Francis Donnat

L'intelligence artificielle, un danger pour la vie privée ?

« There was a time when humanity faced the universe alone and without a friend. »

Isaac Asimov[1]

Très tôt, il est apparu nécessaire de réguler les traitements automatisés de données personnelles afin de protéger le droit au respect de la vie privée. Le législateur français, précurseur il y a plus de quarante ans avec la loi du 6 janvier 1978, a posé les grands principes qui régissent encore aujourd'hui la matière.

Les récents progrès foudroyants de l'intelligence artificielle invitent manifestement à repenser la législation applicable tant les problématiques issues de l'IA semblent relever d'un autre ordre. Si un algorithme est une suite d'instructions permettant, à partir d'éléments donnés, d'obtenir un résultat, le développement de nouveaux types d'algorithmes apprenants (ou *machine learning*) représente une rupture qualitative : s'écartant du déterminisme des algorithmes classiques, l'algorithme apprenant présente la particularité d'être conçu de telle sorte qu'il peut découvrir lui-même les opérations à suivre. Cette capacité à apprendre « seul » implique un degré croissant d'autonomie de l'algorithme par rapport à son programmeur. Là où celui-ci définissait en amont l'enchaînement des instructions que l'algorithme classique suit, il entraînera désormais l'algorithme apprenant avec des exemples. Là où la machine suivait des

1. *I, Robot*, New York (N. Y.), Gnome Press, 1950, introduction.

étapes prédéfinies, elle explorera et définira désormais elle-même le chemin à suivre ; figé dans un cas, le processus est évolutif dans l'autre ; il était relativement prédictible, il l'est moins ; déterministe dans le premier cas, l'algorithme devient probabiliste dans le second, et la machine, de serve qu'elle était par rapport au programmeur, s'affranchit progressivement de lui. Quelles que soient ces différences, substantielles, les enjeux restent toutefois les mêmes pour les humains : il s'agit de préserver notre capacité à protéger le droit au respect de la vie privée et à la dignité. À cet égard, si la portée et l'utilité des règles qui existent déjà ne doivent pas être minimisées, l'essor de l'IA est à l'origine de problématiques nouvelles auxquelles la législation devra s'adapter.

La régulation existante

96

Les algorithmes et les données personnelles qu'ils utilisent sont déjà régulés par les règles de droit en vigueur[2].

A. Il en va tout d'abord ainsi des données personnelles dont se nourrissent les algorithmes. Le règlement général sur la protection des données (RGPD), dans la ligne des grands principes établis dans la loi française Informatique et libertés de 1978, pose à cet égard un certain nombre de règles. Les données collectées doivent être « adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées » (art. 5, § 1). C'est le principe dit de minimisation des données. De plus, elles doivent être exactes et tenues à jour, et traitées de façon à garantir une sécurité appropriée.

Les droits des internautes et, plus généralement, de toute personne dont les données personnelles sont collectées sont en outre garantis. Les personnes ont tout d'abord le droit élémentaire de savoir si des données à caractère personnel les concernant font ou non l'objet d'un traitement (art. 15, § 1). La personne concernée a par ailleurs le droit d'avoir accès à ces données et aux informations concernant les finalités du traitement, les catégories de données concernées, leur durée de conservation ou les critères utilisés pour déterminer cette durée, ainsi que les destinataires auxquels elles ont été ou seront communiquées. Elle a enfin le droit d'exiger que les données inexactes ou incomplètes soient corrigées ou, dans certains cas, effacées. Le droit à l'effacement de données à caractère

2. Pour une présentation plus exhaustive, cf. Francis Donnat, *Droit européen de l'internet*, Paris, LGDJ, 2018, p. 65 et suiv.

personnel est notamment ouvert lorsque les données ne sont plus nécessaires au regard des finalités pour lesquelles elles ont été collectées, lorsque la personne concernée retire son consentement au traitement et qu'il n'existe pas d'autre fondement juridique à celui-ci, ou encore lorsqu'elles ont fait l'objet d'un traitement illicite (art. 17). La collecte, la conservation et la qualité des données personnelles collectées sont ainsi déjà fermement encadrées.

