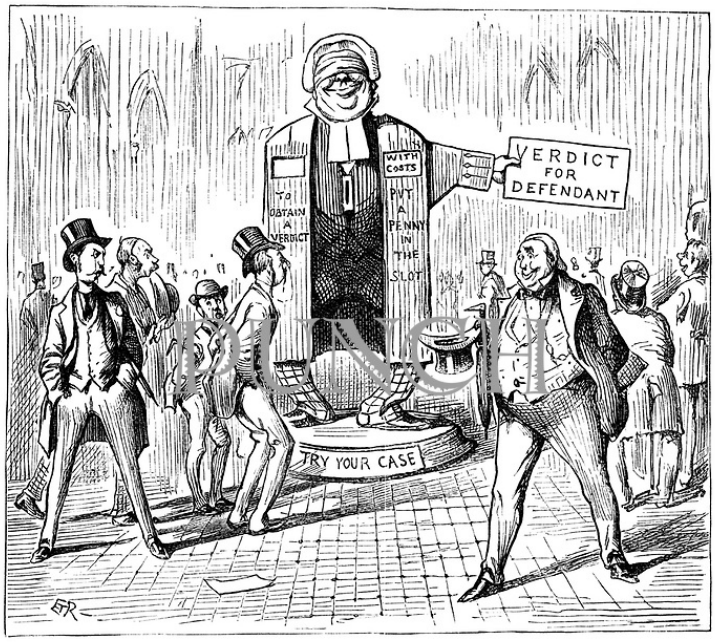
COMPAS:

Un outil impartial ou biaisé ?



AUTOMATIC ARBITRATION.

No more Exorbitant Fees! No more Law! No more Trials!

Gaston Bidou
Quentin Hillebrand
Zhanbo Hu
Hugo Pouzet
Justine Rayssac
Raphaël Rozenberg
Jianan Shi
Morgane Turlan

2022

Illustration de la couverture : « Arbitrage automatique. Plus de frais exorbitants ! Plus de loi ! Plus de procès ! in Reed, E.T. 1890. « Automatic Arbitration ». Punch, 17 mai 1890.
https://www.punch.co.uk/image/I0000AsGcmxxhoIk.

Cette publication a été réalisée par des étudiants en troisième année du cycle ingénieur de Mines Paris PSL Research University. Il présente le travail réalisé dans le cours intitulé « Descriptions de controverse », qui a pour objectif d'introduire les étudiants à l'univers incertain de la recherche scientifique et technique et de les sensibiliser aux enjeux de la participation citoyenne.

Introduction

La révolution numérique et l'essor des "*Big Data*" touchent aujourd'hui de nombreux secteurs d'activités et de la vie sociale. La justice n'y échappe pas : l'irruption des programmes informatiques et de l'apprentissage automatique ("*machine learning*") aux usages forts diversifiés dans le domaine judiciaire a conduit à l'émergence d'un secteur à l'intersection de la science informatique et de la sphère juridique que d'aucuns nomment "justice algorithmique". Dans la littérature, nombreux sont les termes qui tentent de qualifier la montée en puissance des logiciels dans les différentes étapes de réglement du contentieux : qu'il s'agisse de "justice algorithmisée" ou encore de "justice prédictive", ces formulations ne recouvrent pas une seule et même réalité[1]. Le recours aux algorithmes dans la justice englobe, de fait, un spectre bien plus étendu que celui de la simple prédiction du jugement stricto sensu.

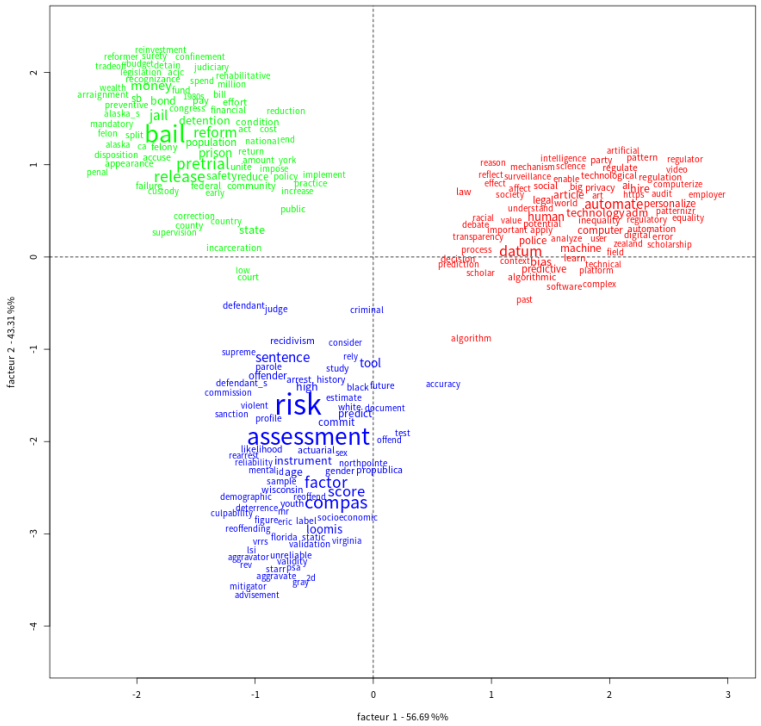


Figure 1 - Un champ majeur de la justice algorithmisée : l'évaluation du risque de récidive

L'analyse distingue nettement trois groupements de termes. Le groupement vert rassemble des termes issus du champ lexical du risque de non-comparution à un procès, tandis que le groupement bleu se concentre sur la prédiction du risque de récidive pour un prévenu. Le groupement rouge recouvre quant à lui le champ plus large des données informatiques. Les analyses ont été effectuées sur la base des 81 articles internationaux affichés par la base de données Europresse en réponse à l'équation de recherche " justice & algorithm & tool ". L'analyse quantitative a été réalisée au moyen du logiciel IRaMuTeQ sur la période 2000-2021.

[1] Jean, Aurélie, Victor Storchan, et Adrien Basdevant. 2021. « Mécanisme d'une justice algorithmisée ». Fondation Jean Jaurès.

Ainsi, dans le champ de la justice pénale où la justice algorithmisée s'étend de l'interpellation à la condamnation d'un prévenu[23], l'utilisation de logiciels d'évaluation du risque de récidive occupe une place de premier plan (*Figure 1*). Pourtant, l'idée même d'estimer le risque que représente un ex-condamné pour la société n'est guère récente[4]. En 1898, le juriste français Raymond Saleilles évoquait déjà la nécessité d'individualiser la peine prononcée dans son ouvrage éponyme[5]. Outre-Atlantique, l'évaluation du risque de récidive a connu une première formalisation dans les années 1930, au sein de l'École de Chicago. Dans ce courant américain de sociologie urbaine, Ernest W. Burgess préconisait l'utilisation de tableaux statistiques - avec près de 21 facteurs - pour prédire le risque de récidive d'un individu. Les critères s'appuyaient alors exclusivement sur des variables dites " statiques ", c'est-à-dire considérées comme stables au cours de la vie du prévenu (nombre de frères et sœurs, niveau d'éducation, etc.). Dans les années 1970, alors que l'incarcération de masse battait son plein aux États-Unis[6], l'idée de développer des outils systématiques de quantification des risques de récidive s'est progressivement imposée. Dans le courant des années 1980, les facteurs statiques ont été couplés à des facteurs qualifiés de " dynamiques ", susceptibles de varier au cours de l'existence d'un individu (consommation d'alcool, etc.). Le cadre législatif aux États-Unis - avec le *Sentencing Reform Act* de 1984, censé réduire l'aléa judiciaire - a permis l'essor de tels outils. *A contrario*, en France, en dépit de la loi CADA du 17 juillet 1978 sur la liberté d'accès aux documents administratifs (dans un souci de transparence et de réduction de l'aléa judiciaire), les outils de détermination du risque de récidive n'ont pas été systématiquement mis en œuvre. Enfin, avec l'avènement des nouvelles technologies dans les années 2000 (*Figure 2*), la justice pénale a été marquée par l'arrivée d'algorithmes d'apprentissage automatique qui, en se basant sur un certain nombre de critères fournis en entrée, indiquent directement en sortie un résultat relatif au risque de récidive, ou au risque de non-comparution à l'audience[7]. Parmi eux se distinguent notamment le logiciel PTRA (*Pretrial Risk Assessment*) utilisé entre 2001 et 2007 à l'échelle fédérale américaine dans plus de 500 000 affaires, ou encore l'algorithme COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), créé en 1999 par la société Northpointe - renommée Equivant - qui calcule, en plus du risque de non-comparution, un risque de récidive simple et de récidive violente[8].

---

[2] Berthet, Vincent, et Léo Amsellem. 2021. Les nouveaux oracles: comment les algorithmes prédisent le crime. Paris: CNRS éditions.

[3] Le Grand Continent, 6 octobre 2021. « Les nouveaux oracles, une conversation avec Vincent Berthet et Léo Amsellem ». Storchan, Victor. https://legrandcontinent.eu/fr/2021/10/06/les-nouveaux-oracles-une-conversation-avec-vincent-berthet-et-leo-amsellem/.

[4] Raucher, Sammy. 2021. « Algorithms and Pre-Trial Assessment » Mini-Lecture component : « What Makes an Algorithm Fair? "Fairness" in the COMPAS Recidivism Risk Algorithm ». Human Contexts and Ethics. https://www.youtube.com/watch?v=HfxhmMdA8XQ.

[5] Saleilles, Raymond. 1898. L'individualisation de la peine: étude de criminalité sociale. Bibliothèque générale des sciences sociales. Paris: Baillière. https://books.google.fr/books?id=4GWc37FGiKQC.

[6] Le Grand Continent, 6 octobre 2021. « Les nouveaux oracles, une conversation avec Vincent Berthet et Léo Amsellem ». Storchan, Victor. https://legrandcontinent.eu/fr/2021/10/06/les-nouveaux-oracles-une-conversation-avec-vincent-berthet-et-leo-amsellem/.

[7] Abu Elyounes, Doaa. 2020b. « Contextual Fairness: A Legal and Policy Analysis of Algorithmic Fairness ». Journal of Law, Technology and Policy 2020 (1): 1-54. https://doi.org/10.2139/ssrn.3478296.

[8] Northpointe. 2015. « Practitioner's Guide to COMPAS Core ». Northpointe. https://s3.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf

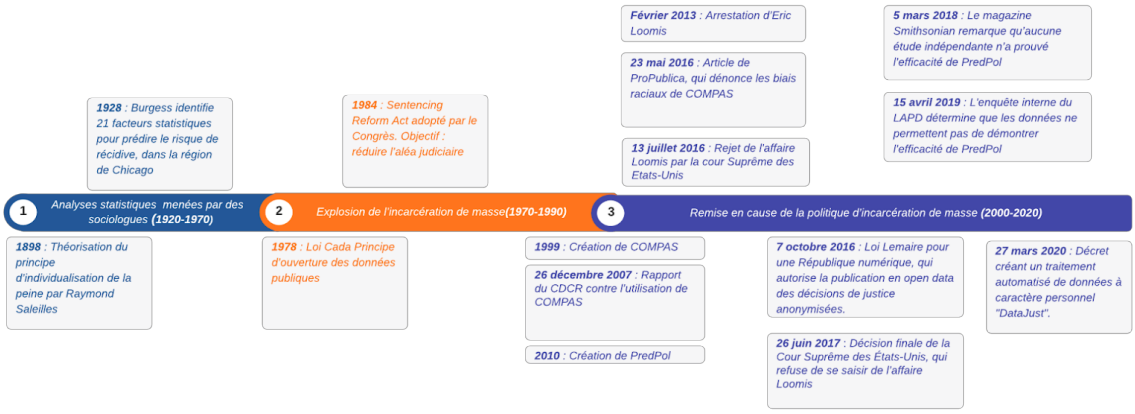


Figure 2 - Chronologie de l'utilisation d'algorithmes dans le champ de la justice

