'algorithme, un outil de défiance envers le pouvoir judiciaire

L'usage des outils de prévision algorithmique de la dangerosité se répand de manière exponentielle aux États-Unis³⁷. Ils sont présentés comme le perfectionnement de la fonction juridictionnelle. Cette modernisation est bien

l'expertise peut être mise en perspective par une contre-expertise, le résultat de l'algorithme doit pouvoir être discuté sur la base des facteurs pris en considération et de leur pondération.

36 Le principe de légalité des peines reconnaît « au seul législateur compétence pour déterminer les comportements exposant leur auteur à une peine ainsi que la nature et le quantum de cette dernière », E. DREYER, *Droit pénal général*, Paris, LexisNexis, 3^e éd., coll. « Manuel », 2014, p. 15. Lorsqu'il est appliqué strictement, le principe de légalité façonne l'idéal d'un « juge-machine » dont les excès de la Terreur ont été les témoins, J.-M. CARBASSE, « Légalité des délits et des peines », in D. ALLAND et S. RIALS, *Dictionnaire de la culture juridique*, Paris, Puf, 2003, p. 921. L'objectif de ce principe est, au-delà de la sécurité juridique, de réduire au maximum l'arbitraire des juges en déterminant *a priori* les sanctions encourues pour chaque infraction. L'algorithme prédictif se voit assigner un objectif similaire puisque derrière se profile la volonté d'éliminer l'imprévisibilité, l'aléa judiciaire et assurer l'égalité entre les justiciables. En outre, on peut soutenir que l'algorithme détermine lui-même les comportements qui sont de nature à entraîner le prononcé d'une certaine peine. Il détermine ainsi autant les comportements ou les faits que la peine encourue. Par exemple, le fait de vivre dans tel quartier, d'avoir tant d'enfants, d'avoir telle profession (qui quoique légale peuvent constituer le fondement de la peine) entraîne une peine ou un traitement conseillé par l'algorithme.

37 J. K. ELEK, R. K. WARREN et P. M. CASEY, *Using Risk and Needs Assessment Information at Sentencing: Observations from Ten Jurisdictions*, National Center for State Courts, 2015, p. 65-66.

souvent imposée par les autorités politiques, ce qui laisse supposer que l'objectif poursuivi en filigrane consiste à limiter la marge de manœuvre du juge.

De plus en plus de législations étatiques prévoient que l'analyse de la dangerosité des individus soit réalisée à partir d'instruments actuariels. En Tennessee par exemple, une réforme de 2016 prévoit que « [...] lors de la préparation des rapports présentenciels, le service correctionnel doit inclure des informations identifiant les risques et les besoins du défendeur, tels que déterminés par l'utilisation d'un instrument d'évaluation actualisé, ainsi que les programmes de traitement recommandés pour répondre aux risques et aux besoins et améliorer les chances du défendeur de réussir sa réinsertion dans la communauté »³⁸. Dans l'État du Wisconsin, une disposition prévoit que les services juridictionnels « ciblent les délinquants qui présentent un risque moyen ou élevé [...] de récidive, tel que déterminé par des instruments d'évaluation du risque actualisé, fiables et objectifs que le ministère a approuvés »³⁹. Le ministère de la Justice du Wisconsin a ainsi généralisé l'usage de l'outil COMPAS dans l'ensemble des juridictions de l'État⁴⁰.

Cette diffusion s'explique en outre par de fortes incitations de droit souple à l'échelle fédérale. Il s'agit de guide et de rapport de recommandation en faveur de l'utilisation des outils algorithmiques produits par des organismes judiciaires tels que le *National Center for State Courts* dirigé par la *Conference of Chief Justices* et la *Conference of State Court Administrators*. Cet organisme a produit en 2011 et 2015 deux rapports encourageant fortement l'usage des outils algorithmiques⁴¹. Leur usage est également recommandé par le *National Institute of Corrections*, une agence de l'exécutif américain relevant du ministère de la Justice⁴² ainsi que par l'*Administrative Office of the United States Courts Office of Probation and Pretrial Services*⁴³.

Ces recommandations reposent sur la volonté de mesurer avec plus de précision le risque de récidive des justiciables, de déterminer leur peine et les aménagements les plus adaptés à leur réinsertion sociale comme la semi-liberté et le placement sous surveillance électronique. Derrière ces objectifs humanistes, se profile de manière plus pragmatique la volonté de justifier l'optimisation carcérale et de maîtriser les flux criminels et, par ce biais, le pouvoir d'appréciation du juge.

L'infiltration du *New public management* en droit pénal soutient cette tendance. Il s'agit de calquer le fonctionnement du système judiciaire sur le modèle de l'entreprise privée favorisant l'efficacité et l'optimisation entre le peu de ressources et

38 Code pénal du Tennessee, 2016, § 41-1-412 (b) : « *In preparing presentence reports in accordance with § 40-35-207, the department of correction shall include information identifying the defendant's risks and needs as determined through the use of a validated assessment instrument, along with recommended treatment programs to address the risks and needs and enhance the defendant's opportunity for successful reentry into the community.* »

39 Code pénal du Wisconsin, article 301.068 : « *The community services target offenders at a medium or high risk for revocation or recidivism as determined by valid, reliable, and objective risk assessment instruments that the Department has approved.* »

40 V. Ministère de la Justice du Wisconsin, *Becky Young Community Corrections Recidivism Reduction FY13 Report*, octobre 2013 {Source en ligne} URL : <http://doc.wi.gov/Documents/WEB/ABOUT/OVERVIEW/Reentry%20Unit/Becky%20Young%20Recidivism%20Reduction%2013%20Report.pdf> (consulté le 10 janvier 2020).

