ChatGPT ou la question de l'autorité Description

Alexandre Joux

ChatGPT ou la question de l’autorité

Description

Dis-moi qui tu es et comment tu fonctionnes ?

Mis en ligne le 30 novembre 2022 en version test, le chatbot conversationnel ChatGPT d’OpenAI a surpris tout le monde. Avec ChatGPT, il y aura un avant et un après dans la relation du grand public à l’intelligence artificielle (IA). Pour deux raisons : le service lui-même, bluffant ; sa capacité à s’améliorer très rapidement puisque ChatGPT a été utilisé par 100 millions d’utilisateurs en deux mois, un rythme d’adoption extrêmement élevé (le nombre d’utilisateurs est passé à 200 millions début mai 2023). Or, en termes d’intelligence artificielle et de prédiction statistique, plus un service à un nombre élevé d’utilisateurs, plus il affine ses résultats. Les futurs concurrents de ChatGPT, qu’il s’agisse de Bard de Google ou des IA chinoises, partiront donc avec un handicap.

> Avec chatgpt en accès libre, tout un chacun peut désormais utiliser l’intelligence artificielle à sa guise

ChatGPT est bluffant parce qu’il s’adresse pour la première fois au grand public. L’IA dite « générative », celle qui produit du texte original en réponse à une question, existe déjà depuis quelques années. Mais elle n’était jamais sortie des laboratoires et des milieux experts. Ainsi, les outils d’OpenAI, la société qui édite ChatGPT, étaient déjà proposés sur GitHub, la plateforme open source dédiée au développement que Microsoft a rachetée en 2018 (voir *La rem* n°49, p.78). Le service Copilot, lancé en 2021, permet de générer automatiquement du code informatique avec l’aide de ChatGPT. Mais le développeur sur GitHub n’est pas l’internaute *lambda*. Avec ChatGPT en accès libre, tout un chacun peut désormais utiliser l’intelligence artificielle à sa guise. ChatGPT est une IA dite « générale » : c’est en quelque sorte un socle commun pour des outils d’IA qui peuvent se décliner de mille manières. Parce qu’elle n’est pas conçue pour faire une chose en particulier, on peut donc lui demander de faire nos devoirs, de faire de la publicité, de faire des résumés de réunion, d’expliquer des résultats biologiques. Son potentiel semble infini.

Ces intelligences artificielles sont des supercalculateurs qui ne comprennent rien, mais apprennent toujours plus

À cette première rupture, qui est avant tout stratégique, parce que le service espère atteindre très vite une taille critique qui lui donnera un avantage statistique et un avantage concurrentiel, s'ajoute une seconde rupture que le terme « ChatGPT » synthétise. GPT veut dire *« Generative Pre-Trained Transformer »*. Il s'agit d'un type d'intelligence artificielle entraînée sur d'immenses jeux de données textuelles pour « reproduire » l'équivalent du langage humain à partir de l'identification de régularités statistiques, ce que les spécialistes de l'IA appellent des *« large language models »*. Une intelligence artificielle de ce type produit donc mathématiquement des phrases, avec un degré de perfection élevé, sans rien y comprendre : les performances de Google Translate ou de DeepL reposent sur ce type de calcul, même si ces services ne génèrent pas un texte original et proposent seulement de traduire un texte existant. Ces services de traduction ainsi que l'IA générative reposent sur une approche particulière de l'intelligence artificielle dite « apprentissage automatique profond » ou *« deep learning »*, qui s'inspire du fonctionnement des réseaux neuronaux (voir *La rem* n°40, p.91). Mais ce sont des intelligences non intelligentes, au sens où l'intelligence humaine est capable de réflexivité, d'engagement, de critique, de conscience. Ces intelligences artificielles sont des supercalculateurs qui ne comprennent rien, mais apprennent toujours plus des nouveaux calculs qu'elles produisent sur les textes auxquels elles ont accès. Comme il s'agit de calculs gigantesques, ces IA sont toutes installées sur des services cloud aux immenses capacités, des *hyperscalers*, à l'instar d'Azure pour Microsoft, qui héberge ChatGPT, ou du cloud d'Amazon. D'ailleurs, cette dépendance aux *hyperscalers* américains devrait en toute logique conduire les futurs grands services d'IA générative à en dépendre, quand ils ne passeront pas directement sous le contrôle de Microsoft, de Google ou d'Amazon.