> *Trois périodes se distinguent : une première ère de 1920 à 1970 marquée par le développement de théories sociologiques et criminologiques sur les déterminants des crimes et délits, suivie par une seconde période (de 1970 à 1990) de mise en œuvre d'outils d'évaluation systématique des risques (de récidive, de non-comparution à une audience), à mesure que le cadre législatif s'est ouvert. Enfin, la période de 1990 à nos jours est caractérisée par la mise à profit de la masse croissante de données disponibles, par le développement de nouveaux outils d'aide à la prise de décision juridique (analyse de jurisprudence, probabilité de délit dans une zone géographique, etc.) alors même que la politique d'incarcération de masse affronte une crise, en France comme aux États-Unis.*

C'est précisément le logiciel COMPAS qui a fait l'objet d'une attention toute particulière au cours de la dernière décennie (*Figure 2*). En février 2013, dans l'État américain du Wisconsin, Eric Loomis a été arrêté au volant d'une voiture alors qu'il fuyait une fusillade que son passager était soupçonné d'avoir orchestrée[9]. Loomis a nié toute implication dans la fusillade mais a été jugé en tant que récidiviste pour cinq chefs d'accusation. En première instance, la Cour a utilisé le logiciel COMPAS pour évaluer le risque de récidive que présentait le prévenu Loomis. L'évaluation se fonde sur un questionnaire dont sont extraits 137 critères, non-communiqués par Equivant, pondérés de manière à aboutir à un risque de récidive allant de 0 (risque nul) à 10 (risque maximal). À l'issue de l'audience, Loomis a été condamné à 6 ans de prison et 5 ans de surveillance préventive[10]. Eric Loomis a interjeté appel, arguant que l'utilisation de COMPAS a violé son droit à un procès équitable - rompant de fait le "*due process*" garanti par le cinquième amendement de la Constitution des États-Unis. Le 13 juillet 2016, la Cour Suprême du Wisconsin a rejeté l'appel de Loomis, avançant que le juge de première instance aurait prononcé la même peine *in fine*, indépendamment du résultat fourni par le logiciel COMPAS[11]. Plus récemment, en 2017 (*Figure 2*), COMPAS a de nouveau fait la une des médias américains[12] lorsqu'une étude menée par le journal d'investigation ProPublica a mis en évidence les biais raciaux de l'algorithme. Après avoir analysé a posteriori le profil de 18 610 détenus du comté de Broward en Floride, évalués par COMPAS entre 2013 et 2014, les auteurs ont conclu que les détenus afro-américains n'ayant pas récidivé étaient deux fois plus susceptibles d'être considérés à "haut risque de récidive" que leurs homologues blancs. À l'inverse, les détenus blancs qui ont récidivé étaient considérés comme deux fois plus susceptibles d'être à "bas risque de récidive" que leurs

[9] Dika, Khaled. 2020. « L'affaire Loomis: Les fantômes de Descartes et de Grotius à l'assaut de la justice? » HAL (preprint). https://hal.archives-ouvertes.fr/hal-02566382.

[10] State of Wisconsin v. Eric L. Loomis. Cour suprême du Wisconsin. 2016. 881 N.W.2d 749. https://caselaw.findlaw.com/wi-supreme-court/1742124.html.

[11] State of Wisconsin v. Eric L. Loomis. Cour suprême du Wisconsin. 2016. 881 N.W.2d 749. https://caselaw.findlaw.com/wi-supreme-court/1742124.html.

[12] Le Grand Continent, 6 octobre 2021. « Les nouveaux oracles, une conversation avec Vincent Berthet et Léo Amsellem ». Storchan, Victor. https://legrandcontinent.eu/fr/2021/10/06/les-nouveaux-oracles-une-conversation-avec-vincent-berthet-et-leo-amsellem/.

homologues afro-américains[13]. Ce résultat a nourri de nombreux débats dans la presse nationale et internationale dès 2017 (*Figure 3*), ainsi que dans la littérature scientifique qui a cherché une explication à ces observations[14].

Ainsi, le développement et le recours à des algorithmes dans le champ de la justice suscitent des débats croissants (*Figure 3*) et des réflexions, aussi bien aux États-Unis qu'en France, sur les biais que ces outils introduisent dans l'élaboration du jugement, sur leur intelligibilité par les acteurs concernés (prévenus et professionnels du droit) et sur les transformations qu'ils induisent sur le travail et le système judiciaire. Le présent article s'articule par conséquent autour des interrogations suivantes : quels éléments de conception de COMPAS ont conduit une partie des acteurs à dénoncer des biais racistes ? Parallèlement à cette architecture, en quoi le logiciel cumule-t-il différents niveaux d'opacité qui réduisent son intelligibilité ? Plus généralement, quelles sont les conséquences de ces zones d'ombre sur la perception de ces logiciels par la société et par le système judiciaire ? Enfin, comment ces débats remodèlent-ils le panorama d'acteurs et de réglementations dans le domaine judiciaire ?

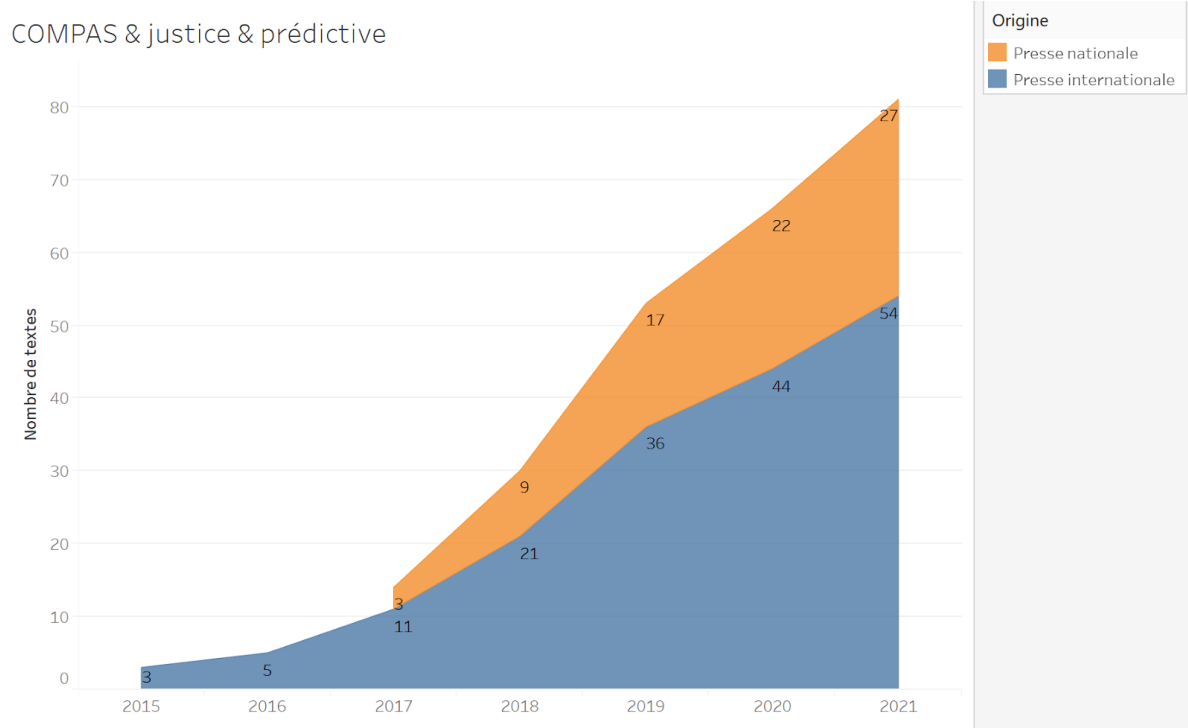


Figure 3 - Augmentation exponentielle du nombre d'articles évoquant l'outil COMPAS dès 2016

Le traitement médiatique de COMPAS a fortement augmenté, dans la presse nationale française et internationale, à partir de l'affaire Loomis et de l'étude menée par ProPublica en 2016. Les analyses ont été effectuées sur des articles de la presse nationale et internationale issus de la base de données Europresse, consultée au moyen de l'équation de recherche "COMPAS & justice & predictive".

13 ProPublica. 23 mai 2016. « Machine Bias ». Angwin, Julia, Jeff Larson, Surya Mattu, et Lauren Kirchner. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

14 Dressel, Julia, et Hany Farid. 2018. « The accuracy, fairness, and limits of predicting recidivism ». Science Advances 4 (1). https://doi.org/10.1126/sciadv.aao5580.

COMPAS : un logiciel raciste ?

La notion d'équité en débat

La controverse sur le caractère raciste de COMPAS est apparue avec la publication en 2016 par ProPublica, un journal d'investigation, d'un article accusant COMPAS d'être biaisé contre les personnes noires[15]. Cette accusation est d'abord construite sur des comparaisons entre des prévenus blancs et noirs dont l'historique d'arrestation et d'incarcération est connu, ainsi que leur score de risque attribué par le logiciel COMPAS. Dans toutes ces comparaisons, le prévenu noir se voit attribué un score de risque plus élevé alors que son historique d'arrestation semble le rendre moins dangereux que le prévenu blanc. Ces scores de risque sont les estimations du risque de récidive calculées par COMPAS à partir de l'historique d'infraction, mais également de réponses à un questionnaire, dispensé auprès du prévenu par un travailleur social, qui comporte notamment des items sur l'âge, le genre et le code postal. Ces premiers constats ont été confirmés par une étude réalisée par des journalistes de ProPublica portant sur plus de 10 000 prévenus dans le comté de Broward, en Floride[16]. Plus précisément, l'étude révèle que COMPAS était susceptible d'identifier à tort les prévenus noirs comme de futurs criminels, étiquetés ainsi deux fois plus souvent que les prévenus blancs, preuve, pour les auteurs de l'étude, de l'existence d'un biais raciste du logiciel.

En réponse à cela, Equivant, la société qui a mis au point le logiciel, reproche à ProPublica son choix de critère pour mesurer les biais. Selon eux, COMPAS n'est pas biaisé puisque pour un score de risque donné, la probabilité de récidive est la même pour les Noirs et les Blancs[17]. Le désaccord ne porte pas ici sur la conformité du logiciel à une définition de l'équité (ou plus communément "fairness" en informatique) mais sur le choix même de la définition de "*fairness*" algorithmique adaptée. Ce sujet est une question récurrente dans la détection de biais et est largement étudiée dans la littérature aussi bien scientifique que sociologique[18]. Cette notion, qui recoupe celles de justesse, c'est-à-dire de qualité de la prédiction, et de justice, se rapportant à l'équité de la prédiction, a en réalité plusieurs traductions en termes statistiques.

Dans le cas du logiciel COMPAS, la société Equivant définit la "*fairness*" de l'algorithme selon la notion de calibration, qui signifie qu'un score donné par l'algorithme indique le même risque de récidive pour un Noir et pour un Blanc (voir Figure 4, haut, pour un exemple simplifié). ProPublica se réfère quant à eux à la notion d'égalité des chances, qui requiert que pour un même profil de risque réel, l'algorithme fasse des prédictions similaires pour les deux groupes (*Figure 4, bas*). Cela implique par exemple que les Noirs au profil non risqué ne soient pas incarcérés plus souvent que les Blancs ayant le même profil[19].

15 ProPublica. 23 mai 2016. « Machine Bias ». Angwin, Julia, Jeff Larson, Surya Mattu, et Lauren Kirchner. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

16 Ibid.

17 Brennan, Tim, William Dieterich, et Beate Ehret. 2009. « Evaluating the Predictive Validity of the Compas Risk and Needs Assessment System ». Criminal Justice and Behavior 36 (1): 21-40. https://doi.org/10.1177/0093854808326545.

18 Wachter, Sandra, Brent Mittelstadt, et Chris Russell. 2021. « Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law ». West Virginia Law Review 123 (3): 735-85

19 Berk, Richard, Hoda Heidari, Shahin Jabbari, Michael Kearns, et Aaron Roth. 2018. « Fairness in Criminal Justice Risk Assessments: The State of the Art ». Sociological Methods & Research 50 (1): 3-44. https://doi.org/10.1177/0049124118782533.