B. Il en va de même du traitement dont ces données personnelles peuvent faire l'objet. Les algorithmes doivent être considérés comme un traitement de données, cette notion étant définie de façon très large par l'article 4 du RGPD comme visant « toute opération ou tout ensemble d'opérations effectuées ou non à l'aide de procédés automatisés et appliquées à des données ou des ensembles de données à caractère personnel ». Cinq séries de règles essentielles peuvent être citées. En premier lieu, le traitement de données doit être licite (art. 6) : il l'est si la personne concernée a consenti au traitement de ses données personnelles pour une ou plusieurs finalités spécifiques ou si le traitement est « nécessaire » aux fins légitimes poursuivies par le responsable du traitement. Le traitement de données doit, en deuxième lieu, être loyal et transparent, ce qui implique que les informations relatives au traitement soient facilement accessibles et aisément intelligibles. Autrement dit, toute personne a le droit d'obtenir des informations sur la façon dont l'algorithme fonctionne. Troisièmement sont en principe interdits, sauf exceptions bien identifiées, les traitements de données dites sensibles, c'est-à-dire qui révèlent l'origine raciale ou ethnique de la personne, ses opinions politiques, ses convictions religieuses ou philosophiques ou son appartenance syndicale, de même que ses données génétiques, concernant sa santé ou son orientation sexuelle (art. 9). Quatrièmement, le RGPD consacre le droit de la personne à s'opposer, dans certains cas et « pour des raisons tenant à sa situation particulière », à un traitement de données ou à son profilage (art. 21). Enfin, disposition essentielle, l'article 22 du RGPD, dans la ligne de ce que prévoit déjà la loi française de 1978, consacre le droit de la personne concernée à ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, produisant des effets juridiques la regardant ou l'affectant de manière significative. Toute personne a ainsi le droit à ce que les décisions importantes à son sujet ne soient pas prises par un algorithme de façon automatisée, mais le soient par un être humain. Cette disposition vise notamment le profilage, défini de façon large à l'article 4 du RGPD comme toute forme de traitement

automatisé de données consistant à les utiliser « pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique ».

On le voit à cette rapide énumération des règles actuellement en vigueur : elles garantissent déjà un niveau élevé de protection pour l'individu. C'est au regard de ce niveau de protection qu'il faut se demander si le développement des algorithmes apprenants est susceptible de constituer une menace.

Ce que l'IA change

Le fonctionnement de ces algorithmes et les données qu'ils requièrent soulèvent en effet des problématiques bien spécifiques au regard des règles visant à protéger la vie privée.

A. S'agissant tout d'abord des données utilisées, l'IA soulève plus particulièrement trois types d'enjeux nouveaux. L'exigence de qualité et de pertinence des données revêt tout d'abord une importance essentielle en matière d'IA. En effet, la capacité de l'algorithme à apprendre correctement sera directement influencée par le choix des données qui lui seront fournies lors de la phase d'apprentissage. Microsoft en avait fait l'amère expérience en 2016 en constatant que son robot conversationnel Tay, nourri par les conversations des réseaux sociaux, s'était rapidement mis à tenir des propos racistes et misogynes, obligeant les ingénieurs à le débrancher. Cet exemple malheureux montre bien que le risque de biaiser irrémédiablement l'IA est grand en cas de mauvaise sélection des données dont elle a besoin lors de sa phase d'apprentissage, sans même parler des hypothèses de négligence ou, pire encore, de malveillance dans le choix de ces données. Il sera, au final, impossible de savoir si le biais ou la reproduction à l'infini d'une discrimination provient de l'algorithme, chaque jour plus autonome, ou des données qui lui ont été fournies à l'origine. L'algorithme bénéficiant par ailleurs – à tort – dans l'opinion publique d'une image d'impartialité ou d'objectivité quasi scientifique, le biais n'en sera que plus difficile à distinguer.

La question de la quantité des données livrées à l'IA est également délicate. D'un côté, le principe de minimisation, consacré par le RGPD, implique de limiter la collecte de données personnelles à ce qui est

strictement nécessaire au regard des finalités pour lesquelles elles sont traitées. Mais, de l'autre, il pourrait être opportun de livrer à l'IA autant de données que possible afin, d'une part, de la rendre plus efficace et, d'autre part, d'éviter les biais statistiques qui pourraient découler d'un mauvais choix dans l'échantillonnage de données. Ainsi que la Commission nationale de l'informatique et des libertés (CNIL) le relève dans la synthèse des débats publics menés sous son autorité sur le sujet, il n'est pas interdit de se demander si le développement de l'IA ne doit pas conduire à repenser l'équilibre général de la législation au regard du principe de minimisation[3].

Enfin, les règles issues du RGPD et la protection qui en découle ne visent que les données « personnelles ». Si celles-ci sont certes définies très largement comme toute information concernant une personne physique identifiée ou identifiable, il n'en reste pas moins que le RGPD ne vise en aucun cas les autres types de données, celles qui ne revêtent pas un caractère personnel – comme certaines données financières ou boursières, ou encore des données de santé agrégées au niveau d'une population déterminée –, qui sont pourtant largement utilisées par certains algorithmes. Or l'exploitation systématique à grande échelle de ces données par des algorithmes chaque jour plus autonomes peut avoir des répercussions sociales et économiques d'envergure. Il s'agit là d'un véritable angle mort de la législation sur lequel il conviendrait de se pencher rapidement.