À ce dispositif de type *« Generative Pre-Trained Transformer »*, qui produit des phrases comparables à celles des humains, a été ajoutée une dimension conversationnelle, le « chat » de ChatGPT. Un chatbot, un robot conversationnel, apprend à connaître son utilisateur, se souvient de lui (il faut s'identifier pour utiliser ChatGPT), afin de lui fournir des réponses adaptées, c'est-à-dire qui prennent en compte le « contexte », ce qui suppose là encore une intelligence toute statistique. Se produit alors le « miracle » ChatGPT qui donne le sentiment qu'une intelligence, très proche de celle de l'homme, s'engage pour de bon dans une conversation. C'est une impression, un « ressenti » crédible, donc très efficace, même s'il ne faut surtout pas anthropomorphiser ChatGPT.

Cette IA générative et conversationnelle, qui repose sur un modèle général, offre donc des perspectives inédites exceptionnelles pour proposer une panoplie de services et elle est en même temps extrêmement dangereuse car son apparence humaine peut facilement convaincre les plus naïfs et les fins qu'on lui assigne peuvent être éminemment problématiques. D'où le titre de cet article : « ChatGPT ou la question de l'autorité. » Si ChatGPT semble faire autorité quand il donne des réponses péremptoires aux questions qui lui sont posées, ses failles sont nombreuses qui nécessitent une réflexion sur le statut de cette IA et, surtout,

sur ceux qui la programment et l'utilisent.

La version de ChatGPT dévoilée en novembre 2022 est la version 3.5 (une version 4, encore plus performante, est commercialisée depuis mars 2023). L'histoire de son développement remonte à 2015 quand quelques entrepreneurs de l'internet et quelques visionnaires technophiles décident de fonder OpenAI en y investissant 1 milliard de dollars. Il s'agit de donner naissance à une IA générale *« au bénéfice de l'ensemble de l'humanité »*.

> Ses failles sont nombreuses qui nécessitent une réflexion sur le statut de cette ia et, surtout, sur ceux qui la programment et l'utilisent

À l'origine, OpenAI est donc une fondation à but non lucratif qui mise sur l'open source. Parmi ses fondateurs, on trouve Elon Musk et Peter Thiel, déjà associés à l'époque de PayPal et libertariens convaincus, Greg Brockman, alors directeur technique de Stripe et désormais président d'OpenAI, ou encore Sam Altman, à la tête de l'incubateur de start-ups Y Combinator, désormais PDG d'OpenAI. Trois ans plus tard, en 2018, Elon Musk se retire sans que les raisons de ce départ soient réellement connues. Au même moment, les coûts d'OpenAI s'envolent du fait des besoins en puissance de calcul pour entraîner son IA. Sam Altman change les statuts d'OpenAI qui devient, en 2019, une société à but non lucratif, ce qui permet d'accueillir de nouveaux investisseurs. Microsoft entre à son capital et apporte d'un coup 1 milliard de dollars et un partenariat avec Azure. Puis 2 milliards de dollars en 2021. Microsoft propose désormais 10 milliards de plus pour renforcer encore l'association approfondie avec OpenAI. Sam Altman, pour OpenAI, et Satya Nadella, le PDG de Microsoft, ont donc une certaine autorité sur la société qui héberge ChatGPT. Ils décideront à l'avenir du futur commercial de ChatGPT, comme de l'éthique que celle-ci devra suivre.

Comment comptes-tu révolutionner la recherche en ligne et l'internet en général ?

ChatGPT relance la compétition sur le marché de la recherche en ligne. Intégré à Bing depuis mars 2023 en version béta, puis ouvert à tous le 4 mai 2023, le service Bing Chat permet à Microsoft de menacer Google sur son cœur d'activité historique, l'établissement de listes d'adresses URL et la facturation de liens sponsorisés associés à une requête en particulier. Google a, en effet, été à l'origine d'une révolution de l'internet en imposant ses propres critères d'autorité aux sites web : une adresse URL et le contenu associé, pour une requête donnée, est d'autant plus pertinente qu'elle est souvent citée et souvent cliquée.

> Chatgpt relance la compétition sur le marché de la recherche en ligne

Il s'agit là du *page rank* de Google que ChatGPT ignore. Plus besoin d'une liste exhaustive des sources disponibles sur internet pour un sujet donné, plus besoin d'un classement pertinent de ces sources réalisé grâce à une puissante statistique sur des milliards de requêtes (Google en reçoit chaque jour 9 milliards), mais une réponse unique à une question unique. Autant dire que ChatGPT révolutionne la recherche en ligne parce qu'elle l'émancipe du référencement et, en tant que chatbot, autorise des réponses hors écran : si Microsoft a raté le rendez-vous avec l'internet des smartphones, il ne compte pas manquer l'internet de demain où les interfaces seront de moins en moins les écrans et de plus en plus la voix et les gestes humains. Or, en basculant dans l'oralité, l'internet exigera des réponses uniques. Les récentes annonces de Microsoft ne sont donc que les prémices d'une probable révolution des outils de recherche en ligne, comme des outils de productivité.