41 J. K. ELEK, R. K. WARREN et P. M. CASEY, *op. cit.*, 2015 ; J. K. ELEK, R. K. WARREN et P. M. CASEY, *Guiding Principles on Using Risk and Needs Assessment Information in the Sentencing Process*, *National Center for State Courts*, 2011.

42 <https://nicic.gov/selecting-and-using-risk-and-need-assessment> (consulté le 28 janvier 2020)

43 *Administrative Office of the United States Courts Office of Probation and Pretrial Services, An Overview of the Federal Post Conviction Risk Assessment*, septembre 2011 https://www.uscourts.gov/sites/default/files/pcra_sep_2011_0.pdf (consulté le 21 janvier 2020).

la masse de délinquants, à l'aide d'indicateurs statistiques élaborés par les ministères ou même directement par des entreprises privées. Les juges sont incités à recourir aux instruments algorithmiques pour l'efficacité de leur mission. Selon Alain Supiot, cela couvrirait une déshumanisation du travail car on demanderait « désormais moins aux hommes d'obéir à des prescriptions que d'atteindre des objectifs dont la réalisation est évaluée au regard d'indicateurs chiffrés »⁴⁴. L'office du juge se rapprocherait ainsi de la fonction de manager optimisant les peines et les traitements, la préoccupation d'individualisation des sanctions passant en second plan. Ces transformations s'accordent avec la prégnance de l'utilitarisme dans les politiques judiciaires américaines. La vision instrumentale et économique du droit favorise le raisonnement en termes de « bilan "coûts-avantages" dépouillé de considérations morales »⁴⁵. C'est pourquoi, la prééminence des évaluations statistiques se présente comme un excellent moyen de contraindre les juges en les décourageant de déroger aux règles de droit souple.

Ces incitations se rapprochent des fameuses *sentencing guidelines* que l'institution judiciaire américaine impose au juge depuis plus d'une trentaine d'années sans véritable possibilité d'y déroger en pratique. Elles consistent en des règles non contraignantes uniformisant une politique pénale. Développées depuis 1987⁴⁶, ces lignes directrices calibrent les sanctions de manière extrêmement précise à partir de facteurs comme le casier judiciaire, la gravité des faits, leur nature, la culpabilité subjective ou le type de préjudice. Basées sur les statistiques de la pratique juridictionnelle, ces *guidelines* ne s'imposaient pas aux juges qui les appliquaient au début par mimétisme de leurs contemporains⁴⁷. Aujourd'hui, les dérogations aux *guidelines* sont devenues exceptionnelles puisque leur application dans une décision rend l'appel irrecevable et ce dernier est automatiquement recevable si le juge s'est écarté des créneaux fixés⁴⁸. Seules les circonstances aggravantes ou atténuantes peuvent justifier un ajustement de la peine à condition que le juge motive l'écart et respecte une fourchette stricte⁴⁹. L'algorithme évaluant la dangerosité d'un individu permet, à l'instar des *sentencing guidelines*, de canaliser la marge d'appréciation du juge. Si ce dernier souhaite déroger au résultat de l'algorithme, il devra développer de manière significative les justifications de sa dérogation ce qui l'en dissuade nécessairement.

Il existe aux États-Unis une tendance très forte à l'encadrement du pouvoir d'appréciation du juge. L'outil algorithmique apparaît comme un outil supplémentaire pour contrôler davantage l'office du juge, promouvoir son impartialité et permettre l'harmonisation du taux d'incarcération. Bien que les algorithmes prédictifs de dangerosité soient un outil nouveau, ils répondent à un

44 A. SUPIOT, *La gouvernance par les nombres : cours au Collège de France, 2012-2014*, Nantes, Institut d'Études Avancées de Nantes, 2015, p. 216.

45 A. GARAPON et I. PAPADOPOULOS, *op. cit.*, p. 260.

46 À partir du *Sentencing Reform Act* adopté par le Congrès américain en 1984, les juges sont vivement incités à se conformer aux *guidelines* fédérales, des États fédérés, qui déterminent la peine à appliquer en fonction de critères de plus en plus précis.

47 A. GARAPON et I. PAPADOPOULOS, *op. cit.*, p. 272.

48 *Ibid.*, p. 273. R. F. WRIGHT, « Sentencers, Bureaucrats, and the Administrative Law Perspective on the federal sentencing guidelines », *California Law Review*, vol. 79, n° 1, 1991, p. 67-69 ; K. STITH et J. CABRANES, *Fear of judging : sentencing guidelines in the federal courts*, Chicago, University of Chicago Press, 1998.

49 La Cour suprême américaine déclara inconstitutionnel l'application automatique des *guidelines*. Toutefois, les juges doivent obligatoirement les consulter avant d'imposer une peine et justifier une éventuelle dérogation par un raisonnement éclairé (Cour suprême des États-Unis, 12 janvier 2005, *Booker v. United States*).

objectif ancien tenant à la maîtrise de la prise de décision judiciaire. En réalité, cette lutte est aussi ancienne que l'apparition du pouvoir judiciaire autonome du pouvoir politique, le second cherchant à endiguer la liberté du premier. Le pouvoir politique s'appuie aujourd'hui sur les outils du numérique pour contrôler l'office du juge. Cependant, ces objectifs ne semblent pas assumés officiellement et les résultats générés par l'algorithme n'étant pas produits directement par les autorités politiques, on peut douter de leur légitimité. Ce secteur représente un marché aux intérêts financiers colossaux pour les entreprises privées qui s'y engagent, étant donné le nombre de décisions rendues à l'échelle des États-Unis. Particulièrement perméable au lobbysme, les institutions américaines n'hésitent pas à exposer les compétences régaliennes aux tentations de la privatisation, alors même que la fiabilité de ces outils algorithmiques n'a pas été démontrée⁵⁰.