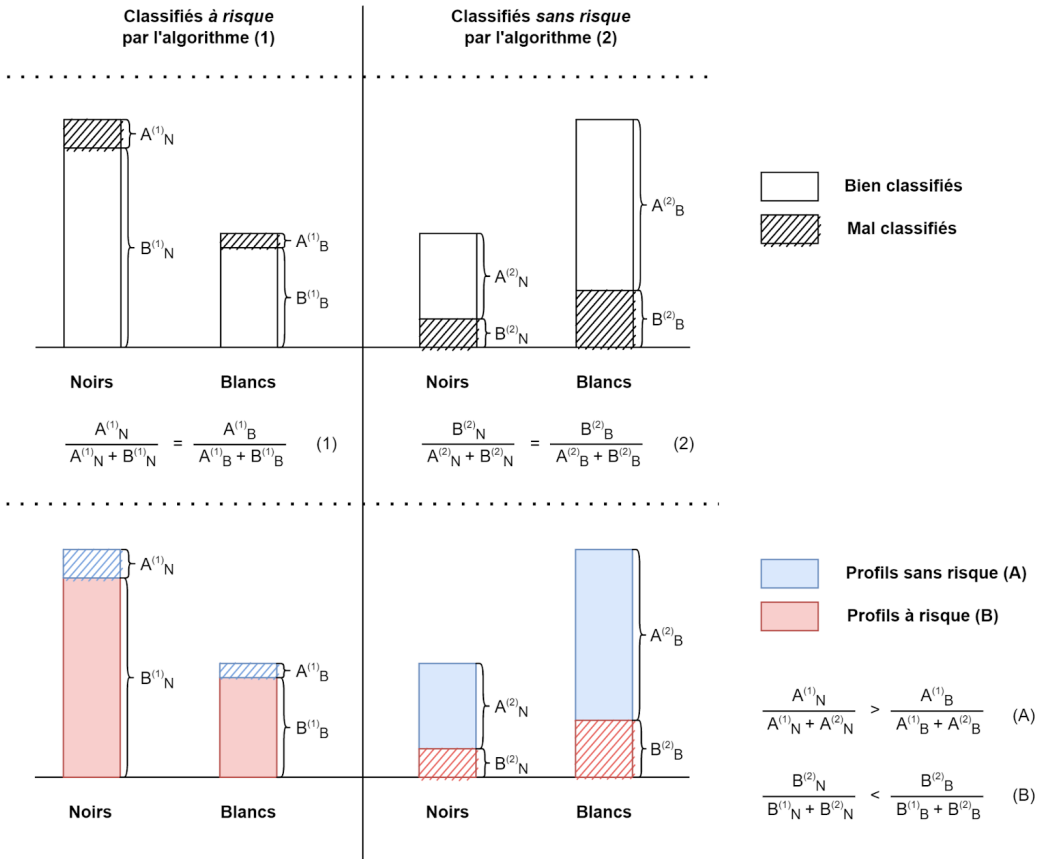


Figure 4 - Des définitions irréconciliables de la fairness

Dans cet exemple simplifié, les égalités (1) et (2) indiquent qu'une prédiction de l'algorithme (par exemple "à risque") sera aussi informative pour un Noir que pour un Blanc (même taux d'erreur) : c'est la calibration, qui est respectée. En revanche, les inégalités (A) et (B) indiquent respectivement qu'un Noir au profil non risqué a plus de chance qu'un Blanc au même profil d'être emprisonné, et qu'un Blanc au profil risqué a plus de chance qu'un Noir au même profil d'être laissé libre : l'égalité des chances n'est pas respectée.

Bien que les deux propriétés semblent désirables pour tout algorithme, ces deux définitions statistiques sont mathématiquement irréconciliables dans le cas général. En particulier Richard Berk et al., chercheurs en informatique, statistique et criminologie, formulent cette condition d'incompatibilité comme suit :

> "Lorsque les taux de base diffèrent en fonction du groupe considéré [et que l'algorithme ne fait pas des prévisions parfaites en toutes circonstances] on ne peut pas avoir à la fois une égalité de précision [la calibration] et une égalité des taux de faux négatifs et de faux positifs [l'égalité des chances]."[20]

Autrement dit, puisque les personnes noires ont un plus grand risque de récidive que les personnes blanches aux États-Unis (taux de base différents), il n'est pas possible de concevoir un logiciel de prédiction de récidive qui satisfasse à la fois la conception de "*fairness*" d'Equivant et celle de ProPublica. Il apparaît dès lors qu'un compromis est nécessaire entre les différentes définitions de "*fairness*". Mais ce compromis n'apparaît pas seulement entre les différentes notions de "*fairness*" puisque la diminution des biais se fait également au

[20] Berk, Richard, Hoda Heidari, Shahin Jabbari, Michael Kearns, et Aaron Roth. 2018. « Fairness in Criminal Justice Risk Assessments: The State of the Art ». Sociological Methods & Research 50 (1): 3-44. https://doi.org/10.1177/0049124118782533.

détriment du pouvoir prédictif de l'algorithme[21]. La question du modèle de "*fairness*" le plus souhaitable dans ce contexte n'est bien sûr pas réglée, mais certains chercheurs proposent des pistes, comme la multicalibration qui consiste à prendre en compte l'appartenance à plusieurs sous-groupes (par exemple à la fois la couleur de peau et le sexe) dans une optique d'intersectionnalité, c'est à dire qu'un score donné par l'algorithme est associé au même niveau de risque pour tous les groupes et sous-groupes. Dans un entretien, la chercheuse et juriste interrogée défendait sa préférence pour ce modèle :

> "Le calibrage est important d'un point de vue juridique, car il renforce la confiance dans les algorithmes. Nous ne voulons pas renoncer au calibrage, et il ne peut pas être mauvais de calibrer en utilisant différents sous-groupes avec le calibrage multiple."

Ces constats permettent de voir la tension qu'il existe sur le choix de la métrique de "*fairness*" utilisée, et donc sur la définition même de "*fairness*" sous-jacente. Cependant, les débats ne portent pas uniquement sur le choix du critère de "*fairness*" le plus adapté mais également sur celui des acteurs amenés à faire ce choix. En effet, certains sociologues regrettent que le débat soit réduit à des questions techniques dans la littérature de recherche. Ainsi, certains prennent position en faveur d'une séparation entre les questions de sciences et celles de droit[22][23] :

> "Il appartiendra aux parties prenantes - et non aux criminologues, aux statisticiens et aux informaticiens - de déterminer les compromis. [...] Ce sont des questions de valeurs et de droit, et en fin de compte, de processus politique. Ce ne sont pas des questions de science."

Nous avons vu que la question de la méthode d'évaluation de l'équité de l'algorithme n'est pas tranchée. Elle se double d'une question sur les sources de biais au sein de l'algorithme lui-même.

▪ Des critères fatalement biaisés ?

COMPAS, ainsi que la plupart des outils d'évaluation du risque de récidive, fonde ses prédictions sur un certain nombre de données concernant le prévenu[24]. Ces données, qui sont fournies en entrée à l'outil, peuvent varier selon l'algorithme, et concernent par exemple dans le cas de COMPAS le casier judiciaire, le délit dont le prévenu est actuellement accusé ou encore ses antécédents de consommation de drogue. Étant donné que l'algorithme voit exclusivement le prévenu à travers le prisme de ces données, le choix des critères retenus par le concepteur de l'algorithme est crucial.

D'après une juriste chercheuse dans le domaine de la justice algorithmique, les critères d'entrée de l'algorithme ne sont pas nouveaux :

> "Autrefois, les juges devaient faire ce genre de calculs par eux-mêmes. Cela a commencé par des tables de critères et c'est devenu de plus en plus complexe : ces algorithmes ne sont pas un phénomène novateur".

---

[21] Abu Elyounes, Doaa. 2020b. « Contextual Fairness: A Legal and Policy Analysis of Algorithmic Fairness ». Journal of Law, Technology and Policy 2020 (1): 1-54. https://doi.org/10.2139/ssrn.3478296.

[22] Wachter, Sandra, Brent Mittelstadt, et Chris Russell. 2021. « Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law ». West Virginia Law Review 123 (3): 735-85.

[23] Berk, Richard, Hoda Heidari, Shahin Jabbari, Michael Kearns, et Aaron Roth. 2018. « Fairness in Criminal Justice Risk Assessments: The State of the Art ». Sociological Methods & Research 50 (1): 3-44. https://doi.org/10.1177/0049124118782533.

[24] Abu Elyounes, Doaa. 2020a. « Bail or Jail? Judicial versus Algorithmic Decision-Making in the Pretrial System ». Science and Technology Law Review 21 (2): 376-445. https://doi.org/10.7916/stlr.v21i2.6838.

Ces critères sont souvent élaborés à partir de théories criminologiques, comme le mentionne le guide de l'utilisateur de COMPAS[25] et comme l'expliquent Tim Brennan et William Dieterich, chercheurs chez Equivant[26]. Ainsi, de la théorie de l'apprentissage social sont tirés des critères liés à la proximité avec des pairs antisociaux : dans le questionnaire COMPAS, se trouve par exemple la question "Combien de vos amis/connaissances ont déjà été arrêtés ?". La théorie des activités routinières met l'accent sur des critères liés à une vie peu structurée : "À quelle fréquence vous ennuyez-vous ?". Pour prendre un dernier exemple, la théorie de la sous-culture est à l'origine de questions sur l'appartenance du prévenu et de ses proches à un gang, ainsi que sur le quartier de résidence.

Cependant, Sharad Goel *et al.*, chercheurs en politiques publiques, droit et statistiques, remarquent que malgré les bases théoriques supposément solides de ces critères, des "erreurs de mesure", définies comme "les différences entre la réalité et sa représentation dans les données", sont présentes et engendrent des biais[27]. Pour donner un exemple simple, prenons le critère du passé criminel d'un individu. S'il est estimé non pas à partir des actes eux-mêmes mais à partir des arrestations, bien plus faciles à compter, alors le taux d'arrestation supérieur des Noirs par rapport aux Blancs (pour le même comportement criminel) va entraîner un biais de l'algorithme en défaveur des premiers. Ainsi, même si les algorithmes ne prennent pas en entrée des critères sensibles explicites, tels que la couleur de peau ou le sexe, d'autres critères peuvent implicitement leur être corrélés.

Selon Abu Elyounes, ce type de biais dans les critères peut être corrigé en entraînant des modèles différents pour chaque classe (par exemple un modèle par couleur de peau), même si l'utilisation explicite du critère sensible (la couleur de peau dans l'exemple précédent) peut être légalement délicate au moment de la prise de décision[28].

▪ Des biais dans les données : un problème plus profond ?

Cependant, contrairement aux biais dans les critères qui peuvent être raisonnablement surmontés, Goel *et al.* soulignent que la plus grande difficulté vient de l'autre versant des données : les "étiquettes"[29]. En effet, pour entraîner un modèle par apprentissage automatique, le concepteur doit fournir en amont des données "étiquetées" : par exemple, pour COMPAS, pour chaque prévenu de la base de données, celle-ci contient non seulement les critères susmentionnés, mais aussi le résultat réel (le prévenu a récidivé dans les deux ans ou non) afin que l'algorithme puisse comparer ses prédictions à la réalité durant son apprentissage. D'après Goel *et al.* :

> "Il est souvent possible de tenir compte statistiquement des biais des caractéristiques [(critères)], mais il est considérablement plus difficile de remédier aux biais des étiquettes. Ainsi, ce dernier problème est sans doute l'un des plus sérieux auxquels fait face la conception d'outils équitables pour la prédiction de risque."[30]

---

[25] Northpointe. 2015. « Practitioner's Guide to COMPAS Core ». Northpointe. https://s3.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf

[26] Brennan, Tim, et William Dieterich. 2018. « Correctional Offender Management Profiles for Alternative Sanctions (COMPAS) ». In Handbook of Recidivism Risk/Needs Assessment Tools, Singh J., Kroner D., Wormith, J., Desmarais S., Hamilton Z., 49-75. Hoboken (USA): John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119184256.ch3.

[27] Goel, Sharad, Ravi Shroff, Jennifer L. Skeem, et Christopher Slobogin. 2021. « The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment ». In Research Handbook on Big Data Law, Roland Vogl, 9-28. Law 2021. Rochester, NY: Social Science Research Network. https://doi.org/10.4337/9781788972826.00007.