Après Copilot sur GitHub, la première annonce de Microsoft a en effet concerné Bing, son moteur de recherche. Bing Chat propose encore des adresses URL mais également, dans un encadré à droite de l'écran, un texte généré automatiquement. Il faudra en effet que les internautes utilisent les moteurs de recherche différemment, notamment en posant des questions plutôt qu'en listant des mots-clés, pour que les IA génératives puissent être parfaitement exploitées. Microsoft entend signer par cette annonce la fin de la recherche classique incarnée par Google depuis le tout début des années 2000 et qui ne s'est pas vraiment renouvelée, Google proposant au mieux des réponses directes à des requêtes simples avec son *knowledge graph*, sa base de données de faits, lieux et dates. Reste que pour l'instant Google Search est à peu près fiable, à l'inverse des IA génératives qui ont ce défaut de forger des réponses quoi qu'il arrive, même quand elles n'en disposent pas. C'est d'ailleurs la raison pour laquelle Google s'est jusqu'ici refusé à mettre à la disposition du grand public son propre système d'IA générative, LaMDA (*langage model for dialogue application*).

Sous la pression de Microsoft, Google a toutefois présenté un chatbot à son tour, baptisé Bard, le 8 février 2023, le lendemain de l'annonce de Microsoft. Le jour de cette présentation, faite sur YouTube depuis Paris, Bard a fourni des réponses erronées. Et il n'est pas possible de reprocher à Google son manque d'expertise : Bard comme ChatGPT sont la conséquence de recherches partagées dès 2017 par les ingénieurs de Google qui ont publié l'un des articles fondateurs consacrés au *deep learning*, moment à partir duquel l'IA va faire d'immenses progrès. Meta avait été confronté à la même déconvenue en octobre 2022 lorsqu'il avait présenté son modèle de langage Galactica qui s'est vite mis à raconter n'importe quoi et fut retiré au bout de trois jours. Microsoft l'a été à son tour avec Bing Chat. Le 17 février 2023, le groupe a en effet annoncé qu'il allait limiter à 50 questions par jour, dont 5 par session, le nombre d'interactions avec Bing Chat car, au-delà d'un certain temps d'échange, se produit un phénomène d'hallucination de l'IA. Concrètement, l'intelligence générative répond n'importe quoi. Des journalistes testeurs sont en effet parvenus à pousser Bing Chat à se déclarer amoureux…

Microsoft a également multiplié les annonces en lien avec l'intégration de ChatGPT dans ses logiciels destinés aux professionnels. L'outil va être intégré à Teams pour proposer automatiquement des comptes rendus de réunion en visioconférence. Il va être intégré à la suite bureautique Office pour proposer des

réponses à des e-mails (Outlook) ou synthétiser des données chiffrées (Excel) ou générer automatiquement des présentations (PowerPoint). Cette offre sera aussi appelée Copilot, comme sur GitHub, mais dédiée à l'augmentation de la productivité des utilisateurs d'Office. Après la recherche en ligne, c'est donc aussi au monde ordinaire du travail salarié que s'attelle Microsoft, celui qui repose sur la production de tâches intellectuelles plutôt standardisées. Et Microsoft espère le révolutionner. Goldman Sachs a ainsi estimé à 300 millions les emplois qui pourraient être supprimés par ChatGPT. En effet, l'IA générative a ceci de particulier qu'elle peut prendre en charge des tâches intellectuelles basiques pour lesquelles les humains, mal formés ou mal concentrés, sont bien moins efficaces. D'ailleurs, Goldman Sachs estime que ce sont les fonctions administratives et support qui seront les plus concernées par cette concurrence des robots. Les premiers signes de cette évolution se font jour : en mai 2023, IBM a gelé les embauches sur les postes où l'humain peut être remplacé par des IA.

