

ÉTUDES ET COMMENTAIRES



À l'ère du numérique, ce que le criminel pourrait apprendre au civil en l'état ¹

par Jean-Paul Jean, Président de chambre honoraire à la Cour de cassation,
Secrétaire général de l'AHJUCAF (Association des hautes juridictions de cassation
ayant en partage l'usage du français)

L'essentiel > La profusion d'articles techniques et juridiques relatifs à la justice dite prédictive souligne les zones d'incertitude des bouleversements en cours à l'ère du numérique. Les bases de données judiciaires constituent un « bien commun », dont la valeur augmente sous l'effet conjugué de l'accroissement de la demande des usagers et des offres de produits ciblés dans un marché très concurrentiel. La communauté des juristes doit dégager rapidement des principes directeurs fondés sur la qualité et la hiérarchisation de la jurisprudence par ceux qui la produisent, en cohérence avec les finalités protégées par la loi des modes de traitement par algorithmes sur des masses de décisions de plus en plus considérables. Quelques leçons tirées de l'histoire de l'usage statistique des données pénales par la puissance publique au titre de la justice actuarielle peuvent aider à mieux évaluer certains risques, et à déterminer, dans le champ du droit civil, le cadre d'un usage raisonné de ces outils d'analyse à l'ère du numérique.

2

Pourquoi ce titre évidemment inspiré de l'adage « le criminel tient le civil en l'état » ? Pour essayer de réfléchir en toute liberté, à partir de connaissances acquises depuis plusieurs années dans le domaine pénal, face à l'étonnement provoqué par la frénésie actuelle de publications autour des bouleversements annoncés de la justice civile par le traitement des bases de données juridiques par les algorithmes, et cette « disruption numérique ».

A-t-on déjà assisté, dans les revues juridiques, à une telle production doctrinale concentrée sur un même sujet que celui de la justice « dite prédictive », comme si chacun devait se positionner dans une fébrilité parfois consubstantielle à celle des réseaux sociaux.

Il y a moins de deux ans, ce qui paraît très vieux à l'ère du numérique, nous avons conçu au SDER (service des études,

de la documentation et du rapport), le colloque de la Cour de cassation intitulé « La jurisprudence dans le mouvement de l'*open data* ² ». Dans cette même Grand-chambre, il avait aussi été question essentiellement de justice civile. Il s'agissait alors de mettre à plat la véritable révolution initiée par le recours au traitement des données jurisprudentielles par algorithmes. Les actes de ce colloque publiés peu après dans une édition spéciale du JCP téléchargeable gratuitement depuis le site de la Cour de cassation ont posé clairement nombre de termes du débat ³. Quelques mois plus tard, un colloque de l'INHESJ (Institut national des hautes études sur la sécurité et la justice) était centré sur les mêmes questions concernant la justice pénale ⁴.

Nous sommes maintenant déjà à un autre moment. Il semble que le débat technique a été largement documenté par les « *juristes geeks* ». Pour observer et essayer de comprendre cette

(1) Ce texte reproduit l'intervention de l'auteur lors du colloque Le juge et le numérique : un défi pour la justice du XXI^e siècle, qui s'est tenu le 8 juin 2018, à la Cour de cassation. (2) Colloque du 14 oct. 2016. Programme et interventions accessibles en vidéo sur le site de la Cour de cassation. (3) JCP suppl. n° 9, du 27 févr. 2017. (4) INHESJ 27 juin 2017, Sécurité et justice, les défis de l'algorithme, accessible en vidéo, <https://inhesj.fr/evenements/videotheque/securite-et-justice>.

communauté en avance d'un savoir spécifique, il faut parfois, tel un anthropologue, s'inscrire dans une approche d'observation bienveillante.

Quant aux démarches d'anticipation (et non pas de prédiction...) des évolutions inéluctables, elles paraissent parfois bien fragiles, au regard du nombre de variables à intégrer portant non seulement sur les évolutions technologiques, mais aussi sur les stratégies des acteurs du marché du traitement des bases de données. Les bases de données juridiques publiques constituent, en effet, un trésor - un « bien » - commun, dont la valeur augmente rapidement ; ce, sous l'effet conjugué de l'accroissement de la demande, suscitée par les nouvelles offres de produits ciblés proposés par les *start-ups* de la *Legal-Tech* et les éditeurs juridiques, alors que se pose par ailleurs le problème de la disponibilité de certaines matières premières (les différentes sources de jurisprudence) difficilement accessibles ou bien réservées à quelques utilisateurs institutionnels et à ceux qui ont les moyens de les acheter.

Fort heureusement d'éminents juristes, tels le professeur Loïc Cadiet dans son rapport sur l'*open data* des décisions de justice⁵, posent en langage clair les balises nécessaires, sous l'éclairage de principes juridiques et éthiques qui me semblent essentiels et qu'il revient à la communauté des juristes de dégager rapidement comme principes directeurs.

Parmi ces principes à rappeler sans cesse, notamment à ceux qui ne viennent pas du monde du droit, il convient de souligner qu'au regard du droit à la protection des données personnelles un traitement de données doit toujours être attaché à une finalité précise, principe qui régit depuis l'origine l'action de la Commission nationale de l'informatique et des libertés (CNIL). Le règlement européen sur la protection des données personnelles (RGPD) prévoit en ce sens dans son article 5 un principe clair de « minimisation des données » : « les données à caractère personnel doivent être (...) adéquates, pertinentes et limitées à ce qui est nécessaire pour les finalités pour lesquelles elles sont traitées »⁶.

