

s les plus profondes t réalité »



Pour moi, la science-fiction fonctionne comme un laboratoire d'idées, où les auteurs explorent les limites du possible

”

rithmique, la manipulation de l'information ou encore l'érosion de la vie privée. Ces problèmes sont certes beaucoup plus terre-à-terre, mais infiniment plus réels.

D'un autre côté, la science-fiction joue un important rôle d'alerte. Il vaut mieux pécher par excès de prudence que d'être pris au dépourvu. Le problème survient lorsque les visions dystopiques paralysent le débat constructif ou servent de prétexte à l'absence de réglementation : « De toute façon, nous ne pouvons rien faire contre une superintelligence. » En ce sens, ni l'optimisme exagéré ni la technophobie radicale n'offrent de solutions aux problèmes émergents.

Des études montrent que les robots peuvent reproduire les préjugés des humains... D'où vient cet effet et comment le limiter ?

J'ai travaillé sur cette question, notamment à travers l'exemple des préjugés à l'égard des professions prestigieuses. C'est un problème fondamental : l'IA apprend à partir de données produites par des humains ; elle hérite donc de tous nos préjugés. Par exemple, si les femmes apparaissent moins souvent comme PDG dans les jeux de données d'apprentissage, l'algorithme en déduira que cette fonction est « masculine ». Il ne s'agit pas de malveillance de la part de la machine, mais du reflet de notre société.

Pour limiter cet effet, des mesures doivent être prises à plusieurs niveaux : constituer des équipes diversifiées pour concevoir les systèmes d'IA, mettre en place des audits algorithmiques capables de détecter les discriminations, et intégrer des techniques

de « débiaisage » dans le processus d'apprentissage. Mais le plus crucial est la transparence : nous devons savoir à partir de quelles données un système a été formé et comprendre ses limites.

A l'ère des modèles génératifs, le contrôle de l'information imaginé par les classiques est-il plus proche que jamais ?

L'IA générative est une arme à double tranchant. D'une part, elle démocratise la création de contenu : tout le monde peut devenir « auteur ». D'autre part, elle permet une manipulation à une échelle sans précédent. *Deepfakes*, textes générés par l'IA, images ou vidéos synthétiques : la frontière entre vérité et fiction devient de plus en plus floue. Le contrôle orwellien de l'information est techniquement possible, mais paradoxalement plus difficile à instaurer dans un monde où chacun dispose d'outils de création.

Le problème n'est pas qu'un acteur unique contrôle la vérité, mais que la vérité se noie dans un océan de contenus générés. Cela nécessite un nouveau type de vigilance : non seulement face à la censure, mais aussi face au flot de désinformation.

Quels changements positifs, dans le domaine scientifique et social, attendez-vous de l'IA au cours de la prochaine décennie ?

Je vois le plus grand potentiel dans la démocratisation de l'accès à la connaissance et aux outils. L'IA, en tant que tuteur personnel, peut révolutionner l'éducation : chaque enfant pourrait bénéficier d'un enseignant virtuel capable de s'adapter à son rythme d'apprentissage. Mais avant d'en arriver là, nous pourrions introduire un apprentissage par mémorisation basé sur la « répétition espacée », adapté individuellement à chaque élève. Les outils existent déjà. Ce qui manque encore, c'est la mise en œuvre et le soutien de l'Etat.

En médecine, l'IA facilitera le diagnostic dans les régions dépourvues de spécialistes et contribuera à réduire les files d'attente. Dans le domaine scientifique, l'IA accélère déjà les découvertes, qu'il s'agisse de la conception de nouveaux médicaments ou de la modélisation du climat.

Mais le changement le plus significatif pourrait être d'ordre social. Bien utilisée, l'IA pourrait nous libérer de certaines tâches monotones, nous laissant davantage de temps pour la créativité, les relations humaines et le développement personnel. C'est une opportunité de redéfinir ce que signifie « être humain » au XXI^e siècle. Le problème, bien sûr, est que là où nous espérons une automatisation chargée des corvées domestiques pour nous libérer du temps pour l'art, nous avons conçu des systèmes capables de créer de l'art, ne nous laissant parfois plus que le balai.

ABONNÉS



Littérature et IA : des visions pessimistes, mais aussi optimistes
A lire sur notre site.



Dune
FRANK HERBERT
Pocket
11,95 euros



L'âge de diamant
NEAL STEPHENSON
Le Livre de Poche
640 p.
9,90 euros



La servante écarlate
MARGARET ATWOOD
Robert Laffont - Pavillons
576 p.
12,50 euros

réglementation « Nous ne pouvons pas laisser l'IA fonctionner comme une mystérieuse prophétie »

KA.KI

Dans *Frankenstein*, Mary Shelley avertissait que le créateur était responsable du monstre. Quel type de contrôle est possible dans le cas de l'IA ?

La métaphore de Frankenstein est particulièrement pertinente : nous créons quelque chose que nous ne comprenons pas totalement. Les modèles contemporains d'IA sont des « boîtes noires » : nous savons comment les entraîner, mais pas toujours pourquoi ils prennent certaines décisions. Le contrôle reste possible, mais il doit être pensé de manière systémique : responsabilité juridique des concepteurs, normes de sécurité, mécanismes d'arrêt d'urgence, audits réguliers. L'« IA explicable », c'est-à-dire des systèmes capables d'expliquer leurs décisions, est également essentielle. Nous ne pouvons pas laisser l'IA fonctionner comme une mystérieuse prophétie.

