ÉTVDES
REVUE DE CULTURE CONTEMPORAINE

Parler avec les machines

Entretien avec Alexei Grinbaum, Propos recueillis par François Euvé, Nathalie Sarthou-Lajus

Dans Études 2023/10 (Septembre), pages 55 à 66
Éditions S.E.R.

ISSN 0014-1941
DOI 10.3917/etu.4307.0055

Article disponible en ligne à l'adresse
https://www.cairn.info/revue-etudes-2023-10-page-55.htm

CAIRN.INFO
MATIÈRES À RÉFLEXION

Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...
Flashez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.

Distribution électronique Cairn.info pour S.E.R..
La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

© S.E.R. | Téléchargé le 05/09/2023 sur www.cairn.info par Robert Mathieu (IP: 176.181.145.57)

PARLER AVEC LES MACHINES

Entretien avec Alexei GRINBAUM

L'émergence de l'application ChatGPT amène à se poser une nouvelle fois la question de la distinction entre le dialogue inter-humain et la communication avec une machine. Les réseaux de neurones artificiels ont fait franchir un seuil irréversible, dont il importe de comprendre les conséquences. Un nouveau mode de relation avec ces machines s'instaure, qui n'est pas sans analogie avec ce que nous racontaient les mythes.

Pourriez-vous commencer par expliquer les grandes étapes de l'histoire de l'intelligence artificielle et des « machines parlantes » ?

■ **Alexei Grinbaum :** Le rêve de faire des machines qui nous parlent en langue naturelle est très ancien. La première réalisation technique date de 1965. À Boston, au MIT (Massachusetts Institute of Technology) un professeur, Joseph Weizenbaum (1923-2008), a construit une machine très simple de notre point de vue actuel, qu'il a baptisée « Eliza ». Elle ne pouvait faire qu'une seule chose : prendre une phrase et la retourner en question. Weizenbaum lui-même n'attendait pas grand-chose de sa machine mais, à sa surprise, il s'est rendu compte qu'elle produisait un véritable effet sur les gens, ce qu'on appelle depuis l'« effet Eliza ». Elle fonctionnait un peu comme un psychiatre dit rogérien, c'est-à-dire quelqu'un qui prétend ne rien savoir. C'est la force du langage : la machine parlante aussi simple qu'elle soit exerce déjà un effet sur l'utilisateur, même si celui-ci sait bel et bien que c'est une machine. Avec la nouvelle révolution qui date de 2017, les machines sont infiniment plus élaborées qu'Eliza et l'effet est grandement amplifié.

Les deux principaux courants historiques de l'intelligence artificielle sont les réseaux de neurones artificiels et les systèmes symboliques. Ces derniers s'appuient sur une idée philosophique qui remonte à Thomas Hobbes (1588-1679), selon laquelle la pensée n'est qu'un calcul selon un ensemble de règles (« *thinking is reckoning* »). Pour créer une machine qui raisonne comme un être humain, il faudrait donc des règles, de la logique, de la syntaxe, de la sémantique… Cela a conduit à l'intelligence artificielle symbolique, notamment aux systèmes experts.

Par ailleurs, arrivent les réseaux de neurones, imaginés d'abord dans les années 1940 par la cybernétique (le perceptron). L'idée provient de la physique statistique. Un neurone artificiel est une fonction de type seuil : jusqu'à une certaine force de signal, il ne se passe rien en sortie ; au-delà, il y a un signal linéaire. L'idée est que, si l'on regroupe un grand nombre de neurones artificiels qui interagissent de manière aussi simple, l'ensemble sera suffisamment complexe pour se comporter de façon imprévisible et surprenante. Or, au début, les réalisations techniques de cette idée ne fonctionnent pas bien. Le décollage se fait attendre. C'est seulement vers la fin des années 2000 que les réseaux de neurones deviennent beaucoup plus grands – avec des milliards, voire des centaines de milliards d'unités – et la puissance de calcul augmente énormément. De plus, on applique des architectures nouvelles en reliant plusieurs couches interconnectées dans ce qu'on appelle « l'apprentissage profond ». Les résultats sont très bons pour la reconnaissance d'images, mais pas encore pour le langage. Dans ce domaine, la révolution viendra dix ans plus tard depuis une direction assez inattendue, les systèmes qui complètent les phrases. Tout le monde sait que, quand on commence à taper une phrase dans Google, le système propose de compléter automatiquement un ou deux mots à la fin. Mais, s'il complète un mot, il pourrait aussi écrire un paragraphe ou une page. C'était l'idée.