La finalité des bases de données juridiques est de permettre une meilleure connaissance et une meilleure transparence des règles de droit et de leur application à des circonstances de fait. Les demandes légitimes des citoyens de connaissance de leurs droits, de prévisibilité et de sécurité juridiques, nécessitent une large diffusion des jurisprudences pour d'utiles rapprochements. Mais d'autres usages se développent. Ainsi, dans notre système de *Civil Law*, la collecte massive d'informations sur la personne des juges, des greffiers ou des avocats, permettant de générer des statistiques sur l'ensemble de leur activité professionnelle et de leurs « performances », voire de leur système de valeurs (nombre d'affaires par an, moyenne

de temps de traitement, résultats, motivation, etc.), n'intègre pas cette finalité et peut constituer un détournement de finalité. Le débat se pose différemment aux États-Unis, avec une conception plus large du droit à l'information, l'élection des juges qui, en outre, statuent en juge unique, ce qui a pour conséquence une totale transparence sur la personnalisation des décisions.

Compte tenu de tous les éléments déjà évoqués par ailleurs au cours du colloque, ce propos sera concentré autour de deux problématiques :

- Une rapide remise en perspective historique en droit criminel ;
- Une approche en termes d'analyse économique de la justice en tant que système, pour traiter de cette justice dite prédictive au regard des stratégies judiciaires des principaux acteurs.

I - Ce que le criminel peut apprendre au civil en l'état

Le terme « prédiction », concept central, ne paraît pas adapté quand il est utilisé par les civilistes, puisqu'il correspond à un autre sens, plus ancien, au pénal.

En effet, les algorithmes, quelle que soit leur puissance actuelle, demeurent nécessairement réduits aux interactions opérées dans la phase d'apprentissage, comme dans les modèles de « *machine-learning* », et traitent donc des informations passées, avec effet cumulatif. La prévision statistique repose sur des échantillons représentatifs. Elle repose ici sur une masse de plus en plus considérable d'informations passées soutenant des calculs de probabilité avec des marges d'erreur sans doute de plus en plus faibles si tous les biais sont exclus et si tous les éléments de contexte sont intégrés.

Le terme de justice « prédictive » paraît erroné, car il s'agit d'une simple traduction du terme américain « *prediction* ». Le terme de justice « actuarielle » paraît plus pertinent, s'inscrivant dans la ligne des travaux du criminologue américain Bernard Harcourt - « *against prediction* » -, qui a beaucoup enrichi les débats franco-français il y a une dizaine d'années sur les méthodes d'évaluation en matière de justice pénale⁷.

A - Rappel historique

Dans nombre d'articles traitant de la justice dite « prédictive », sont cités les travaux d'Alan Turing datant des années cinquante⁸. Mais la justice actuarielle pénale a une histoire encore plus longue, appliquée plus spécialement au calcul du risque de récidive et aux politiques de *sentencing*.

(5) L. Cadiet, *L'open data des décisions de justice*, rapp. à M^{me} la garde des Sceaux, nov. 2017. (6) Règl. UE 2016/679 du Parlement européen et du Conseil du 27 avr. 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la Dir. 95/46/CE (règl. général sur la protection des données). (7) B. Harcourt, *Against prediction ; Profiling, policing and punishing in an actuarial age*, University of Chicago Press, 2006. (8) A. Turing, *Computing machinery and Intelligence*, 1950.

Ce débat a été posé en France à propos des peines planchers⁹. Par-delà la sanction aggravée encourue par le récidiviste, c'est le risque de récidive que l'on a tenté d'évaluer, en intégrant une approche actuarielle fondée sur les probabilités statistiques que connaissaient depuis longtemps les États-Unis¹⁰.

Il peut être simplement rappelé ici que c'est un professeur de sociologie de l'université de Chicago, Ernest Burgess, de sensibilité démocrate, qui, en 1928, avec les meilleures intentions du monde, a voulu objectiver des catégories d'individus à risques supérieurs de récidive, à partir d'une comparaison portant sur des groupes de détenus libérés de trois prisons différentes. Il voulait éliminer les discriminations de fait portant essentiellement sur la couleur de peau. S'appuyant sur ces travaux, la pénologie actuarielle, à partir d'indicateurs dotés de coefficients, il définit des catégories et donc des profils socio-psychologiques d'individus plus ou moins à risque¹¹, qui vont donc fonder une décision accordant ou non une libération conditionnelle (*parole*).

La criminologie actuarielle se fonde sur des probabilités statistiques, sur des actuaires, comme pour les assurances qui évaluent les risques à partir de données objectivées et qui, en conséquence, exigent des surprimes pour un jeune conducteur ou une personne en surpoids. Cette méthode appliquée « scientifiquement » au risque de récidive pour un condamné, donc à une anticipation sur son comportement futur, s'oppose au principe d'individualisation fondant la décision d'un juge, même si, dans les deux cas, le passé pénal constitue un élément d'appréciation déterminant. C'est dans le même mouvement, et avec là encore les meilleures intentions fondées sur le principe d'égalité (qui s'est retourné contre les jeunes noirs avec notamment l'incarcération massive des usagers-revendeurs de crack-cocaïne), qu'a été instauré dans les modèles anglo-américains le système de *sentencing* pour la détermination des peines.

Ces tables de *sentencing* sont des guides pour le juge, très indicatifs mais avec des conséquences plus ou moins automatiques dès la qualification retenue. Cela doit nous aider à réfléchir sur tous les dispositifs, nécessaires sans doute, d'aide à la décision du juge, à condition de bien distinguer les types de décision auxquelles l'apport de ces méthodes scientifiques peut être utile.

Il existe là une différence essentielle avec la matière civile. En effet, dans le domaine pénal, si les outils d'analyse s'appuient

aussi sur les algorithmes, l'accès aux données et aux modes de traitement relève strictement de la puissance publique et des agents étatiques. L'accès de la personne mise en cause, de son avocat, ne peut concerner que le dossier individuel. Mais, dans l'avenir, l'accès devra aussi porter, au nom des droits de la défense, sur les choix méthodologiques de l'analyse dite « prédictive » qui est utilisée. Ce débat a été soulevé aux USA dans une affaire *Bloomis*, portant sur les discriminations relatives au genre et à la race.