Malheureusement, le problème tient au fait que nous tirons de plus en plus de bénéfices de ces modèles. Les arrêter devient donc « non rentable », alors même que nous courons le risque d'assister à la destruction soudaine de la civilisation.

La loi européenne sur l'IA est-elle l'équivalent moderne des lois d'Asimov ?

L'AI Act constitue un pas dans la bonne direction, mais la comparaison avec les lois d'Asimov souligne l'ampleur du défi. Les lois de la robotique étaient simples et universelles. L'AI Act, lui, s'étend sur des centaines de pages de réglementations qui tentent de couvrir une infinité de scénarios possibles. C'est indispensable dans un cadre juridique, mais est-ce suffisant ?

Le principal problème réside dans le rythme des évolutions : alors que la loi met des années à être élaborée, la technologie évolue en quelques mois. L'AI Act risque donc de devenir obsolète avant même d'être pleinement appliqué. Nous avons besoin d'un cadre plus flexible, de règles qui puissent s'adapter rapidement aux nouveaux défis sans qu'il soit nécessaire de tout réécrire. C'est pour cette raison que j'ai proposé, dans les médias spécialisés, l'introduction d'une « quatrième loi de la robotique » : une réglementation exigeant au minimum l'identification claire des agents IA.

Mais n'est-ce pas un simple ilot européen dans un océan de chaos juridique ? La Chine et les Etats-Unis semblent beaucoup moins soucieux des réglementations. Sommes-nous donc condamnés à un développement incontrôlé de l'IA ?

C'est une menace réelle : la « course à l'armement » technologique dans le domaine de l'IA pourrait conduire à un abaissement des normes de sécurité. Pourtant, l'Europe dispose d'un atout : un immense marché qu'aucune entreprise ne peut ignorer. L'« effet Bruxelles » a fait du RGPD une norme mondiale *de facto*. Le problème est que la mise en œuvre de cette norme laisse à désirer, tout comme la réglementation sur les cookies. En réalité, la coopération internationale

est essentielle. Nous avons besoin d'un « traité de non-prolifération » pour les IA dangereuses, et de normes de sécurité mondiales. C'est bien sûr difficile en cette période de tensions géopolitiques, mais l'alternative, c'est-à-dire un Far West numérique, est trop dangereuse. J'espère qu'un jour, les Etats-Unis et l'Europe sauront former un front commun.

Vous soutenez l'idée que l'IA doit toujours révéler qu'elle est une IA. Comment cela fonctionnerait-il, concrètement ?

C'est absolument fondamental pour maintenir la confiance de la société. Une première étape pourrait être l'usage de « filigranes » numériques, des marqueurs invisibles pour l'utilisateur intégrés dans les contenus générés par l'IA. Certaines entreprises, comme Google avec son modèle Veo, commencent à les introduire. Ensuite, chaque interaction avec un *bot* devrait commencer par une mention claire précisant que nous n'avons pas affaire à un être humain. Concrètement, cela pourrait prendre la forme de marquages standardisés – à l'image du label « E » pour les additifs alimentaires –, d'API obligatoires permettant d'identifier l'origine du contenu, et de sanctions en cas de tromperie.

La technologie existe déjà. Ce qui manque, c'est la volonté politique et la mise en œuvre effective. Il ne s'agit pas seulement d'une question technique, mais aussi d'un respect fondamental de l'autonomie des individus.

Alors que les experts en débattent depuis des années, nous ignorons encore, par exemple, qui devrait être tenu responsable en cas d'erreur d'une voiture autonome.

Il faut accepter que le droit aura toujours un temps de retard sur la technologie. La solution consiste à créer un cadre adaptatif, avec des règles générales de responsabilité plutôt que des dispositions détaillées. Il devrait, par principe, toujours y avoir un « humain dans la boucle » capable d'assumer cette responsabilité : qu'il s'agisse du fabricant, du propriétaire ou de l'opérateur.

Mais cette présence humaine peut parfois n'être qu'un fusible illusoire. Si l'on m'indique, par exemple, qu'un drone a localisé un terroriste avec 95 % de certitude et que j'ai vingt secondes pour valider sa neutralisation, la décision est en réalité déjà prise depuis longtemps. En réalité, l'essentiel devrait être d'inverser la charge de la preuve : ce n'est pas à la société de prouver qu'une IA est dangereuse, mais au créateur de démontrer que son système est sûr. Enfin, il faut envisager des mécanismes comme des assurances obligatoires ou des fonds d'indemnisation.

Nous ne pouvons pas stopper le progrès, mais nous pouvons en atténuer les effets négatifs. « Heureusement », nous connaissons les erreurs des algorithmes depuis des années. En cas de négligence manifeste, la responsabilité incombe évidemment au fabricant. Mais il existe des cas où la faute n'est attribuable à personne en particulier. Il en ira de même avec l'IA : l'essentiel est que ces systèmes soient plus fiables que la moyenne des êtres humains.