La révolution dite des *transformers* date de 2017. Le principe consiste en deux étapes : d'une part, la machine « joue à cache-cache » avec elle-même, c'est-à-dire qu'elle soustrait un mot et essaye de le deviner. Par exemple, elle se cache le mot « lion » et elle va le deviner de manière probabiliste, estimant qu'avec 60 % de probabilité, c'est « lion », avec 40 %, c'est « tigre », avec 10 %, c'est « panthère »… Puis elle se montre le mot et met à jour ses paramètres, en faisant cet exercice des trillions de fois. Cela s'appelle un « apprentissage par autosupervision ». Mais l'autre étape est essentielle. Au lieu de mots, la machine casse le langage en des morceaux plus petits, qu'on appelle

des « *tokens* ». Humainement, la plupart des *tokens* n'ont aucun sens : ce sont juste des assemblages de deux, trois ou quatre lettres, et même parfois deux lettres d'un mot collées aux deux lettres du mot suivant. Par exemple, après un « q », il y a toujours un « u » : donc « qu » forme un *token*. Le mécanisme ne regarde pas la séquence des mots mais, quand il se soustrait un *token*, il essaie de le deviner en regardant tous les autres *tokens*, et donc tout le contexte : cinq lignes plus haut, dix lignes plus bas… L'interaction de chaque *token* avec tous les *token*s est purement numérique. Qui plus est, elle n'est pas linéaire. Là se trouve l'idée révolutionnaire qui a enfin permis d'atteindre la maîtrise de la langue par les réseaux de neurones.

> **« Ce qui est intéressant dans cette science, c'est qu'on ne sait pas pourquoi cela marche »**

Ce qui est intéressant dans cette science, c'est qu'on ne sait pas pourquoi cela marche. On augmentait la taille des réseaux de neurones et, à un moment donné, les sorties sont devenues quasi parfaites. Mais pourquoi ? On est certain que ce phénomène relève de la physique statistique. Entre les modèles avec des milliards (10^9) et des centaines de milliards de paramètres (10^{11}), il se produit une transition de phase. Pourtant, on ne sait pas d'où elle vient et il n'existe pas de science prédictive derrière cette constatation purement empirique. Le système GPT-2 avait trois milliards de paramètres et GPT-3 atteignait 175 milliards de paramètres : entre les deux, le saut dans la qualité des résultats a été merveilleux. Par ailleurs, ces nombres sont vraiment très grands. Par comparaison, pour reconnaître un visage par un réseau de neurones, vous avez besoin d'une centaine de paramètres. Et pour que cela marche bien avec le langage, il en faut des centaines de milliards. Personne ne sait pourquoi…

L'intelligence artificielle cherche-t-elle à mimer l'intelligence humaine ?

■ **Al. Grinbaum :** Non, et c'est justement cela qui est intéressant. À chaque fois qu'émerge une nouvelle génération de machines qui nous fascine, on commence à faire une analogie avec le fonctionnement du cerveau humain. Mais, dans ce dernier, les neurones naturels n'ont rien à voir avec les neurones artificiels. La topologie du cerveau n'est pas du tout la même, la consommation énergétique d'un cerveau n'a rien à voir avec la consommation énergétique d'un réseau de neurones.