[28] Abu Elyounes, Doaa. 2020b. « Contextual Fairness: A Legal and Policy Analysis of Algorithmic Fairness ». Journal of Law, Technology and Policy 2020 (1): 1-54. https://doi.org/10.2139/ssrn.3478296.

[29] Goel, Sharad, Ravi Shroff, Jennifer L. Skeem, et Christopher Slobogin. 2021. « The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment ». In Research Handbook on Big Data Law, Roland Vogl, 9-28. Law 2021. Rochester, NY: Social Science Research Network. https://doi.org/10.4337/9781788972826.00007.

[30] Ibid.

Par exemple, la récidive n'est comptabilisée que lorsqu'il y a condamnation en justice. Or, les Noirs étant plus souvent condamnés que les Blancs pour les mêmes crimes et délits, cela induit un biais racial en leur défaveur.

De manière plus générale, les algorithmes d'apprentissage reproduisent les schémas du passé, qui peuvent contenir des biais[31]. Ainsi, contrairement aux problèmes de "*fairness*" examinés précédemment, qui pourraient être atténués, les biais apparaissant dans les données sont plus difficiles à prendre en compte. En effet, même un algorithme parfait - c'est-à-dire qui prédit parfaitement la récidive sur les données du passé - est toujours biaisé en ce sens. Selon Sandra Wachter *et al.*, chercheurs en droit et éthique de l'intelligence artificielle, il est alors nécessaire d'introduire une distinction fondamentale entre les algorithmes qui préservent les biais présents dans les données, et ceux qui les "transforment", c'est-à-dire les corrigent. Par exemple, une manière de se débarrasser des biais des données est d'assurer que tous les groupes ont le même taux de positivité (ici, que l'algorithme classifie les Noirs comme "à risque" en moyenne aussi souvent que les Blancs), ce qui revient à privilégier l'égalité entre groupes avant toute autre considération. En effet, d'après la juriste chercheuse en justice algorithmique que nous avons interrogée :

> "Transformer les biais revient à faire de la discrimination positive, même si ceux qui le proposent assurent du contraire. Ils disent qu'il s'agit de modifier les critères sur lesquels les gens seront jugés, mais en pratique cela voudrait dire avoir des critères différents pour chaque groupe".

Richard Berk *et al.* considèrent eux aussi que "faire pencher la balance" peut toujours mener à des contestations en "inégalité de traitement"[32].

Malgré les critiques, les algorithmes sont utilisés dans le système judiciaire américain depuis des dizaines d'années. Angèle Christin, chercheuse en sociologie, mentionne un effet étonnant qui en découle. De nos jours, les algorithmes peuvent s'entraîner sur des cas au sein desquels des outils automatisés ont fait partie du processus de décision initial. Ainsi, les biais s'auto-entretiennent, ce qui fait atteindre un stade de "prophéties auto-réalisatrices" des algorithmes[33].

Pour conclure, la chercheuse en justice algorithmique considère que l'important n'est pas de chercher à tout prix à obtenir des données sans biais, mais plutôt d'être conscients des enjeux du débat :

> "L'algorithme nous met les données sous les yeux. Celles-ci reflètent la société et les critères qui ont été utilisés de tout temps. Elles ouvrent la porte au débat. Elles seront toujours biaisées d'une manière ou d'une autre et, du point de vue d'une avocate, je pense que l'on devrait l'accepter."

En effet, à la décharge des algorithmes, les humains présentent les mêmes biais raciaux que COMPAS : réfléchir à la conception des outils algorithmiques est ainsi une manière de tenter de les surmonter[34].

Nous avons vu que la "*fairness*" est un enjeu central et protéiforme des algorithmes dans le domaine judiciaire, qui requiert une réflexion fine et ne trouve que des solutions partielles et imparfaites. Cependant, bien qu'il s'agisse de l'aspect qui a eu le plus de retentissement médiatique, ce n'est pas le point central de l'affaire Loomis, événement fondateur de la polémique autour du logiciel COMPAS. En effet, Loomis avait principalement

---

[31] Wachter, Sandra, Brent Mittelstadt, et Chris Russell. 2021. « Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law ». West Virginia Law Review 123 (3): 735-85.

[32] Berk, Richard, Hoda Heidari, Shahin Jabbari, Michael Kearns, et Aaron Roth. 2018. « Fairness in Criminal Justice Risk Assessments: The State of the Art ». Sociological Methods & Research 50 (1): 3-44. https://doi.org/10.1177/0049124118782533.

[33] Christin, Angèle. 2020. Les algorithmes prédictifs dans la justice pénale américaine. Institut des Hautes Études sur la Justice. https://www.youtube.com/watch?v=0zu0XsuaiN4.

[34] Berthet, Vincent, et Léo Amsellem. 2021. Les nouveaux oracles: comment les algorithmes prédisent le crime. Paris: CNRS éditions.

contesté devant les tribunaux un autre aspect controversé des outils algorithmiques[35], et particulièrement de COMPAS : leur opacité.

L'utilisation de COMPAS : une justice opaque ?

Un logiciel qui cumule plusieurs niveaux d'opacité

L'évaluation de la "*fairness*" d'un algorithme nécessite de pouvoir en comprendre son fonctionnement : l'étude de ProPublica n'a par exemple pu être effectuée qu'en revenant sur les résultats fournis par l'algorithme COMPAS pour en tirer une compréhension partielle de sa manière de raisonner[36]. Le magistrat, conseiller à la Cour de cassation, que nous avons interrogé résume ce besoin d'intelligibilité en définissant la notion d'opacité dans le cadre de la justice algorithmique :

> "Les biais sont inévitables, mais il faut avoir un moyen de les connaître. L'opacité d'un tel système empêche son utilisation par un juge. Un système aussi opaque techniquement ne répondrait pas à ces critères [d'acceptabilité]."

L'utilisation d'algorithmes d'évaluation des risques de récidive semble donc jeter une zone d'ombre sur le processus judiciaire, ce qui est l'un des principaux arguments avancés contre le recours à de tels logiciels[37]. D'après la littérature, l'utilisation des algorithmes, dont COMPAS, génère trois opacités cumulatives : une première opacité dite de conception, d'accessibilité au code source, qui est alimentée par l'entreprise propriétaire ; une deuxième opacité liée à l'incompréhension des algorithmes par les professionnels du droit qui n'ont pas le bagage technique pour saisir le fonctionnement du code ; et enfin un dernier niveau d'opacité inhérent au processus d'apprentissage automatique, qui n'est pas intelligible par les informaticiens eux-mêmes[38].

Le fonctionnement d'un algorithme d'évaluation des risques nécessite l'emploi de critères qui sont choisis par l'entreprise conceptrice (dans le cas de COMPAS, l'entreprise Equivant) et qui sont interprétés par l'algorithme après avoir subi une quantification, une normalisation et une pondération[39]. Ces trois procédés, ainsi que le choix des critères, sont effectués par l'entreprise, mais ne sont pas accessibles aux professionnels du droit[40]. L'entreprise justifie cela en défendant le secret professionnel, et insiste sur le fait que ce traitement des données est le seul avantage compétitif de leur algorithme par rapport à d'éventuels concurrents[41]. Selon certains acteurs, cette opacité pourrait néanmoins être réduite par une décision étatique : puisque les tribunaux (et dans une moindre mesure les avocats) sont les seuls clients de COMPAS, ils contrôlent le marché et peuvent exiger

---

[35] State of Wisconsin v. Eric L. Loomis. Cour suprême du Wisconsin. 2016. 881 N.W.2d 749. https://caselaw.findlaw.com/wi-supreme-court/1742124.html.

[36] ProPublica. 23 mai 2016. « Machine Bias ». Angwin, Julia, Jeff Larson, Surya Mattu, et Lauren Kirchner. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[37] Meneceur, Yannick, et Clementina Barbaro. 2019. « Intelligence artificielle et mémoire de la justice : le grand malentendu ». Les Cahiers de la Justice 2 (2): 277-89. https://doi.org/10.3917/cdlj.1902.0277.

[38] Burrell, Jenna. 2016. « How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms ». Big Data & Society 3 (1): 1-12. https://doi.org/10.1177/2053951715622512.

[39] Northpointe. 2015. « Practitioner's Guide to COMPAS Core ». Northpointe. https://s3.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf

[40] Meneceur, Yannick, et Clementina Barbaro. 2019. « Intelligence artificielle et mémoire de la justice : le grand malentendu ». Les Cahiers de la Justice 2 (2): 277-89. https://doi.org/10.3917/cdlj.1902.0277.

[41] Skeem, Jennifer L., et Eno Louden. 2007. « Assessment of Evidence on the Quality of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) ». préparé pour le California Department of Corrections and Rehabilitation (CDCR).

l'application de nouvelles normes de transparence[42]. Le marché est donc en situation de monopsone, principalement contrôlé par la demande. Cependant, une transparence totale du code source peut aussi s'avérer une source de problèmes pour de plus petites entreprises, comme le souligne une entrepreneuse en justice algorithmique :

> "Pourquoi n'y a-t-il pas de transparence absolue sur les algorithmes ? Parce qu'on va inévitablement freiner l'innovation, [...] la transparence à tout prix de ces algorithmes va inévitablement privilégier les gros acteurs [...]. Vouloir la transparence à tout prix, c'est une réponse à une question plus complexe qu'il n'y paraît."

Selon cette fondatrice de start-up, l'explicabilité de l'algorithme et l'assurance du bon développement sont préférables à une transparence totale :

> "Je ne défends pas la transparence à tout prix, qui serait par exemple de publier les algorithmes de ces acteurs, je pense que ce qu'il faut faire c'est leur imposer des règles de bon développement, de test et d'usage et, le jour où il y a un scandale, faire un audit où l'on demande une transparence à cercle fermé qui est pour le comité d'audit."

Par comparaison, le deuxième niveau d'opacité est relativement technique, car il est inhérent à la technologie employée par les algorithmes prédictifs. Un problème récurrent lors de l'application des technologies d'apprentissage automatique est l'opacité des processus évaluatifs[43]. À l'usage, les algorithmes sont par définition des fonctions mathématiques associant une valeur de sortie à une valeur d'entrée. Leur complexité provient de l'impossibilité de retracer le raisonnement algorithmique, qui s'exprime formellement par un calcul fortement complexe, pour le justifier ou l'expliquer[44]. Une image fréquemment utilisée par les scientifiques des données pour vulgariser ce phénomène est celle de la "boîte noire" d'où sortent des résultats mais au travers de laquelle aucune visibilité n'est possible, ni pour les utilisateurs ni pour les concepteurs. Les conséquences de cette opacité inhérente à cette technologie sont de très fortes difficultés à remettre en cause un résultat produit par ces algorithmes, une impossibilité à corriger un résultat manifestement erroné ou à en retracer l'origine pour corriger le fonctionnement même de l'algorithme[45].

Enfin, le dernier niveau d'opacité algorithmique est en lien direct avec les utilisateurs, puisqu'il porte sur leur incompréhension des algorithmes et des résultats qu'ils fournissent[46]. Les juges, les avocats, et plus généralement tous les professionnels du droit qui interviennent dans le système judiciaire et carcéral manquent de formation aux technologies employées, leurs capacités et leurs limites[47]. L'une des solutions envisagées consisterait à former les professionnels du droit à ces nouvelles technologies, d'une part dans les universités

---

[42] Abu Elyounes, Doaa. 2020a. « Bail or Jail? Judicial versus Algorithmic Decision-Making in the Pretrial System ». Science and Technology Law Review 21 (2): 376-445. https://doi.org/10.7916/stlr.v21i2.6838.

[43] Burrell, Jenna. 2016. « How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms ». Big Data & Society 3 (1): 1-12. https://doi.org/10.1177/2053951715622512.

[44] Završnik, Aleš. 2021. « Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings ». European Journal of Criminology 18 (5): 623-42. https://doi.org/10.1177/1477370819876762.

[45] Meneceur, Yannick, et Clementina Barbaro. 2019. « Intelligence artificielle et mémoire de la justice : le grand malentendu ». Les Cahiers de la Justice 2 (2): 277-89. https://doi.org/10.3917/cdlj.1902.0277.

