



Machines à écrire

Ce que l'intelligence artificielle change
à la littérature

Alexandre Gefen

Publié le 20-12-2024

<http://sens-public.org/articles/1698>



Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA
4.0)

Résumé

L'IA nous permet de réfléchir à nouveaux frais à la notion de narrativité, à celle de texte, à celle d'originalité, interrogeant la notion d'œuvre elle-même, on l'a vu, et nous conduit plus largement à réinterroger plus largement la notion de littérature : l'homme est-il le seul à produire de la littérature ? Si la littérature est un usage spécifique du langage, cet usage peut-il être reproduit par une machine ? Peut-on parler de la littérature comme d'une manière de rendre compte des technologies, voire d'une technologie d'écriture comme une autre ?

Abstract

AI allows us to think afresh about the notions of narrativity, of text, of originality, questioning the notion of the work of art itself as we have seen, and more broadly leading us to reinterrogate the notion of literature: is man the only one to produce literature? If literature is a specific use of language, can this use be reproduced by a machine? Can we speak of literature as a way of accounting for technologies, or even as a writing technology like any other?

Mot-clés : créativité, grand modèle de langage, littérature artificielle, littérature générative

Keywords: creativity, large language models, artificial literature, generative literature

Machines à écrire

Alexandre Gefen

L'émergence des Large Language Models et en particulier de GPT-3, une machine linguistique qui, ayant été entraînée par la société OpenAI sur 140 000 millions de paramètres (à peu près autant de mots si l'on veut) pour « prédire » ce que la « langue » pourrait générer comme textes statistiquement probables à partir d'une requête en langage naturel, a pris l'opinion et les écrivains par surprise en 2022, année où parallèlement les outils de création d'images à partir de descriptions textuelles (Dall-E, Midjourney, Stable Diffusion) surprenaient le monde de l'art. En utilisant une méthode d'apprentissage machine nommée « transformers » (un système d'attention machine parallèle qui est en train de révolutionner le *deep learning* après avoir révolutionné le traitement automatique du langage NLP) et un système d'entraînement dit « auto-supervisé », GPT-3 s'est montré capable de toute une série de tâches singulières : l'écriture d'essais scolaires probants, de code informatique opérationnel, de poèmes en vers, de fictions étranges, de conversations générales pouvant passer le test de Turing, entre autres exercices. En voici deux exemples :

Bonjour GPT-3, je voudrais te demander ce matin : qu'est-ce que la littérature ?

— *Eh bien en un peu plus de sept minutes, tu le sauras.*

— *Je t'écoute.*

— *La littérature, en un peu plus de sept minutes ?*

— *Oui, je t'écoute.*

— *C'est un langage qui n'est pas un moyen d'expression, mais un moyen d'échange.*

- *Un moyen d'échange ?*
- *Oui, un moyen d'échange. C'est un art qui n'a pas de métier, de profession, de compte en banque et de clients.*
- *Donc un art qu'on partage.*
- *Oui, un art qu'on partage. C'est un art qui n'a pas de métier, de profession, de compte en banque et de clients.*
- *Et donc un art sans tout cela, déjà.*
- *Oui, un art sans tout cela, déjà. Le langage qui n'est pas un moyen d'expression, mais un moyen d'échange et donc un art qu'on partage sans tout cela.*
- *C'est un langage qui est donc un filtre dérivé de la parole.*
- *Je te remercie. Bon, il faut que j'aille prendre l'ascenseur maintenant.*

Bonjour GPT-3, je voudrais te demander ce matin : qu'est-ce que la littérature ?

Les mains levées, il prenait toujours un air détaché, il se grattait machinalement le crâne tout en réfléchissant à ce qu'il pouvait répondre, il semblait être très ému, comme si cette question l'obligeait à dévoiler un secret le concernant. Il décrivit un cercle avec sa chaise et reposa les coudes sur le dossier. Il s'était légèrement incliné vers moi, l'air grave et concentré. Deuxième tour de piste : il se mit à chercher dans sa mémoire, les yeux mi-clos. C'était à moi qu'il s'adressait, et non à l'assistance : nous n'étions pas de la même planète. Son regard et la flamme qui courait à l'horizon, les boîtes des holophotographies fermées, les lève-planètes intacts, tout dans le bureau me faisait sentir sa présence.