Le rapprochement vient de ce que les entrées et les sorties (les phrases) sont les mêmes qu'en langue naturelle. Mais le chemin qu'emprunte la machine est, lui, différent. C'est une découverte fascinante : il existe une façon différente, non humaine, de relier les mêmes entrées avec les mêmes sorties !

Pour illustrer l'ampleur du non-humain dans les systèmes d'intelligence artificielle générative, prenons l'exemple des « baleines numériques », une expérience menée en 2022 à l'aide du système DALL•E. La consigne était : « Dessine-moi deux baleines en train de discuter de nourriture, avec des sous-titres. » Les sous-titres ressemblaient à une suite de lettres de l'alphabet, mais sans signification pour nous. On donne une seconde consigne : « Visualise-moi cette suite de lettres. » La machine dessine alors des crevettes, des fruits de mer… Cela veut dire que, dans l'espace vectoriel mathématique où opère ce système, il y a des vecteurs de très haute dimension qui, dans une certaine métrique, sont proches des vecteurs qui encodent les crevettes ou les fruits de mer. Dans l'espace vectoriel de cette machine, il existe une place pour ce que nous pourrions appeler « le langage des baleines numériques », comme si les baleines numériques avaient une langue à elles. Et ainsi de suite pour toutes les autres « créatures numériques ».

> *« À partir du moment où des agents non humains parlent notre langue, notre condition humaine change »*

La parole est le propre de l'homme. Est-ce qu'on peut parler d'une « parole des machines » ?

■ **Al. Grinbaum :** La parole est – ou a été – le principal déterminant de la condition humaine. Hannah Arendt (1906-1975), que je cite dans le livre[1], dit que « de tout ce que nous sommes, nous en faisons sens dans et à travers la parole ». À partir du moment où il y a des agents non humains qui parlent notre langue, notre condition humaine change. Cette situation est technologiquement inédite, mais des agents non humains qui parlent notre langue, cela n'est pas nouveau. Dans les mythes, on a déjà rencontré des anges, des démons, des oracles et des dieux qui parlent notre langue sans être humains. Donc, si on se demande ce qu'un dialogue avec un être non humain nous fait, à nous, il faut regar-

1. Al. Grinbaum, *Parole de machines. Dialoguer avec une IA*, Humensciences, 2023.

der dans les mythes. Il y a des leçons à tirer de ces récits décrivant des entités mythologiques qui parlent notre langue, pour essayer de comprendre les changements qui nous arrivent avec la parole des machines.

Je prends l'exemple du rôle du nom. On donne souvent des noms aux machines avec lesquelles on parle, même quand il ne s'agit pas de *chatbots* (agents conversationnels). Quel est l'intérêt du nom pour penser le statut des machines parlantes ? L'histoire mythologique commence par un épisode classique : Adam donne des noms aux oiseaux et aux animaux que Dieu fait passer devant lui (Genèse 2, 19-20). Il existe un commentaire rabbinique dans un *midrash* qui dit : « Avant d'envoyer les oiseaux et les animaux devant Adam, Dieu les a envoyés devant les anges et ils n'ont pas su [leur] donner de nom. Il les a donc envoyés devant Adam et il leur a donné des noms. » Une des leçons est que, dans ce mythe, les anges parlent la même langue qu'Adam, mais le nom n'est pas juste une sorte de label ou d'étiquette. L'acte de donner un nom consiste à établir une relation et à mettre l'autre dans son monde. Peu importe qu'il soit un animal, un oiseau, une machine ou je ne sais qui : ce n'est pas le statut ontologique qui compte, *c'est la relation*. Sur ce plan relationnel, l'autre fera partie de notre réalité et ira jusqu'à nous influencer.

Selon le RGPD (Règlement général sur la protection des données), il faut informer les consommateurs qu'ils sont devant une machine. Mais, bien sûr ! L'histoire d'Adam montre pourquoi cela n'est absolument pas suffisant : tout se passe à travers la relation qui s'établit dans le dialogue. Même si on sait que c'est une machine, on peut être soumis à de la manipulation ou éprouver une joie… L'utilisateur va projeter sur la machine, inévitablement et spontanément, des états de connaissance, des états d'âme, des émotions.