L'affaire « *Loomis vs Wisconsin* »

En 2013, à l'issue d'un accord avec l'autorité de poursuite, Eric L. Loomis a accepté de plaider coupable pour deux chefs de prévention (tentative de fuite lors d'un contrôle routier et conduite d'un véhicule sans l'accord du propriétaire). Eric L. Loomis a été condamné à une peine de six ans d'emprisonnement pour avoir conduit un véhicule volé utilisé lors d'une fusillade. La juridiction dans l'État du Wisconsin ayant prononcé cette peine a notamment relevé qu'il avait été identifié, sur la base de l'évaluation COMPAS (Correctionnal Offender Management Profiling for Alternative Sanctions), comme un individu présentant un haut risque de récidive. La question de la violation du *right to due process* a été transmise à la Cour suprême du Wisconsin qui a rejeté les prétentions d'Eric L. Loomis, en posant des limites à l'utilisation du logiciel COMPAS dans la fixation d'une peine, mais en considérant qu'en l'espèce les droits de la défense n'avaient pas été violés. En octobre 2016, Eric L. Loomis a saisi la Cour suprême des États-Unis de la question suivante : la prise en compte par un tribunal du logiciel COMPAS pour la fixation d'une peine constitue-t-elle une violation du *right to due process* dès lors, d'une part, que la nature de COMPAS empêche le défendeur de remettre en question l'exactitude et la validité scientifique de l'évaluation du risque et, d'autre part, que les évaluations de COMPAS prennent en considération la race et le genre ? L'État du Wisconsin a déposé un mémoire concluant au rejet de la demande, de même que l'État fédéral, intervenant en tant qu'*amicus curiae*. La Cour suprême des États-Unis décidera souverainement d'examiner ou non l'affaire¹².

L'application d'une analyse statistique de probabilité à une situation individuelle de culpabilité ou d'évaluation du risque de récidive pose un problème de fond. Dans tout processus d'aide à la décision, au pénal comme au civil, le même débat se pose quant à l'exactitude et la validité scientifique de l'analyse effectuée par le logiciel d'intelligence artificielle, tant au niveau des données traitées qu'à celui des modalités du traite-

(9) J.-P. Jean, Les politiques criminelles face à la récidive, in *Les récidivistes, Représentations et traitements de la récidive XIX^e-XXI^e siècle* (ss la dir. J.-P. Allinne et M. Soula, préf. P. Truche), PUR, déc. 2010, p. 263-276 ; *Ibid.* Récidive : évolutions législatives et politique pénale, évaluation. Expertise pour la conférence de consensus, Ministère de la Justice, févr. 2013. (10) Pour une approche critique, B. Harcourt, *Against Prediction, Profiling, policing and punishing in an actuarial age*, préc. ; Slingeneyer, La nouvelle pénologie, une grille d'analyse des transformations des discours, des techniques et des objectifs dans la pénalité, *Revue Champ pénal/penal field*, <http://champpenal.revues.org/2853> ; L. Delannoy-Brabant, Quelles évolutions des politiques de traitement du crime à l'ère de la « nouvelle pénologie ? », Centre d'analyse stratégique, note de veille n° 106, juill. 2008. (11) Ces profils évoluent selon les époques, les cultures et l'état des connaissances. Ainsi, à Chicago, dans les années 1930, Burgess distinguait huit catégories fleurant bon Jack Kérouac, dont il comparait le taux de récidive au taux moyen de celui des sortants de prison : les *hobos* (chemineaux), *never-do-well* (bons à rien), les citoyens moyens, les alcooliques, les gangsters, les immigrants récents, les garçons de ferme et les toxicomanes. (12) Mémoires consultables sur le site de *l'Harvard Law Review*, mis en ligne le 10 mai 2017 <https://harvardlawreview.org/2017/03/state-v-loomis/>.

ment et de tous les biais possibles. Mais cela n'exclut en rien l'utilisation de ces outils en appui de décisions de politique publique, d'options stratégiques publiques ou privées relatives à la tenue d'un procès, les modalités de celui-ci ou la fixation de normes de références pour certains contentieux de masse ou (et) présentant des caractéristiques similaires.

Un traitement par une institution publique avec des normes rigoureuses applicables et contrôlables constitue en soi un niveau de garanties scientifiques. En matière civile, les risques paraissent plus élevés, car des recherches et analyses sont effectuées par des entreprises privées pour des usages et des clients privés. Des biais méthodologiques peuvent être acceptés pour aboutir au résultat souhaité. L'on sait que malheureusement cela peut arriver même dans la recherche scientifique pour que des travaux présentent un intérêt afin d'être publiés.

B - Du bon usage de la justice actuarielle

Sous réserve de garanties scientifiques et déontologiques exigeantes, il existe un bon usage de la justice actuarielle, au pénal comme au civil. En matière d'indemnisation du préjudice, par exemple, les actuaires des compagnies d'assurance s'appuient depuis toujours sur cette méthode qui seule permet d'effectuer un calcul de risque par prévision en s'appuyant sur les données passées, de provisionner les sommes nécessaires à ce risque et de fixer les cotisations au regard des facteurs de risque de leurs souscripteurs. La méthode scientifique a fait sa preuve empirique. Il appartient alors au législateur de fixer les limites pour faire jouer la solidarité, éviter les discriminations, limiter les surprimes.

Là encore, pour les données de santé tout particulièrement, qui sont tout aussi sensibles que les données judiciaires, sinon plus, l'enjeu pour les assureurs en concurrence est celui de l'accès aux données personnelles par-delà les données épidémiologiques générales. Si des limites juridiques ne sont pas fixées, le seul intérêt économique prévaudra, qui est de discriminer les individus en fonction des facteurs de risque qu'ils présentent. Le parallèle est évident avec la justice pénale ou civile, et la nécessité de distinguer, d'une part, les données générales anonymisées en *open source* et, d'autre part, les données personnelles dont il faut renforcer la protection.