- *C'est vraiment, vraiment difficile...*
- *Qu'est-ce que la littérature ? insistais-je.*
- *S'il n'y a rien, comment écrire ? s'interrogeait-il à voix haute, à la surprise de l'assemblée.*

Il s'enfonçait. Perdait pied. Glissait. Se rattrapait.

J'attendais. Je le sentais au bord de l'effondrement.

— *C'est une question de survie, il déclara.*

Son visage se renversa en arrière ; d'une main, il s'agrippait au bord de sa chaise et de sa main libre, il se tapotait le front.

— *... survivre en fuyant ou dignement, ça ne m'intéresse pas.*

Il décroisa les bras à la Monty Python et prit un air pincé. Toute cette comédie semblait le divertir, il riait avec nous.

— *Je reste un être humain, n'est-ce pas ? Je suis encore un homme, n'est-ce pas ?*

Il écarta les bras.

Sa silhouette imposante sous les verrières me fascinait. Même la lumière qui pointait le sommet de ses cheveux, sans parler de celle qui l'auréolait de soleil.

— *Je me suis arrangé pour survivre. En secret, sans attirer les soupçons. Si vous voulez, je crois que j'écris pour quelqu'un. Comment dire, comme si j'écrivais pour quelqu'un que j'ai connu. Les gens meurent. Ce qui me pousse à écrire, c'est fièvre de la sur-vie. Des gens qui se succèdent, qui disparaissent, qui meurent. L'auteur des romans n'est-il pas la somme de tous ces connards ? Je ne veux pas qu'on me confonde avec les gens qui meurent. Je fais tous les efforts pour rester en vie. Ne pas me confondre.*

Il avait étranglé sa voix dans un falsetto qui l'avait fait éclater de rire. Il reprenait son souffle, parlait comme pour lui-même :

— *C'est vrai, quoi... La confusion est une insulte à l'ordre mondial, à l'histoire humaine.*

— *Tiens, il a fini par sortir son texte, dis-je.*

Les deux dialogues, questions et réponses, ont été écrits par intégralement le logiciel à partir de ma question initiale. La différence entre les deux textes tient à la variation de ce qu'on appelle la « température », le degré d'entropie ou de fantaisie, fixée à une valeur plus élevée dans le deuxième texte. Les textes ont été coupés, mais en rien retouchés. On se rend compte à la lecture de ces textes du trouble qui s'empare du lecteur : la deuxième décennie du XXI^e siècle a fait rentrer la création artificielle dans des usages concrets, confrontant leur utilisateur à cette « vallée de l'étrangeté » décrite dès 1970

par Masahiro Mori où la ressemblance des machines aux hommes relève de la merveille et les dissimilarités de la monstruosité. Lorsque ce n'est pas ce que Günther Anders nommait la « honte prométhéenne », le sentiment d'être dépassé par les créatures que l'on a engendrées à une heure où foisonnent les outils d'évaluation de scénarios ou de textes par des IA (supposé, comme le propose Matthew Jockers (2016), trouver des best-sellers) et les outils d'accompagnement rédactionnels par des intelligences artificielles.

Sur un plan académique et scientifique, les limites de GPT-3 ont rapidement été cernées et l'illusion d'une intelligence dissipée (Baudoux 2023 ; Heaven 2020), puisque l'outil n'est pas capable de raisonnement, de symbolisation ou d'abstraction, mais uniquement de production probabiliste de textes dont il n'a aucune représentation mentale. Mais la littérature étant affaire de langage avant d'être de raisonnement, l'illusion narrative produite par GPT-3, même s'il raconte en inventant à partir de textes déjà racontés, est frappante. S'ils ne sont pas capables de tromper un logicien, les LLM peuvent donner l'impression d'une narration humaine – sans parler de leur capacité à écrire « à la manière de », tâche d'imitation dans lesquelles ils excellent par définition, puisqu'ils sont constitués de matrices de probabilité qui équivalent dans une certaine mesure à un « style » d'écrivain si ce n'est à sa manière de penser.