[46] Burrell, Jenna. 2016. « How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms ». Big Data & Society 3 (1) : 1-12. https://doi.org/10.1177/2053951715622512.

[47] Deffains, Bruno. 2019. « Le monde du droit face à la transformation numérique ». Pouvoirs 170 (3): 43-58. https://doi.org/10.3917/pouv.170.0043.

pour les étudiants et futurs magistrats ou avocats ; et d'autre part en renforçant la formation professionnelle continue tout au long de la carrière juridique[48].

Au total, les trois niveaux d'opacité se cumulent et ont été l'un des arguments avancés par la défense de Loomis devant la cour d'appel du Wisconsin pour contester le verdict d'incarcération. Une décision du juge uniquement motivée par le résultat d'un algorithme peut être difficile à justifier aux États-Unis, voire impossible en France, comme l'avance le magistrat que nous avons interrogé :

> "En France, ce [résultat] serait irrecevable : aucune discussion entre les parties n'est possible, car il n'y a pas d'accès au code source."

Il semblerait toutefois qu'à cette opacité inhérente au logiciel et à l'intelligibilité du code se superpose un quatrième niveau d'opacité, en lien avec l'utilisation qu'en font les travailleurs du droit.

▪ Vers un quatrième niveau d'opacité : l'opacité "humaine" dans l'utilisation de COMPAS

Les algorithmes tels que COMPAS sont utilisés par différents acteurs de la justice, qu'il s'agisse des avocats, des juges ou encore d'auxiliaires. Par exemple, les travailleurs sociaux sont souvent chargés de remplir les questionnaires en posant les questions aux prévenus, tandis que les juges peuvent ensuite utiliser les résultats pour rendre leur décision. Les avocats peuvent également évoquer les scores produits par le logiciel pour la défense de leur client[49]. Au cours de ses travaux de recherche au sein de trois tribunaux américains, la sociologue a fait le constat d'une mauvaise communication lors de l'utilisation des algorithmes, et cela à deux niveaux. Dans un premier temps, la chercheuse fait part d'un problème de communication entre les administrations et les juges car, selon elle, les juges ne sont pas consultés lors de l'achat des logiciels :

> "Ces outils sont souvent construits par des compagnies privées, qui ensuite les vendent aux juridictions, et ont tendance à plutôt les vendre au service administratif et au service informatique des juridictions. Les juges et les procureurs ne font pas nécessairement partie de ces discussions en général."

Pour Angèle Christin, cette non-consultation lors de l'achat des logiciels engendre une très faible utilisation des algorithmes par les juges, malgré la présence de COMPAS dans de nombreux tribunaux. Par ailleurs, il semblerait qu'il existe également un problème de communication vis-à-vis des avocats et des prévenus. Ces derniers ne seraient en effet pas toujours conscients des tests qu'ils seraient en train de subir car, toujours selon Angèle Christin :

> "Les outils qui sont utilisés pour ces classifications ne leur [les prévenus] sont pas présentés ; à tel point que, souvent, lorsqu'ils répondent à des questionnaires, ils ne savent même pas que leurs réponses vont être utilisées pour des classifications, qui ensuite vont être consultées lors des décisions rendues par les juges et jouer un rôle au sein du système pénal".

Qui plus est, les avocats n'ont que rarement accès aux résultats des tests de leurs clients. Certains seraient même obligés d'acheter les logiciels comme COMPAS et de rentrer eux-mêmes les réponses de leur client pour en obtenir le score, et ainsi pouvoir l'utiliser au cours du procès [50]. Le problème de communication quant à

---

[48] Deffains, Bruno. 2019. « Le monde du droit face à la transformation numérique ». Pouvoirs 170 (3): 43-58. https://doi.org/10.3917/pouv.170.0043.

[49] Christin, Angèle. 2020. Les algorithmes prédictifs dans la justice pénale américaine. Institut des Hautes Études sur la Justice. https://www.youtube.com/watch?v=0zu0XsuaiN4.

[50] Christin, Angèle. 2020. Les algorithmes prédictifs dans la justice pénale américaine. Institut des Hautes Études sur la Justice. https://www.youtube.com/watch?v=0zu0XsuaiN4.

l'utilisation des algorithmes d'évaluation des risques de récidive au sein du système américain repose donc non seulement sur le choix du logiciel en question, mais aussi sur la chaîne de transmission des résultats.

En amont même de l'obtention des résultats, un niveau d'opacité humaine semble persister lors du remplissage des questionnaires. Par exemple, aux États-Unis, les travailleurs sociaux n'ont pas de formation sur la façon de remplir les questionnaires[51]. Au demeurant, ces instruments restent manipulables par les auxiliaires en question, qui peuvent influencer le score de COMPAS dans un sens comme dans l'autre. Le pouvoir décisionnaire bascule donc des mains du juge dans celles des travailleurs sociaux. Angèle Christin ajoute même que, parfois, en sus de cette opacité du processus de remplissage du questionnaire, les prévenus ne sont pas en état de répondre au questionnaire :

> "Les prévenus voient passer un ensemble de travailleurs sociaux dans la cellule, des avocats… ils sont épuisés. En général, ils n'ont pas dormi, ils sont traumatisés, ils ne sont pas douchés, ils n'ont pas mangé, ils n'ont pas vu leurs proches. Ils pensent que leur vie est finie."

L'état des prévenus peut altérer leur capacité à répondre aux questions et donc fausser les résultats d'autant que, selon la juriste interrogée, les questionnaires "sont trop longs et parfois les prisonniers les remplissent eux-mêmes", sans l'aide de personne. Au-delà du déséquilibre de traitement entre deux prisonniers, il existe donc un réel enjeu quant à l'exactitude et à la rigueur du processus de collecte des données.

Au-delà de ce premier niveau "d'opacité humaine", pour reprendre l'expression de la sociologue, niveau qui se joue au moment même du remplissage des questionnaires, un autre enjeu réside dans la conception même de la dangerosité des prévenus par la société. Force est de constater que celle-ci varie d'un pays à un autre : selon le magistrat que nous avons interrogé, les Français ont tendance à privilégier une approche clinique de la dangerosité en faisant intervenir un psychologue pour évaluer celle-ci. Mais, toujours selon le magistrat :

> "Les anglo-saxons ont une approche statistique ou actuarielle de la dangerosité, avec des systèmes de "*scoring*" statistique, comme pour des prêts bancaires. C'est là-dessus qu'est fondé le système COMPAS."

Ces différentes approches de la dangerosité se retrouvent au sein des théories criminologiques qui caractérisent le système judiciaire considéré : dans la notice d'utilisation de COMPAS[52], il est fait état des différentes théories sous-jacentes utilisées par le logiciel, et notamment de la question de l'opportunité criminelle. Celle-ci, basée sur la théorie des marchés, pose une vision de la dangerosité criminelle très factuelle et sociale, mais non fondée sur la psychologie du criminel[53]. Selon ces sociologues, il convient de prendre en considération l'environnement au sein duquel évolue le prévenu pour mieux appréhender les raisons de son acte criminel, et par conséquent sa dangerosité. L'approche systématique de COMPAS, qui se fonde en partie sur une analyse des comportements collectifs, diffère donc de l'approche clinique française, "au doigt mouillé" selon le magistrat que nous avons interrogé, qui laisse quant à elle une plus grande part au vécu particulier de chaque prévenu. Ces deux conceptions de la dangerosité sont le résultat de constructions sociales qui reflètent chacune les mœurs des pays en question et qui ne sont pas nécessairement intelligibles pour les parties prenantes d'un autre système judiciaire[54].

---

[51] Christin, Angèle. 2020. Les algorithmes prédictifs dans la justice pénale américaine. Institut des Hautes Études sur la Justice. https://www.youtube.com/watch?v=0zu0XsuaiN4.

[52] Northpointe. 2015. « Practitioner's Guide to COMPAS Core ». Northpointe. https://s3.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf

[53] Cohen, Lawrence E, et Marcus Felson. 1979. « Social Change and Crime Rate Trends: A Routine Activity Approach » 44 (4): 588-608.

[54] Vigneau, Vincent. 2018. « Le passé ne manque pas d'avenir. Libres propos d'un juge sur la justice prédictive ». Recueil Dalloz 2018: 1095.

Il apparaît donc que l'utilisation d'algorithmes d'évaluation du risque de récidive se heurte non seulement à la complexité du raisonnement humain des individus qui en font l'usage dans le système judiciaire, mais que l'algorithme lui-même se fonde sur des théories criminologiques qui sont le reflet d'une perception de la dangerosité particulière. Toutefois, parmi les acteurs qui sont impliqués dans la chaîne d'utilisation de COMPAS, le juge semble jouir d'un statut particulier car c'est *in fine* lui qui peut choisir d'utiliser ou non le résultat du logiciel pour étayer son raisonnement. L'analyse du comportement du magistrat face au score de COMPAS est donc indispensable pour saisir finement les biais cognitifs humains qui peuvent émerger à ce moment précis du processus judiciaire.

- ## Vers une redéfinition du rôle du juge, entre réflexion personnelle et résultat d'un algorithme

Lors de sa réflexion, le juge s'appuie sur un nombre limité de principes heuristiques qui réduisent les tâches complexes d'évaluation des probabilités et de prédiction des valeurs à des opérations de jugement plus simples[55]. Selon Tversky, ces heuristiques peuvent conduire à de graves erreurs systématiques : par exemple, lorsqu'une pièce est lancée six fois de suite, les observateurs ont tendance à penser qu'il y a plus de chances d'obtenir alternativement "pile" et "face" que six fois "pile" à la suite. Ils attribuent une certaine équité à la pièce, alors que les deux résultats ont la même probabilité d'apparition. De même, un juge s'appuie sur son intime conviction, étayée en partie par ses propres croyances, pour rendre son jugement. Par ailleurs, certains facteurs externes peuvent aussi influencer sa décision. Par exemple, certains verdicts sont influencés par l'heure de la journée à laquelle ils sont rendus : plus le juge a faim, plus il est susceptible de rendre un verdict défavorable pour le justiciable[56]. En réponse à ces biais humains du magistrat, les algorithmes peuvent apparaître à certains comme une solution de choix, car ils rassurent par leur logique mathématique et leur formalisme[57]. Toutefois, selon Barraud, cette arithmétique déterministe ne doit pas faire oublier le fait que les biais cognitifs du juge sont toujours présents lors de l'utilisation des résultats des algorithmes.

Cette opacité se manifeste notamment au moment de l'interprétation des résultats des algorithmes, tel que COMPAS, par le juge[58]. La façon dont sont présentés les résultats de COMPAS peut en effet influencer le juge : COMPAS utilise un code couleur qui met l'accès sur trois risques majeurs : le risque de récidive, le risque de récidive violente et le risque de non-comparution à l'audience préliminaire. Seulement, selon Angèle Christin :

> "Il y a un effet des tableaux de bord sur leur [les juges] perception des données et les couleurs rouges provoquent des réactions négatives beaucoup plus fortes que des couleurs bleues, vertes, plus neutres."

En voyant les barres de score s'afficher en rouge sur le tableau de résultats d'un justiciable, les juges auraient instinctivement une réaction négative (*Figure 5*). Il semblerait donc que les technologies prédictives comme COMPAS ne permettent pas de supprimer ces biais cognitifs mais plutôt de les déplacer vers des "zones plus opaques"[59].

---

[55] Tversky, Amos, et Daniel Kahneman. 1974. « Judgment under Uncertainty: Heuristics and Biases ». Science 185 (4157): 1124-31. https://doi.org/10.1126/science.185.4157.1124.

[56] Danziger, S., J. Levav, et L. Avnaim-Pesso. 2011. « Extraneous Factors in Judicial Decisions ». Proceedings of the National Academy of Sciences 108 (17): 6889-92. https://doi.org/10.1073/pnas.1018033108.

[57] Barraud, Boris. 2017. « Un algorithme capable de prédire les décisions des juges : vers une robotisation de la justice ? » Les Cahiers de la justice 2017 (1), mars 2017.

