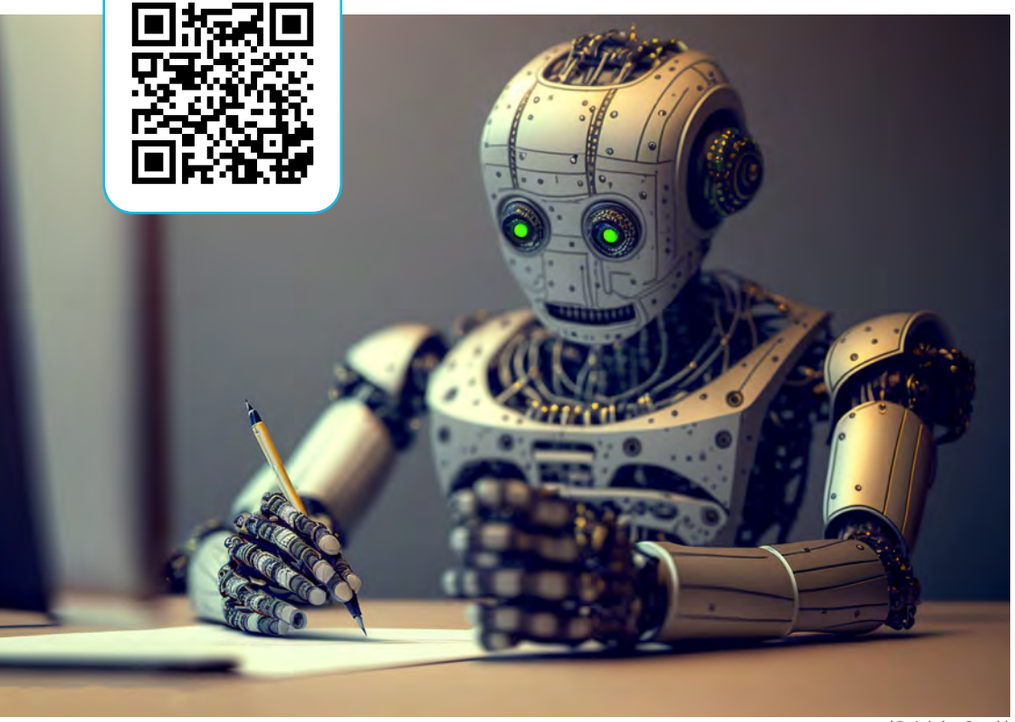
ChatGPT

intelligence artificielle générative ou dégénérative ?

C'est le nouveau dilemme. L'IA générative est-elle amie ou ennemie ? Nous fera-t-elle perdre des neurones ? Si ces questions sont déjà biaisées par leur formulation, la diffusion au grand public de ChatGPT aura au moins eu le mérite de vulgariser cette technologie qui était il y a quelques mois encore peu connue en dehors des cercles d'initiés. Retour sur le dernier événement #IADATES du 13 juin dernier,* coorganisé par l'Institut EuropIA, la Maison de l'intelligence artificielle et le Département des Alpes-Maritimes.

par Magali Chelpi-den Hamer



(© Adobe Stock)

Connaissez-vous le *nudging* ? Laurence Devillers, professeur en IA à l'Université Paris-Sorbonne et chercheuse au Laboratoire interdisciplinaire des Sciences du numérique de Saclay, travaille sur cette pratique au quotidien avec des doctorants qui étudient l'incitation à faire changer d'avis les gens au moyen de chatbots. Parler avec un robot génère en effet de l'affect et l'algorithme du robot peut être entrainé à détecter la variation des émotions de son interlocuteur et ajuster ses réponses en conséquence, parfois pour l'amener à un comportement souhaité. Si l'IA n'a pas inventé les techniques de manipulation (qui est un travers bien humain pour le coup), la diffusion sans limites des chatbots dans les usages quotidiens questionne de plus en plus. Le *nudge* digital est communément utilisé dans les approches marketing avec l'intention derrière - toujours humaine - de modifier les perceptions des prospects pour les amener à concrétiser un acte d'achat. Les *nudges* s'appuyant sur les sciences comportementales et la manipulation des émotions pour orienter le choix des individus vers une direction souhaitée, une autre application évidente est le champ politique pour orienter les comportements de vote.