En tant qu'il est devenu un instrument incontournable dans le procès pénal, l'algorithme prédictif de dangerosité réduit la marge de manœuvre du juge. Il représente la défiance du pouvoir politique envers le pouvoir judiciaire. Il semble que la méfiance envers ce dernier s'appuie aujourd'hui sur une confiance irrationnelle en l'initiative privée. Une nouvelle forme de discours prescriptif s'impose au juge pénal fortement incité à prendre en compte le résultat de l'algorithme. Le principe de légalité pénale se trouve ainsi concurrencé par l'usage de cet algorithme prédictif. Celui-ci est imposé tant par le haut par les instances judiciaires que par le bas en raison de l'effet subjectif de l'algorithme sur la prise de décision du juge.

B.- Le résultat de l'algorithme perçu comme contraignant pour le juge

Il est impossible de quantifier l'influence exacte du résultat de l'algorithme sur la décision du juge car cela supposerait de sonder le cœur de chaque juge pour connaître la motivation réelle de leurs décisions. Richard Posner publie en 1993 un article visant à répondre à la question de savoir ce que les juges cherchent à maximiser dans leur décision⁵¹. Selon lui, les juges demeurent inéluctablement motivés par la recherche de popularité et de prestige au sein de la sphère des juristes mais aussi de la population en général. L'opinion publique influence à cet égard le processus décisionnel du juge⁵². L'outil algorithmique par ses caractéristiques intrinsèques permet au juge pénal de justifier objectivement sa décision. À défaut de rechercher le prestige, en suivant les indications de l'algorithme, le juge se préserve à tout le moins d'une prise de risque liée notamment à une minoration de la dangerosité d'un individu et de la remise en cause de sa décision par les cours supérieures.

S'il est délicat d'attester que l'argument tiré du résultat de l'algorithme est décisif dans le processus décisionnel subjectif, il est possible de déterminer les raisons pour lesquelles cette donnée est susceptible d'avoir une influence telle qu'elle devient contraignante pour le juge.

50 J. ANGWIN, J. LARSON, S. MATTU et L. KIRCHNER, « Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks », *ProPublica*, 23 mai 2016 (Source en ligne) URL : <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (consulté le 28 janvier 2020).

51 R. POSNER « What do judges and Justices maximize? The same thing everybody else does », *Supreme Court Economic Review*, vol. 3, 1993, p. 1-41.

52 I. FASSASSI, *La légitimité du contrôle juridictionnel de la constitutionnalité des lois aux États-Unis. Étude critique de l'argument contre-majoritaire*, Paris, Dalloz, 2017, p. 611-612.

L'instrument algorithmique se présente comme un outil calculant un niveau de dangerosité assorti du traitement le plus adapté à sa réinsertion sociale. Le niveau de dangerosité s'opère à partir d'un calcul mathématique de probabilité.

Il semblerait que le langage chiffré des probabilités soit particulièrement persuasif aux yeux du juge. En effet, le résultat de l'analyse ne peut apparaître que neutre et objectif puisqu'il émane d'un système informatique, se contentant d'appliquer certaines formules mathématiques à partir d'une masse de données considérables présumée représentative de la réalité. Il n'existe *a priori* aucune raison de douter de l'exactitude du résultat calculé par un ordinateur⁵³.

Le juge sera enclin à respecter l'indication proposée par l'algorithme car, pour reprendre les termes de Günther Anders connu pour sa critique philosophique de la technologie, il est devenu commun aujourd'hui de ne considérer « comme "objectif" que les jugements prononcés par des objets »⁵⁴. Le juge pourrait avoir tendance à considérer le résultat de l'algorithme comme une donnée objective et rationnelle directement applicable. C'est bien souvent le sort des connaissances scientifiques, lorsqu'elles sont convoquées dans un champ extérieur à leur domaine d'action : selon Danièle Bourcier et Monique De Bonis, elles perdent « en distanciation et en démarche critique mais elles gagnent sur le terrain de la croyance – dans laquelle s'ancre une grande partie de nos processus de décision – parce que la croyance n'est pas nécessairement fondée en raison »⁵⁵.

Il existerait pourtant des raisons de douter de la fiabilité du résultat de l'algorithme en témoignent certaines études mentionnées en dernière partie. Une mise en garde des juges sur les failles de cet outil paraît dès lors indispensable. Seule la Cour suprême du Wisconsin a jusqu'à présent jugé utile d'indiquer que le rapport présentiel doit contenir les précautions d'usage et d'interprétation précisant que certaines études ont soulevé des soupçons à l'égard de la fiabilité des outils algorithmiques quant à la surestimation du risque de récidive à l'égard de délinquants appartenant à une minorité⁵⁶.

Une comparaison entre l'analyse algorithmique de dangerosité et l'expertise psychiatrique renforce les inquiétudes relatives à la portée du résultat de l'algorithme. Lorsqu'il est confronté à un expert de sciences sociales comme la psychologie clinique, le juge a tendance à demander des expertises complémentaires afin de croiser les résultats obtenus par les cliniciens⁵⁷. Cela est moins vrai des expertises relevant des sciences exactes que le juge reprend généralement sans discuter comme argument d'autorité⁵⁸. En dépit du critère poppérien de falsifiabilité, ces dernières sont, dans le théâtre judiciaire, moins sujettes aux contre-expertises. À l'inverse, les expertises cliniques de dangerosité sont souvent placées au cœur du débat contradictoire car elles sont empreintes d'une subjectivité évidente.

