

« Il est urgent de réguler les apprentis sorciers de l'IA »

Freiner les développements pour réfléchir aux enjeux de l'intelligence artificielle (IA) est une nécessité, appuie Laurence Devillers (Sorbonne). Mais gare aux arrière-pensées commerciales de cette initiative.

ENTRETIEN
PHILIPPE LALOIX

Les intelligences artificielles se développent-elles trop vite ? C'est l'avis de 1.377 scientifiques, experts et acteurs de l'intelligence artificielle (IA) qui, dans une lettre ouverte publiée mercredi, plaident pour une pause de six mois dans les développements de ces outils, comparables à ChatGPT. Parmi eux, des entrepreneurs de la Silicon Valley aussi, comme Elon Musk, patron de SpaceX, Tesla et Twitter, les cofondateurs de Skype, de Pinterest. Ou Steve Wozniak (Apple). Leur crainte commune : que les IA viennent bouleverser le fonctionnement de notre société sans que nous y soyons suffisamment préparés.

Pour Laurence Devillers, professeure en intelligence artificielle à l'Université Paris-Sorbonne, « ce mélange des genres est insupportable ». Méfions-nous des arrière-pensées commerciales de cette pétition. Oui, il faut appuyer sur la pédale de frein, « pour identifier les enjeux éthiques et normaliser l'IA », mais « cela doit rester une affaire de chercheurs, pas de business ».

Comment expliquer que ce sujet soit à ce point clivant, opposant ceux qui affichent un enthousiasme sans limite (Sam Altman parle de « tsunami technologique ») et les pires prophètes de l'apocalypse ?

On observe en effet des opinions assez extrêmes qu'il serait urgent de pouvoir pondérer. Il ne s'agit ni d'avoir peur ni d'imaginer que c'est l'invention du siècle. Il y a trois grands sujets. Un, la norme qui doit s'appliquer aux industriels pour essayer d'éviter les dérives. Deux, la loi. C'est l'AI Act, le texte en cours de discussion à la Commission européenne prévu en 2025. Et enfin, l'éthique. Pour moi, l'éthique, c'est un outil. Il ne s'agit pas de dire « c'est bien ou mal ». Il s'agit de comprendre en quoi une voiture autonome qui prend des décisions à votre place est souhaitable, en quoi il y a une discrimination, une amplification des stéréotypes. Ou en quoi ces modèles de langage peuvent générer une parole qui n'a pas de sens.

Il faut démystifier l'IA, la remettre à sa juste place, soit un outil d'informatique statistique ?

Exactement. C'est un objet socio-technique. Il faut expliquer comment il fonctionne, mais aussi comprendre que l'on projette sur ces machines des capacités qu'elles n'ont pas. Par anthropomorphisme.



On leur donne trop de pouvoir ?

C'est encore plus subtil que ça. On n'est pas forcément dupes mais, à travers ces IA, on recherche un miroir. Pour quelqu'un qui n'a pas l'habitude de manipuler ces objets, les productions de ChatGPT peuvent sembler très intelligentes. Ce qui peut même le placer dans un état de vulnérabilité, en le persuadant qu'il est moins intelligent que cette machine.

On a parfois le sentiment que la machine s'emballe, voire même que leurs créateurs ne la maîtrisent plus vraiment...



A chaque fois, l'accueil, c'est de l'amener à faire de l'humain. Or, ils n'ont qu'une connaissance livresque, sans contexte, sans histoire, sans notre ADN, notre vécu... Ce n'est pas de la superintelligence humaine, c'est un supercalculateur



production totalement hallucinante. Qui peut alimenter la désinformation, de la manipulation politique. Ou poser des risques pour la santé mentale, jusqu'à pousser quelqu'un à se suicider (NDLR, comme cela aurait été tragiquement le cas pour un jeune Belge). L'IA balance des stéréotypes. Si vous lui

dites « je vais me suicider », l'IA est incapable de comprendre le sens de ce mot. De l'autre côté, les humains, parce qu'ils ont des biais cognitifs, ou parce qu'ils sont engagés avec une machine qui semble les écouter, vont l'écouter. Et donc croire ce que dit la machine.

Alors que l'on ne peut pas la croire...

