

ALEXANDRE
GEFEN

VIVRE AVEC CHATGPT

Séduire, penser, créer,
se cultiver, s'enrichir...

**L'intelligence artificielle
aura-t-elle réponse à tout ?**

Vivre avec ChatGPT

Du même auteur

La Littérature, une infographie, CNRS Éditions, 2022.

La littérature est une affaire politique, Éditions de l'Observatoire, 2022.

L'Idée de la littérature. De l'art pour l'art aux écritures d'intervention, José Corti, 2021.

Réparer le monde : La littérature française face aux XXI^e siècle, José Corti, 2017.

Inventer une vie : La fabrique littéraire de l'individu, Les Impressions nouvelles, 2015.

Alexandre Gefen

Vivre avec ChatGPT

L'intelligence artificielle
aura-t-elle réponse à tout ?

ISBN : 979-10-329-2977-3
Dépôt légal : 2023, mai
© Éditions de l'Observatoire/Humensis, 2023
170 *bis*, boulevard du Montparnasse, 75014 Paris

Note aux lecteurs : ce livre a-t-il été écrit avec ChatGPT ?

Cet essai vise à réfléchir avec les outils de l'histoire des idées, des sciences et des techniques, et, avec un grain de philosophie, à la manière dont ChatGPT s'est imposé en quelques mois dans nos vies. Son objectif est double : mener une réflexion critique informée, mais aussi accompagner les nouveaux usages, pour permettre au lecteur de mieux comprendre et de mieux apprivoiser ChatGPT – en maîtrisant ce qu'on appelle « l'art du prompt » : l'art de rédiger les meilleures requêtes.

Derrière le logiciel offert récemment au grand public par la société californienne OpenAI, j'ai essayé de réfléchir aux questions posées par les outils d'intelligence artificielle (IA) contemporains et en particulier ce qu'on appelle les grands modèles de langage et les intelligences artificielles génératives dont ChatGPT est un représentant, comme DALL-E (développé aussi par OpenAI) ou Midjourney pour les images.

Une des fictions actuelles sur les IA consiste à leur attribuer la capacité d'écrire intégralement des livres entiers. Qu'en pense ChatGPT lui-même ? Posons tout de suite une question qui pourrait nous éviter bien des amalgames.

Mon chat, es-tu capable d'écrire un livre ?

En tant que modèle de langage, je suis capable de générer du texte et donc, en théorie, je pourrais écrire un livre. Cependant, il est important de noter que mes capacités sont limitées par les données que j'ai été entraîné sur, ainsi que par les instructions et les paramètres fournis par mes utilisateurs.

En d'autres termes, si on me donne suffisamment de données de haute qualité sur un sujet particulier et que les paramètres de génération de texte sont correctement configurés, il est possible pour moi de générer un texte de qualité suffisante pour être considéré comme un livre. Cependant, cela nécessiterait un certain niveau d'interaction et de collaboration avec un utilisateur humain pour que le résultat final soit cohérent et satisfaisant.

En fin de compte, bien qu'il soit théoriquement possible pour moi d'écrire un livre, il est important de rappeler que la qualité et la pertinence du contenu dépendent de nombreux facteurs et ne peuvent être garanties que par une interaction humaine étroite et une évaluation attentive.

Pour des raisons sur lesquelles je reviendrai, ce rêve (ou ce cauchemar) n'est vraiment pas encore tout à fait d'actualité, hormis des cas très spécifiques, ou si l'on entend par livres de courts textes illustrés ou de brèves synthèses médiocres. Aucune des tentatives de vendre ces textes sur les plateformes d'autopublication n'a encore conduit à la production d'un best-seller, le plus grand succès à ce jour n'ayant pas dépassé un bénéfice de 100 dollars sur Amazon. Mais cela ne veut pas dire que ChatGPT ne puisse pas être un outil d'aide à la création, ni que certains textes brefs écrits par des IA ne puissent faire illusion.

Parler des IA génératives – et en particulier des IA rédactionnelles comme ChatGPT – sans les utiliser

eût été en revanche peu pédagogique. La philosophie générale de ce livre est de promouvoir une relation de cohabitation avec les IA, cohabitation critique, mais aussi bienveillante. J'ai fait le choix d'utiliser parfois ChatGPT, en identifiant strictement sa contribution par une typographie différente ; manière d'en montrer les pouvoirs au lecteur, mais aussi les usages et les limites. Aucun des textes produits par ChatGPT (nous avons utilisé la version 3 puis, à sa sortie, la version 4) que l'on trouvera dans ce livre n'a été retouché ou corrigé. À défaut d'être une pythie ayant des réponses universelles à toutes nos interrogations, ChatGPT est, on le verra, un utile copilote et un étonnant commensal.

