Sous réserve de modifications

Allocution de Patricia Kosseim, commissaire à l'information et à la protection de la vie privée de l'Ontario – Symposium canadien sur la protection de la vie privée de 2023 de l'IAPP
26 mai 2023

Lorsque j'ai écrit mon billet de blogue *La vie privée et l'humanité à la croisée des chemins*, il y a plus d'un an, j'ai comparé l'avenir de l'intelligence artificielle (IA) à d'autres menaces existentielles comme le réchauffement de la planète et la possibilité d'une guerre nucléaire. J'ai affirmé que de plus en plus, l'IA va au-delà de la prédiction du comportement humain avec une précision quasi parfaite, pour influencer ce comportement d'une manière qui met en péril notre capacité à agir. Comme l'a dit Daniel Solove, « les algorithmes ne font pas que prédire l'avenir : ils le créent ».

J'ai ajouté que les algorithmes peuvent influer sur les parcours éducatifs que nous réservons (ou non) à nos enfants, les emplois qu'ils sont susceptibles d'obtenir, voire de postuler, les peines que l'on impose et l'évaluation des demandes de libération conditionnelle, la personne que nous décidons d'épouser ou pour qui nous voulons voter, et les décisions importantes que nous prenons pour notre santé et d'autres aspects de notre vie en fonction de notre patrimoine génétique et de probabilités statistiques.

J'ai écrit au sujet de l'influence qu'exercent les plateformes de médias sociaux et les fils de nouvelles personnalisés sur l'information à laquelle nous sommes exposés et qui renforce nos convictions, tendances, attitudes et préjugés en fonction de ce que nous lisons et écrivons, des choses que nous aimons et de nos amis.

Et cela, c'était avant que ChatGPT ne devienne accessible au grand public. Depuis le lancement de ce système, en mars, le monde a pris conscience de la puissance de l'IA générative et des grands modèles de langage pour créer de l'information synthétique, y compris de la désinformation, et nous avons vu des exemples concrets de la façon dont les hypertrucages peuvent déstabiliser la vérité et altérer notre compréhension du monde qui nous entoure.

En tant qu'organismes de réglementation, nous devons nous garder d'être alarmistes, d'exagérer les risques ou de susciter des craintes, de peur de paraître déconnectés de la réalité. Nous ne pouvons pas crier au loup sans risquer de perdre notre crédibilité, notre force de persuasion et notre pertinence. Est-il donc exagéré de dire que l'IA met la protection de la vie privée et l'humanité à la croisée des chemins?

Les récits de science-fiction bien connus où des robots envahissent le monde remontent à 1921, dans une pièce de théâtre du dramaturge tchèque Karel Čapek appelée *Rossumovi univerzální roboti* (R.U.R. dans sa version française), longtemps avant que l'on puisse même imaginer l'existence d'ordinateurs, de robots ou même de l'IA.

Cependant, plus récemment, des experts scientifiques de renommée mondiale, qui sont aux premières loges du développement de l'IA et de ses risques éventuels, ont sommé l'alarme.

Dans une entrevue à la BBC en 2014, Stephen Hawking a mis en garde contre le développement éventuel d'une IA complète, qui évoluerait par elle-même et se redéfinirait de plus en plus vite, ce qui « pourrait mettre fin à la race humaine ».

Il y a cinq ans, le Pew Research Centre a sollicité l'avis de près d'un millier d'experts en technologie, chefs d'entreprise, décideurs et militants sur les incidences éventuelles des systèmes d'IA et a publié ses conclusions dans un rapport intitulé *Artificial Intelligence and the Future of Humans* (« L'intelligence artificielle et l'avenir de l'humanité »). La plupart des experts interrogés s'inquiètent profondément des menaces que les technologies de l'IA font peser à long terme sur l'autonomie et la capacité d'action des êtres humains, ainsi que sur ce qu'ils appellent les éléments essentiels de l'être humain.

Geoffrey Hinton, que l'on appelle parfois le parrain de l'IA, a quitté récemment son emploi chez Google afin de pouvoir s'exprimer librement sur les risques et les avantages des systèmes d'IA. Selon lui, il est tout à fait raisonnable de se préoccuper de ces questions actuellement. Nous nous approchons du moment où les ordinateurs pourront se perfectionner d'une manière qui échappe à notre contrôle, ce qui pourrait « signifier la fin de l'humanité ». Il s'inquiète de l'incidence de l'intelligence artificielle sur l'humanité, en particulier lorsqu'elle est utilisée par des autocraties à des fins malveillantes. Il réclame la signature de traités internationaux portant notamment sur les armes létales autonomes. Selon lui, si nous ne faisons pas preuve de bon sens dans l'utilisation des technologies de l'IA, il n'est pas inconcevable que l'IA anéantisse l'humanité.

Yoshua Bengio, Elon Musk et d'autres technologues éminents ont signé une lettre ouverte intitulée *Pause Giant AI Experiments* (« Interrompons les expériences d'IA de grande envergure »). Cette lettre réclame un moratoire immédiat de six mois sur l'élaboration de systèmes d'IA plus puissants que le GPT-4, pendant lequel les chercheurs pourraient s'employer à rendre les systèmes d'IA actuels « plus précis, sécuritaires, interprétables, transparents, robustes, alignés, dignes de confiance et loyaux ». Elle cite l'un des principes d'Alisomar sur l'IA, selon lequel [traduction] « l'IA avancée pourrait constituer un changement profond dans l'histoire de la vie sur Terre; il y a lieu de planifier et de gérer son avènement avec soin et avec des ressources suffisantes ».