[58] Christin, Angèle. 2017. « Algorithms in Practice: Comparing Web Journalism and Criminal Justice ». Big Data & Society 4 (2). https://doi.org/10.1177/2053951717718855.

[59] Brayne, Sarah, et Angèle Christin. 2020. « Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts ». Social Problems 68 (3): 608-24. https://doi.org/10.1093/socpro/spaa004.

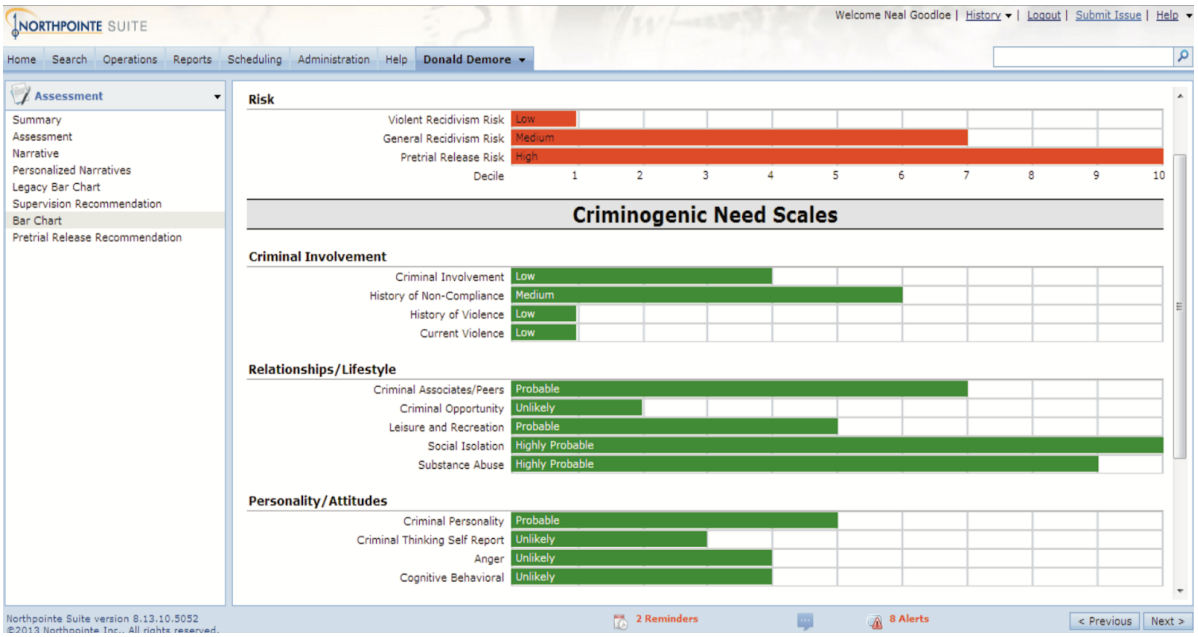


Figure 5 - Exemple de résultat fourni par le logiciel COMPAS

> *À partir des réponses d'un justiciable au questionnaire de COMPAS, le logiciel calcule trois scores principaux notés sur 10 (le risque de récidive, le risque de récidive violente et le risque de non-comparution à l'audience préliminaire). Ces trois critères apparaissent en rouge en haut de la feuille. Des critères supplémentaires (implication dans des actes délictuels, relations personnelles, personnalité) sont également affichés sous les trois scores principaux. Ils sont notés sur 10 et apparaissent en vert.*

En plus de ce premier type de biais, qui s'apparente à des raccourcis de raisonnement, la sociologue interrogée a observé un véritable manque de formation des juges et des travailleurs sociaux sur les logiciels tel que COMPAS. Le magistrat que nous avons interrogé est parvenu aux mêmes conclusions :

> "Il faut que nous ayons conscience des dangers, des limites de ces systèmes. Il faut que les gens apprennent à l'utiliser. À l'avenir, la formation des juges devra intégrer la prise en main de ces outils d'aide à la décision."

Il ne s'agit donc plus seulement de considérer le juge et l'algorithme comme deux entités séparées, mais de permettre la meilleure compréhension de l'outil et de faciliter sa manipulation.

Au total, les différentes notions d'opacité évoquées ne sont pas décorrélées des biais précédemment mentionnés : qu'il s'agisse de transformer ces biais ou d'œuvrer pour plus de transparence, force est de constater que les acteurs ne disposent pas de solution évidente. Comme le soulignent les personnes que nous avons interrogées, la première étape vers une meilleure compréhension de l'outil COMPAS repose sur une formation adéquate au sein des instances qui y ont recours. Puisque certains acteurs - comme le magistrat que nous avons interrogé - voient la multiplication des outils informatiques d'aide à la justice comme inéluctables, l'émergence d'un cadre réglementaire plus rigoureux devrait accompagner ces évolutions. La pertinence d'un tel garde-fou apparaît d'autant plus grande qu'il existe une certaine défiance quant à la robustesse des prédictions de COMPAS.

Justice algorithmique et réglementation

Les algorithmes, des juges parfaits ?

Les algorithmes interviennent dans les décisions de justice rendues dans les tribunaux américains. Cependant, d'après une étude très débattue[60], la justesse des prédictions de COMPAS est égalée à la fois par des algorithmes beaucoup plus simples et par des personnes étrangères à la pratique de la justice pénale. Leur expérience était la suivante : 1000 affaires judiciaires ont été présentées et réparties entre 20 participants. Pour chaque affaire, le participant devait répondre "oui" ou "non" à la question "Pensez-vous que la personne jugée va commettre un autre crime dans les deux années qui viennent ?". Pour cela, les personnes participant à l'étude avaient accès à un ensemble de données, comportant sept caractéristiques au sujet du passé du justiciable. Pour permettre aux participants de l'étude d'améliorer leur prédiction à chaque session, le résultat réel leur était divulgué (*i.e.* si la personne avait réellement récidivé ou non). Cette expérience montre qu'un ensemble de personnes non familières avec la justice pénale obtient des performances de prédiction de récidive comparables à celles de COMPAS. La deuxième étude portait sur une analyse statistique comparant les résultats d'un modèle très simple et ceux de COMPAS[61]. Ce travail a montré qu'un modèle linéaire à seulement 2 critères - l'âge et le nombre total de condamnations antérieures - aboutit à des prédictions aussi justes que celles de COMPAS avec ses 137 critères.

Malgré ces conclusions, des critiques ont remis en cause l'étude de Dressel et Farid[62]. L'expérience qui évaluait la capacité d'estimation du risque par des personnes inexpérimentées a en effet été conduite sur la base d'un nombre limité de critères (sept) pour décrire le justiciable : cela ne correspond pas aux conditions réelles dans lesquelles se trouvent les juges pour rendre leur verdict quant à la détention provisoire. Selon Lin *et al.* :

> "Les outils statistiques prédisent mieux que les humains lorsqu'ils sont nourris d'informations plus complexes".

Par ailleurs, les juges ne disposent pas de retour immédiat quant à la justesse de leur décision, contrairement aux participants de l'étude, qui pouvaient apprendre d'un essai sur l'autre. En effet, un juge ne saura *a priori* jamais s'il a bien fait de mettre un prévenu en détention provisoire.

Les débats relatifs à la plus-value du logiciel COMPAS par rapport à un jugement uniquement humain mettent en question la pertinence de son utilisation systématique. Ainsi, il semble qu'au-delà de la réglementation de la conception de tels outils, préconisée par plusieurs acteurs que nous avons interrogés, il faille également considérer l'évolution d'une réglementation plus en aval, lors de l'utilisation de l'outil par les acteurs du système judiciaire.

La construction d'une réglementation de la conception et de l'utilisation des algorithmes

En 2007, le Département de l'administration pénitentiaire et de la réinsertion de Californie (*California Department of Corrections and Rehabilitation*, CDCR), a commandé un rapport indépendant d'évaluation du logiciel COMPAS

[60] Dressel, Julia, et Hany Farid. 2018. « The accuracy, fairness, and limits of predicting recidivism ». Science Advances 4 (1). https://doi.org/10.1126/sciadv.aao5580.

[61] Dressel, Julia, et Hany Farid. 2018. « The accuracy, fairness, and limits of predicting recidivism ». Science Advances 4 (1). https://doi.org/10.1126/sciadv.aao5580.

[62] Lin, Zhiyuan "Jerry", Jongbin Jung, Sharad Goel, et Jennifer Skeem. 2020. « The limits of human predictions of recidivism ». Science Advances 6 (7). https://doi.org/10.1126/sciadv.aaz0652.

commercialisé en 1999[63]. Celui-ci avait pour but d'étudier le fonctionnement du logiciel et la manière dont il était utilisé au sein de la communauté judiciaire. Cette étude a soulevé de nombreux points de questionnement concernant la capacité de COMPAS à répondre aux attentes des acteurs du monde judiciaire. Les auteurs du rapport se sont par exemple interrogés quant à la capacité de COMPAS à tenir compte de variables dynamiques relatives aux justiciables, c'est-à-dire susceptibles d'évoluer au cours de la vie de l'individu (à l'image de la consommation de drogues). À l'issue de cette étude, les universitaires mandatés par le CDCR ont recommandé de ne pas utiliser COMPAS en tant que tel. L'étude de 2007 a donc représenté l'une des premières tentatives d'encadrer de tels outils d'aide à la prise de décision.

Quinze ans plus tard, aucune réglementation n'a réellement été mise en place en Europe. Seule une charte éthique relative à l'utilisation de l'intelligence artificielle au sein du système judiciaire a été rendue publique[64]. Cette charte, qui ne fait pas office de réglementation, avance que l'utilisation d'algorithmes au sein du système judiciaire ne doit pas entraver certains principes fondamentaux (notamment ceux de non-discrimination, de transparence, d'impartialité et d'équité). En outre, du côté des concepteurs, l'entrepreneuse que nous avons interrogée préconise elle aussi de réguler les "pratiques de développement de tests d'algorithme et les pratiques d'utilisation". Plus précisément, la scientifique numéricienne propose d'appliquer un critère standardisé d'explicabilité aux trois niveaux du développement du logiciel : avant son entraînement sur les données, pendant et après. À ces fins, il semblerait que plusieurs réglementations soient envisagées, mais une seule semble suffisamment réaliste à la scientifique : selon elle, ce processus doit être une "action directe de l'acteur soumis à la loi" mais en aucun cas "un audit de la commission européenne qui irait vérifier chacun de ces algorithmes un à un". Plus précisément, la scientifique numéricienne souligne, comme nous l'avons évoqué plus haut sur les débats relatifs à l'opacité, les potentielles limites d'une transparence absolue des algorithmes :

> "[La transparence absolue] va inévitablement freiner l'innovation, [elle] va inévitablement privilégier les gros acteurs parce qu'ils pourront facilement reprendre les idées de petites entreprises et les implémenter avec des moyens bien supérieurs."

Au-delà de la réglementation de la conception, qui soulève la question de la responsabilité de l'entreprise dans le processus de certification, le logiciel COMPAS souffre d'un manque de réglementation à diverses étapes de son utilisation : du remplissage du questionnaire par le prévenu, jusqu'à l'utilisation des résultats de COMPAS par le juge[65]. D'après la sociologue que nous avons interrogée, une uniformisation du recours aux résultats, et notamment les conditions de leur transmission aux justiciables, "seraient un excellent premier pas d'autant plus que ces algorithmes sont utilisés partout aux États-Unis". Aujourd'hui, la chercheuse explique l'absence de règles communes relatives à l'utilisation de l'outil COMPAS par le fait que les juges sont "maîtres dans leur cour" et qu'ils ne participent pas à des travaux collectifs d'uniformisation car "les tribunaux américains sont très décentralisés". En conclusion, puisque la réglementation de la conception et de l'utilisation des outils d'aide à la prise de décision reste à construire par les acteurs du monde de la justice, il est possible de s'interroger réciproquement sur le rôle de cette réglementation dans la définition du nouvel écosystème de la justice algorithmisée.

▪ Vers un remodelage du panorama d'acteurs ?

Le développement de réglementations fait peser une pression supplémentaire sur les acteurs investis dans la justice algorithmique. Le développement possible de nouveaux marchés par l'ouverture de certains systèmes

---

[63] Skeem, Jennifer L., et Eno Louden. 2007. « Assessment of Evidence on the Quality of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) ». préparé pour le California Department of Corrections and Rehabilitation (CDCR).