Un rapide rappel, en matière historique, peut aussi illustrer le changement d'échelle en train de s'opérer, qui fait passer la justice pénale dite prédictive du stade artisanal au stade industriel.

C - Le jury populaire et le calcul de probabilités

L'intégration de calculs de probabilité dans la logique du jugement pénal intéresse au premier chef le moment essentiel du

procès, celui du verdict sur la culpabilité, encore plus dans le cadre d'un jury populaire. Par-delà les probabilités statistiques, chaque décision va concerner une situation particulière, un individu condamné ou acquitté. Si le principe fondamental est celui de la présomption d'innocence, toute décision, lorsque les faits sont niés, relève d'une conviction, d'un arbitrage entre le risque de condamner un innocent et celui de laisser un coupable en liberté, en privilégiant l'une au l'autre hypothèse grâce à un calcul de probabilités. En régime de *Common Law*, il s'agit du *reasonable doubt*.

La question n'intéresse pas que les juristes et les criminologues, et le calcul de probabilité est utilisé depuis longtemps comme support de lois procédurales. Elle a depuis plus de deux siècles fait l'objet de débats passionnés entre scientifiques, philosophes et politiques, ainsi que le rappelait l'épidémiologiste Daniel Schwartz dans la revue *Pénombre* en 1995¹³. Condorcet le premier a proposé en 1785 une formule permettant, sous certaines hypothèses, de calculer la probabilité de condamner un innocent¹⁴. Laplace, en 1830, précisant cette formule, l'a appliquée aux jugements en cour d'assises, où les jurys comprenaient douze citoyens, la condamnation étant prononcée à la majorité (sept voix contre cinq). Laplace estimait que le risque d'erreur était alors de 1/4, valeur manifestement intolérable ! À l'arrivée des libéraux aux affaires, en 1830, la proportion a été portée à huit voix contre quatre ; mais, même avec cette modification, le risque d'erreur restait encore très élevé, et Arago exposait à la Chambre des Députés que, « sur huit hommes qui montent à l'échafaud, il y en a un d'innocent » ! Et proposait de rabaisser le risque à 1/16... Chacun pourrait calculer le risque d'erreur judiciaire « acceptable » au regard d'une réforme du jury... Les multiples articles consacrés à la « statistique des jugements » correspondent aujourd'hui aux débats sur la justice prédictive, pour partie, sur les mêmes bases, celles d'une probabilité statistique : « il est de notre nature, soutenait Condorcet, de ne pouvoir juger que sur des probabilités, il n'est donc pas injuste de condamner un innocent, pourvu que l'on soit assuré qu'il y a une grande probabilité que la décision rendue soit exacte ».

Et aujourd'hui c'est toujours en cours d'assises qu'est utilisée cette justice « prédictive » d'artisan, grâce aux moteurs de recherche sur internet.

En effet, nombre d'avocats de la défense prennent déjà tout renseignement sur les jurés par les informations directement accessibles sur Internet, en particulier celles accessibles via les réseaux sociaux, soit pour les récuser, soit pour choisir les arguments auxquels certains pourraient être sensibles. Peut-être certains magistrats du ministère public font-ils de même... On peut mesurer l'évolution des informations disponibles grâce à ces nouvelles sources en rappelant que, dans l'affaire *Buffet/Bontemps* condamnés à mort, Robert Badinter

(13) D. Schwartz, Le test statistique et le jugement de culpabilité, Revue *Pénombre* n° 8, déc. 1995. (14) Marquis de Condorcet, Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix, Gallica, site BNF.

avait récusé sur intuition un juré qui lui a dit ensuite être le président local de la Ligue des droits de l'homme, opposé par principe à la peine de mort¹⁵.

Peut-être un traitement par algorithmes d'arrêts de cour d'assises, avec des critères de distinction fondés sur le sexe, l'âge, le métier, le domicile des jurés (les quatre éléments connus, outre les informations personnelles trouvées sur les réseaux sociaux), et le *quantum* des condamnations ou le pourcentage d'acquittements pourrait-il permettre une approche probabiliste ? Mais les praticiens savent les limites de l'exercice en termes de représentations sociales (ainsi la distinction homme/femme n'est pas pertinente concernant la sévérité dans les affaires de viols). Et il ne faut jamais oublier dans toute analyse statistique la distinction entre une simple corrélation et un lien effectif de causalité.

En matière correctionnelle, la justice probabiliste d'artisan existe aussi. C'est celle de l'avocat expérimenté qui connaît bien « sa » juridiction, les juges, les jurisprudences locales selon les présidents, et qui va adapter sa plaidoirie, accepter une procédure de composition pénale ou plaider-coupable plutôt que de faire prendre à son client le risque d'une peine supérieure à l'audience publique.

Au Canada, il existe une devise pour apprécier la compétence d'un avocat : « *A good lawyer knows the law, an excellent lawyer knows the judge* ». Cela ne veut pas dire que l'avocat connaît personnellement le juge, mais qu'il connaît sa jurisprudence, sa façon d'appréhender un contentieux et qu'il va donc chercher à le convaincre par des arguments pertinents auxquels il sait que ce juge est sensible.

À terme, avec l'appui des algorithmes, en matière civile comme en matière pénale, le changement d'échelle dans cette connaissance des juges et de la jurisprudence va être considérable. Cette évolution doit être évoquée dans ses effets négatifs comme dans ses effets positifs, ainsi que ses conséquences en termes de stratégies judiciaires de différents acteurs, notamment dans le domaine des contentieux de masse où l'on est déjà passé de cette justice d'artisan à une justice dite prédictive traitée de façon « industrielle ».