> L'ia générative questionne la réalité de cette intelligence collective humaine que l'internet aurait fait émerger

Mais c'est là considérer l'IA générative comme un *alter ego* de l'humain. Or sa fonction – et c'est le cas avec Copilot sur GitHub – est d'abord d'utiliser le potentiel de l'IA pour augmenter l'humain, d'où le risque anthropomorphique si on confond le moyen avec la fin et le risque aussi de substitution de l'humain par l'IA. Au moins ces outils rappellent-ils une chose : la force de l'intelligence humaine, celle qui sous-tend les métiers de l'« information », repose sur la réflexivité et la créativité. Le même argument est valable aussi pour les IA génératives d'images : elles ne remplacent les graphistes que lorsqu'il s'agit de faire de la mise en image ou de la mise en page sans prétention. Se pose alors de nouveau, mais autrement, la question de l'autorité. L'IA pourrait bien permettre de discriminer, entre des statuts irremplaçables et d'autres substituables, ceux pour lesquels l'apport humain est souvent faible comparé à la machine. L'IA générative questionne la réalité de cette intelligence collective humaine que l'internet aurait fait émerger et dont Wikipédia serait le parangon. Cette intelligence-là est celle que les calculs de l'IA générative ne peuvent pas remplacer.

Doit-on se méfier de toi ? Ou de ceux qui te contrôlent ?

Pour ceux qui ont un regard plutôt pessimiste sur l'intelligence humaine, l'IA générative est perçue comme menaçante car elle peut se substituer à l'homme pour une multitude de tâches à faible valeur ajoutée, ces tâches ayant toutefois d'autres fonctions que la seule production d'information, par exemple des fonctions de communication, d'accompagnement, de sociabilité. En s'y substituant, l'IA générative pourrait supprimer ces fonctions associées, les modalités de sociabilité qu'elles rendent possibles. Le 22 mars 2023, une semaine après le lancement de ChatGPT 4, une lettre ouverte publiée sur le site de l'institut Future of life, et signée par plus de mille experts et personnalités, dont Elon Musk et Steve Wozniak, demandait un moratoire de six mois sur toutes les recherches liées à l'IA, le temps d'en imaginer la régulation dans l'intérêt de

l'humanité.

Plusieurs problèmes sont soulignés en lien avec la possibilité, pour l'IA, de devenir véritablement compétitive face à l'humain. À cet égard, le regard anthropomorphique sur l'IA domine puisque la lettre parle d'*« esprits non-humains »* qui pourraient déborder l'humanité en nombre et en puissance. Ces décisions-là ne sauraient être laissées aux seuls entrepreneurs de l'IA, d'où la demande d'une régulation par les autorités publiques qui assignerait aux systèmes d'IA des objectifs légitimes : la lettre parle d'IA *« précise, sûre, interprétable, transparente, solide, harmonisée, digne de confiance et loyale »* (traduction de *« accurate, safe, interpretable, transparent, robust, aligned, trustworthy, and loyal »*).

> Un moratoire de six mois sur toutes les recherches liées à l'ia, le temps d'en imaginer la régulation dans l'intérêt de l'humanité

Ici, la question de l'autorité est centrale. Qui contrôle l'IA ? Qui a autorité sur son développement ? L'ingénieur ou le politique ? Les signataires de la lettre optent pour le politique parce qu'ils considèrent que l'intelligence artificielle, aussi mathématique soit-elle, n'est pas neutre. C'est ce que soulignait déjà Dominique Cardon quand il s'interrogeait, dès 2015, sur la transparence des algorithmes : *« S'il n'est guère possible d'enquêter sur les variables versatiles des algorithmes, il est en revanche décisif de demander à ceux qui les fabriquent de rendre publics les objectifs qu'ils leur donnent. »* Les algorithmes – l'IA générative est une sorte de super-algorithme – viennent répondre à des besoins et sont conçus ainsi. S'il n'est pas possible de comprendre comment ils calculent, surtout lorsqu'ils s'autonomisent avec le *deep learning*, il est souvent possible de savoir pourquoi ils calculent, dans quel but, même pour les IA générales. Ainsi, ChatGPT vise des réponses pertinentes à des questions humaines : il a donc un objectif qui relève, non de l'IA, mais d'une définition de la « pertinence pour l'humain » qui repose sur un jugement de valeur. D'ailleurs, lors de la présentation de ChatGPT 4, la moindre récurrence des propos offensants a été soulignée, par comparaison avec ChatGPT 3.5. Les propos n'étant offensants que pour ceux qu'ils offensent, ChatGPT prend parfaitement en considération l'objectif qui est le sien, celui d'être un chatbot performant parce qu'il doit fidéliser ses utilisateurs, donc ne pas les heurter.

