



ITW DE N. NEVEJANS
PROPOS RECUEILLIS PAR
NICOLAS GAMBARDILLA
8602C

PROTÉGER LA SANTÉ, LA SÉCURITÉ ET LES DROITS FONDAMENTAUX LE RÈGLEMENT EUROPÉEN SUR L'IA

Je m'appelle Nathalie Nevejans, je suis professeure de droit privé à l'université d'Artois, titulaire de la chaire d'enseignement et de recherche "IA responsable", un projet résolument interdisciplinaire, avec des aspects informatiques, éthiques et légaux. J'ai aussi créé le Diplôme Universitaire "Responsable de l'Éthique de l'Intelligence Artificielle" qui forme les futurs acteurs de la conformité éthique des systèmes d'IA. Je me suis initialement intéressée à la robotique, il y a une quinzaine d'années. J'ai publié en 2017 le premier livre de recherche sur le droit et l'éthique de la robotique civile¹. Cet ouvrage m'a permis d'être reconnue sur le plan national, européen et même international, et de rencontrer de très nombreux experts. Puis il y a eu la belle rencontre avec le monde de la santé. J'ai vraiment découvert l'IA en 2018. Je peux dire aujourd'hui que la robotique, l'IA et la santé sont vraiment mes centres d'intérêts principaux en tant que juriste.

Qu'est-ce que le règlement européen sur l'intelligence artificielle ?

C'est un texte dont le but est d'encadrer les risques générés par les systèmes d'IA en matière de santé, de sécurité et de droits fondamentaux des citoyens de l'UE. Le règlement sur l'IA a été adopté définitivement le 13 juin 2024 et publié en juillet de la même année. Il a une valeur obligatoire et est d'application directe. Cela veut dire que tous les États membres sont obligés de l'appli-

quer ; ils n'ont pas le choix. Il est donc à distinguer d'une directive qui doit être transposée dans le droit national pour pouvoir être appliquée. C'est important quand on sait que la France est souvent en retard pour transposer les directives européennes. Nous sommes dans la même situation que pour le règlement général sur la protection des données (RGPD). Le règlement sur l'IA est un texte destiné à promouvoir une IA de confiance et centrée sur l'humain, tout en protégeant l'innovation et encourageant le développement de l'économie européenne. Le texte essaye de trouver un équilibre entre la protection des personnes et la protection de l'innovation des entreprises, un exercice difficile. Ce texte concerne à la fois la mise sur

le marché ou en service des systèmes d'IA, et l'utilisation professionnelle de ces systèmes. Le fabricant, par exemple d'un robot doté d'IA, doit respecter ce texte s'il veut le vendre sur le marché européen, et ce même s'il est chinois ou américain. Par conséquent, on peut considérer que ce texte protège les entreprises européennes contre la concurrence internationale, en empêchant cette dernière d'avoir un accès plus facile à un marché européen de près de 500 millions d'habitants. Quant à l'utilisateur professionnel, il devra également respecter le règlement sur l'IA². Il s'agit, par exemple, d'un hôpital qui voudrait utiliser un système d'IA dans le cadre des soins.



Que nous dit ce texte ? Que rend-il obligatoire ? Qu'interdit-il ?

Ce texte impose des obligations de conformité, qui sont décrites dans le règlement. Il s'applique à tout le monde sauf au domaine militaire. Quand une entreprise développe un système d'IA, elle va devoir se poser une première question : « Est-ce que mon système d'IA va entraîner des risques ? » Le règlement propose une hiérarchie des risques.

Certains sont considérés comme inacceptables. Par exemple, des systèmes utilisés pour la notation sociale (*social scoring*) ou bien qui recourent à une technique subliminale pour influencer et pour amener des personnes à prendre des décisions qu'elles n'auraient pas prises autrement. Dans ces cas de figure, la mise sur le marché comme l'utilisation sont interdites. Notons que cela ne s'applique pas à la recherche. Les entreprises ont le droit d'avoir des services R&D qui réfléchissent sur des sujets dont la mise sur le marché est impossible à ce moment-là. Il y a également quelques exceptions à l'interdiction, par exemple dans le domaine pénal, mais c'est vraiment à la marge.