B. Les caractéristiques particulières des algorithmes apprenants soulèvent quant à elles deux séries d'enjeux majeurs. En premier lieu, la question de la transparence et de l'explicabilité du fonctionnement de l'algorithme, déjà délicate s'agissant des algorithmes déterministes, devient redoutable s'agissant des algorithmes apprenants. Si le propre de ces algorithmes est précisément d'être en mesure de développer leurs propres critères de fonctionnement, et d'acquérir ainsi une autonomie croissante au fur et à mesure de leur apprentissage solitaire, comment s'assurer que la façon dont les résultats ont été obtenus reste compréhensible ? Le fait que leur mode de fonctionnement peut parfois devenir difficilement explicable même pour le propre codeur qui en est à l'origine fait, *a fortiori*, fortement douter de la capacité de l'utilisateur de cet algorithme à en comprendre le fonctionnement. L'explicabilité du mécanisme d'apprentissage constitue

3. « Comment permettre à l'Homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle », CNIL.fr, décembre 2017, p. 38 et 40.

ainsi un enjeu majeur, faute de quoi le principe de transparence du traitement des données risque fort de rester lettre morte.

Seconde problématique particulièrement compliquée à régler vis-à-vis des algorithmes apprenants : celle du droit à ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé. Là encore, comment garantir que le pouvoir de recommandation et de prédiction des algorithmes apprenants reste cantonné à une aide à la prise de décision humaine, sans la remplacer ? Nombreux sont déjà les exemples où l'on pense pouvoir dire que le diagnostic algorithmique est plus fiable que l'appréciation faite par l'homme. À l'avenir, avec une efficacité croissante des algorithmes, il sera de plus en plus difficile et rare d'oser s'écarter de leur préconisation. Dans les faits, le droit à ne pas faire l'objet d'une décision individuelle automatisée risque fort d'être mis à bas, ce qui ne sera par ailleurs pas sans soulever de redoutables questions de partage de responsabilité entre l'être humain, décideur apparent mais qui se sera senti en situation de compétence liée par rapport à un algorithme présumé infaillible, le codeur à l'origine de l'algorithme et ce dernier.

Ainsi, les enjeux liés au développement des algorithmes apprenants sont énormes et appellent, si l'on souhaite maintenir les grands principes de protection de la vie privée et de la dignité, des évolutions dans la régulation.

Adapter les règles aux progrès de l'IA

Dans la ligne des règles déjà inscrites dans le RGPD, qu'il convient de prolonger tout en les adaptant, plusieurs pistes sont envisageables. La première serait de mettre en place une législation spécifique aux données non personnelles. Si la protection de la vie privée de l'individu justifie que les données personnelles soient régulées, les enjeux collectifs de société peuvent justifier qu'une forme de régulation soit instaurée sur l'usage qui est fait des données non personnelles et plus particulièrement sur celles utilisées pour nourrir les algorithmes apprenants. Il pourrait être demandé que ces données soient de qualité, pertinentes et non biaisées afin notamment d'éviter que l'algorithme ne reproduise à l'infini les biais ou les discriminations présents dans les données qui ont servi à son apprentissage. Une exigence de traçabilité des données utilisées pourrait également être requise, de sorte que leur origine soit documentée. Sur ces derniers points (absence de biais et traçabilité), la base de données servant à l'apprentissage de l'IA ferait potentiellement l'objet, dans le respect du secret industriel et des droits de la propriété intellectuelle,

d'un audit. Un certain degré d'information de l'utilisateur et de transparence sur la nature des données utilisées devrait, par ailleurs, être exigé.

Le droit pourrait en deuxième lieu évoluer pour rendre le mode de fonctionnement des algorithmes apprenants plus transparent et intelligible. À l'instar du droit de chacun de savoir si ses données personnelles font ou non l'objet d'un traitement automatisé, la première des choses serait de consacrer le droit de savoir si l'on s'adresse à une IA conversationnelle ou à un humain. Lorsque le test de Turing[4] ne sera plus qu'une formalité pour la plupart des IA, aucune ne devra pouvoir faire croire à l'utilisateur qu'elle est un humain.

En matière d'intelligibilité et d'explicabilité (*explainable AI*, ou XAI), et même s'il est illusoire de croire qu'une parfaite intelligibilité pour tous est envisageable, il est nécessaire d'améliorer l'explicabilité des modèles, la qualité des interfaces utilisateurs et la compréhension du mode de fonctionnement de l'IA. Par exemple, on peut imaginer que l'utilisateur devrait pouvoir se voir proposer un outil de visualisation lui permettant d'appréhender les ressorts de fonctionnement de l'algorithme et, le cas échéant, de le tester sous diverses configurations.