Les ingénieurs qui ont développé ChatGPT lui ont donc inculqué des principes bien assurément humains, même si l'IA ne les comprend pas. Outre son entraînement sur de larges jeux de données, qui n'ont pas été vérifiés (la production web jusqu'en 2021 où l'on trouve tout et n'importe quoi), ChatGPT a en effet été corrigé par des humains jusqu'à parvenir à des propos cohérents et acceptables. C'est ce qui explique le ton plutôt neutre de ChatGPT et ses réponses assez approximatives sur des questions qui prêtent à controverse, alors même qu'on attendrait d'une personne une sorte d'engagement. C'est que ChatGPT s'est entraîné sur des jeux de questions/réponses qu'on lui a présentés comme corrects parce qu'établis par des humains. L'IA générative a ensuite produit ses propres réponses qui ont été corrigées par des humains. Ce retour humain, ce *feedback* a consisté à classer par ordre de pertinence les différentes réponses de l'IA à une question. À

force de *feedback*, ChatGPT a mieux appris à décider quelle réponse apporter en priorité à quelle question. Et le résultat est sans appel.

Nous avons eu l'occasion, en février 2023, dans le cadre d'un dispositif pédagogique, de réaliser une expérimentation avec une promotion de master 1 dans un enseignement sur les théories des sciences de l'information et de la communication. Les étudiants devaient choisir un auteur et s'interroger sur l'une de ses notions clés. Ils formulaient ainsi une question précise en rapport avec l'approche de l'auteur. Cette question a été posée à ChatGPT en français et en anglais (selon *Le Figaro* du 13 mars 2023, ChatGPT s'est entraîné sur des textes anglais pour 46 % de son corpus et sur des textes français pour moins de 5 % de son corpus). Très vite, nous avons constaté qu'il fallait poser la même question dans deux langues différentes à partir de deux sessions distinctes pour éviter que les réponses en français et en anglais ne soient trop proches. En effet, si un texte littéraire existe tout à la fois en anglais et en français pour un même auteur, ChatGPT privilégiera par défaut la langue dans laquelle la question est posée, mais il tient compte de la réponse précédente dans une autre langue si la même question est posée dans la même session. On risque alors d'avoir une presque-traduction de la réponse précédente, ChatGPT faisant en sorte de ne jamais répondre de la même manière à la même question. En séparant les sessions, la réponse en français repose plutôt sur des textes français et la réponse en anglais plutôt sur des textes anglais, ce que nous avons déduit des résultats qui nous ont été proposés.

La nécessité de savoir quels jeux de données sont utilisés pour entraîner les intelligences artificielles

L'objectif pédagogique était de mettre en lumière la construction culturelle des réponses de ChatGPT. Dans notre exercice, ChatGPT nous donne un bon aperçu de la nature de la pensée anglo-saxonne, du moment et de l'importance qu'elle accorde à l'idée de dispositifs de domination non conscients. En voici un exemple. Les étudiants ont posé la question suivante, formulée dans leurs termes, à ChatGPT : *« Quel est le rapport entre l'approche de l'ethnoscape d'Arjun Appadurai et l'appropriation culturelle ? »* La réponse en français insiste sur les échanges entre cultures, les phénomènes d'appropriation culturelle et l'intérêt de la diversité, même si elle finit par souligner – ce qui n'est pas le propos d'Appadurai – que *« l'appropriation culturelle peut également poser des problèmes lorsque les éléments culturels sont utilisés sans respect pour leur contexte d'origine ou sans la permission de la communauté culturelle concernée »*. La réponse en anglais est bien plus directe et c'est probablement elle qui a influencé la conclusion moralisatrice de la réponse française, en lien donc avec le corpus de textes sur lequel ChatGPT a été entraîné. En effet, elle repose dans son ensemble sur une critique de l'appropriation culturelle, dont les limites sont dénoncées dès le début de la réponse, à rebours de la perspective positive d'Arjun Appadurai quand il s'intéresse à la diversité culturelle et au rôle de l'imagination contre la globalisation uniformisante : *« Ethnoscape highlights the movement of people and cultures across national borders, leading to the mixing and altering of cultures. This cultural movement often results in the creation of new cultural forms resulting from the interaction between different cultures, thus increasing cultural diversity. However, this can also lead to cultural conflicts, particularly when dominant cultures adopt cultural elements belonging to marginalized cultures without recognition or*

respect for their origin. » La réponse en anglais se termine par un bréviaire moralisateur qui n'a rien à voir avec la nature des propos d'Arjun Appadurai mais qui nous plonge dans le langage du politiquement correct de nombreux Américains : *« Therefore, understanding the relationship between ethnoscape and cultural appropriation requires recognizing the importance of power dynamics and structural inequalities that shape cultural interactions. It also involves acknowledging the need for cultural sensitivity, respect, and dialogue when engaging with different cultural forms and expressions. By recognizing and addressing these issues, we can work towards creating more just and equitable societies that celebrate cultural diversity and promote cultural exchange and dialogue while avoiding harmful cultural appropriation. »*

Cet exemple est significatif. Sur un sujet délicat, en lien avec l'idée d'identité culturelle, dont Arjun Appadurai n'est pas le meilleur des défenseurs puisqu'il s'oppose à toute forme d'essentialisation, ChatGPT propose une approche qui trahit le raisonnement de l'auteur pour dire comment penser tout en faisant la part belle au relativisme. Ce « comment penser » lui a été inculqué et c'est en quelque sorte la signature de ChatGPT (toutes les réponses, sur tous les auteurs concernés, finissent par ce petit bréviaire).