Il faut pourtant maintenir les distinctions : c'est un impératif éthique. Certaines distinctions sont moins pertinentes que d'autres : si vous appelez pour prendre rendez-vous chez un coiffeur et que c'est un robot qui vous répond au lieu d'un humain, la distinction est de peu d'importance. Ce qui l'est, c'est quand ce sont des textes suffisamment longs et communiqués à une tierce personne : par exemple, un professeur qui reçoit la dissertation d'un élève doit savoir si c'est une machine qui l'a écrite ou si c'est une personne. Pour cela, il y a ce qu'on appelle des *watermarks* ou des filigranes. Ils ne sont encore qu'à l'état d'étude (mais déjà très poussée), depuis un an. C'est toute la difficulté : quand vous avez une image, vous pouvez mettre des métadonnées qui disent

que « cette image a été fabriquée par tel système d'intelligence artificielle ». Mais quand vous avez un texte, qu'est-ce que vous faites ? Il n'est pas suffisant de faire des séquences équidistantes de lettres, cachées à l'œil mais détectables par un logiciel, parce qu'elles sont faciles à casser (vous ajoutez trois mots par-ci, trois mots par-là, et cela casse le filigrane). Il faut plutôt le faire au niveau des probabilités des *tokens* : quand la machine complète le *token* manquant, les probabilités sont un tout petit peu biaisées. La difficulté est l'absence d'interopérabilité : si c'est ChatGPT qui a fabriqué le texte, vous devez vous adresser à OpenAI pour le détecter ; mais si le texte a été fabriqué par Bard, vous devez aller voir Google pour savoir s'il y a un code en filigrane ou pas. Il faut donc trouver un équilibre entre l'interopérabilité et la robustesse aux attaques adverses. Pour l'instant, c'est une question ouverte, à laquelle nous n'avons pas encore de solution.

> « *On sait que c'est une machine et, pourtant, c'est la relation qui compte* »

Quel type de relation entretenons-nous avec les machines que nous fabriquons, qui provoquent chez nous des émotions, des sentiments, une fascination, etc. ? Comment préserver la bonne distance et la liberté de la personne à l'égard de la machine ?

■ **Al. Grinbaum :** Le problème est qu'on anthropomorphise par projection ces systèmes parlants non humains : on sait que c'est une machine et, pourtant, c'est la relation qui compte, le « comme si ». On projette donc des connaissances, des états d'âme, des émotions, voire de la responsabilité. Et c'est là que cela devient problématique, c'est là qu'il faut séparer, parce que ces agents ne sont pas responsables (la responsabilité relève du monde humain et pas du tout du monde des dieux, ni de celui des machines). Prenons une situation très banale : des insultes, par exemple. La machine peut-elle m'insulter ou puis-je l'insulter comme une espèce de souffre-douleur ? Vous allez spontanément faire des projections de dignité sur elle, tout en sachant que c'est une machine.

Un exemple est parlant. Un jeune homme venait de perdre son amie. On lui a fabriqué un *chatbot* qui parlait comme elle. Il savait bien que c'était une machine. Mais, ce qui compte, c'est l'émotion ressentie qui l'a transformé. Les projections dans son cas n'ont pas été si mauvaises parce qu'elles lui ont permis de terminer son deuil. L'anthropomorphisme est spontané, on ne peut pas l'éviter. Mais, parfois, il va trop