II - Justice actuarielle et stratégies judiciaires

L'économie du système pénal s'est totalement recomposée en France depuis vingt ans¹⁶ et ce phénomène va s'accélérer avec l'immense saut technologique en cours. Dans ce cadre, se développe en conséquence une « justice au tarif », quasi automatique pour les contentieux de masse (radars, flash aux franchissements de feux rouges, repérage des automobilistes dangereux...), avec un juge en position de recours en cas de

contestation ou de nécessaire individualisation. C'est aussi grâce au temps et aux moyens ainsi dégagés par l'apport de la technologie que pourra par ailleurs s'exercer, pour les affaires complexes ou discutées, une justice pénale de qualité avec le temps suffisant consacré aux affaires qui le nécessitent.

Cette tendance s'accélère en Europe, ainsi que le met en évidence le dernier rapport de la CEPEJ¹⁷. On assiste désormais à une utilisation généralisée des amendes forfaitaires, des ordonnances pénales et des transactions pour les contentieux de masse en matière délictuelle : 30 % de transactions pénales aux Pays-Bas, d'ordonnances pénales en Suisse. Il en est de même pour les procédures alternatives, *alternative dispute resolution* - ADR -, la France atteignant déjà un taux supérieur à 50 % de procédures alternatives aux poursuites sous l'égide du parquet pour les affaires avec auteur identifié.

A - La « matière première » du pénal n'est pas celle du civil

La matière première pénale traitée par les algorithmes est constituée des données issues des différentes phases de la procédure, depuis l'enquête jusqu'à la phase d'exécution des peines.

Dans le domaine civil - ou plutôt « non-pénal » -, il en va différemment. La matière première soumise aux algorithmes est essentiellement constituée de la jurisprudence, des arrêts, domaine déjà largement accessible. Cette matière première va être de plus en plus travaillée et transformée par la logique des algorithmes, avec les logiciels de type *machine-learning* traitant des masses de décisions disponibles de plus en plus volumineuses, avec des résultats impressionnants.

Concernant les décisions pénales, les données publiées en France sont moins facilement accessibles au nom de la protection de la vie privée et de la difficulté d'anonymiser toutes les données directement ou indirectement nominatives (nom, adresse, mais aussi téléphone, carte bancaire, immatriculation du véhicule, éléments de vie privée dont vie sexuelle, et une masse d'informations factuelles qui intéressent toutes un traitement par algorithmes...), ainsi que des questions liées aux amnisties - jusqu'en 2007 - et aux procédures de réhabilitation.

Mais la question des données traitées ne se réduit pas à celle de leur accès. Elle pose aussi celles de leur qualité, voire de leur labellisation spécifique, de leur hiérarchisation, toutes opérations à effectuer à la source, par le producteur. La jurisprudence ne peut se réduire à une masse informe de décisions de justice retraitée par les algorithmes. La jurisprudence, et chacun de ses sous-ensembles par type de juridictions et de contentieux, intègre des hiérarchies, des interactions et des

(15) P. Robert-Diard, Choisir ses jurés sur internet, JCP 10 oct. 2016, n° 41, p. 1841. (16) J.-P. Jean, Le système pénal, La Découverte, coll. Repères, 2008. (17) Systèmes judiciaires européens : efficacité et qualité, Rapp. de la CEPEJ sur 47 pays, éd. du Conseil de l'Europe, Les études de la CEPEJ n° 27, 2018 ; J.-P. Jean, Les transformations de la justice pénale en Europe au regard du rapport 2018 de la CEPEJ, AJ pénal 2019. 52.

chronologies subtiles. Et les données factuelles intégrées dans les décisions de justice ne peuvent pas être toutes disponibles, car elles sont tout à fois diffusées et protégées, pour répondre à des finalités précises fixées par la CNIL dans le cadre de la loi de 1978.

Contrairement à ce que voudraient certains, le débat ne peut donc pas se réduire à une approche quantitative (plus j'ai de données dans la base de traitement et plus les résultats sont pertinents), mais doit au minimum se doubler d'une approche qualitative (quelles données accessibles, quelles hiérarchies, quelles pertinences méthodologiques ?), ce qui en matière d'analyse de jurisprudence soulève des problèmes d'interprétation souvent complexes.

B - La justice dite prédictive pour les contentieux de masse

Le débat sur la justice pénale face aux algorithmes se pose de façon tout à fait différente selon les types de contentieux. Il est évident qu'elle tendra d'abord à se développer dans les contentieux de masse, avec certains effets qui peuvent être positifs, d'autres qu'il faudra éviter ou réduire.

La matière disponible. Dans un avenir qui est difficilement « prévisible », plus de 800 000 décisions pénales (données 2016) pourraient être à terme anonymisées chaque année en France avant d'être mises en *open source*, soit :

- 102 000 arrêts pénaux de cour d'appel (44 000 arrêts correctionnels, 36 000 arrêts de la chambre de l'instruction, 22 000 arrêts des chambres de l'application des peines) ;
- 3 000 arrêts de cours d'assises ;
- Environ 612 000 jugements et ordonnances des tribunaux correctionnels (dont 150 000 ordonnances pénales et 71 000 ordonnances d'homologation de comparution sur reconnaissance préalable de culpabilité (CRPC)) ;
- Environ 100 000 décisions du juge des libertés et de la détention.

Il faut ajouter le jugement des contraventions par le tribunal de police (47 000), notamment celles relevant du contentieux routier.

On peut rappeler aussi, pour mémoire, les autres contraventions de la compétence des juridictions de proximité (384 000) et les amendes forfaitaires majorées (13 millions).