Avec les ia génératives, les sources disparaissent

D'où la nécessité de demander à ses promoteurs d'expliciter les choix qui ont été les leurs dans la programmation de l'IA. D'où la nécessité aussi de savoir quels jeux de données sont utilisés pour entraîner les intelligences artificielles. Car, si ChatGPT dit comment penser, il peut aussi induire en erreur : programmé pour ne pas donner crédit aux *fake news*, si celles-ci sont trop nombreuses dans les textes à partir desquels il a été entraîné, certaines finiront par passer à travers les mailles du filet. Enfin, se pose la question, au-delà des failles des IA génératives elles-mêmes, des risques liés à leur utilisation par des individus malveillants. Ce sont toutes ces raisons qui conduisent les signataires de la lettre à demander une régulation de toute urgence et un moratoire. Ils ne sont pas les seuls. Le 4 mai 2023, Kamala Harris, la vice-présidente américaine, a invité les responsables des groupes américains les plus en pointe sur l'IA pour discuter de ses dangers. Quant à l'Union européenne, si elle travaille sur une régulation de l'intelligence artificielle depuis 2021 avec l'IA Act, elle n'a pas encore tranché la question du statut des IA génératives et de leur insertion, ou non, dans la liste des IA considérées comme à haut risque.

Il sera, quoi qu'il arrive, très difficile de décider ce qu'est une IA *« accurate, safe, interpretable, transparent, robust, aligned, trustworthy, and loyal »* comme le demandent les signataires du courrier et d'imposer cette vision aux ingénieurs de l'IA. Quand les médias de masse étaient suspectés, les chercheurs ne sont jamais parvenus à définir précisément ce à quoi pouvait correspondre la responsabilité sociale des médias et des journalistes. Il n'y a pas de raison qu'on y parvienne mieux aujourd'hui, car ces critères sont socialement, politiquement, culturellement déterminés. Ils sont à l'image de ceux qui les défendent collectivement, ce que montre le test réalisé avec nos étudiants. Mais une chose a changé. Quand les médias étaient suspectés d'influencer l'opinion publique, il était possible de discuter l'information qu'ils

produisaient parce que celle-ci, établie par des journalistes, devait s'appuyer sur des sources, une méthode d'établissement des faits : autant d'exigences qu'il était possible de vérifier et d'opposer aux professionnels de l'information. Avec les IA génératives, les sources disparaissent : trop nombreuses, trop retravaillées, elles ne sont plus citées. Le risque de relativisme est donc beaucoup plus marqué parce qu'il relèverait ainsi d'une sorte de déterminisme technologique, de l'acceptation du renoncement au contrôle dans la fabrication des énoncés des IA génératives. Il faut alors s'en remettre à la seule autorité de l'algorithme : il faut le croire parce que, dans la durée, force est de constater la cohérence de la plupart de ses réponses, même s'ils s'entraînent sur des corpus où se logent aussi des *fake news*, des affirmations fausses, des contre-vérités. Il faut donc accepter de ne pas pouvoir facilement identifier les erreurs ou jugements que l'intelligence artificielle glisse en fait dans ses réponses en apparence neutres, surtout sur des sujets complexes ou controversés. À l'évidence, ChatGPT est autant sophiste qu'ingénieur, mais probablement pas philosophe. De ce point de vue, la question de l'auteur, soulevée dès les premiers ouvrages publiés avec l'aide de ChatGPT, est tout à fait intéressante. Si l'auteur est responsable de ses écrits, ChatGPT peut-il l'être ?

Es-tu un auteur à part entière ?

Très vite, le potentiel de ChatGPT a été exploité. Dès le mois de février 2023, soit à peine plus de deux mois après sa mise à disposition, des livres écrits avec l'aide de ChatGPT étaient mis en ligne sur Amazon. Ces livres sont cosignés par l'auteur qui donne ses consignes à ChatGPT et par le ou les services d'IA. En effet, en plus de ChatGPT, les couvertures des livres sont aussi générées par des outils issus de l'intelligence artificielle comme Midjourney ou encore DALL-E, ce dernier appartenant à OpenAI. Ces livres ne sont pas des chefs-d'œuvre, mais ils existent et rencontrent un public qui s'en satisfait parce qu'ils relèvent du divertissement. Le potentiel commercial de ChatGPT dans les industries culturelles, dans la publicité, dans l'*entertainment* est donc bien réel. Or, ces secteurs ont en commun la particularité de revendiquer leur dimension créative et donc de faire appel à des auteurs humains. L'auteur est le créateur de son œuvre. L'est-on encore quand on s'en remet en partie à ChatGPT ?