PREMIÈRE PARTIE

L'invasion des perroquets stochastiques : comprendre ChatGPT

Chapitre 1

La Silicon Valley de l'étrangeté

C'est un grand *open space*, où tout a été fait pour rendre l'effort ludique et pour repousser les limites de la productivité personnelle. À côté d'énormes écrans de Mac sous-utilisés, des distributeurs de bonbons et de gros poufs participent d'une coolitude typique de la Silicon Valley. Nous sommes au printemps 2016 au Center for Spatial and Textual Analysis de l'université Stanford, en Californie, à quelques stations en « Caltrain » de San Francisco. C'est l'une des plus prestigieuses universités du monde, mais nous pourrions être tout aussi bien dans une des innombrables entreprises de *deeptech* naissantes, et pas sur un campus universitaire. Ici, comme dans les GAFAM Google et Apple dont les sièges sociaux sont situés à quelques kilomètres, les références à la contre-culture californienne et l'ostentation des richesses font bon ménage, composant cette formule magique de la réussite technologique de la côte ouest : disponibilité quasi infinie de fonds grâce à une économie fondée sur la capitalisation, apologie de l'inventivité sans filet, comme en témoigne tristement l'abondance des *mobile homes* dans le voisinage. C'est la formule du capitalisme créatif, avec ce mélange entre la vie personnelle et le travail, l'idéologie de la confiance en soi et l'acceptation institutionnalisée

de l'insécurité, le monde des start-up et de la recherche. C'est ici même, à Stanford, qu'il y a quelques années, Steve Jobs, fondateur et patron d'Apple, avait donné son discours testamentaire, méditation sur la mort et la prise de risque, qui se terminait par cette formule : « Restez affamés. Restez dingues. » C'est ici aussi que le fondateur de la société derrière ChatGPT, Sam Altman, a commencé ses études en informatique avant de les abandonner pour créer une start-up.

En 2016, à Palo Alto la vaniteuse, les Tesla Model X commencent à s'imposer. L'idée d'un véhicule électrique redevient à la mode, même si leur système d'aide à la conduite est encore sommaire – il faudra encore attendre quelques mois pour que de nouveaux processus Nvidia permettent à la voiture de changer de voie ou de se garer par elle-même avec l'aide d'un nouveau système de Lidar, un type de radar permettant à un ordinateur de se représenter un espace en trois dimensions. En juillet, pour la première fois dans l'histoire de l'autonomie, un accident a causé un premier mort dans une collision avec une semi-remorque. « La remorque blanche pourrait avoir été confondue par le système de pilotage avec un panneau d'affichage, sachant que le ciel était très lumineux et qu'il y avait un contre-jour », reconnaît Tesla. Mais nous faisons confiance aux ascenseurs, le pilotage de nos avions est presque totalement automatisé : pourquoi ne pas accepter le risque supposément beaucoup plus faible d'une voiture pilotée par une intelligence artificielle ? Voilà ce qu'on se dit ici, avant même de penser à qui payera en cas d'accident.

Dans cette grande salle ouverte vingt-quatre heures sur vingt-quatre, où je me suis rendu avec une équipe de chercheurs français venue d'une vieille Sorbonne

dont les codes semblent appartenir à une autre planète, le décalage horaire brouille le temps. À quelques mètres se situe la porte du mythique Literary Lab fondé par Franco Moretti, génial théoricien de la *world literature* et fondateur des humanités numériques, qui eut le premier l'intuition de considérer la culture comme un ensemble de données qui pouvaient s'analyser à distance par des algorithmes, en faisant l'économie de la lecture humaine. Puis il a fui en Europe pour une affaire de mœurs... Si nous étions venus de France sans sweat-shirts à capuche marqués à l'effigie de notre université comme cela se fait aux États-Unis, nous n'étions pas arrivés les mains vides : nous apportions d'immenses ensembles de textes numérisés, des milliers d'ouvrages représentant une bonne partie de la critique littéraire du XIX^e siècle. Même si le mammoth nommé Bibliothèque nationale de France avait eu du mal à y croire, le projet Gallica de numérisation de livres et de périodiques avait pris de l'ampleur, autorisant les chercheurs à constituer les premiers big data de la culture, d'abord en retranscrivant manuellement les textes, tâche d'abord menée par des amateurs passionnés comme l'excentrique bibliothécaire de la bibliothèque municipale de Lisieux, Olivier Bogros, qui a constitué en France la première bibliothèque numérique accessible à tous sur le Web. Ces premières et déjà vieilles bibliothèques numériques sont les piliers des intelligences artificielles d'aujourd'hui, et en cherchant bien, leur mémoire se trouve enfouie dans la mémoire culturelle des outils d'IA que nous utilisons pour rédiger à notre place des sujets de cours ou des *punchlines* pour réseaux sociaux