53 L'étude *ProPublica* a montré que des doutes sérieux pouvaient néanmoins demeurer : J. ANGWIN, J. LARSON, S. MATTU et L. KIRCHNER, « Machine Bias : There's software used across the country to predict future criminals. And it's biased against blacks », *ProPublica*, 23 mai 2016 [Source en ligne] URL : <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (consulté le 28 janvier 2020).

54 G. ANDERS, *Die Antiquiertheit des Menschen*, Munich, Beck, 1956, traduction française par C. David, *L'obsolescence de l'homme. Sur l'âme à l'époque de la deuxième révolution industrielle*, Paris, Ivrea, 2002, p. 79.

55 D. BOURCIER et M. DE BONIS, *op. cit.*, p. 14.

56 Cour suprême du Wisconsin, 13 juillet 2016, *State v. Loomis*.

57 S. M. SMYTH, *op. cit.*, p. 6.

58 L. DUMOULIN, « L'expertise judiciaire dans la construction du jugement », *Droit et société*, n° 44-45, 2000, p. 218-219.

La plupart repose sur une approche psychanalytique : le clinicien recherche des critères comme « l'évolution de la personnalité du sujet, le rapport de celui-ci aux faits, le degré de reconnaissance de ces faits, le degré de contrainte éventuelle, son "vécu surmoïque", sa capacité à appréhender le retentissement psychologique de la victime et le rapport du sujet à la loi »⁵⁹.

L'expertise algorithmique de dangerosité se trouve à mi-chemin entre sciences sociales et sciences exactes. Tout en ayant pour objet l'étude des comportements humains, elle applique des formules mathématiques et des chiffrages informatiques qui la revêtent de l'apparence d'objectivité propre aux expertises de sciences exactes. Dès lors, on peut craindre que le juge soit incité à reprendre le résultat tel quel dans sa décision sans le soumettre à la discussion contradictoire. L'analyse algorithmique de dangerosité semble ainsi demeurer plus persuasive que les expertises psychiatriques traditionnelles.

En plus de prétendre constituer un élément de connaissance objectif, le calcul de probabilité a pour caractéristique d'inciter à un certain comportement. Il ne se contente pas de décrire ou d'établir des faits comme la plupart des méthodes statistiques, mais suggère, dans le contexte juridique, l'adoption d'un certain comportement. Alain Supiot pousse au bout cette logique pour estimer que « les premières applications du calcul des probabilités n'ont pas en effet seulement visé à établir certains faits, mais aussi à décider de la règle qui doit les régir », c'est la raison pour laquelle, « le calcul des probabilités déborde [...] le domaine de l'être pour investir celui du devoir être. Il ne sert plus seulement à décrire mais à prescrire »⁶⁰. Par une confusion entre discours scientifique et discours du jugement⁶¹, le calcul de probabilité se rapprocherait ainsi d'un discours prescriptif en tant qu'il vise à imposer certains « standards applicables aux comportements des individus »⁶² et fournir au juge une aide à la décision. Il porte en lui « en puissance [...] un verdict, les résultats de la décision »⁶³ pour reprendre les termes de Danièle Bourcier et Monique De Bonis. L'algorithme intervient non plus en vue de fournir au juge un argument au soutien de ce qu'il tient à prouver, mais le résultat auquel il doit parvenir. Cette inversion du raisonnement pourrait occasionner une modification profonde de la manière d'administrer la justice car la logique de l'algorithme repose sur « une rationalité d'ordre technique, instrumentale, fondée sur des considérations d'efficacité et de nécessité »⁶⁴, tandis que le juge répond à des considérations de justice voire d'équité.

La Cour suprême du Wisconsin a tenté d'endiguer la portée que le juge pourrait vouloir conférer au résultat de l'algorithme en indiquant dans sa décision de

59 S. CRAMPAGNE, *L'évaluation de la dangerosité dans le cadre de l'expertise psychiatrique pénale*, Thèse, Université de Grenoble, 2013, p. 45.

60 A. SUPIOT, *op. cit.*, p. 148. Dans les pages précédentes, l'auteur explique que « la statistique élabore des énoncés qui échappent à la réflexivité du langage et acquiert par là même une puissance dogmatique particulière. Cette puissance symbolique des nombres a été renforcée par l'usage de ce qu'on appelle la numérotation arabe [...]. À la différence des chiffres romains ou grecs, qui véhiculaient les traces sémantiques d'un alphabet donné, les chiffres indo-arabes ont, entre autres vertus (notamment celle d'abolir la distance entre écriture et calcul), une universalité sans égal. [...] Avec la généralisation de l'emploi des chiffres indo-arabes, les nombres ont acquis une force normative universelle », A. SUPIOT, *op. cit.*, p. 141.

61 D. BOURCIER et M. DE BONIS, *op. cit.*, p. 21.

62 *Ibid.*, p. 90.

63 *Ibid.*, p. 51.

64 *Ibid.*, p. 90.

2016 sur l'affaire « Loomis » que ce dernier ne pouvait être un élément décisif de sa décision⁶⁵.

Se saisissant du développement des nouvelles technologies, les institutions judiciaires américaines mettent à la disposition du juge pénal une source d'information d'un nouveau type. Non seulement les juges sont fortement incités, voire contraints de recourir à cet outil pour déterminer le niveau de dangerosité d'un individu, mais en plus le résultat produit par l'algorithme présente une forte dimension prescriptive de par son fonctionnement.

Une atteinte au principe de légalité des délits et des peines peut ici être caractérisée puisque le juge est contraint de mobiliser une donnée extérieure à la légalité pénale pour déterminer l'intensité de la peine. Le résultat de l'algorithme s'imposant au juge en tant que discours prescriptif dans la détermination de la peine, la loi n'est plus l'unique source prévoyant les infractions et les sanctions.