> Le risque de relativisme est donc beaucoup plus marqué

Paradoxalement, cette invention venue des États-Unis pourrait bien donner au droit d'auteur européen, à l'inverse du copyright anglo-saxon, et contre toute attente, une nouvelle légitimité. Outre les droits commerciaux attachés au droit d'auteur et aux droits voisins, le droit d'auteur européen consacre aussi un droit moral de l'auteur sur son œuvre qui sanctuarise sa création. Il faut d'ailleurs en faire la preuve : dans la politique française de quotas audiovisuels, il a fallu inventer la notion d'« œuvre patrimoniale » pour distinguer entre les œuvres qui font l'objet d'un travail de création par un scénariste et un réalisateur et les productions sur mesure, sans originalité, bas de gamme. Dès lors, n'est pas substituable ce qui relève d'une forme d'engagement, d'où émerge l'idée d'auteur, celui qui assume ses propos, ses textes, ses images. Cet engagement moral est ce qui distingue à l'évidence le travail d'un humain sur son œuvre et la productivité

des IA génératives. Après le succès de la post-vérité et l'oubli des faits dans un contexte de relativisme informationnel favorisé par les réseaux sociaux, l'IA générative sera peut-être à l'origine d'une époque de post-auctorialité et d'oubli des engagements individuels. Ces derniers ressurgiront quand même, moulinés, statistiquement rétablis sous la forme du sens commun le plus normé et le plus acceptable sur le moment. Les intelligences individuelles, celles qui parviennent à transgresser règles et normes pour penser autrement, ne devraient pas être menacées par l'IA mais bien plutôt par le conformisme intellectuel que l'IA risque de renforcer ou de radicaliser.

Pour l'instant, la régulation de l'IA est presque inexistante et laisse ces questions en suspens. Les premiers débats sur le statut des productions de l'IA portent sur la question des données personnelles et sur la question des droits voisins, posant autrement la question de l'autorité sur les textes ainsi produits. En ce qui concerne les données personnelles, celles donc associées à un être humain, la réglementation s'applique, notamment le RGPD (règlement général sur la protection des données) dans l'Union européenne. Le 31 mars 2023, l'Italie a même demandé à ChatGPT de suspendre son service faute de donner à ses utilisateurs toutes les informations nécessaires sur les moyens de protéger leurs données personnelles d'une exploitation par l'IA. ChatGPT a rendu publiques ses règles dans la foulée, et le service est rouvert en Italie depuis le 28 avril 2023. Mais, au-delà du cas italien, la question plus fondamentale est celle de l'accès aux conversations humaines qui relèvent de propos personnels, même si certains sont accessibles dans l'espace public en ligne. Pour imiter les réponses humaines, l'IA a besoin de réelles conversations et de vraies réactions : nos réactions aux photos, notre interactivité sur Twitter, sur Reddit, sur LinkedIn, etc. Or, les services qui hébergent ces interactions humaines ont des devoirs de modération ou de signalement, même s'ils n'ont que rarement le statut d'éditeur. La question se posera donc pour les IA génératives d'obtenir ou non l'accord préalable des utilisateurs de leurs services pour s'entraîner, et d'obtenir également l'accord préalable des services eux-mêmes pour exploiter les échanges qu'ils hébergent.

> L'ia générative sera peut-être à l'origine d'une époque de post-auctorialité et d'oubli des engagements individuels

Sur ce dernier point, les enjeux sont en apparence plus simples, car ils relèvent du droit voisin, donc de l'exploitation économique indirecte par les services d'IA de stocks de données contrôlés par des tiers. Twitter a ainsi menacé ChatGPT de poursuites par l'intermédiaire d'Elon Musk. Reddit envisage de rendre payant l'accès à ses données pour toutes les utilisations massives dont les IA ont besoin. Les éditeurs de presse envisagent eux aussi de faire valoir leurs droits voisins. Des plaintes ont par ailleurs été déposées contre Microsoft et OpenAI qui reprochent au service Copilot de plagier potentiellement le code sur lequel il s'est formé. Dans certains cas, les services d'IA acceptent le principe d'une rémunération : OpenAI paie ainsi la banque d'images Shutterstock pour que DALL-E accède à son stock et l'exploite pour générer de nouvelles images. La contrainte commerciale pourrait bien ici jouer un rôle très problématique de régulation par le marché. Parce que les IA sont à l'image des stocks de données sur lesquels elles s'entraînent, si ces