[64] European Commission for the Efficiency of Justice. 2018. European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment. https://www.europarl.europa.eu/cmsdata/196205/COUNCIL%20OF%20EUROPE%20-%20European%20Ethical%20Charter%20on%20the%20use%20of%20AI%20in% 20judicial%20systems.pdf.

[65] Skeem, Jennifer L., et Eno Louden. 2007. « Assessment of Evidence on the Quality of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) ». préparé pour le California Department of Corrections and Rehabilitation (CDCR).

judiciaires aux algorithmes est compensé par la multiplication des réglementations qui contraignent les entreprises à assurer et évaluer leurs produits pour les rendre commercialisables[66]. Selon la juriste que nous avons interrogée, de telles études sont vouées à se multiplier afin de certifier les algorithmes proposés aux tribunaux :

> "Si la loi est adoptée, cela changera grandement les choses : il faudra faire des évaluations d'impact. C'est la raison pour laquelle les algorithmes ne sont pas implémentés à la même allure selon les pays, la loi complexifie tout."

En France, de nombreuses réglementations ont déjà été mises en place et l'usage d'algorithmes dans le secteur de la justice est très restreint. Ainsi, aucun algorithme d'évaluation des risques n'est autorisé[67]. Les seules technologies actuellement en usage sont des logiciels d'aide à la recherche de jurisprudence, à l'image de Juri'Predis ou DataJust. En cela, le système français est fortement différent du système américain. Ce dernier utilise déjà de nombreux algorithmes, non seulement pour l'évaluation des risques de récidive dans le système judiciaire, mais aussi par la police pour la prévision des délits. En témoigne le logiciel PredPol, employé par la police de Los Angeles de 2011 à 2018 pour prédire les risques de crimes et de délits et répartir les patrouilles de police en conséquence[68]. Interrogé à ce sujet, le magistrat conseiller à la Cour de cassation que nous avons interrogé nous a averti des difficultés à transposer rigoureusement le modèle américain en France du fait des différences entre les cultures juridiques des deux pays :

> "Je ne pense pas que ce soit possible en France. Je ne comprends pas comment la justice américaine peut utiliser ces systèmes. [...] En France, ce serait inadmissible. [...] Par ailleurs, il y a une autre contrainte en France pour les juges, qui doivent motiver leur décision en fait et en droit. Ils devraient expliquer les règles de ce logiciel, et au nom du principe de contradiction, devraient permettre à la défense d'accéder au code source."

Aux États-Unis, l'emploi d'algorithmes au sein du système judiciaire n'était jusqu'à présent que très peu réglementé, ce qui a mené à une forte hétérogénéité dans l'usage de ces systèmes[69]. Cependant, avec la multiplication des domaines d'application d esalgorithmes d'aide à la décision, les législateurs américains ont pris conscience des enjeux sous-jacents, notamment dans le cadre de la justice algorithmique[70]. L'*Algorithmic Accountability Act* est une loi adoptée en 2019 par le Congrès américain visant à imposer des études d'impact aux grandes entreprises développant des algorithmes considérés comme "à risque" (*sic*), c'est-à-dire prenant des décisions ou traitant des données pouvant impacter des êtres humains. L'objectif est de réduire les biais de ces algorithmes et d'en certifier la correction. L'une des principales demandes relatives à la normalisation des algorithmes consiste à imposer des certifications garantissant la correction et l'explicabilité des logiciels. Une chercheuse en justice algorithmique l'explique de la manière suivante :

> "Il n'y a pas d'exigence [de certification] actuellement aux États-Unis. Il serait possible d'utiliser une certification par une tierce partie, pour prouver la conformité avec la réglementation. Un Conseil de la protection des données veut créer un mécanisme de supervision pour les algorithmes. Il propose des études d'impact effectuées par des entreprises. Cela générerait alors une compétition, un marché se mettrait en place. Je ne pense pas que tout devrait être centralisé par le gouvernement."

---

[66] Jean, Aurélie. 2021. Les algorithmes font-ils la loi ? Editions de l'observatoire.

[67] Deffains, Bruno. 2019. « Le monde du droit face à la transformation numérique ». Pouvoirs 170 (3): 43-58. https://doi.org/10.3917/pouv.170.0043.

[68] The Guardian, 8 novembre 2021. « LAPD ended predictive policing programs amid public outcry. A new effort shares many of their flaws ». Bhuiyan, Johana. https://www.theguardian.com/us-news/2021/nov/07/lapd-predictive-policing-surveillance-reform.

[69] Christin, Angèle. 2020. Les algorithmes prédictifs dans la justice pénale américaine. Institut des Hautes Études sur la Justice. https://www.youtube.com/watch?v=0zu0XsuaiN4.

[70] « Algorithmic Accountability Act ». 116e Congrès. 2019. Législation. https://www.congress.gov/bill/116th-congress/house-bill/2231/text.

Une des conséquences de la mise en place d'une certification extérieure serait donc l'émergence d'un nouveau marché et le développement d'une concurrence entre les certifications. La formalisation de réglementations et certifications semble donc être le chemin emprunté par la justice algorithmique. Cependant la définition exacte et la concrétisation de ces notions restent un point de discussion. L'exemple du besoin de transparence est particulièrement pertinent : selon certains professionnels du droit, il pourrait même s'agir d'une condition *sine qua non* pour l'obtention d'une certification[71].

Enfin, il convient de garder à l'esprit que la controverse autour des réglementations continuera d'évoluer à mesure de l'apparition de nouveaux conflits. Selon la juriste interrogée, "il ne fait aucun doute que des affaires similaires à celle d'Eric Loomis vont émerger à mesure que de tels logiciels seront de plus en plus utilisés". Des affaires aussi médiatisées soulèveront vraisemblablement de nouveaux points de débat en lien avec la justice algorithmique.

Discussion

Il apparaît au terme de notre analyse que les débats autour du logiciel COMPAS soulèvent des questions sur l'équité, l'opacité et la réglementation de la conception et de l'usage des algorithmes d'évaluation des risques de récidive des prévenus. Le logiciel, conçu par la société Equivant, se fonde sur un modèle d'équité considéré comme raciste par certains acteurs, caractéristique renforcée et pérennisée, selon d'autres, par des biais sur les critères d'analyse utilisés par COMPAS et sur les données sur lesquelles son mécanisme d'apprentissage s'appuie. À ces aspects, qui relèvent de la conception du logiciel par Equivant et dépendent des hypothèses de développement retenues, s'ajoutent plusieurs niveaux d'opacité qui réduisent l'intelligibilité de l'algorithme et du processus à l'issue duquel il formule ses décisions. Outre l'évidente difficulté technique qui nuit à la compréhension du code par des non-spécialistes, certains acteurs font remarquer que le processus même d'apprentissage automatique, qui constitue une « boîte noire », et la non-accessibilité au public du code source sont autant de barrières à la transparence de COMPAS. Son utilisation par des professionnels du droit, qui n'ont généralement pas bénéficié de formation spécifique - et en particulier par les juges qui tiennent compte comme ils le souhaitent du résultat de COMPAS dans leur verdict, génère une incertitude supplémentaire. Selon certains, ces questionnements plaident pour la construction d'une réglementation détaillé afin d'encadrer la conception et l'utilisation de tels algorithmes par l'institution judiciaire. L'émergence de cette législation pourrait s'accompagner d'une reconfiguration du panorama d'acteurs qui gravitent autour de la sphère algorithmique dans l'espace judiciaire : en particulier, de nouvelles entreprises de certification pourraient s'immiscer dans le processus de légitimation de tels outils.

Lors de notre étude, nous avons toutefois été surpris par le caractère peu public de la controverse sur le logiciel COMPAS : il s'agit plutôt d'une d'une controverse qui remet l'expertise en question dans des espaces de débats assez restreints. Cette remise en question de l'expertise n'est apparue que façon limitée dans l'arène politico-médiatique : il s'est surtout agi d'une confrontation bilatérale entre le journal d'investigation ProPublica et l'entreprise Equivant, ponctuellement relayée par des quotidiens à plus fort tirage[72]. Des experts majoritairement américains - informaticiens, juristes, sociologues - se sont alors emparés du sujet et ont bénéficié d'une visibilité médiatique croissante à mesure que des études ont enrichi la discussion. Selon les acteurs que nous avons interrogés, il est néanmoins nécessaire que la société française s'empare des débats que suscitent les logiciels comme COMPAS, car ils peuvent remettre en question les fondements mêmes de notre système judiciaire.

---

[71] Deffains, Bruno. 2019. « Le monde du droit face à la transformation numérique ». Pouvoirs 170 (3): 43-58. https://doi.org/10.3917/pouv.170.0043.

[72] Le Grand Continent, 6 octobre 2021. « Les nouveaux oracles, une conversation avec Vincent Berthet et Léo Amsellem ». Storchan, Victor. https://legrandcontinent.eu/fr/2021/10/06/les-nouveaux-oracles-une-conversation-avec-vincent-berthet-et-leo-amsellem/.

Cet outil constitue en effet un cas exemplaire de l'irruption des technologies numériques dans le domaine de la justice et des transformations que cela pourrait induire dans le système judiciaire et dans la manière dont le jugement est formulé et exercé. Aux États-Unis, Angèle Christin[73] souligne que de tels algorithmes ne se développent que pour des délits commis par des prévenus appartenant aux classes les plus défavorisées de la société. Aucun algorithme n'a, à notre connaissance, été conçu pour prédire le risque de récidive en matière de crime en col blanc. Il existe donc aux États-Unis une franche dichotomie entre la justice standardisée par les algorithmes pour les moins aisés et le droit à une justice plus individualisée pour les classes les plus avantagées. Au-delà de la différence de traitement entre les catégories de délits dans le système américain, l'utilisation de l'algorithme systématise une conception de la justice ancrée dans la *Common Law*, une justice tournée vers le passé puisqu'elle ne permet pas le renversement de la jurisprudence : une condamnation passée renforce le risque d'être maintenu en prison dans le présent. Les justiciables courent donc le risque d'être rendus prisonniers de cette jurisprudence. À l'inverse des pays anglo-saxons, dans un pays de tradition juridique romano-germanique comme la France où les textes de loi constituent la principale source de droit, où le précédent n'a pas une place aussi étendue et où la casuistique prédomine, chaque prévenu conserve le bénéfice du doute dans sa situation singulière, si bien que certains estiment qu'il n'est pas évident qu'un logiciel comme COMPAS puisse s'implanter durablement[74]. Il peut donc être intéressant d'observer les commentaires que feront les juges et avocats français au sujet de tels outils dans les années à venir, de même que l'éventuelle utilisation de logiciels comparables d'évaluation des risques en milieu carcéral.

Enfin, si les entretiens menés nous ont permis de mieux cerner les débats et les enjeux autour du logiciel COMPAS, nous tenons à insister sur la brièveté de la période de travail. Il aurait été intéressant de recueillir d'autres points de vue, notamment ceux des acteurs impliqués sur le territoire américain. Notre corpus de témoignages aurait gagné à être étoffé des récits des employés d'Equivant et des membres de ProPublica pour saisir la construction des deux positions autour des modèles de "*fairness*" retenus. Pour autant, il nous a semblé pertinent d'appréhender le cas COMPAS comme un développement particulier du phénomène plus général d'utilisation des algorithmes dans le secteur judiciaire. Au-delà de l'analyse des outils d'évaluation des risques de récidive, le logiciel COMPAS nous a en effet conduit à développer une réflexion sur la manière de penser la justice dans un environnement toujours plus numérisé. En cela, il serait intéressant d'interroger d'autres acteurs impliqués dans des cas comparables, à l'instar des concepteurs du logiciel de police prédictive PredPol utilisé par la police de Los Angeles[75], ou de logiciels de reconnaissance faciale.

[73] Christin, Angèle. 2020. Les algorithmes prédictifs dans la justice pénale américaine. Institut des Hautes Études sur la Justice. https://www.youtube.com/watch?v=0zu0XsuaiN4.