Le principe de légalité des délits et des peines a pour finalité d'encadrer le pouvoir d'appréciation du juge. La rigueur de ce principe a néanmoins été tempérée par le développement du principe d'individualité des peines, desserrant l'étau de la contrainte textuelle au profit d'un ajustement du juge. L'utilisation de l'algorithme porte atteinte au principe de légalité en créant une norme extra-juridique concurrente, sans pour autant valoriser la marge d'appréciation du juge. Bien au contraire, le principe d'individualité des peines se trouve amoindri par l'usage de l'algorithme, ce qui rend d'autant moins légitime l'atteinte portée au principe de légalité pénale.

II.- L'INDIVIDUALISATION DES PEINES DÉLÉGUÉE À UN ALGORITHME

Si l'atteinte au principe de légalité était justifiée par une revalorisation du principe d'individualisation des peines, cela aurait été compréhensible dans la mesure où ces deux principes jouent l'un pour l'autre le rôle de contrepoids. Il semble toutefois que le principe d'individualisation ne gagne rien au développement des algorithmes de dangerosité, contrairement à ce qu'en disent leurs promoteurs. À ce stade crucial de la prise de décision du juge, ce dernier voit sa marge d'appréciation se réduire car non seulement il ne dispose pas des moyens de vérifier, ni *a fortiori* de soumettre au débat contradictoire, les données de l'individualisation appliquées par l'algorithme (A). S'ajoute à ces contraintes, le renoncement prévisible du juge à l'individualisation des peines (B).

A. Le transfert contraint des éléments d'individualisation à l'algorithme

L'individualisation des peines relève en principe de la compétence du juge, encore faut-il que celui-ci dispose des moyens de l'appliquer. Lorsque l'usage de l'algorithme devient obligatoire, le juge est paradoxalement susceptible de perdre la maîtrise de l'individualisation des peines. Il n'est plus en capacité de maîtriser les éléments de discussions le conduisant à l'individualisation de la peine car il est placé dans l'impossibilité de satisfaire son obligation de contrôle des processus de production des vérités scientifiques (méthodes, instruments de mesures, catégories, concepts...) ⁶⁶.

65 Cour suprême du Wisconsin, 13 juillet 2016, *State v. Loomis*.

66 D. BOURCIER et M. DE BONIS, *op. cit.*, p. 27.

Cela s'explique principalement par la confidentialité des brevets industriels qui dissimulent les formules algorithmiques utilisées par les programmes de prédiction⁶⁷. Le processus de calcul du risque de récidive peut pourtant être d'une importance majeure pour le juge dans son travail d'individualisation de la peine. La dissimulation des méthodes algorithmiques risque de biaiser l'individualisation des peines.

Les brevets commerciaux constituent à ce titre une entrave à la vérification de la validité scientifique de l'analyse. D'après la Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires, les autorités devraient pouvoir certifier le fonctionnement de l'algorithme et rendre compte de leur processus décisionnel, sans quoi leur responsabilité ou *accountability* pourrait être mise en jeu⁶⁸. Or, ni le juge, ni le défendeur ne peuvent vérifier intégralement l'intégrité intellectuelle et l'intangibilité du système.

Les rapports des expertises cliniques utilisées traditionnellement contiennent systématiquement deux parties relatives pour l'une à l'observation, à la constatation ou aux calculs et pour la seconde à l'interprétation et l'explication des résultats. Ainsi le rapport du clinicien détaille l'ensemble des opérations menées et donne matière à discussion des résultats et de l'interprétation. Dans le cas de l'analyse algorithmique de dangerosité, la phase d'observation est totalement opaque. Le juge ne dispose que du résultat de l'interprétation c'est-à-dire, en l'occurrence, du niveau de dangerosité et la recommandation de traitement. Le diagnostic qui permet le passage du savoir existant à l'analyse du cas concret est occulté. L'analyse algorithmique apparaît ainsi bien moins inintelligible que l'expertise clinique.

Cette atteinte à la transparence de la procédure pénale entrave le principe du contradictoire et l'égalité des armes. À partir du moment où le processus d'individualisation des peines est, pour partie, transféré à un algorithme, le juge dispose d'une plus faible marge de manœuvre pour assurer le respect du contradictoire et de l'égalité des armes en matière d'individualisation de la peine. La procédure semble ainsi lui glisser des mains. Il se trouve dans l'impossibilité de permettre un déploiement optimal des arguments de la défense⁶⁹.

L'affaire *Loomis v. State of Wisconsin* pourrait ici nous permettre de comprendre les difficultés que rencontrent les défendeurs et devant lesquelles le juge se trouve démuní. L'un des griefs d'Eric Loomis repose sur son droit au *due process*. Il conteste le fait de ne pas avoir pu mettre en doute le taux de dangerosité que l'algorithme avait calculé à son égard. Le prévenu n'a eu accès qu'au taux de dangerosité sans précisions sur la façon dont ce taux a été obtenu, en particulier sur la manière dont les différents facteurs ont été mis en balance. En effet, la société *Northpointe* avait refusé de les dévoiler lors du procès en raison de la confidentialité des brevets industriels⁷⁰. Le docteur Thompson dont l'expertise a été sollicitée par l'avocat de la défense atteste que la Cour elle-même a très peu d'informations sur le

67 La plupart des outils algorithmiques récents sont conçus par des entreprises privées, d'autres par l'institution judiciaire elle-même. Le problème de la confidentialité industriel ne se pose pas de la même manière dans ces deux cas. Seuls les algorithmes proposés par des entreprises privées sont concernés par les développements qui suivent.