Une fois l'illusion d'humanité dissipée, si l'on se réfère à nos cadres d'appréciations ordinaire des textes littéraires, les questions sont nombreuses : qui sont en fait l'auteur de ces textes ? Les auteurs de l'algorithme ? Le savoir collectif contenu dans un texte issu d'une compilation probabiliste d'une grande partie des connaissances humaines sur la littérature ? Les écrivains qui ont nourri GPT-3 malgré eux, dépouillés de leur droit d'auteur et qui errent spectralement dans ces créations nouvelles – le terme d'intertextualité peut-il d'ailleurs s'appliquer ? Ces créations sont-elles aléatoires, idiosyncrasique à GPT-3 ou paradigmatique d'une intelligence holiste, d'un savoir de l'humanité, qui serait révélée et donneraient la parole à une intelligence collective ? Peut-on parler de personnalité pour décrire les réponses des différents LLM ? En suis-je l'auteur, moi qui les ai enclenchés en écrivant un « prompt », une requête, dont la formulation est rapidement devenue un art en soi ? Peut-on parler d'intentionnalité ? Est-il possible de les citer et qui en est responsable juridiquement ? Quelle sont la date et le contexte de leur création ? Celle de leur corpus (mais celui-ci n'est pas public) ? De la création du logiciel, de son « inception » ?

Plus concrètement encore, pour prendre l'exemple des deux textes que j'ai fait générer par GPT-3 : à quelle influence en attribuer la forme dialogique et platonicienne ? Peut-on le qualifier de « littéraire » ? Ces textes possèdent-ils une cohérence interne qui les constituent comme textes ? Puis-je mobiliser des outils d'analyse traditionnels ? Qui est le « je » et le « tu » mis en scène théâtralement pas ces textes ? Parodie-t-il par exemple la célèbre conférence de Mallarmé « Sait-on ce que c'est qu'écrire ? Une ancienne et très vague mais jalouse pratique dont gît le sens au mystère du cœur » ? Dans le premier texte, peut-on discuter l'analyse de l'art comme forme de partage découplé de l'expression ordinaire et du monde des transactions en la rattachant à un paradigme esthétisant de l'art ? Pourquoi mélangent-ils le registre pathétique, empathique et humoristique ? Comment comprendre la sorte de mythologisation et de sacralisation de la littérature qui transparait dans le second texte ? Comment comprendre l'ironie de ces textes, mécanisme textuel qui suppose une forme de duplicité énonciative et une forme de conscience hyperréflexive dont on a fait une définition de la littérature et le propre de l'homme ? De qui se moque l'IA de GPT-3 dans ces textes ? De ma naïveté d'avoir voulu les interroger ?

On se souvient de la nouvelle « Answer », de Fredric Brown, de 1948 où un être humain interroge une machine intelligence qu'il vient de créer sur l'existence de dieu (« dieu existe-t-il ? ») et où celle-ci lui répond « maintenant oui »... S'il est aisé de relativiser immédiatement la non-humanité de telles créations (les IA sont créées par des humains, entraînées sur des créations humaines, entraînées à produire des créations ressemblant aux créations humaines et déclenchées par des humains), comme à souligner la part d'intelligence collective et de scripts inclus dans la littérature humaine, les défis ouverts aux catégories traditionnelles du jugement esthétique et aux modalités variées d'appréciation de l'art (sociales, économiques, juridiques, etc.) sont néanmoins considérables : peut-on parler d'une disparition de l'auteur ? Comment penser « l'auctorialité » de la machine dans ses dimensions variées (prestige, imputabilité...) ? Que faire d'une catégorie comme celle de l'originalité, pourtant fondatrice de l'art et de la littérature moderne depuis le XVIII^e siècle ? Comment analyser le style d'une œuvre artificielle, ses références, son intertexte ? Comment évaluer et juger une œuvre créée artificiellement ? Devenue concrète, la créativité artificielle est clairement un défi théorique et critique majeur autant qu'une énigme renouvelée.