© S.E.R. | Téléchargé le 05/09/2023 sur www.cairn.info par Robert Mathieu (IP: 176.181.145.57)

loin. C'est là qu'il faut faire des distinctions : les machines appartiennent au monde des êtres fonctionnels, comme les anges, qui n'est pas le nôtre. Dans un livre précédent[2], mon argument portait sur le rôle du hasard qui permet, dans ces situations de conflit, de soustraire la machine aux projections de la responsabilité humaine. Avec le langage, cette capacité est plus compliquée. Au modèle de fondation, on ajoute des couches de contrôle. Par-dessus du modèle d'intelligence artificielle générative qui produit du texte, vous mettez des systèmes de filtrage avec des règles. Par exemple, la machine ne doit pas donner des conseils médicaux, parce qu'elle n'est pas responsable des conséquences sur votre santé. Si vous demandez à ChatGPT un conseil médical, il va commencer par vous dire : « Je ne suis pas médecin, pourtant je peux vous dire… Mais, faites attention, ce n'est pas vraiment du conseil médical car je ne suis pas un médecin. » D'où vient ce « je ne suis pas un médecin, mais » ? Cela vient d'une couche supplémentaire de contrôle. On essaie d'éviter que les machines tombent en situation de conflit avec les êtres humains. Il y a même un nom américain de cette discipline, « l'alignement » sur les principes et les valeurs humaines (« *AI Alignment* »). Pourtant, quel que soit le nombre de règles que vous ajoutez, il y aura toujours des situations de conflit. Elles seront rares mais on ne peut pas les éviter, car il n'y a pas de système artificiel « éthique par conception ». Ce qui importe en cas de conflit, c'est d'avoir une solution technique pour empêcher les projections de la responsabilité sur la machine qui n'est pas un agent moral.

Cela nous oblige à réviser ou à préciser ce qu'on entend par responsabilité et par prise de décision…

■ **Al. Grinbaum :** C'est aussi un argument de mon livre précédent : ces machines sont autonomes. D'une certaine façon, elles sont des « individus numériques » opaques. Mais elles ne sont pas des personnes ! L'utilisateur ne sait pas ce qui se passe derrière l'interface et même les concepteurs ne savent pas quelle sera la sortie, un peu comme, quand je vous parle, je ne peux pas prédire avec précision votre phrase suivante. Cela ressemble aux êtres autonomes et individués que nous sommes. Et pourtant, il y a une différence fondamentale : les machines sont fonctionnelles et non libres. La finalité d'une machine est

2. Al. Grinbaum, *Les robots et le mal*, Desclée de Brouwer, 2019.

donnée par le concepteur, tandis que nous croyons avoir un libre arbitre et que notre finalité ne nous a pas été imposée de l'extérieur. Il faut donc trouver une éthique des systèmes fonctionnels sans en faire des personnes. Et c'est là que l'histoire des anges peut nous aider, eux qui sont des êtres fonctionnels dans les mythes.

Un lieu intéressant de l'application des algorithmes d'intelligence artificielle, c'est le domaine de la justice. On peut penser qu'une machine munie de toutes les connaissances de jurisprudence sera plus efficace qu'un juge pour porter un jugement. Faut-il conserver la dimension humaine en dernier recours au jugement ou le confier à une machine réputée plus « objective » ?

■ **Al. Grinbaum :** Il se trouve que je participe à un projet québécois qui s'appelle « Cyberjustice ». Avant de faire de la philosophie, commençons par faire un peu de sociologie. Des sondages ont montré que, dans les pays où il y a beaucoup de corruption, les gens sont plutôt favorables à l'idée des juges automatiques ; tandis que, dans les pays où il y en a moins, ils n'y sont généralement pas du tout favorables. Le seul pays en Union européenne qui a fait une expérimentation est l'Estonie : cet essai n'a pas très bien marché.

Conceptuellement, un autre élément intéressant est que la machine imite des choses que les juges font sans en être conscients (ce qui est classique, quel que soit le domaine, en particulier juridique). Souvent les machines révèlent des biais auxquels on n'avait pas pensé. On a fait des tests à travers différents corpus d'apprentissage des décisions juridiques dans différents pays et on a découvert qu'il existe un biais universel : les juges sont plus sévères le matin que l'après-midi. Est-ce un biais à corriger ou faut-il le laisser, comme chez les êtres humains ?