68 Commission européenne pour l'efficacité de la justice (CEPEJ), *Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement*, Strasbourg, 3-4 décembre 2018, p. 12.

69 Il est rare que le défendeur ait l'opportunité de contester l'analyse puisque si le score est transmis à l'avocat, les calculs permettant la transformation des données sous-jacentes en taux de récidive ne sont pas révélés. J. ANGWIN, J. LARSON, S. MATTU et L. KIRCHNER, *op. cit.*

70 Cour suprême du Wisconsin, 13 juillet 2016, *State v. Loomis*.

fonctionnement de l'algorithme et en particulier « sur la manière dont COMPAS compare le passé individuel avec celui d'un échantillon de population »⁷¹. Il ajoute que « la cour ne connaît même pas la composition de cet échantillon, s'agit-il de la population du Wisconsin, de New York, de Californie... Il y a une grande quantité d'informations que la Cour n'a pas. La Cour n'est pas suffisamment informée alors qu'elle doit utiliser ces graphiques pour prononcer une peine »⁷².

Que reste-t-il au requérant pour contester le taux de dangerosité qui lui est attribué ? La Cour suprême du Wisconsin, saisie en dernier ressort de cette affaire, a estimé qu'en dépit des arguments de Loomis, son droit au *due process* n'était pas entravé. Elle explique que le guide d'utilisation COMPAS élaboré par la société Northpointe lui permettait de connaître les facteurs pris en compte, que les facteurs dynamiques étaient finalement peu nombreux (ce qui laisserait moins de place à un résultat inattendu) et que Loomis pouvait somme toute utiliser les questions qui lui ont été posées par l'algorithme pour se défendre⁷³.

Ces éléments ne sont guère convaincants car premièrement, le guide d'utilisation de COMPAS évoque dix-sept facteurs d'analyse pour la plupart imprécis (la personnalité, l'environnement social, les occasions favorisant la commission d'infractions) ou peu pertinents (les difficultés financières, l'instabilité résidentielle, désœuvrement)⁷⁴. Il est possible de s'interroger sur la façon dont l'algorithme apprécie ces facteurs dans chacun des cas particuliers ainsi que sur le poids qu'il accorde à ces derniers respectivement. L'opacité des facteurs et de leur pondération empêche le juge de procéder à leur vérification et de les soumettre au débat contradictoire.

Deuxièmement, face à un tel instrument, il apparaît délicat pour le requérant de fournir la preuve de sa non-dangerosité et ainsi de construire sa défense, étant donné qu'il n'a pas accès à ce qui permet à l'algorithme de parvenir aux conclusions. Les droits de la défense doivent permettre au justiciable de contester effectivement les conclusions d'un expert. Il ne semble pas à ce titre incongru d'assimiler l'analyse algorithmique à une expertise judiciaire et de considérer comme souhaitable la soumission de cette analyse aux mêmes exigences processuelles.

La protection qu'offre le droit des brevets dans ces cas de figure pourrait à l'avenir tomber. En effet, dans l'État de New York, une entreprise propriétaire d'outil de police prédictive a été contrainte de communiquer certains mécanismes de fonctionnement⁷⁵. Cela pourrait s'étendre aux outils d'évaluation de la dangerosité utilisés par le juge pénal. Une autre lueur d'espoir pourrait apparaître au travers de la création d'audit. Il s'agit d'un expert de confiance désigné par les autorités publiques dont la mission est de contrôler dans l'intérêt public sans publier l'algorithme. Ainsi, en France, la CNIL propose de mettre en place une plateforme nationale d'audit des algorithmes. La surveillance et le test des algorithmes pourraient aussi être effectués par des cabinets d'audit privés, étant donné l'ampleur de la tâche et le coût que cela représente⁷⁶.

Il est clair ici que le secret industriel ainsi que la charge de la preuve de l'innocuité lorsque celle-ci est établie par un algorithme rendent particulièrement

71 *Idem.*

72 *Idem.*

73 *Idem.*

74 Northpointe, *Practitioners Guide to COMPAS*, 17 août 2012, p. 29-43.

75 Cour suprême de l'État de New York, 27 décembre 2017, *Brennan Center for Justice v. New York Police Department*.

76 P. BERTAIL, D. BOUNIE, S. CLÉMENTON et P. WAELBROECK, *op. cit.*, p. 18.

complexe la défense du prévenu et provoquent ainsi une rupture dans l'égalité des armes et du principe du contradictoire. Le juge est dessaisi de sa mission d'organiser la discussion lors du procès puisque le résultat de l'algorithme ne peut se discuter que de manière extrêmement limitée. La défense ne pourra contre-argumenter à l'aide d'un outil aussi persuasif et incontestable sur le plan technique que l'algorithme prédictif.

La motivation de la décision du juge ne peut pas demeurer hermétique à ces évolutions du procès. Elles laissent craindre le renoncement du juge à l'individualisation des peines.

B.- Le risque de renoncement du juge à l'individualisation des peines

Le juge étant dans l'incapacité de réguler le débat contradictoire en ce qui concerne l'individualisation de la peine, il sera plus enclin à s'appuyer de manière décisive sur l'algorithme et ainsi à délaissé volontairement son pouvoir d'appréciation à ce stade primordial de sa décision. Deux autres raisons viennent appuyer cette inquiétude.