Surtout pas. Je crois que les fournisseurs de ces IA ne sont pas forcément dans cette manipulation. Ce sont des conséquences des milliards de données traitées dans les systèmes. Ce ne sont pas forcément les algorithmes qui sont en cause, ni les humains qui les conçoivent, mais le marketing que certains construisent. Et qui tend à nous faire croire que ces chatbots sont intelligents alors qu'ils n'ont rien d'intelligent. Elon Musk ou d'autres sont des oiseaux de mauvais augure qui, au départ de leur stature de milliardaire, nous donnent des conseils scientifiques. C'est inadmissible. Je ne peux pas donner beaucoup de crédit à ce que disent ces messieurs.

Pourquoi, dans ce cas, signent-ils cette pétition réclamant un moratoire de six mois sur les recherches ? Peut-on contraindre des scientifiques ou des entreprises privées à appuyer sur le bouton « pause » ?

Ce sont de bonnes questions : est-ce que l'on peut arrêter. Et qui décide de cela ? On voit arriver une pétition des Etats-Unis alors qu'en Europe, on travaille ferme sur le sujet. Quelle est la volonté réelle de ces différents acteurs, à la fois des chercheurs de renom, des transhumanistes ou des personnes qui marchandent ces sujets (comme Elon Musk).

C'est une affaire de prise de pouvoir. On est d'accord pour dire qu'il faut freiner les développements. Moi, je travaille sur la normalisation et l'éthique de ces sujets. On veut mettre en œuvre des mécanismes pour que les industriels se sentent responsables de ce qu'ils font. J'appartiens à un comité, le CNPEN (Comité national pilote d'éthique numérique), où on essaie d'identifier les enjeux d'éthique essentiels : l'absence de vérité, ses paroles non sémantiques, la manipulation de l'utilisateur.

Les puissances de calcul nécessaires pour déployer ces systèmes sont immenses. Seuls les géants du numérique peuvent y arriver. En France et en Eu-

rope, depuis 2021, on travaille sur Bloom, une IA générative ouverte et participative, qui inclut un millier de chercheurs, avec le soutien du CNRS, du ministère de l'Enseignement supérieur et de la Recherche, des partenaires privés...

Donc, on assiste, en ce moment, à une guerre des normes, où les Américains voudraient passer devant, les Chinois aussi. Il y a une guerre sur le terrain de la régulation. Pourquoi y a-t-il si peu d'Européens dans les signataires ? Posez-vous la question. Tout le monde doit se soutenir. Mais pas en mélangeant recherche et profit.

Si ces chatbots sont à ce point non fiables, pourquoi les développe-t-on ? Seront-ils un jour vraiment utiles pour l'humain ?

Oui, si on les démystifie. Il ne faut pas que tout le monde se mette à coder, mais puisse accéder à un niveau d'éducation sur les concepts qui sont derrière et que l'on raisonne sur l'éthique. Pour éviter d'être bluffé par des personnes qui viennent nous apporter la bonne soupe. Les chercheurs doivent garder la main.

Ce qui est intéressant avec ces outils, c'est leur capacité d'aller détecter les signaux faibles dans de grands ensembles de données que notre cerveau n'est pas capable de traiter. C'est le cas, par exemple, pour détecter des cancers du sein sur des radiographies. Ils peuvent être une source de créativité sauf qu'on les utilise à mauvais escient pour plagier. A chaque fois, l'accueil, c'est de l'amener à faire de l'humain. Or, ils n'ont qu'une connaissance livresque, sans contexte, sans histoire, sans notre ADN, notre vécu... Ce n'est pas de la superintelligence humaine, c'est un supercalculateur.

Est-ce qu'OpenAI a commis une grave erreur en lâchant ChatGPT entre les mains du grand public (dans le but d'entraîner son modèle de langage à grande échelle) ?

Non. Cela a permis, pour la première fois, que chacun se rende compte des dangers de l'IA. Que l'Europe puisse travailler sur une meilleure compréhension des risques, c'est une très bonne chose. Il faut que les fournisseurs de données qui alimentent ces outils soient responsables, à l'inverse de ce qui se passe sur les réseaux sociaux.

« On assiste, en ce moment, à une guerre des normes, où les Américains voudraient passer devant, les Chinois aussi. » © D.R.

Laurence Devillers

Technophile, Laurence Devillers conjugue robotique et linguistique. Chercheuse en intelligence artificielle, professeure à Sorbonne Université (Laboratoire interdisciplinaire de sciences numériques), elle fut l'une des premières à travailler sur l'informatique affective. Autrice de nombreuses publications et de 2 essais (*Des robots et des hommes*, édité chez Plon, 2017, et *Les robots "émotionnels"*, L'observatoire, 2020), elle est membre du Comité national de pilotage de l'éthique numérique (CNPEN), accompagnant le gouvernement français sur ces questions. PH.L.