Au vu des enjeux éthiques évidents, on ne peut pas faire l'économie d'observer les interactions hommes-Chatbot pour en comprendre l'évolution, voire anticiper certains risques. C'est tout l'objet de la Chaire interdisciplinaire HUMAAINE qui regroupe des experts en IA, mais aussi des linguistes, des économistes du comportement, des juristes, des philosophes et même un théologien. Les membres du Comité national pilote d'éthique du numérique veillent également au grain. Lucky nous, Laurence Desvillers fait partie de ces deux instances.

L'intelligence humaine est-elle en danger ?

« Nous avons une intelligence qui n'a rien à voir avec celle des machines. On ressent. On a des valeurs morales. On a un appétit de vie. On a des intentions. Tout ce que n'a pas la machine. Un enfant va mettre sa main sur le feu une fois, il va se brûler et en même temps il apprend de ce contexte qui va être intégré dans son histoire personnelle. Il ne met normalement pas sa main sur le feu une deuxième fois. Une machine ne fonctionne pas comme cela. Elle va devoir répéter plusieurs fois une expérience. La machine n'est pas intuitive. Le goût de la pomme est encyclopédique. Elle n'a pas de notion de distance. Elle peut sans ciller vous

Have you heard of *nudging*? Laurence Devillers, professor of AI at the University of Paris-Sorbonne and researcher at the Laboratoire Interdisciplinaire des Sciences du Numérique in Saclay, works on this practice on a daily basis with doctoral students who are studying how chatbots can encourage people to change their minds. Talking to a robot generates emotions, and the robot's algorithm can be trained to detect variations in the emotions of the person it is talking to and adjust its responses accordingly, sometimes leading them to adopt a desired behaviour. While AI has not invented the techniques of manipulation (which is a very human failing, by the way), the unlimited spread of chatbots in everyday use is raising more and more questions. Digital *nudges* are commonly used in marketing approaches with the underlying intention - always human - of changing the perceptions of potential customers to encourage them to make a purchase. Since *nudges* are based on behavioural science and the manipulation of emotions to steer people's choices in a desired direction, another obvious application is in politics, to steer voting behaviour.

In view of the obvious ethical issues at stake, we must observe human-chatbot interactions in order to understand how they evolve and even to anticipate certain risks. This is the aim of the interdisciplinary HUMAAINE Chair, which brings together AI experts, linguists, behavioural economists, lawyers, philosophers and even a theologian. The members of the national digital ethics steering committee are also keeping a close eye on things. Luckily for us, Laurence Desvillers is a member of both these bodies.

Is human intelligence in danger?

"We have an intelligence that has nothing to do with machines. We feel. We have moral values. We have an appetite for life. We have intentions. Everything the machine doesn't have. A child will put its hand in the fire once, it will burn itself and at the same time it learns from this context that will be integrated into its personal story. The child doesn't normally put its hand in the fire a second time. A machine doesn't work that way. It will have to repeat an experiment several times. The machine is not intuitive. The taste of an apple is vast. It has no notion of distance. It can safely ask you to jump out of the window because it doesn't affect its repository. A machine only analyses the context of the questions you ask. It reasons in 0 and 1 in the choice of its answer. This modus operandi that is completely different from the modus operandi of a human intelligence must remain in our minds."