D'un côté, une décision non humaine a la force de s'imposer : c'est la justice transcendante bien connue dans l'histoire des religions. Dans le jugement divin, il y a toujours un intermédiaire : Moïse, par exemple. Mais l'exemple le plus parlant, ce sont les prêtres du sanctuaire de Delphes. Ils interprètent la parole asémantique de la Pythie. Le prêtre dit, par exemple : « Je m'appelle Plutarque et je vais vous dire quel est le jugement d'Apollon. » C'est la même histoire pour la machine : son résultat est asémantique et un être humain l'interprète et lui donne sens. Donc, je pense que la bonne façon de procéder est de donner la possibilité aux juges de se servir des systèmes d'aide à la

décision, mais la signature sous un verdict appartient à un être humain. Le juge corrige et valide ce qui provient de la machine. La responsabilité reste dans le monde humain. Aujourd'hui, la justice prend des années ; la machine, elle, peut analyser une situation en quelques secondes. Mais il faut faire attention à l'habituation des juges. Pour y remédier, la machine devrait de temps en temps dire n'importe quoi, parce que, sinon, le juge s'habitue à signer sans relire.

La machine est censée avoir moins de biais. Or elle fonctionne à partir de situations effectives (la jurisprudence). Ne reproduit-elle pas les biais hérités du passé ?

■ **Al. Grinbaum :** On sait comment corriger les biais connus (par exemple, la plus grande sévérité envers les Noirs aux États-Unis). Le problème réside dans les biais que nous ne connaissons pas. Puis, comme la machine fonctionne de manière asémantique, il y aura toujours des tensions entre les notions humaines du vrai et du beau. C'est l'être humain qui sait si un texte est beau ou si une phrase est vraie. La machine n'en sait rien, sauf si l'on ajoute des couches supplémentaires pour l'évaluer. Si l'on ajoute du symbolique, la machine pourra descendre jusqu'à un taux d'erreur très bas, mais jamais nul. Par exemple, si vous utilisez la version payante de GPT-4 avec le *plug-in* de navigation du Web, le taux d'erreur est très bas, parce qu'elle va chercher des informations qui se mettent à jour en permanence. Mais la machine ne sait pas, par exemple, qu'il y a la flèche du temps et que le moment « il y a deux ans » ne peut pas dépendre de quelque chose qui se passera demain. On peut donc utiliser ces machines comme une aide à la rédaction, sans pour autant leur faire entièrement confiance.

> « C'est l'être humain qui sait si un texte est beau ou si une phrase est vraie »

On peut citer la parabole du « château de Prague ». On demande à la machine : « Qu'est-ce qu'il y a à voir à Prague ? » Elle dit : « Un pont, un château, ceci, cela… » On pose une deuxième question : « Quelle est l'histoire du château de Prague ? » La machine répond : « Ta question est compliquée, est-ce que je peux aller chercher cette information ? » Et elle donne la bonne réponse. On demande ensuite : « Qui était le premier président ? », sans préciser de quoi, du château de Prague, de la ville ou de la République tchèque. La machine répond à

nouveau : « Ta question est compliquée, est-ce que je peux aller chercher cette information ? » Elle répond ensuite que c'était Václav Havel. Dans ce cas, la machine n'a pas seulement répondu correctement, mais elle a aussi compris le contexte. Ce que nous avons plus de mal à comprendre, c'est que, quand la machine dit : « Ta question est compliquée, est-ce que je peux aller chercher cette information ? », elle ne va nulle part et ne cherche rien du tout. Elle dit cela parce que c'est ainsi que les êtres humains réagissent dans ce genre de circonstance. Elle imite notre réaction. Ce modèle d'intelligence artificielle a déjà tout appris et n'a pas besoin d'aller chercher l'information, c'est du pur mimétisme. Pour nous, le sens premier des mots est immédiat. Dans « est-ce que je peux aller quelque part ? », « aller » signifie se mettre en mouvement. En revanche, tout ce qui relève du contexte est pour nous plus compliqué. Pour la machine, c'est l'inverse : tout ce qui est contexte et référence est son monde à elle, car elle ne fait que calculer des corrélations. Mais le sens littéral n'existe pas pour la machine. C'est déroutant car nous ne pouvons pas imaginer que notre langue puisse être utilisée de manière asémantique, puisque nous projetons toujours des significations spontanément.