Ensuite, il y a le cœur du règlement, ce qu'on appelle les systèmes d'IA à haut risque (pour la sécurité, la santé, et les droits fondamentaux des personnes). L'entreprise développant un système d'IA, doit être capable de savoir dans quelle catégorie il se place. Le règlement

contient un système de classification du haut risque. L'annexe III contient une liste de systèmes qui sont, par définition, à haut risque. Par exemple, dans le domaine de l'enseignement, un système d'IA utilisé pour détecter les élèves qui auraient un mauvais comportement, qui risqueraient de tricher, est considéré à haut risque. On retrouve encore le domaine de la démocratie, le domaine de la justice, ou encore le domaine de l'emploi, donc des ressources humaines. Si le système d'IA ne figure pas dans cette liste, à l'image des systèmes d'IA utilisés en santé, il peut encore être à haut risque s'il remplit cumulativement deux conditions, conformément à l'article 6 du règlement. D'une part, s'il relève d'un autre texte européen visé à l'annexe I. Par exemple, on trouve dans cette annexe, les textes concernant les machines ou encore les dispositifs médicaux. D'autre part, si cet autre texte européen impose qu'un tiers évalue la conformité, par exemple de la machine ou du dispositif médical. Donc un dispositif médical, comme un robot doté d'IA, qui interviendrait sur un patient serait nécessairement considéré comme à haut risque car 1/ c'est un dispositif médical relevant du Règlement relatif aux dispositifs médicaux et 2/ un tiers doit évaluer la conformité de ce dispositif au règlement sur les dispositifs médicaux pour les classes de risque IIa (risque potentiel modéré/mesuré), IIb (risque potentiel élevé/important) et III (risque élevé), sachant que le tiers

n'intervient pas pour la classe I (risque faible).

On trouve plus bas la catégorie des systèmes d'IA qui imposent aux fournisseurs et utilisateurs professionnels la fourniture d'une information à la personne qui utilisera le système d'IA en raison de risques spécifiques liés à la nature particulière du système, comme un *chatbot* ou des *deep fakes*³ (article 50). Par exemple, pour un *chatbot* dans le domaine de la santé, il faudra délivrer à l'utilisateur les informations adéquates pour être sûr qu'il sait qu'il discute avec une machine et non un humain.

Enfin, le système peut être à risque minime ou nul. Dans ce cas, on n'applique pas le règlement sur l'IA. L'UE a souhaité réglementer *a minima*, et a donc prévu dès l'origine que la majorité des systèmes d'IA seraient à risques minimes ou nuls. Si c'est un système d'IA relevant du domaine de la santé, seules les règles relevant du règlement sur les dispositifs médicaux et celles du RGPD seront alors applicables.

Quid d'un système qui prédit les meilleurs traitements ou au contraire des traitements à éviter, car susceptibles d'entraîner des événements indésirables ?

Si on regarde ce qui vient d'être dit sur les conditions de l'article 6, on sait que ce système relève de l'annexe I, puisqu'il est un dispositif médical. Mais il faut

Interview de Nathalie NEVEJANS, Professeure de droit privé, titulaire de la chaire IA Responsable de l'université d'Artois.

1. Nathalie Nevejans (2017) *Traité de droit et d'éthique de la robotique civile*. LEH Édition. 1229 pages ISBN13 : 978-2-84874-668-5

2. L'article 2, § 12 dispose également : « Le présent règlement ne s'applique pas aux systèmes d'IA publiés dans le cadre de licences libres et ouvertes, sauf s'ils sont mis sur le marché ou mis en service en tant que systèmes d'IA à haut risque ou en tant que systèmes d'IA qui relèvent de l'article 5 ou de l'article 50. » L'article 5 concerne les systèmes à risque inacceptables, qui sont interdits, et l'article 50 les systèmes qui imposent une obligation d'information, comme les *chatbots*.

3. Technique de synthèse multimédia reposant sur l'IA qui peut être utilisée pour créer des infox et des canulars malveillants.



aussi que l'évaluation de conformité d'un tiers soit obligatoire pour la mise sur le marché de ce dispositif. Ici, on peut penser qu'on serait dans un cas de figure montrant un certain risque pour le patient, donc un dispositif médical plutôt de type IIa ou IIb, voire III qui exige l'évaluation de conformité d'un tiers, car il pourrait provoquer le décès de patients à la suite de recommandations inadéquates. On peut en déduire que ce dispositif médical serait également considéré comme un système d'IA à haut risque au sens du règlement sur l'IA. Son fournisseur devra remplir toutes les obligations de conformité, qui sont assez lourdes, notamment en matière de données, de gouvernance, d'existence d'un humain aux manettes (comme le médecin qui va contrôler les résultats pour la prise de décision).