stocks se raréfient, alors les IA risquent bien de s'entraîner de plus en plus sur les dernières données en libre accès qui ne seront pas nécessairement les meilleures. De ce point de vue, la question de l'accès aux données par les IA renouvelle les enjeux qui ont été posés par la recherche en ligne, entre défenseurs d'un internet ouvert et promoteurs d'écosystèmes fermés, entre la recherche générique du premier Google Search et les recommandations personnalisées dans des univers plus fermés comme ceux d'Instagram ou de TikTok.

Sources :

- Dominique Cardon, *À quoi rêvent les algorithmes ?*, Le Seuil, 2015, p. 61.
- Hortense Goulard, « ChatGPT marque un tournant pour l'intelligence artificielle », *Les Échos*, 13 décembre 2022.
- Samir Touzani, Marina Alcaraz, « Microsoft va intégrer ChatGPT à Teams », *Les Échos*, 3 janvier 2023.
- Leïla Marchand, « ChatGPT versus Google : la guerre des moteurs de recherche est lancée », *Les Échos*, 9 janvier 2023.
- Caroline Beyer, Paule Gonzalez, Stéphane Kovacs, Jean-Marc Leclerc, « ChatGPT : cette intelligence artificielle qui fascine et inquiète », *Le Figaro*, 25 janvier 2023.
- Florian Dèbes, « Google en embuscade sur l'IA générative », *Les Échos*, 6 février 2023.
- Chloé Woitier, « Microsoft et Google sifflent le départ de la grande bataille de l'IA », *Le Figaro*, 8 février 2023.
- Raphaël Balenieri, Florian Dèbes, Leïla Marchand, Marina Alcaraz, « L'écosystème de la recherche en ligne se prépare à la révolution ChatGPT », *Les Échos*, 9 février 2023.
- Ingrid Vergara, « Le cloud, une arme majeure dans la grande bataille de l'IA », *Le Figaro*, 9 février 2023.
- Chloé Woitier, « Microsoft met Google en difficulté dans la nouvelle bataille des moteurs de recherche », *Le Figaro*, 16 février 2023.
- Hortense Goulard, « Microsoft veut reprendre le contrôle de l'IA du nouveau Bing », *Les Échos*, 21 février 2023.
- Claudia Cohen, « Sur Amazon, l'inquiétant business des livres écrits par ChatGPT », *Le Figaro*, 23 février 2023.
- Hortense Goulard, « La propriété intellectuelle, un défi majeur pour le développement de l'IA », *Les Échos*, 1er mars 2023.
- Alice Develey, « ChatGPT est-il écrivain ? », *Le Figaro*, 13 mars 2023.
- Hortense Goulard, « Avec GPT-4, OpenAI veut creuser l'écart dans la révolution de l'IA », *Les Échos*, 16 mars 2023.
- Marina Alcaraz, « Microsoft intègre l'IA à ses logiciels de bureautique », *Les Échos*, 20 mars 2023.
- « Pause Giant AI Experiments : on open letter », publiée en ligne le 22 mars 2023 : https://futureoflife.org/open-letter/pause-giant-ai-experiments
- NeIla Beyer, « ChatGPT et l'IA menacent 300 millions d'emplois dans le monde », *Les Échos*, 29 mars 2023.

- Marina Alcaraz, « Intelligence artificielle : le cri d'alarme d'Elon Musk et des experts de la tech », *Les Échos*, 30 mars 2023.
- Ingrid Veragara, « L'intelligence artificielle va-t-elle supprimer 300 millions d'emplois », *Le Figaro* , 2 mai 2023.
- Chloé Woitier, « IBM gèle les embauches sur des postes qu'il estime remplaçables par les IA », *Le Figaro*, 3 mai 2023.
- Chloé Woitier, « La Maison-Blanche convoque les géants de l'IA pour les appeler à la responsabilité », *Le Figaro*, 4 mai 2023.
- Florian Dèbes, « Après le succès fulgurant de ChatGPT, des sites web réclament leur part », *Les Échos*, 5 mai 2023.
- Chloé Woitier, « L'Europe se lance en pionnière dans la régulation de l'intelligence artificielle », *Le Figaro*, 9 mai 2023.

Categorie

1. Articles & chroniques

date créée
30 août 2023
Auteur
alexandrejoux