Les avantages. On voit déjà l'intérêt du recours aux algorithmes pour les jugements et ordonnances des tribunaux correctionnels. Connaître les jurisprudences et les analyser, c'est à terme, nécessairement, les rapprocher au nom de l'égalité des citoyens devant la loi. C'est donc, pour les infractions répétitives, aboutir à une « barémisation » qui existe déjà en fait dans nombre de juridictions, au siège comme au parquet, voire

commune aux deux, de façon implicite ou de fait. Ainsi, en matière pénale, les barèmes ou « tarifs » de référence existent, officieux mais systématiques dans les contentieux routiers, ne serait-ce que pour éviter les écarts trop importants entre les juges uniques pour des faits similaires. Au parquet aussi, pour les validations de transaction, les taux de poursuites selon le type et les quantités de drogues, etc.

Cette connaissance des jurisprudences nationales et locales constitue une information essentielle pour favoriser le travail des avocats et des magistrats. Elle n'est en rien antinomique avec l'individualisation des peines, bien au contraire. C'est sur cette base de référentiels documentés que pourra se développer une économie du procès à niveaux différents au regard de l'infraction commise et de la procédure choisie¹⁸.

C - Une justice plus prévisible

La connaissance du niveau des peines moyennes prononcées au niveau national pour de mêmes faits, la connaissance de la jurisprudence par juridiction, ne peuvent que favoriser la justice acceptée ou négociée, à la différence de la *justice imposée* et de l'aléa du procès traditionnel :

- *La justice acceptée*, c'est l'ordonnance pénale (exécutoire en l'absence d'opposition). C'est (indirectement) aussi la composition pénale.
- *La justice négociée* c'est la comparution sur reconnaissance préalable de culpabilité et tous les systèmes explicites ou implicites de *guilty-plea*. L'avocat joue alors un rôle de conseil éclairé auprès de son client pour accepter ou non à la fois la procédure proposée et négocier les modalités et le niveau de sanction.

Le point central du système est celui de la qualité de la défense pour tout le secteur de la justice négociée. L'analyse économique est là encore utile. Ainsi, aux États-Unis, des chercheurs ont montré comment des condamnations qui sont autant d'erreurs judiciaires ont été acceptées par des personnes préférant une peine apparemment réduite plutôt que d'encourir le risque de la peine maximale automatique en cas de procès aboutissant à une déclaration de culpabilité, ces prévenus étant généralement conseillés par des avocats indemnisés au titre du *legal aid* qui font tourner rapidement les affaires.

Les statistiques nationales et locales aujourd'hui donnent aux magistrats les informations utiles sur les délais et niveau de sanction et il n'y a aucune raison pour que ces données quantitatives ne soient pas accessibles publiquement. La publication des quanta de condamnations pour des catégories précises d'infractions répétitives, des taux d'indemnisation largement diffusés favorisent un mode de règlement transactionnel, l'acceptation d'une peine négociée ou non, sur des bases claires et objectives. Il s'agit d'un outil de transparence démocratique, par-delà la seule connaissance empirique que possèdent

(18) M. Massé, J.-P. Jean et A. Giudicelli, *Un droit pénal postmoderne ?*, PUF, 2009.

les avocats locaux, ce qui pose un problème d'égalité dans la défense et de la qualité de la défense négociée.

Cette barémisation est essentielle en matière civile dans les contentieux de l'indemnisation. Mais, en France, contrairement aux évolutions qui ont marqué tous les pays comparables, des années ont été inutilement perdues dans l'introduction raisonnée des barèmes et des méthodes normalisées de calcul dans l'aide à la décision des juges civils pour certains contentieux dits « de masse ». Dès 2000, s'appuyant sur des expériences en juridictions, le travail piloté par la Mission de recherche droit et justice avec la Direction des affaires civiles et du sceau, des praticiens et des économistes avait abouti à un rapport et des propositions sur des modes de calcul des pensions alimentaires intégrant les prestations sociales, pouvant servir de référence indicative objective face aux injustices de l'aléa judiciaire¹⁹.

Il aura fallu ensuite attendre dix années pour que la chancellerie diffuse par circulaire aux magistrats une table de référence indicative²⁰. Mais la Cour de cassation, par sa première chambre civile, dans un arrêt du 23 octobre 2013²¹, estimant que les juges ne pouvaient fonder leur décision sur une circulaire, par une interprétation discutable d'une référence seulement indicative, a cassé une décision de la cour d'appel d'Angers. Une certaine confusion a ainsi été ajoutée aux débats, freinant le mouvement vers une rationalisation accrue de ce type de décisions. Dans les travaux préparatoires à cet arrêt, on ne trouve aucune étude d'impact ou analyse des éléments de fond faisant référence aux travaux scientifiques alors disponibles. Au vu de sa seule rédaction, l'arrêt est diversement interprété. Aujourd'hui, on pourrait espérer qu'un tel arrêt aurait fait l'objet d'une étude d'impact et d'une « motivation enrichie »²² pour lever toute difficulté d'interprétation compte tenu de son importance pratique majeure pour la justice quotidienne.

D - Le nécessaire aléa du procès traditionnel

Concernant non plus les infractions reconnues et les contentieux répétitifs, mais les infractions contestées ou complexes, s'appliquent la procédure de jugement public et le débat contradictoire, avec l'aléa judiciaire qui en résulte nécessairement. Le juge est un garant, un recours dans les hypothèses de justice acceptée ou négociée. Pour les infractions contestées ou complexes, il est un juge de plein exercice, en première ligne, qui dispose du temps nécessaire à l'exercice de sa fonction, lors du débat contradictoire à l'audience, ainsi que des outils d'aide à la décision, dont fait partie l'accès aux bases de données de jurisprudence.

S'il n'existe pas une volonté politique forte de rééquilibrage dans l'accès aux traitements de données par algorithmes, ces nouveaux outils techniques vont favoriser encore plus ceux qui bénéficient en priorité de la justice négociée, à savoir les délinquants économiques. D'où un marché pour les start-ups, avec les conseils d'avocats spécialisés et les moyens de connaître jurisprudences et *quantum* des peines.