Qu'est-ce que ce règlement change pour le citoyen lambda ?

Comme ce règlement va s'appliquer automatiquement, les citoyens seront protégés lorsqu'ils utiliseront n'importe quel produit contenant de l'IA mis sur le marché dans l'Union européenne. Si les autorités s'aperçoivent qu'un système d'IA n'est pas conforme, il y aura des sanctions, qui peuvent être extrêmement lourdes puisqu'elles peuvent être calculées à partir du chiffre d'affaires mondial annuel du fournisseur.

La question sera de savoir si le règlement sera à même de lutter contre les risques ? En principe, oui. Mais en pratique, le règlement veut trouver un équilibre entre la protection des personnes et la protection de l'innovation et de l'économie. Il pose un certain nombre de règles protectrices des personnes, mais il ne va pas très loin. Par exemple, dans le cas d'un *chatbot* pour la santé mentale qui ne serait pas un système d'IA à haut risque, si la seule obligation pour le

fournisseur est d'informer l'utilisateur qu'il discute avec un *chatbot*, est-ce suffisant ? Est-ce que cela sera adapté à tout le monde ? On voit là les limites du texte qui devra être amélioré, peut-être en le combinant avec une approche éthique. On peut imaginer des personnes, des hôpitaux, des développeurs de systèmes d'IA en santé, qui appliquent le règlement mais se disent que cela ne suffit pas dans le cas de la santé mentale, et qui appliquent alors une couche supplémentaire de protection, pour les jeunes, pour les personnes vulnérables, pour les personnes âgées, par exemple en vérifiant que l'utilisateur a quelqu'un dans son entourage pour discuter des limites d'usages, qu'il est bien conscient qu'il s'agit qu'une machine et qu'elle n'a pas de sentiments, etc.

Comment les fournisseurs accueillent-ils le règlement ?

Certaines entreprises sont effrayées par la réglementation et ont très peur pour le développement et l'innovation. Alors qu'il faut voir le règlement comme un accompagnement des entreprises, d'abord pour donner confiance aux utilisateurs, et c'est d'ailleurs la vision de l'Union européenne. En effet, si les utilisateurs ont confiance en un système IA, ils vont l'utiliser, ils vont l'acheter ou encore ils ne vont pas hésiter à y recourir. L'Europe et les entreprises européennes ont une carte à jouer dans le développement de systèmes d'IA sûrs. Cela pourrait même constituer un message publicitaire : « Nous faisons de l'IA de confiance et vous pouvez aveuglément utiliser nos systèmes ». Un second impact à prendre en compte est l'"effet Bruxelles". Cet effet avait été constaté avec le RGPD. De nombreux pays n'appartenant pas à l'UE se sont rendu compte que le RGPD protégeait les données des citoyens, et ces derniers faisaient de ce fait plus confiance

à certains médias, certains outils. Ils se sont donc emparés du RGPD et l'ont traduit dans leurs législations respectives. Je pense qu'on aura de façon similaire une exportation de notre règlement IA à l'international. Pour preuve, dès que l'Europe a annoncé en 2021 qu'elle allait légiférer, plusieurs grands pays ont commencé à travailler sur une réglementation similaire, comme la Grande-Bretagne, l'Australie, la Chine, l'Afrique du Sud, les États-Unis, et le Canada.

Il y a donc un règlement. Mais quels sont les moyens dont dispose l'État ou l'Union européenne pour le faire respecter ?

Un certain nombre d'organes vont être créés ou sont déjà en train d'être créés pour justement pouvoir encadrer, entourer, aider les entreprises et les utilisateurs d'IA. Par exemple, au niveau européen, il y a le Bureau de l'IA déjà créé et qui a tenu sa première réunion. D'autres groupes d'experts en IA vont venir expliquer le fonctionnement du règlement, préparer des documents destinés aux entreprises, préparer des éléments de *soft law*.