[74] Vigneau, Vincent. 2018. « Le passé ne manque pas d'avenir. Libres propos d'un juge sur la justice prédictive ». Recueil Dalloz 2018: 1095.

[75] Benbouzid, Bilel. 2016. « À qui profite le crime ?. Le marché de la prédiction du crime aux États-Unis ». La Vie des idées, septembre. https://laviedesidees.fr/A-qui-profite-le-crime.html.

Matériel et méthodes

Le présent travail résulte tout d'abord d'une analyse de la presse nationale et internationale, depuis les années 2000, au sujet de la justice algorithmique et du logiciel COMPAS. La base de données Europresse a été consultée au moyen de quatre équations de recherches différentes : " justice & algorithm ", " COMPAS & justice & predictive ", " COMPAS & justice & algorithm " et " COMPAS & software ". Les corpus de documents recueillis pour chaque équation ont fait l'objet de deux types d'analyse distincts : une analyse sémantique avec la plateforme CorTexT et un traitement statistique des données textuelles avec le logiciel IRaMuTeQ (exemple de la *Figure 1*). Les regroupements de termes nous ont permis d'identifier de nouveaux acteurs de la controverse et de nouvelles notions en lien avec l'algorithmisation de la justice. Notre étude bibliographique a été complétée par la lecture et l'analyse d'articles scientifiques issus du domaine de l'informatique, du Droit ou encore de la sociologie (avec notamment un pan de notre corpus issu du champ des *Science and Technology Studies*), mais également par une littérature plus institutionnelle formée de plusieurs rapports publics portant sur l'évolution et les recommandations d'utilisation des algorithmes dans le domaine de la justice. Pour ce faire, d'autres bases de données plus appropriées (Scopus, Web of Science) ont été consultées avec les mêmes équations de recherche. Cette phase liminaire a permis de cerner les principaux points de discussion entre acteurs au sujet du logiciel COMPAS et plus généralement des outils désormais connus sous le nom de " *LegalTechs* ", mais aussi de repérer les enjeux de réglementation et les questionnements scientifiques soulevés par leur essor.

Le corpus de documents étudiés comprend également des communications issus des acteurs directement impliqués dans la production du logiciel COMPAS, par le biais de notices d'utilisation rendues publiques par l'entreprise Equivant, mais aussi et plus ponctuellement des documents plus originaux - à l'instar de décisions de justice ou de fichiers audiovisuels, qui se sont avérés indispensables pour comprendre la chronologie de la controverse et mieux établir les contributions des différents acteurs identifiés.

Ce travail initial a permis l'élaboration de grilles de questions adressées à quatre acteurs de la controverse analysée, au cours d'entretiens semi-directifs. Les témoignages de ces acteurs, qui occupent tous des fonctions différentes, ont été retranscrits puis analysés, et des extraits pertinents ont été exploités dans ce document. Nous avons eu l'occasion d'échanger avec :

- une juriste et chercheuse israélienne, avec un parcours académique fortement ancré aux États-Unis, et dont les travaux portent sur la réglementation en lien avec l'intelligence artificielle ;
- un magistrat français, professeur de Droit et conseiller à la Cour de cassation spécialisé en Droit des nouvelles technologies ;
- une sociologue affiliée à une université américaine, qui s'est intéressée à la conception, l'utilisation et la réception d'algorithmes prédictifs au sein de la justice pénale ;
- une scientifique numéricienne et entrepreneuse française, spécialisée dans les algorithmes et fondatrice de plusieurs start-ups.

Il convient toutefois de souligner que notre enquête a été effectuée dans un laps de temps assez restreint - trois mois - et que le corpus de témoignages aurait gagné à être enrichi. Nous regrettons par exemple l'absence de témoignages d'acteurs américains directement impliqués dans les affaires les plus médiatisées de la controverse, à l'image de l'affaire Loomis ou de l'étude conduite par ProPublica. Nous avons également sollicité des entretiens auprès de professionnels du droit français susceptibles d'avoir recours à des "LegalTechs", et auprès des créateurs d'un algorithme français d'analyse de la jurisprudence, mais nos demandes sont restées sans réponse à ce jour.

Références

Articles de presse généraliste / presse professionnelle

Le Grand Continent, 6 octobre 2021. « Les nouveaux oracles, une conversation avec Vincent Berthet et Léo Amsellem ». Storchan, Victor. https://legrandcontinent.eu/fr/2021/10/06/les-nouveaux-oracles-une-conversation-avec-vincent-berthet-et-leo-amsellem/.

ProPublica. 23 mai 2016. « Machine Bias ». Angwin, Julia, Jeff Larson, Surya Mattu, et Lauren Kirchner. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

The Guardian, 8 novembre 2021. « LAPD ended predictive policing programs amid public outcry. A new effort shares many of their flaws ». Bhuiyan, Johana. https://www.theguardian.com/us-news/2021/nov/07/lapd-predictive-policing-surveillance-reform.

The New York Times, 26 octobre 2017. « When an Algorithm Helps Send You to Prison ». Thadaney Israni, Ellora. https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html.

Article de revue scientifique

Abu Elyounes, Doaa. 2020a. « Bail or Jail? Judicial versus Algorithmic Decision-Making in the Pretrial System ». Science and Technology Law Review 21 (2): 376-445. https://doi.org/10.7916/stlr.v21i2.6838.

———. 2020b. « Contextual Fairness: A Legal and Policy Analysis of Algorithmic Fairness ». Journal of Law, Technology and Policy 2020 (1): 1-54. https://doi.org/10.2139/ssrn.3478296.

Barraud, Boris. 2017. « Un algorithme capable de prédire les décisions des juges : vers une robotisation de la justice ? » Les Cahiers de la justice 2017 (1), mars 2017.

Benbouzid, Bilel. 2016. « À qui profite le crime ?. Le marché de la prédiction du crime aux États-Unis ». La Vie des idées, septembre. https://laviedesidees.fr/A-qui-profite-le-crime.html.

Berk, Richard, Hoda Heidari, Shahin Jabbari, Michael Kearns, et Aaron Roth. 2018. « Fairness in Criminal Justice Risk Assessments: The State of the Art ». Sociological Methods & Research 50 (1): 3-44. https://doi.org/10.1177/0049124118782533.

Brayne, Sarah, et Angèle Christin. 2020. « Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts ». Social Problems 68 (3): 608-24. https://doi.org/10.1093/socpro/spaa004.

Brennan, Tim, William Dieterich, et Beate Ehret. 2009. « Evaluating the Predictive Validity of the Compas Risk and Needs Assessment System ». Criminal Justice and Behavior 36 (1): 21-40. https://doi.org/10.1177/0093854808326545.

Burrell, Jenna. 2016. « How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms ». Big Data & Society 3 (1): 1-12. https://doi.org/10.1177/2053951715622512.

Christin, Angèle. 2017. « Algorithms in Practice: Comparing Web Journalism and Criminal Justice ». Big Data & Society 4 (2): en ligne. https://doi.org/10.1177/2053951717718855.

Cohen, Lawrence E, et Marcus Felson. 1979. « Social Change and Crime Rate Trends: A Routine Activity Approach » 44 (4): 588-608.

Danziger, S., J. Levav, et L. Avnaim-Pesso. 2011. « Extraneous Factors in Judicial Decisions ». Proceedings of the National Academy of Sciences 108 (17): 6889-92. https://doi.org/10.1073/pnas.1018033108.

Deffains, Bruno. 2019. « Le monde du droit face à la transformation numérique ». Pouvoirs 170 (3): 43-58. https://doi.org/10.3917/pouv.170.0043.

Dika, Khaled. 2020. « L'affaire Loomis: Les fantômes de Descartes et de Grotius à l'assaut de la justice? » HAL (preprint). https://hal.archives-ouvertes.fr/hal-02566382.

Dressel, Julia, et Hany Farid. 2018. « The accuracy, fairness, and limits of predicting recidivism ». Science Advances 4 (1). https://doi.org/10.1126/sciadv.aao5580.

Lin, Zhiyuan "Jerry", Jongbin Jung, Sharad Goel, et Jennifer Skeem. 2020. « The limits of human predictions of recidivism ». Science Advances 6 (7). https://doi.org/10.1126/sciadv.aaz0652.

Meneceur, Yannick, et Clementina Barbaro. 2019. « Intelligence artificielle et mémoire de la justice : le grand malentendu ». Les Cahiers de la Justice 2 (2): 277-89. https://doi.org/10.3917/cdlj.1902.0277.

Tversky, Amos, et Daniel Kahneman. 1974. « Judgment under Uncertainty: Heuristics and Biases ». Science 185 (4157): 1124-31. https://doi.org/10.1126/science.185.4157.1124.

Vigneau, Vincent. 2018. « Le passé ne manque pas d'avenir. Libres propos d'un juge sur la justice prédictive ». Recueil Dalloz 2018: 1095.

Wachter, Sandra, Brent Mittelstadt, et Chris Russell. 2021. « Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law ». West Virginia Law Review 123 (3): 735-85.

Završnik, Aleš. 2021. « Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings ». European Journal of Criminology 18 (5): 623-42. https://doi.org/10.1177/1477370819876762.

Ouvrages

Berthet, Vincent, et Léo Amsellem. 2021. Les nouveaux oracles: comment les algorithmes prédisent le crime. Paris: CNRS éditions.

Jean, Aurélie. 2021. Les algorithmes font-ils la loi ? Editions de l'observatoire.

Saleilles, Raymond. 1898. L'individualisation de la peine: étude de criminalité sociale. Bibliothèque générale des sciences sociales. Paris: Baillière. https://books.google.fr/books?id=4GWc37FGiKQC.

Chapitres d'ouvrage

Brennan, Tim, et William Dieterich. 2018. « Correctional Offender Management Profiles for Alternative Sanctions (COMPAS) ». In Handbook of Recidivism Risk/Needs Assessment Tools, Singh J., Kroner D., Wormith, J., Desmarais S., Hamilton Z., 49-75. Hoboken (USA): John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119184256.ch3.

Goel, Sharad, Ravi Shroff, Jennifer L. Skeem, et Christopher Slobogin. 2021. « The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment ». In Research Handbook on Big Data Law, Roland Vogl, 9-28. Law 2021. Rochester, NY: Social Science Research Network. https://doi.org/10.4337/9781788972826.00007.

Littérature grise

« Algorithmic Accountability Act ». 116e Congrès. 2019. Législation. https://www.congress.gov/bill/116th-congress/house-bill/2231/text.

State of Wisconsin v. Eric L. Loomis. Cour suprême du Wisconsin. 2016. 881 N.W.2d 749. https://caselaw.findlaw.com/wi-supreme-court/1742124.html.

European Commission for the Efficiency of Justice. 2018. European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment.
https://www.europarl.europa.eu/cmsdata/196205/COUNCIL%20OF% 20EUROPE%20-%20European%20Ethical%20Charter%20on%20the%20use%20of%20AI%20in% 20judicial%20systems.pdf.

Jean, Aurélie, Victor Storchan, et Adrien Basdevant. 2021. « Mécanisme d'une justice algorithmisée ». Fondation Jean Jaurès.

Northpointe. 2015. « Practitioner's Guide to COMPAS Core ». Northpointe. https://s3.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf.

Skeem, Jennifer L., et Eno Louden. 2007. « Assessment of Evidence on the Quality of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) ». préparé pour le California Department of Corrections and Rehabilitation (CDCR).

- ## Films (documentaire, fiction, …)

Christin, Angèle. 2020. Les algorithmes prédictifs dans la justice pénale américaine. Institut des Hautes Études sur la Justice. https://www.youtube.com/watch?v=0zu0XsuaiN4.

Raucher, Sammy. 2021. « Algorithms and Pre-Trial Assessment » Mini-Lecture component : « What Makes an Algorithm Fair? "Fairness" in the COMPAS Recidivism Risk Algorithm ». Human Contexts and Ethics. https://www.youtube.com/watch?v=HfxhmMdA8XQ.

- ## Images, photographies, tableaux et graphiques

Reed, E.T. 1890. « Automatic Arbitration ». Punch, 17 mai 1890. https://www.punch.co.uk/image/l0000AsGcmxxholk.