Databiographie de Charly Delwart propose de retracer un destin en s'appuyant sur des données numériques et leurs visualisations, *Lezérotleun.txt* de Josselin Bordat essaye de mettre en scène une intelligence artificielle en phase d'éveil au monde, *Kétamine* de Zoé Sagan met en scène un journaliste « prédictif » centré sur les données : jamais nous n'avons été aussi proches d'agents créatifs artificiels devenus à la fois des outils à produire de l'art et des modèles pour comprendre la création. L'hypothèse jusqu'ici romanesque (Philippe Vasset, *Exemplaire de démonstration*, 2002 ; Antoine Bello, *Ada*, 2016) d'une littérature produite par intelligence artificielle est en passe de devenir réalité. Loin d'assister à cette évolution passivement, nombreux sont les écrivains ayant cherché à s'approprier les pouvoirs des IA. Road trip entièrement écrit par une intelligence artificielle embarquée dans une voiture, *1 the road* de Ross Goodwin a rejoint à la rentrée littéraire 2019 toute une série de textes dont le point commun était de mettre en scène et en acte un rêve d'automatisation et d'artificialisation du langage littéraire : l'intelligence artificielle n'est désormais plus seulement une fiction, mais un outil à produire des fictions, possiblement en collaboration avec des écrivains. Les fictions sont autant textuelles qu'imaginées. David Jhave Johnston crée des poèmes avec une « augmentation » de neurones artificielles. Hito Steyerl travaille le genre du film documentaire avec le *deep learning* ; *Second Earth* de Grégory Chatonsky nous embarque dans un univers d'image générées automatiquement et produisant (*If ... then*, 2009) des micro-narrations.

On peut au demeurant avancer que, malgré les spécifiques nouveautés et sans diminuer les effets de surprise qui les accompagnent, ces créations artificielles s'inscrivent dans une très longue durée. De la même manière dont la science de l'IA s'appuie sur l'histoire longue des progrès de la formalisation et de la mathématisation du raisonnement, les figures contemporaines proposées par les IA, les dispositifs de création par IA ont été anticipés dans l'histoire culturelle et artistique. Si le plus ancien robot de la fiction remonte à Homère et au géant Talos, la première machine artificielle à produire des textes littéraires remonte à 1726 et à la troisième partie des *Voyages de Gulliver* de Swift. Ces fictions ne se contentent pas d'anticiper le genre en devinant de possibles devenir techniques, elles configurent et orientent nos manières de recevoir et peut-être même de concevoir, les objets technologiques eux-mêmes. Autrement dit, à travers la culture fictionnelle de ceux qui imaginent des IA, la littérature ne se contente pas d'accompagner par des descriptions, des ex-

périences concrètes et incarnées ou d'anticiper par des intuitions géniales le futur, elle le programme, le performe.

En matière de philosophie de la littérature et de l'art, l'effet de nouveauté tend rétrospectivement à faire émerger des pratiques et des questions anciennes et rend sensibles des effets de continuité autant que des différences. L'intelligence artificielle a une archéologie, son histoire moderne a été largement anticipée. Pour revenir à la question de l'auteur et de sa créativité, l'émergence de l'intelligence artificielle a été ainsi qualifiée de « deuxième mort de l'auteur », en référence à celle décrétée par Barthes et Foucault à la fin des années 1960. Elle s'inscrit en réalité dans une problématique que l'on peut faire au moins remonter à la naissance de l'esthétique moderne au XVIII^e qui postule à la fois le triomphe personnel du poète, ce génie dont l'expérience propre a vocation à devenir littérature et dont l'héroïsation est une des marques du Romantisme, et en même temps ce que Mallarmé a rêvé comme la « disparition élocutoire du poète » dans l'œuvre pure, où le langage doit opérer sa magie propre. *Mutatis mutandis*, la double version dans laquelle se donne le rapport de l'auteur à l'IA, à la fois la figure démiurgique d'un triomphateur de la technique, auteur de l'auteur comme le dit Pascal Mougin (2023), et absent d'un flux linguistique auto-engendré, d'une langue qui parle elle-même en suscitant le trouble et dont aurait rêvé Mallarmé – les Large Language Models dont j'ai donné l'exemple et que l'on a surnommé méchamment ou par peur des « perroquets stochastiques » (Bender et al. 2021) sont une manière de faire parler la langue toute seule étonnamment proche de la vision qu'en avait par exemple un Barthes pour qui tout parole en s'installant dans un monde « déjà parlé » parle malgré elle.