Les utilisateurs pourront déposer une plainte, comme pour le RGPD, mais on ne sait pas encore à quelle autorité. En France, depuis quelques années, la Commission nationale de l'informatique et des libertés (CNIL) se positionne comme l'autorité spécialisée en IA ; elle se veut être l'autorité de référence. Mais pour le moment, il n'y a rien d'officiel. Le règlement a seulement dit que chaque pays devait se doter d'une autorité de contrôle pour vérifier si les systèmes d'IA sont conformes. Dès que ce sera le cas en France, si l'autorité de contrôle était informée d'une violation, elle aurait alors le pouvoir de sanctionner, à la manière dont elle le fait dans le cas du RGPD. ■

NICOLAS GAMBARDILLA
8602C
DIRECTEUR DE RECHERCHE
AU CNRS



QUAND L'IA PARLE DU BAHUT

Les grands modèles de langage, entraînés sur une partie significative des connaissances humaines numérisées et accessibles, prétendent à devenir des assistants universels, capable en particulier de répondre à n'importe quelle question sur n'importe quel sujet (n'oublions pas toutefois que ces modèles ignorent tout des notions même de question et de réponse et ne font que prédire de nouvelles briques de texte en fonction de ce qui précède). Que nous racontent donc les modèles les plus répandus à propos de notre cher bahut ? Nous utiliserons l'invite suivante :

Qu'est-ce que le Prytanée National Militaire de La Flèche et comment a-t-il contribué à la grandeur de la France ? en 2500 caractères

Il est important de réaliser que la moindre modification de cette invite peut entraîner un changement, parfois drastique, de la réponse. Nous avons donc utilisé cette invite verbatim avec tous les outils testés. **À noter que les réponses qui suivent n'ont pas été modifiées ou corrigées et viennent directement des systèmes d'IA.**

Commençons par le modèle qui a tout changé pour le grand public, *ChatGPT*. Ce modèle, développé par la compagnie OpenAI (possédée à 49 % par Microsoft) a été conçu comme un *Chatbot* qui interagirait avec les utilisateurs de manière naturelle en produisant des réponses

qui paraissent avoir été produites par une autre être humain. *ChatGPT* peut être paramétré durant la conversation dans les invites. Par exemple, la "température" permet de préciser à quel point il peut se permettre d'être créatif. Testons-le avec les températures minimale et maximale.

ChatGPT :

Invite + (temperature = 0)

« Le Prytanée National Militaire de La Flèche, fondé en 1604 par Henri IV, est une institution emblématique de l'histoire éducative et militaire de la France. Situé dans la ville de La Flèche, dans la Sarthe, il a d'abord été un collège royal confié à la Compagnie de Jésus avant d'être transformé en école militaire sous la Révolution française en 1808. Cet établissement prestigieux, destiné à la formation des élèves brillants issus de tous milieux sociaux, incarne un modèle d'excellence académique et de discipline militaire.

Le Prytanée se distingue par son double rôle. Sur le plan éducatif, il a maintenu un niveau académique élevé, en dispensant des enseignements variés, allant des sciences aux lettres, tout en mettant l'accent sur la rigueur intellectuelle et morale. Sur le plan militaire, il a contribué à former les futures élites de l'armée française, inculquant des valeurs de service, de patriotisme et de discipline. Parmi ses anciens élèves, on trouve de nombreux officiers, scienti-

fiques, et personnalités ayant marqué l'histoire de France, notamment dans les domaines de la défense et des grandes entreprises publiques.

L'impact du Prytanée sur la grandeur de la France réside dans sa capacité à préparer des générations d'élèves à des responsabilités majeures, que ce soit en temps de guerre ou en période de paix. Par exemple, ses anciens élèves ont joué un rôle crucial dans les grandes campagnes militaires du XIX^e et XX^e siècles, tout en contribuant aux avancées scientifiques et technologiques qui ont renforcé la position de la France sur la scène internationale.

En outre, le Prytanée a également été un vecteur d'égalité sociale, permettant à des jeunes talentueux, quelle que soit leur origine, d'accéder à une éducation de haut niveau et à des carrières prestigieuses. Ce principe républicain d'égalité des chances a consolidé la cohésion nationale et renforcé le sentiment d'appartenance à une France unie et ambitieuse.

Ainsi, le Prytanée National Militaire de La Flèche a non seulement formé des individus brillants, mais a également incarné et promu les valeurs fondamentales de la République française : mérite, excellence et dévouement à la nation. Ces contributions en font une institution clé dans l'histoire de la grandeur et de l'influence de la France. »

4. "Droit souple", ensemble de règles dont la juridicité est discutée, non obligatoires ou non appliquées de manière systématique.