En premier lieu, il sera aisé pour lui d'arguer dans sa motivation que le résultat de l'algorithme propose déjà un traitement individualisé. En effet, les informations concernant l'individu sont elles-mêmes directement traitées par rapport à des facteurs traditionnellement utilisés pour l'individualisation des peines, comme le casier judiciaire de l'individu ou la gravité de l'infraction. En outre, le résultat de l'algorithme est caractérisé par sa portée concrète et individuelle : il porte sur une personne en particulier, placée dans une situation donnée. Le processus de concrétisation manifestant le passage d'un énoncé général et impersonnel à un énoncé individuel et concret s'achève donc en apparence dès l'édiction du rapport de l'algorithme. Le plus haut degré de concrétisation semble ici correspondre au plus haut degré d'individualisation, l'espace laissé à l'interprétation du juge apparaissant comme réduite au minimum.

Cependant, l'individualisation à laquelle procède l'algorithme n'est qu'illusion. Le calcul probabiliste confronte les données relatives à l'individu aux données gémellaires d'un échantillon de personnes, ce qui permet d'identifier le profil statistique du prévenu. La situation de l'individu n'est pas appréciée pour elle-même mais comparée aux comportements passés de plusieurs milliers d'autres individus. Si l'individualisation de la peine se réduit à une comparaison, alors il existe de forts risques pour que celle-ci soit biaisée. Les facteurs pris en compte étant particulièrement englobants, il n'y a qu'un pas pour réduire l'individu à son profil statistique. Le profilage est nécessairement associé à une tendance déterministe que le magistrat Marc Ancel dénonçait vigoureusement au travers de sa philosophie de la défense sociale nouvelle⁷⁷. Dans cette lignée, Clementina Barbaro, co-secrétaire du Comité *ad hoc* sur l'intelligence artificielle du Conseil de l'Europe, dénonce ce procédé d'individualisation en trompe-l'œil car il laisse craindre que « le comportement passé d'un certain groupe de personnes finirait en effet par influencer le sort d'un individu qui, en tant qu'être humain unique, est inséré dans un contexte

77 Cette philosophie « repousse le déterminisme positiviste. Elle n'admet ni le fatalisme biologique dérivé de Lombroso, ni la nécessité sociale de Ferri, ni même cette prédisposition congénitale imposée de manière inéluctable par la nature que certains tenants de la prophylaxie criminelle, certains spécialistes de l'hygiène mentale ou certains adeptes d'une typologie criminelle construite comme une sériation *a priori* des délinquants voudraient parfois faire admettre », in M. ANCEL, *La défense sociale nouvelle*, Paris, Cujas, 1980, p. 185.

social spécifique. Il a reçu une éducation et développé des compétences propres et ne peut être jugé qu'en fonction de son propre degré de culpabilité et des motivations spécifiques qui l'ont conduit à commettre une infraction »⁷⁸. En cela, l'algorithme prédictif semble marquer un retour à la criminologie du XIX^e siècle inscrivant les comportements délictueux dans une logique semblable à celle des lois de la physique⁷⁹.

Au lieu d'apprécier le « libre arbitre » et de favoriser la réinsertion sociale de l'individu concerné, le profilage statistique encourage une application mécanique de la peine et recherche l'optimisation de la population carcérale⁸⁰.

Le juge soumis à cette approche déterministe de la peine ne sert pas les buts de l'individualisation, puisqu'il satisfait la prévention de la dangerosité et fait donc primer la perception de la société sur sa décision plutôt que les besoins réels du prévenu. Cette illusion d'individualisation transforme l'office du juge dont le rôle n'est plus de personnaliser une peine mais de prédire l'avenir. Incité à domestiquer le hasard, le juge cherche désormais moins à comprendre les causes d'un comportement violent qu'à les prédire. La protection de la société semble prendre, sous l'impulsion des outils algorithmiques, le devant face à la réinsertion du prévenu.

Plus inquiétant encore, le fonctionnement de l'algorithme n'est pas aussi fiable et objectif que ses promoteurs le disent⁸¹. Des biais cognitifs et statistiques sont susceptibles de provoquer des résultats discriminatoires, en raison de la qualité des données et de la programmation de leur traitement. L'algorithme peut ainsi présenter des « faux positifs » lorsque le risque de récidive est surestimé ou des « faux négatifs » lorsque celui-ci est sous-estimé. En mai 2016, l'organisation *ProPublica* a montré, après une étude de deux ans auprès du comté de Broward en Floride, que l'outil COMPAS comportait des biais discriminatoires envers la population afro-américaine⁸². La prise en compte de facteurs indépendants de la volonté des individus et hors de leur contrôle tels que le genre, le contexte familial, l'adresse, l'appartenance ethnique, accentue les biais discriminatoires. Ainsi, un individu qui n'a jamais commis d'infraction peut être évalué à haut risque tout simplement parce qu'il est jeune, au chômage et réside dans un quartier défavorisé. Dans ces circonstances, il apparaît délicat d'utiliser l'algorithme à des fins d'individualisation des peines.

En second lieu, le renoncement du juge à l'individualisation pourrait s'expliquer par la difficulté de justifier et légitimer une décision prenant le contre-pied de la prédiction de l'algorithme. Le juge s'engagerait alors dans une lutte contre le déterminisme des profils individuels. Quels sont les moyens de cette lutte ? Si l'algorithme a déjà pris en considération les facteurs traditionnels de l'individualisation des peines (circonstances de l'infraction, la personnalité de son auteur), le juge peut envisager d'en proposer une pondération différente. Par exemple, le juge peut accorder davantage d'importance à la position de mère de famille isolée du prévenu, qu'à son « profil » de récidiviste.

78 https://lestempselectriques.net/index.php/2018/07/30/intelligence-artificielle-et-proces-penal/#_ftnref2 (consulté le 28 janvier 2020).

79 Mentionnons à ce titre les travaux d'André-Michel GUERRY, père de la « statistique morale » à l'origine de la criminologie. Il est l'auteur en particulier d'un *Essai sur la statistique morale de la France*, Paris, Crochard, 1833.