These words from Laurence Devillers are reassuring, and it's not going to happen soon that machines replace women. However, there is no shortage of things to watch out for. First and foremost is the question of parameters. Compared with the current virulent discussions on algorithms (mostly non-European and English-speaking) and on the corpus of data used to train the machines (disconcertingly

demander de sauter par la fenêtre car cela n'a aucune incidence sur son référentiel. Une machine n'analyse que le contexte des questions que vous lui adressez. Elle raisonne en 0 et en 1 dans le choix de sa réponse. Ce mode opératoire qui est complètement différent du mode opératoire d'une intelligence humaine doit rester dans notre esprit. »

Ces propos de Laurence Devillers rassurent et ce n'est pas demain la veille que les machines remplaceront les femmes. Pour autant les points de vigilance ne manquent pas. En premier lieu la question des paramètres. En comparaison des discussions virulentes actuelles sur les algorithmes (majoritairement extra-européens et anglophones) et sur les corpus de données utilisées pour entrainer les machines (d'une opacité déconcertante), la question est rarement soulevée. Or dans les processus d'apprentissage des machines, c'est une question fondamentale, les concepteurs décidant arbitrairement de tout un tas de choses pour paramétrer les IA (la taille de l'historique des prompts, la séquence du token, la taille du lexique, la température qui permet de choisir entre les différentes séquences de mots possibles qui sera employée dans la réponse).

Un deuxième point de vigilance est d'ordre éthique. Jusqu'où peut-on aller sans se perdre ? Même si l'IA le permet techniquement, peut-on tolérer la création d'agents conversationnels basée sur des profils de personnes décédées (qui ont parfois consenti avant leur mort) ? Le 'thanabot' a étrangement le vent en poupe dans certains milieux, ce qui en dit long sur l'évolution morale de nos sociétés et nous ferait presque regretter l'absence de censure. Les questions de propriété intellectuelle sont aussi au cœur des débats. Comment va-t-on savoir quand un contenu est produit par une machine ? Il est possible de mettre des watermarks dans des photographies et des contenus vidéo produits par les IA (les Designers apprécieront), mais c'est plus difficile pour les textes.

Un troisième point de vigilance concerne les corpus. Actuellement, les algorithmes se basent sur tout ce qui est disponible sur la toile. *No limit*. Pour répondre à une question de prompt, les IA vont scanner le net à la recherche de l'éventail des possibles et les sources pourront tout autant être des articles sérieux et des données fiables que des *fake news* et des données manipulées. Peu d'algorithmes sont transparents là-dessus et quand ils le sont (BING s'est mis par exemple à citer ses sources), attention +++ car sous cette illusion de transparence, les références peuvent être fausses car générées à partir de modèles mathématiques. Certains résultats peuvent aussi avoir être discriminants en fonction des données piochées.

Peut-on résister à la colonisation technologique atlantiste ?

Éthique, normes, loi. Penser l'IA au travers de ce triptyque est fondamental. En matière d'IA générative, l'impulsion à ce jour reste largement américaine, même si les Chinois ne sont pas en reste, et le CEO d'Open AI est allé jusqu'à indiquer que si les règles européennes devenaient trop restrictives, les versions suivantes de ChatGPT ne seraient pas rendues accessibles au Vieux Continent. Pour Marco Landi, le président de l'Institut EuropIA, on est revenu au temps des colonies. L'Europe est en train de subir des technologies qui sont développées ailleurs et il est important de réagir vite. Bloom est à date le seul système d'IA générative qui a été conçu en Europe (si l'on exclut AlphaGo qui a été conçu par une entreprise britannique puis racheté par Google). C'est un LLM multilingue qui a été conçu sur le calculateur Jean Zay à Saclay et qui à son lancement était une version ChatGPT 3 (à titre de comparaison, la version ChatGPT qui a été rendue publique pour la première fois en décembre 2022 était une version 3.5).

Un comité technique de normalisation européen dédié à l'IA, CEN-CLC/JTC 21, a été créé il y a quelques années, rassemblant un groupe d'experts ad hoc sur l'IA pour réfléchir à la création de normes correspondant aux besoins du marché européen et développer les documents normatifs correspondants. Son secrétariat est assuré par le Danemark. Curieusement, ce sont les GAFAM et les BATX qui participent aux discussions avec ce comité. Les grands industriels français sont absents, a souligné Laurence Devillers pendant son intervention, même si d'autre plateformes de réflexion existent certainement en inter-industriels.