Pour l'heure, les réseaux de neurones et l'apprentissage profond sont encore dans la pénombre des

laboratoires de recherche, dans ce qu'on nomme encore le « traitement automatique du langage ». Les outils sont peu commodes, lents, dépourvus d'interfaces graphiques conviviales – les développeurs cultivent jusqu'à aujourd'hui les écrans noirs et les polices de caractères des terminaux de la première informatique où tout était lignes de code.

Après avoir fixé un peu au hasard « l'hyperparamètre » de mon outil d'analyse avec l'intuition d'un alchimiste jouant avec ses cornues, j'appuie sur la touche *Entrée*. Dans les entrailles du microprocesseur de mon ordinateur, chaque mot des milliers de textes dont j'ai demandé l'analyse par un algorithme nommé « Word2vec » est analysé dans des milliards de matrices mathématiques qui se combinent et se recombinent, avant que la réponse à ma question « Quels sont les mots qui traduisent le mieux le mot *romantisme* ? » apparaisse :

Word2vec : protestantisme, paganisme, catholicisme, clergé, journalisme, christianisme, symbolisme, jansénisme, socialisme, naturalisme, Cid, christ, Parthénon, rationalisme, cartésianisme, Saint-Esprit, pape, Parnasse, polythéisme, calvinisme, classicisme, bouddhisme, ciel, tiers état, positivisme, globe, choléra, sénat, despotisme

C'est un grand pas pour moi, mais un petit pas technologique pour l'humanité : les collègues américains qui nous accueillait savaient déjà analyser des masses d'opinions, additionner des mots (« roi » – « homme » = « reine », peut calculer un ordinateur), mais pour tout le savoir d'un siècle de professeurs, c'est un saut immense que de se retrouver en équations mathématiques. Pour l'historien de la

littérature à l'ancienne que j'étais avant d'appuyer sur ce bouton, c'est l'entrée dans la « vallée de l'étrangeté », pour reprendre le terme célèbre et déjà ancien (1970) de Masahiro Mori : la découverte d'une vision non humaine de l'univers humain, la manière dont une machine raconte ou dessine le monde de l'homme. Avec Word2vec, qui calcule les mots qui sont statistiquement les plus proches, tout se passe comme si l'inconscient de notre culture s'ouvrait à nous : oui, le « romantisme » est bien une sorte de religion, qui a son « clergé », son « pape » et son « Saint-Esprit », oui, elle affecte son temps comme un « choléra », par l'intermédiaire des « journalistes ». Les historiens du XIX^e siècle – Paul Bénichou, l'auteur du *Sacre de l'écrivain*, le premier – l'avaient bien dit, mais la machine nous le confirme avec l'autorité d'une sibylle. Auparavant, pour savoir ce qu'était ce mouvement littéraire, j'avais deux solutions : soit lire par moi-même les œuvres et toute l'histoire qui les entourait, soit faire confiance à des professeurs et à des érudits qui avaient accompli ce travail pour moi et en tiraient autorité et pouvoir. Avec Wikipédia, la révolution numérique avait accompli un premier tournant en démocratisant le savoir : sans pour autant éliminer l'érudition, l'intelligence collective pouvait redistribuer un peu les cartes. Avec l'intelligence artificielle, le savoir humain peut être digéré et retraduit à l'infini, permettant à chacun de se le réapproprier magiquement. Entraîné sur un savoir plus général que ma bibliothèque du XIX^e siècle, ChatGPT fera, dix ans après, bien mieux que Word2vec, en prenant cette distance et une neutralité *mainstream* qui lui sont propres :

Humain : En quelques mots, peux-tu expliquer au lecteur ce qu'est le romantisme ?

Le romantisme est un mouvement artistique, culturel et intellectuel qui a émergé en Europe au cours des XVIII^e et XIX^e siècles. Il se caractérise par un rejet des idéaux classiques et rationnels de l'époque des Lumières, au profit d'une valorisation de l'émotion, de l'individualisme, de la nature et de l'imagination. Les artistes et écrivains romantiques cherchaient à exprimer des sentiments profonds, souvent liés à l'amour, à la mélancolie, à la solitude, à la nostalgie et à la révolte contre les conventions sociales. Le romantisme a eu une influence considérable sur l'art, la littérature, la musique et la pensée politique et sociale de l'époque, et son héritage continue d'inspirer les créateurs contemporains.