80 N. JEANNE et V. MAQUET, *op. cit.*, p. 234.

81 J. K. ELEK, R. K. WARREN et P. M. CASEY, *op. cit.*, p. 15.

82 J. ANGWIN, J. LARSON, S. MATTU et L. KIRCHNER, *op. cit.*

Afin de relativiser la portée de la prédiction, le juge devrait confronter ce dernier aux éléments qui par nature échappent à l'appréciation de l'algorithme. Il s'agit en particulier des valeurs et des principes intangibles qui ne peuvent se voir attribuer un poids mathématique et doivent donc être écartés des facteurs pris en compte par l'algorithme. Il revient donc au juge de reprendre sa marge discrétionnaire d'interprétation afin de décider de la portée de ces principes pour chaque cas d'espèce. C'est le reproche formulé par Alain Supiot à l'encontre de la doctrine *law and economics* selon laquelle ce qui n'a pas de prix n'a pas d'existence et ce qui ne peut pas être calculé par l'algorithme doit être écarté de l'équation⁸³.

En mobilisant la source algorithmique, le juge doit ainsi fournir un effort significatif pour respecter pleinement son office. S'il est devenu fréquent d'associer l'opération de juger à une obligation de savoir⁸⁴, le juge doit de lui-même distinguer et concilier la neutralité et l'objectivité des connaissances scientifiques avec la souveraineté de son appréciation⁸⁵. Cette liberté d'appréciation fait nécessairement appel à une part de croyance qu'il utilise pour forger son intime conviction. Contrairement à la science qui est évaluée à partir de la vraisemblance ou de la vérité, la croyance s'appuie sur « des connaissances valuées, c'est-à-dire pondérées par des valeurs du bon et du mauvais »⁸⁶. Pour remplir pleinement son rôle, la décision de justice doit puiser dans des considérations de justice et non seulement de justesse, comme le précisent Danièle Bourcier et Monique De Bonis⁸⁷. Par ailleurs, si le juge est habilité à décider en cas d'incertitude c'est parce qu'il est capable de prudence : selon Alain Supiot, le « juris-prudent » sait se scruter lui-même, tenir compte de l'expérience du passé, mesurer la valeur des faits allégués devant lui, et anticiper les effets de sa décision⁸⁸. Par définition, l'algorithme est incapable d'un tel exercice.

Avec la généralisation de l'usage des outils algorithmiques dans le procès pénal, on peut craindre un appauvrissement de la motivation des décisions juridictionnelles, du moins d'après les critères traditionnels d'appréciation de leur qualité. En 2007, Benoît Frydman analysait la mutation des critères d'évaluation de la qualité des décisions de justice en identifiant trois générations de contrôle : la qualité substantielle évaluant la pertinence de l'argumentation du juge a cédé la place à la qualité procédurale de la décision en fonction de la place accordée par celle-ci au débat contradictoire, à l'égalité des armes et au procès équitable, avant que se développe l'approche managériale⁸⁹. Les méthodes du management se trouvent appliquées au travers d'indicateurs chiffrés et d'outils de gestion statistiques des performances. L'utilisation du résultat de l'algorithme dans la motivation du juge illustre ce glissement vers l'optimisation et la rationalisation des décisions de justice. L'efficacité des décisions de justice est ainsi favorisée au détriment de la qualité au sens des deux premières générations (substantielles et procédurales) de Benoît Frydman⁹⁰.

83 A. SUPIOT, *op. cit.*, p. 205-206.

84 D. BOURCIER et M. DE BONIS, *op. cit.*, p. 11.

85 *Ibid.*, p. 13.

86 *Ibid.*, p. 14.

87 *Idem.*

88 A. SUPIOT, *op. cit.*, p. 144.

89 B. FRYDMAN, « L'évolution des critères et des modes de contrôle de la qualité des décisions de justice », in P. MBONGO (dir.), *La qualité des décisions de justice*, Strasbourg, Conseil de l'Europe, 2007, p. 18.

90 Cette tendance pourrait diminuer l'acceptabilité sociale du procès car il est difficile de comprendre la manière dont celle-ci a été prise, J.-B. DUCLERCQ, « Les algorithmes en procès », *RFDA*, n° 1, 2018, p. 131-142.

Une nouvelle source de dégradation du système judiciaire américain semble enclenchée. Au-delà de l'insuffisance des précautions d'usage qui leur sont transmises, certains juges ont pu affirmer que leur expérience d'utilisation de l'algorithme prédictif les avait aidés à « faire confiance à l'outil plutôt que de continuer à croire que leur jugement subjectif était plus précis pour identifier le niveau et les facteurs de risque »⁹¹. C'est un mouvement de défiance envers le jugement subjectif du juge qui est ici à l'œuvre, y compris de la part des juges eux-mêmes qui sont pourtant très conscients du rôle social de représentation qu'ils incarnent en particulier dans les systèmes de *common law*. Qu'elle soit imposée ou provoquée par le juge, la réduction de sa marge d'appréciation pourrait être à l'origine d'un déni de justice. La mobilisation de données extérieures à la légalité pénale pour fixer le niveau de la peine ainsi que la délégation de l'individualisation des peines à un algorithme risquent, à défaut de sérieuses garanties, de priver le justiciable d'un jugement juste et équitable.

91 J. K. ELEK, R. K. WARREN et P. M. CASEY, *op. cit.*, p. 60 : « *Some said that the training helped them trust in the tool rather than continue to believe that their subjective judgment was more accurate in identifying risk level and factors, and to see that rather than taking away their professional judgment.* »