Le recours à des algorithmes de plus en plus performants, dans les versions *premium* des entreprises de la *Legal-tech*, vont donc entraîner une asymétrie d'informations en faveur de ceux qui ont les moyens de les financer et aggraver les inégalités entre justiciables. Les éditeurs juridiques et les *start-ups* peuvent développer des produits de plus en plus sophistiqués à disposition des justiciables aisés, des entreprises et des cabinets d'avocats spécialisés.

Ce recours aux algorithmes va-t-il donc favoriser l'égalité entre citoyens ou aggraver les inégalités devant la justice ? Sans doute les deux en même temps, mais en ne favorisant l'égalité qu'entre les citoyens les plus modestes qui ne peuvent pas financer un niveau élevé d'accès à l'information, de qualité d'expertise et de conseil.

E - Le risque d'accroissement des inégalités entre justiciables

Les leçons du pénal appliquées au contentieux civil montrent que beaucoup, dans le choix des stratégies judiciaires et du résultat du procès, dépend de la qualité de l'accès à l'information, du professionnalisme de l'avocat, et des moyens mobilisés pour la défense ; et ce, plus encore en matière civile où l'initiative revient aux parties.

Les justiciables, individus ou entreprises, qui disposent des moyens suffisants augmentent leur capital de connaissance pour définir des stratégies à tous les stades de la procédure, soit :

- rédiger les contrats et les clauses d'attribution de compétence ;
- choisir son avocat ;
- choisir son arbitre ;
- effectuer une analyse coût/résultat envisagé d'un procès ;
- décider ou non d'une action contentieuse ;
- négocier une transaction ;
- faire du *forum shopping*, choisir sa juridiction, sinon son juge.

Les entreprises de la *Legal-tech* pourront développer à l'envi des produits ciblés pour répondre aux questions de plus en plus

(19) I. Sayn (dir.), Un barème pour les pensions alimentaires ?, Doc. fr., coll. Perspectives sur la justice, 2002. (20) Circulaire de la Direction des affaires civiles et du Sceau CIV/06/10 du 12 avr. 2010 et table de référence indexée. (21) Civ. 1^{re}, 23 oct. 2013, n° 12-25.301, D. 2013. 2518, 2968, entretien A.-M. Leroyer, 2014. 563, chron. C. Capitaine, et 1171, obs. F. Granet-Lambrechts ; AJ fam. 2013. 703, obs. S. Thouret ; RTD civ. 2014. 77, obs. P. Deumier, et 105, obs. J. Hauser. (22) Rapp. de la commission de réflexion sur la réforme de la Cour de cassation, févr. 2017, https://www.courdecassation.fr/institution_1/reforme_cour_7109/reflexion_reforme_8630/.

nombreuses au fur et à mesure de l'importance du contentieux pour le client. Les offres de service vont se multiplier dans un marché sans limite.

Dans le marché du droit, il existe aussi une « niche économique », celle de l'analyse qualitative de haut niveau produite par les universitaires et les avocats spécialisés.

Les débats à venir sur le contenu futur des prestations intellectuelles nécessaires pour convaincre un juge s'annoncent passionnants. Dans la perspective ouverte par les travaux de la commission de réforme de la Cour de cassation, ces débats porteront aussi sur les études d'impact et les incidences des différentes décisions possibles. Quelle sera donc la place, dans le raisonnement et la décision du juge, de l'argument quantitatif, statistique, par rapport à l'argument dit « de pur droit », qui n'est d'ailleurs parfois pas si pur qu'on veut bien le dire... ?

Pour certains contentieux, nous allons donc assister à une évolution de l'argumentation des avocats dans une logique conséquentialiste. L'approche quantitative ne portera pas seulement sur le nombre de décisions allant dans le sens soutenu, mais aussi sur l'impact de la décision en termes de coût, de conséquences systémiques qui pourraient être mise en évidence par d'utiles corrélations.

F - L'open data et le recours aux algorithmes, entre service public et marché

Sans doute, dans une vision prospective, la confrontation entre État et marché s'articulera-t-elle autour de trois secteurs :

- Dans un marché très concurrentiel, les start-ups de la *Legaltech* et les grands éditeurs juridiques vont proposer des produits de plus en plus sophistiqués aux cabinets d'avocats spécialisés et aux services juridiques des entreprises. Ceux-ci vont faire des analyses coûts-bénéfices et des choix financiers ne laissant rapidement sur le marché que les sociétés les plus performantes ;
- L'État et les différentes composantes du secteur public s'adresseront aux publics défavorisés, via les guichets uniques, des plates-formes de services, les politiques des conseils départementaux de l'accès au droit, dans le domaine de l'information juridique, l'aide juridictionnelle, les médiateurs du numérique, les associations spécialisées... ;
- Les particuliers accéderont directement sur internet à des services *low cost* d'information juridique et d'aide aux contentieux sans avocats ou avec avocats à prestation forfaitisée (infractions routières, petits contentieux...).

Face à ces évolutions du marché du droit, que peut donc faire la puissance publique, au nom et au sein de laquelle est pro-

duite et traitée la matière première, composée de l'ensemble des décisions de justice²³ ?

Toutes ces décisions présentent-elles un réel intérêt pour la diffusion de la jurisprudence ? Est-il plus simple de tout traiter de façon quasi automatisée, pour alimenter des bases de données traitées par des algorithmes de plus en plus sophistiqués. Le risque est de multiplier les niches d'information les plus diverses qui intéresseront des clients pour un service payant correspondant à des besoins spécifiques, qui ne seront pas que juridiques.