Le corollaire de cette exigence d'intelligibilité est celui de loyauté de l'algorithme. Ce principe, formulé par le Conseil d'État dans son étude annuelle de 2014 sur le numérique et les droits fondamentaux, doit garantir le fonctionnement de bonne foi et non biaisé d'un algorithme. Le service rendu par celui-ci ne doit pas être altéré ou détourné à des fins étrangères à l'intérêt de l'utilisateur. En outre, afin de répondre à la particularité des algorithmes apprenants qui est de pouvoir développer de façon autonome des modes de fonctionnement, une forme d'audit des résultats produits pourrait être envisagée, ainsi que le recommandent la CNIL[5] et le rapport Villani[6], par des contrôles *ex post* sur la base de tests à l'aveugle réalisés sur des profils fictifs. À l'auditabilité de la base de données utilisée doit pouvoir s'ajouter celle du caractère loyal et non biaisé des résultats fournis par l'IA.

Une dernière piste, encore plus importante que les précédentes, consiste à préserver le principe selon lequel la prise de décision doit pouvoir être imputée à un être humain. Ainsi qu'il a été dit, le RGPD et la loi de 1978 consacrent déjà le droit individuel à ne pas faire l'objet d'une décision

4. Faculté d'une machine à imiter la conversation humaine de telle sorte que l'humain n'est pas en mesure de discerner qu'il ne parle pas à un humain.

5. « Comment permettre à l'Homme de garder la main ? », synthèse citée, p. 63.

6. Cédric Villani, *Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne*, mars 2018, p. 143 (disponible sur LaDocumentationFrançaise.fr).

fondée exclusivement sur un traitement automatisé. Dans cette lignée, la capacité de l'être humain à surveiller et à garder la main sur l'IA doit être absolument garantie. À cette fin, l'obligation doit être posée de prévoir des dispositifs tendant, selon les domaines concernés et leur sensibilité, à préserver l'intervention de l'être humain à tous les stades du processus de décision (*human-in-the-loop*, ou HITL), la capacité de l'être humain à intervenir pendant la conception du système et à veiller à son bon fonctionnement (*human-on-the-loop*, ou HOTL) ou à superviser l'ensemble de l'activité de l'IA (*human-in-command*, ou HIC)[7]. Une forme d'effectivité de ce droit doit être garantie, par exemple une exigence de traçabilité de la décision humaine qui permettra de s'assurer que le rôle de la machine a bien été un rôle de recommandation, et qu'une intervention humaine a bien eu lieu pour valider, modifier ou écarter la
 préconisation algorithmique.

*

On le voit à cet exercice en partie prospectif : les règles issues du RGPD doivent être réaffirmées et, pour certaines, adaptées aux nouveaux enjeux soulevés par l'IA. Garantir le droit de savoir que votre interlocuteur n'est pas un humain mais une IA, celui de savoir et de comprendre ce qu'elle fait de vos données, notamment lorsqu'elle recommande ou prédit quelque chose vous concernant, le droit à ce que cette recommandation ou cette prédiction soit loyale et non biaisée, et enfin le droit à ce que l'IA reste, s'agissant des décisions les plus importantes, cantonnée à un rôle d'aide à la décision, sans remplacer l'être humain, qui doit toujours pouvoir garder la main : tels sont les enjeux auxquels il convient de répondre tant il est vrai que les progrès attendus du développement des algorithmes apprenants ne doivent pas se faire au détriment des valeurs éthiques que portent le droit de l'Union européenne et la législation française.

7. Cf. le rapport du groupe d'experts de haut niveau sur l'éthique de l'intelligence artificielle mis en place par la Commission européenne (*Ethics Guidelines for Trustworthy AI*, avril 2019 ; disponible sur EC.Europa.eu).

RÉSUMÉ

Si le règlement général sur la protection des données (RGPD) s'applique déjà aux algorithmes afin de garantir un haut niveau de protection de la vie privée, le développement d'algorithmes apprenants, toujours plus autonomes, suppose de l'adapter. Plus de transparence et d'intelligibilité dans le mode de fonctionnement de l'algorithme, la garantie que les données utilisées ne sont pas biaisées, que son fonctionnement est loyal et non discriminant, et enfin que son rôle reste limité à une aide à la décision, sans remplacer l'être humain, qui doit toujours pouvoir garder la main : c'est à quoi il faut répondre pour que les progrès attendus de l'IA ne se fassent au détriment des valeurs éthiques portées par le droit de l'Union européenne et la législation française.