Yuval Noah Harari, historien, philosophe et futurologue, a décrit ces grands modèles de langage comme un moyen de « pirater le système d'exploitation de la civilisation humaine » d'une manière qui menace notre survie. Le langage, dit-il, forme « l'essence même de la culture humaine ». Que se passera-t-il, demande-t-il, lorsque l'intelligence non humaine deviendra plus douée que la nôtre pour créer des histoires, de nouvelles idées et de nouveaux artefacts culturels par le biais de « contenus politiques de masse, de fausses nouvelles et d'évangiles pour de nouveaux cultes »? Ou quand elle imitera de mieux en mieux les sentiments et l'intimité des personnes sensibles pour nous manipuler et nous faire changer d'avis et de vision du monde? Tout comme la technologie nucléaire pourrait détruire physiquement la civilisation humaine, les nouveaux modèles d'IA pourraient eux aussi être employés comme une sorte d'arme de destruction massive susceptible d'anéantir notre société. « Je parle ici de la fin possible de l'histoire de l'humanité. Pas la fin de l'histoire; juste la fin de celle qui est dominée par l'humanité. »

Les technologies ont déjà conduit l'humanité au bord du gouffre, et pourtant nous avons délibérément choisi de ne pas faire le grand saut. Par exemple, la communauté internationale a décidé d'interdire le clonage humain; la recherche sur certaines chimères homme-animal pour produire des cellules ou des tissus est autorisée, mais jamais au-delà d'un certain stade de développement.

Comme le dit le dicton en éthique, ce n'est pas parce qu'on peut faire quelque chose qu'on doit nécessairement le faire.

Il y a quelques jours, Sam Altman et ses collègues à Open AI ont publié une déclaration décrivant des systèmes d'IA superintelligents, plus puissants que toute autre invention technologique antérieure. « Étant donné ce risque existentiel, nous ne pouvons nous contenter de simplement réagir. » Comparant la superintelligence à l'énergie nucléaire, ils réclament la création d'un organisme semblable à l'Agence internationale de l'énergie atomique, qui effectuerait des inspections, vérifierait le respect des normes de sécurité et restreindrait le déploiement de capacités dépassant un certain seuil.

Par ailleurs, ils semblent également indiquer qu'il est important de permettre aux entreprises de poursuivre la conception de systèmes d'IA dont les capacités sont inférieures à un seuil élevé, sans qu'une réglementation trop contraignante leur soit imposée.

Pour ma part, je pense que nous devons faire *les deux*. Nous devons agir de toute urgence, sur le plan national et international, pour interdire le déploiement dangereux de systèmes d'IA et définir des limites claires que la société mondiale s'accorde à ne pas dépasser.

Mais il est tout aussi urgent d'agir à l'échelon local pour réglementer le contexte dans lequel les technologies de l'IA peuvent être utilisées pour le bien public, sous certaines conditions et moyennant une surveillance appropriée pour garantir la sécurité, l'équité, la transparence, la responsabilité et la protection de la vie privée. Nous devons disposer d'un ensemble de règles solides pour déterminer quelles sont les utilisations bénéfiques, qui en bénéficiera, qui en sera responsable et qui devra trancher.

Hier, mon bureau et la Commission ontarienne des droits de la personne avons publié une déclaration commune exhortant le gouvernement de l'Ontario à élaborer et à poser des balises efficaces pour l'utilisation des technologies de l'IA dans le secteur public, afin que l'Ontario puisse profiter des avantages de ces technologies d'une manière qui soit éthiquement responsable et durable, et qui bénéficie de la confiance du public.

Le cadre de l'intelligence artificielle de confiance de 2021 du gouvernement et les projets de principes et de lignes directrices connexes, que nous saluons, représentent un bon point de départ, mais il est urgent d'aller de l'avant avec cette initiative et de miser sur cet élan pour établir un ensemble de règles contraignantes, solides et détaillées concernant l'utilisation des technologies de l'IA dans le secteur public. Ces règles sont d'autant plus nécessaires que la *Loi sur l'intelligence artificielle et les données* du gouvernement fédéral, si elle est adoptée, ne s'appliquera pas au secteur public ontarien. Le gouvernement doit agir maintenant pour combler cette lacune.

Bref, pour que nous puissions utiliser des données pour le bien de tous, il faut cesser de les utiliser à des fins malveillantes. Et pour les utiliser pour le bien, il faut disposer de *bonnes* données.

Nous devons interdire les utilisations dangereuses et néfastes de l'IA, y compris des systèmes d'IA générative, qui représentent non seulement une menace physique pour l'humanité, mais qui, par la production délibérée de désinformation, risquent de miner notre cohésion sociale et de défaire le tissu même qui nous unit.

Nous devons mettre fin à ces utilisations malveillantes des systèmes d'IA, tout en créant un espace propice au développement et au déploiement d'utilisations bénéfiques de ces systèmes dans un cadre de gouvernance réglementaire solide. Nous devons poser des balises claires afin que des ensembles de données équitables soient employés pour mettre fin au cycle des préjugés systémiques historiques qui perpétuent la division et l'inégalité dans la société. Ce n'est qu'avec des balises efficaces et des données exactes obtenues de façon juste et équitable que les institutions et organisations pourront utiliser ces données pour le bien commun afin de mériter et de conserver la confiance du public, et de relever les défis les plus urgents de la société.