Mais dès les années 2010, les rudiments de la révolution de ce que l'on appelle les *large language models* (« grands modèles de langage »), dont ChatGPT est le membre le plus fameux, étaient en route avec comme base une idée simple, mais géniale : la sémantique distributionnelle. Un concept fondamental, datant des années 1960, de toutes les intelligences artificielles qui écrivent : le sens des mots peut être déduit par des statistiques complexes des mots qui l'entourent. Entre l'univers des idées et celui des nombres, il existe une passerelle : les probabilités. En construisant l'espace de signification d'un mot et la géométrie de ses relations avec les autres mots d'un même texte, il est possible de donner un équivalent statistique d'un mot par un autre, ce qu'on appelle sa signification « latente, ce qui permet sa traduction. Dès lors, un algorithme peut prédire le mot devant suivre un mot donné, exactement comme le fait ChatGPT, en faisant de nous des nains sur des épaules de géants :

ChatGPT, peux-tu expliquer très simplement au lecteur ce que c'est que la sémantique distributionnelle ?

Bien sûr ! La sémantique distributionnelle est une méthode d'analyse du sens des mots basée sur la façon dont ils sont utilisés dans un contexte. Cette méthode considère que les mots ont un sens en fonction des autres mots qui les entourent. Par exemple, si le mot « chat » est souvent utilisé avec le mot « souris », cela peut indiquer que le mot « chat » a un lien sémantique avec le mot « souris ». La sémantique distributionnelle utilise des méthodes statistiques pour mesurer ces liens sémantiques et créer des représentations vectorielles des mots, qui peuvent être utilisées pour effectuer des tâches comme la recherche d'informations ou la traduction automatique. En somme, la sémantique distributionnelle est une méthode qui permet de comprendre le sens des mots en fonction de leur contexte d'utilisation.

Pourquoi est-elle fondamentale pour l'intelligence artificielle d'aujourd'hui ? Explication simple et rapide s'il te plaît.

La sémantique distributionnelle est importante pour l'intelligence artificielle, car elle permet aux ordinateurs de comprendre le langage humain en analysant la façon dont les mots sont utilisés dans un contexte. Elle permet aux ordinateurs d'interpréter les significations des mots et des phrases et de les utiliser pour des tâches telles que la recherche d'informations, la classification de texte et la traduction automatique. En somme, la sémantique distributionnelle est un élément clé pour que les ordinateurs puissent comprendre et traiter le langage humain, ce qui est essentiel pour de nombreuses applications d'IA.

Ainsi, le sens des mots ne serait qu'une question pragmatique, le langage serait calculable, décryptable à l'infini et, inversement, productible à l'infini. Il n'y aurait plus besoin de lire. Un texte pourrait se

représenter par un ensemble de vecteurs sur lequel on pourrait faire autant d'opérations mathématiques que l'on souhaite. Un siècle de savoir humain, de rhétorique et de grand discours, un mouvement qui avait transformé l'Europe, un cortège de professeurs barbichus et de poètes échevelés semble pouvoir être compris par une machine comme une manière statistique particulière d'organiser des mots.

Table

Note aux lecteurs :	
ce livre a-t-il été écrit avec ChatGPT ?.....	7

PREMIÈRE PARTIE

L'invasion des perroquets stochastiques : comprendre ChatGPT

Chapitre 1. La Silicon Valley de l'étrangeté.....	13
Chapitre 2. Le langage du langage	21
Chapitre 3. Le canard et les automates	29
Chapitre 4. Les algorithmes : à la vie, à la mort	39
Chapitre 5. Ce feu nous brûlera-t-il ?	51
Chapitre 6. La seconde mort de l'auteur	61
Chapitre 7. Le dilemme du tramway	71
Chapitre 8. L'apocalypse des trombones : ChatGPT va-t-il détruire le monde ?	83

SECONDE PARTIE

L'art du prompt : utiliser ChatGPT

Chapitre 9. Accéder à tout le savoir du monde.	97
Chapitre 10. Chatbusiness	107
Chapitre 11. La créativité augmentée	117
Chapitre 12. Comment hacker ChatGPT ?	127
Chapitre 13. Rédiger est-il démontrer ?	137
Chapitre 14. Parler est-il penser ?	147
Chapitre 15. Discuter est-il aimer ?	159
Chapitre 16. Ce que nous apprend l'IA sur nous-mêmes.....	171
Annexe. 20 trucs et astuces pour mieux utiliser ChatGPT dans la vie quotidienne	183
Remerciements	187