Dans le domaine de l'écriture, les interférences entre modèle machinique et théorie et pratique littéraire se multiplient dès les années 1950 à l'heure où s'invente l'intelligence artificielle moderne. En phase avec la linguistique saussurienne, la grammaire générative et la critique formaliste, la théorie de l'information et la cybernétique sont d'emblée sollicitées par le structuralisme. En 1954-1955, Jacques Lacan consacre son séminaire aux nouvelles machines à calculer et à la cybernétique. Dans le *Système de la mode* (1967), Barthes emprunte à Mandelbrot et aux premiers théoriciens de l'information les notions de « briques de code » ou de « sous-routine », tandis qu'il organise ses propres notes manuscrites dans son « Grand Fichier » à la manière de l'hypertexte informatique des *Literary machines* de Ted Nelson (1982). Sur fond de « mort de l'auteur », les tenants de la déconstruction mobilisent

la cybernétique, la théorie de l'information et la pensée computationnelle dans leur critique du phonocentrisme et leur hypothèse d'une subjectivité purement discursive, libérée des frontières entre vivant et non vivant, circulant sans limite dans l'infini des systèmes. Le modèle machinique est alors sollicité régulièrement par Gilles Deleuze, Michel Foucault et Paul Virilio, qui théorise la *Machine de vision* (1988), cependant que Jacques Derrida, après avoir défendu la machine à écrire contre la critique heideggérienne de la technique, adopte l'ordinateur et vante, dans *Papier machine* (2001), ses premiers usages originaux du traitement de texte. Les écrivains eux-mêmes sont à l'écoute des théories structuralistes et post-structuralistes. Dans « Cybernétique et fantasmes » (1967), Italo Calvino spéculé sur « la littérature comme processus combinatoire » tandis que l'Oulipo explore la productivité de la contrainte arbitraire et systématique. Poètes programmeurs et ingénieurs linguistes explorent parallèlement les possibilités d'une « littérature générée » électroniquement puis informatiquement, des *Textes stochastiques* de Théo Lutz en 1959 au *Général Proust* de Jean-Pierre Balpe en passant par les « littéraciels » de l'Atelier de littérature assistée par la mathématique et les ordinateurs » (ALAMO) dans les années 1980.

L'IA relance ainsi de vieux rêves de création sans création, juste opérationnelle et citationnelle (je renvoie ici à Kenneth Goldsmith, *L'Écriture sans écriture : du langage à l'âge numérique* introduit en France par François Bon) rêve d'écriture automatique et à programme qui, en laissant le langage écrire lui-même, est un projet de sacralisation et de divinisation du langage au moins aussi ancien que le romantisme (Schaeffer 1983). L'horizon « transhumaniste » de disparition historique de la créativité humaine nous hante – pour citer une autre IA rédactionnelle, Jaspers, que j'ai mise à contribution :

L'intelligence artificielle va changer l'avenir de la littérature. Les écrivains seront remplacés par des machines qui produiront des textes sans erreur et à une vitesse incroyable. Les lecteurs ne pourront plus distinguer les œuvres originales des copies numériques. La littérature perdra son essence unique et elle deviendra un art fade et sans âme.