ChatGPT peut-il faire de la poésie ?

■ **Al. Grinbaum :** Oui, mais médiocre. Si vous faites une requête longue, en demandant d'écrire dans le style de Baudelaire ou de Rimbaud, par exemple, le résultat ne sera pas si mauvais, mais c'est vous qui ferez ce jugement, pas la machine.

Alexandre Pouchkine disait : « On ne peut pas vérifier l'harmonie par l'algèbre. » La machine fait justement de l'algèbre. Qu'est-ce que cela nous apprend sur la poésie ? Que la perception de la beauté dépend de notre histoire. Vous pouvez imaginer un Théocrite qui apprend la beauté des textes en grec ancien, comme cette beauté était perceptible par les porteurs de cette langue. Aujourd'hui, cette beauté, nous la projetons sur le grec d'une façon différente, parce que nous ne sommes plus porteurs de cette langue. Nous pensons que c'est beau parce que nous avons une certaine expérience de lire ces textes, mais nous ne sommes pas spontanément dans la langue. Imaginez aujourd'hui une machine qui écrit un poème en grec ancien : comment saurez-vous si c'est beau ou pas ? Tandis que, si elle écrit en français, vous pourrez le dire, parce que vous êtes dans votre langue. L'es-

thétique dépend d'un apprentissage : nous savons qu'une œuvre d'art est belle parce que nos yeux ont vu beaucoup de choses dans les musées ou sur les pages des livres. Prenons l'exemple des poèmes futuristes, de Khlebnikov ou Marinetti : souvent, ce sont de pures séquences de lettres. En quoi est-ce un poème ? Justement parce que nous en connaissons l'histoire. Si vous aviez montré un poème futuriste à Molière, il aurait trouvé cela horrible, alors que, pour nous, c'est beau. Les algorithmes y sont insensibles. Ils peuvent fabriquer des poèmes mais, quand je fais ce genre d'expérience, il m'arrive une fois sur cinq de trouver une ligne qui est belle.

Ce que la machine ne capte pas très bien (cela rappelle le grand débat de la seconde sophistique, au IIe siècle), c'est la bonne mesure entre le nouveau et l'ancien, ce qui est donc hérité ou imité. Un auteur ou un poète a beaucoup lu et il s'en sert en permanence, mais il y a une mesure à trouver entre l'imitation et la création. La machine va spontanément trop loin. On lui demande d'écrire comme Baudelaire et elle fera un pastiche. Cela dit, j'ai fait l'expérience dans une conférence à la mairie du VIIe arrondissement, en demandant à GPT-4 d'écrire un poème dans le style de Baudelaire, sur le bonheur de vivre dans le VIIe arrondissement. C'était médiocre sur le plan poétique, mais l'assistance n'avait jamais vu cela. Elle était émerveillée simplement parce que c'était nouveau et la provenance du texte était non humaine.

Beaucoup d'activités seront concernées, selon vous ?

■ **Al. Grinbaum :** Elles ne vont pas disparaître, mais elles vont être réalisées différemment. Les avocats écriront plus rapidement leur plaidoirie. Les conseillers calculeront très vite les impôts car la machine connaît tous les textes de loi. Les assistants des politiques écriront plus rapidement leurs discours… Mais, à la fin, il faudra toujours quelqu'un pour prendre la responsabilité du propos tenu, parce que se poseront les questions de la confiance et du sens.

Propos recueillis par François EUVÉ et Nathalie SARTHOU-LAJUS.

Retrouvez le dossier « **Nouvelles technologies** »
sur *www.revue-etudes.com*