Ou bien, est-il souhaitable et peut-on réellement cibler ce qui constitue la finalité première, la diffusion de la jurisprudence pour aider à la régulation par le droit, et donc sélectionner les décisions ? Le comité des ministres du Conseil de l'Europe, dans une recommandation de 1995 préconisait en ce sens une sélection des décisions afin « d'éviter l'accumulation d'informations inutiles²⁴ ».

Le fait qu'une décision judiciaire soit rendue en audience publique est une chose, qu'elle soit remise aux parties en est une autre, qu'elle soit diffusée publiquement, et notamment sur Internet, en est une troisième ; qu'elle soit retraitée avec des millions d'autres via les algorithmes des moteurs de recherche nous fait entrer dans une nouvelle dimension dont nous ne pouvons pas mesurer aujourd'hui l'ampleur des conséquences.

Dans les nouveaux services proposés, les services publics peuvent conduire des projets. Ainsi aux Pays-Bas un projet récompensé par la CEPEJ « Résolution de conflits en ligne pour les conflits relationnels » consistait-il en une plate-forme en ligne intégrant un dispositif d'accès à l'aide judiciaire, l'orientation vers une procédure négociée et un lien avec le conseil juridique fourni par un avocat spécialisé²⁵.

Mais est-il acceptable en France qu'un employeur puisse vérifier si un candidat à l'embauche a connu des litiges civils ou des condamnations pénales, avec en outre tous les détails : cette personne a-t-elle été condamnée pour conduite en état alcoolique ? A-t-elle déjà attrait un employeur devant le conseil des prud'hommes ? Est-elle mauvais payeur ? Connaît-elle des difficultés familiales susceptibles d'affecter son engagement professionnel ?

Répondre à ces questions, c'est ce que propose aux États-Unis le site internet *checkr.com*, qui offre de simplifier la recherche d'antécédents d'un salarié, ce qui constitue parfois, dans ce pays, une obligation pour l'employeur, susceptible d'engager sa responsabilité si elle n'est pas effectuée. Ce service aboutit à la rédaction d'un rapport après consultation, automatisée ou

(23) La question spécifique des décisions des tribunaux de commerce mériterait, par ailleurs, un traitement particulier tant elle est importante stratégiquement et économiquement. (24) Recommandation n° R (95) 11 du comité des ministres aux États membres relative à la sélection, au traitement, à la présentation et à l'archivage des décisions judiciaires dans les systèmes de documentation juridique automatisés. (25) Bureau d'Aide judiciaire des Pays-Bas (*Raad voor Rechtsbijstand*), projet « Résolution de conflits en ligne pour les conflits relationnels », <https://rm.coe.int/how-to-improve-day-to-day-functioning-of-courts-towards-a-more-efficient/168078ab9e>.

non, des registres de sécurité sociale, des fichiers des agresseurs sexuels, et des décisions de justice pénales et civiles disponibles²⁶.

Au regard des règles européennes, mais aussi tout simplement du système de valeurs qui doit être le nôtre, ce type de service n'est évidemment pas acceptable.

G - Penser les finalités de l'ouverture des bases de données de jurisprudence

La Cour de cassation, dans ses attributions actuelles, comme dans les perspectives tracées par le rapport Cadet, tient et tiendra une place essentielle dans le développement des bases de données de jurisprudence²⁷. Elle veillera au respect du principe fondamental de la protection des données personnelles, un traitement informatisé devant être attaché à une finalité précise au regard du principe européen de « minimisation des données »²⁸.

La Cour de cassation, en tant que producteur de données, apporte également une valeur ajoutée par l'anonymisation mais aussi par l'enrichissement des décisions via la hiérarchisation (PBRI), les sommaires, les mots clés, les rapprochements. Cette valeur ajoutée, qui est aussi celle apportée par la doctrine, le savoir des avocats spécialisés, s'oppose à la

seule approche quantitative du *big data* qui traiterait une masse informe ne permettant plus de distinguer l'important de l'in-signifiant, qui noierait le juriste et le citoyen dans une marée de décisions où un moteur de recherche établirait ses propres hiérarchies, aux finalités parfois discutables.

D'où l'importance du travail qualifié du service public de la justice à la source, sur un original certifié et dont le respect de l'intégrité peut toujours être aisément vérifié, de la hiérarchie des arrêts publiés par un mode d'identification clair et connu des juristes, avec un enrichissement par des sommaires et des titres, les références doctrinales. La diffusion publique de la jurisprudence pour aider à la régulation par le droit implique donc de hiérarchiser et de qualifier, voire de sélectionner les décisions. La France connaît dans d'autres domaines une politique des appellations contrôlées et des sites classés. Il existe aussi des grandes appellations et des grands crus pour les arrêts des Cours suprêmes. Il revient donc à la Cour de cassation et aux Cours suprêmes nationales en Europe de mettre en place, notamment via ECLI, un système « d'étiquetage », en tout cas un mode de repérage clair des arrêts les plus importants qu'une motivation adaptée permettra d'aisément interpréter²⁹.

Les algorithmes ont relevé le défi quantitatif. Il revient à la communauté des juristes de relever le défi qualitatif de la production à la source.

(26) <https://checkr.com/privacy-policy/>. (27) V. la déclaration commune Cour de cassation - Conseil national des barreaux du 20 mars 2019. (28) Art. 5 du Règl. UE 2016/679 du Parlement européen et du Conseil du 27 avr. 2016 susvisés note 5 ; Décis. Cons. constit., 21 mars 2019, consid. n° 85 à 101 ; et art. 33 de la loi de programmation pour la justice, 23 mars 2019, JO 24 mars, distinguant not. publicité et publication des décisions de justice, afin d'élargir leur mise à disposition du public tout en évitant les demandes abusives (29) Il en est de même pour la jurisprudence francophone au sein de l'AHJUCAF sur son site internet <https://www.ahjucaf.org/> et pour la base de données JURICAF <http://www.juricaf.org/>.