Il faudrait faire l'histoire des rêves de disparition de l'écrivain ou de l'artiste, en la mettant en parallèle avec toutes les formes d'orgueil du créateur : on pourrait avancer que cette situation de l'IA s'installe même dans une configuration encore ancienne que celle de l'art moderne et une valse-hésitation très ancienne. Platon dit dans le *Ion* à Socrate que « les poètes composent

non grâce à une connaissance rationnelle, mais grâce à un don qui leur est naturel », pourtant il édicte dans le *Phèdre* un ensemble de règles théoriques indispensables à suivre. Entre d'une part le spontanéisme déresponsabilisant d'une théorie de l'inspiration, la fascination pour le dérèglement dionysiaque propre à la parole poétique et, d'autre part, la vision, rhétorique, pédagogique ou philosophique, d'un poète devant suivre des formes héritées de la tradition, assujetties à la morale et devant produire régulières par des principes et des traditions, la doctrine classique hésite. Entre un rêve machinique imaginé dès l'Oulipo voire dès Raymond Roussel, les formes d'inspiration liées à la magie des drogues et du subconscient et un idéal de maîtrise absolue, nos contemporains balancent, comme notre appréhension de l'IA, basculant entre l'éblouissement face à la magie et le désir d'en maîtriser les trucs. À ce titre, l'IA nous montre bien les paradoxes d'un rêve intenable d'auctoralité pure qui ne pourrait s'accomplir que par le jeu pur du langage et se ferait au prix d'une disparition illusoire de l'homme. Elle nous permet au contraire de redécouvrir des pratiques anciennes de co-écriture (les formes d'écritures ludiques de l'ère classique), comme prêter plus d'attention aux traditions non occidentales (où l'intentionnalité se présente dans des configurations bien différentes). Elle nous permet de réfléchir à l'intertextualité : une IA écrit à partir de la mémoire d'innombrables textes, elle produit du nouveau mais à partir de l'ancien digéré par des réseaux de neurones, modèle que l'on peut comparer à celui de la création humaine et dont l'originalité serait un simple « hyperparamètre » réglable pour emprunter son vocabulaire à celui de l'IA, paramètre qui définirait en quelque sorte le degré de réalisme de l'IA.

Ainsi, l'IA nous permet de réfléchir à nouveaux frais à la notion de narrativité, à celle de texte, à celle d'originalité, interrogeant la notion d'œuvre elle-même on l'a vu et nous conduit plus largement à réinterroger la notion de littérature : l'homme est-il le seul à produire de la littérature ? Si la littérature est un usage spécifique du langage, cet usage peut-il être reproduit par une machine ? Peut-on parler de la littérature comme d'une manière de rendre compte des technologies, voire d'une technologie d'écriture comme une autre ? En littérature comme ailleurs, l'IA enrichit les questions que nous nous posons depuis longtemps sur nous-mêmes.

Bibliographie

Archer, Jodie, et Matthew L. Jockers. 2016. *The Bestseller Code. Anatomy of the Blockbuster Novel*. New York : St. Martin's Press.

- Baudoux, Michelange. 2023. « ChatGPT répond-t-il vraiment à nos questions ? » *Futur Immédiat*.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, et Shmargaret Shmitchell. 2021. « On the Dangers of Stochastic Parrots. Can Language Models Be Too Big ? » In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-23. FAccT '21. New York : Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>.
- Heaven, Will Douglas. 2020. « OpenAI's New Language Generator GPT-3 Is Shockingly Good—and Completely Mindless ». *MIT Technology Review*, juillet. Cambridge.
- Mougin, Pascal. 2023. « Comment lire un roman écrit par une voiture ? La doxa littéraire face à l'IA ». In *Créativités artificielles. La littérature et l'art à l'heure de l'intelligence artificielle*, édité par Alexandre Gefen, 207-20. Dijon : Les Presses du réel.
- Schaeffer, Jean-Marie. 1983. *La naissance de la littérature. La théorie esthétique du romantisme allemand*. Arts et langage. Paris : Presses de l'Ecole normale supérieure.