Depuis 2019, une directive européenne autorise le datamining partout dans le monde pour entrainer les algorithmes développés par les industriels européens. Isabelle Galy, la directrice de la Maison de l'intelligence artificielle, se demande si l'on ne s'est pas tiré une balle dans le pied ici en contribuant à la diffusion des modèles actuels dominants. Pour des usages grand public, probablement. Pour des usages plus ciblés basés sur des corpus fiables et restreints, on peut peut-être encore résister. ●

---

* Panélistes :

Laurence Devillers, professeur en IA à l'Université Paris-Sorbonne et chercheuse au Laboratoire interdisciplinaire des Sciences du numérique. Membre du Comité national pilote d'éthique du numérique (CNPEN).

Jérôme Béranger, GOODALGO, chercheur associé à l'INSERM 1295, CERPOP, équipe BIOETHICS, Université de Toulouse III, fondateur du label ADEL

Alexandre Ozararat, associé du cabinet Grant Thornton, directeur adjoint à l'innovation, directeur du bureau de Nice, expert-comptable et commissaire aux comptes.

opaque), this question is rarely raised. Yet in machine learning processes, this is a fundamental issue, with designers arbitrarily deciding on a whole host of things to parameterise AIs (the size of the prompt history, the sequence of the token, the size of the lexicon, the temperature that enables a choice to be made between the various possible word sequences that will be used in the response).

A second point of vigilance is the ethical aspect. How far can we go without getting lost? Even if AI technically allows it, can we tolerate the creation of conversational agents based on the profiles of deceased people (who have sometimes consented before their death)? The 'thanabot' is strangely in vogue in certain circles, which says a lot about the moral evolution of our societies and almost makes us regret the absence of censorship. Intellectual property issues are also at the heart of the debate. How will we know when content has been produced by a machine? It is possible to put watermarks in photographs and video content produced by AIs (designers will appreciate this), but it is more difficult for text.

A third point to watch out for concerns the corpus. Currently, algorithms are based on everything available on the web. No limits. To answer a prompt question, the AIs will scan the net looking for the full range of possibilities, and the sources may just as well be serious articles and reliable data as fake news and manipulated data. Few algorithms are transparent about this, and when they are (BING, for example, has started quoting its sources), beware +++ because underneath this illusion of transparency, the references may be false because they are generated from mathematical models. Some results may also be discriminating depending on the data selected.

Can we withstand Atlanticist technological colonisation?

Ethics, standards, law. Thinking about AI in terms of this triptych is fundamental. In terms of generative AI, the impetus to date remains largely American, although the Chinese are not to be outdone, and the CEO of Open AI went so far as to indicate that if European rules became too restrictive, subsequent versions of Chat GPT would not be made available on the Old Continent. For Marco Landi, President of the EuropIA Institute, we are back in colonial times. Europe is being subjected to technologies developed elsewhere, and it is important to react quickly. To date, Bloom is the only generative AI system to have been designed in Europe (apart from AlphaGo, which was designed by a British company and then bought by Google). It is a multilingual LLM that was designed on the Jean Say computer at Saclay and which, when it was launched, was a ChatGPT 3 version (by way of comparison, the ChatGPT version that was made public for the first time in December 2022 was a 3.5 version).

A European standardisation technical committee dedicated to AI, CEN-CLC/JTC 21, was set up a few years ago, bringing together a group of ad hoc experts on AI to consider the creation of standards corresponding to the needs of the European market and to develop the corresponding normative documents. Its secretariat is provided by Denmark. Curiously (even dangerously), it is the GAFAMs and BATXs that are taking part in discussions with this committee. The major French manufacturers are absent, as Laurence Devillers pointed out during her speech, even if other inter-industry platforms certainly exist. ●