

Demain, cet article sera écrit par GPT-3

Cassandra Froelicher, Rossiny Longo, Joris Bohren et Nicolas Wunderle

Etudiant-e-s en ingénierie des médias, 1^{ère} année, HEIG-VD

OpenAI développe GPT-3 (Generative Pre-trained Transformer 3), une intelligence artificielle capable de comprendre et produire du texte en plusieurs langues. C'est un modèle de langage qui fonctionne notamment sur la prédiction de séquences de mots. Cette technologie est encore loin d'être parfaite ; néanmoins, nous pouvons déjà observer des applications prometteuses comme : rédiger des articles journalistiques, synthétiser des textes, écrire des programmes dans plusieurs langages informatiques, décrire le fonctionnement de programmes informatiques ou encore converser avec un humain. Mais où en sommes-nous vraiment ?

I. INTRODUCTION

Notre époque où l'intelligence artificielle se développe de plus en plus, celle-ci est capable de réaliser des tâches spécifiques de plus en plus complexes. En 2015, OpenAI voit le jour. Cette entreprise, créée par plusieurs personnalités dont Elon Musk, le fulgurant patron de Tesla, a pour objectif de promouvoir l'intelligence artificielle au bénéfice de toute l'humanité [1]. Cette technologie a un côté novateur, mais a encore beaucoup de progrès à réaliser. Néanmoins, il existe des applications encourageantes, notamment dans le domaine du traitement automatique du langage naturel. Comment les humains et les machines pourraient-ils collaborer dans le domaine de la rédaction ?

II. FONCTIONNEMENT

Une séquence de mots est une suite de mots donnés dans un ordre particulier. La probabilité qu'une séquence soit obtenue dans une langue spécifique est estimée par un modèle de langage statistique dans le but de prédire le mot suivant dans une phrase et ainsi générer du texte écrit [2].

GPT-3 est le troisième modèle publié par OpenAI et se base sur différents principes de l'intelligence artificielle et plus spécifiquement du *deep learning*, à travers plusieurs réseaux neuronaux [3,4].

Le deep learning permet aux machines d'apprendre à travers des données sur lesquelles elles s'entraînent et de prendre une décision par elles-mêmes grâce à des couches d'algorithmes, une structure appelée réseau de neurones qui s'inspire du fonctionnement neuronal humain [5].

L'API (*Application Programming Interface*) autorise une entrée de texte par l'utilisateur, nommée *prompt*, qui décrit les tâches à faire et qui peut être optimisée en indiquant des paramètres pour ajouter un contexte. Ensuite, le modèle génère un texte en sortie, appelé *completion* [3]. Cette procédure est rendue possible grâce aux transformateurs, un réseau neuronal composé d'un encodeur et d'un décodeur. L'encodeur transforme en représentation mathématique la séquence de mots

d'entrée et les liens entre ces mots pour ajouter des éléments de contexte. Le décodeur reçoit cette représentation, estime les probabilités et génère une séquence de mots à la suite d'une prédiction [6].

Le réseau de neurones de GPT-3 a été pré-entraîné sur 45 TB de données textes, et contient 175 milliards de paramètres. Ces derniers sont des poids attribués à un ensemble de données qui déterminent la probabilité de choisir ces données en entrée pour prédire les résultats [7,8]. Les paramètres sont définis durant la phase d'entraînement par GPT-3, donc les caractéristiques qu'elles influencent sont inconnues.

Lorsqu'on donne un texte à GPT-3, il donne un poids à chacun des 175 milliards de paramètres pour générer un texte le plus ressemblant.

III. APPLICATIONS

Les applications de GPT-3 sont incroyablement diverses, grâce à leurs capacités de compréhension du langage courant. En utilisant GPT-3 comme interface avec l'utilisateur, une application peut recevoir des instructions naturelles pour exécuter un programme défini. Les tâches réalisables sont l'identification de groupes de mots selon une classification, comme la reconnaissance d'un texte positif, neutre ou négatif, la transformation de texte comme la traduction ou la synthétisation, la génération de texte grâce aux paramètres comme la rédaction d'une histoire, la complétion de texte et la conversation [3]. OpenAI propose plusieurs exemples d'applications sur son propre site web, comme un créateur de notes à partir d'un texte donné ou un générateur de requêtes SQL pour les bases de données informatiques à partir d'une demande en anglais courant [9].

Beaucoup de développeurs ont créé des applications novatrices qui utilisent les capacités de compréhension de GPT-3 pour effectuer des tâches diverses. Par exemple, une extension pour l'éditeur vectoriel Figma, qui génère des mises en page de sites web à partir d'une simple description [10], un générateur de descriptions d'images pour un article scientifique ou comme texte alternatif pour un site web [11] ou un moteur de recherche de publications scientifiques qui permet de trouver des textes relatifs à une problématique donnée [12].

Malgré le fait que OpenAI soit un laboratoire américain et donc anglophone, GPT-3 est capable de reconnaître et d'écrire du texte dans une multitude de langues. Par conséquent, le système est capable de formuler des traductions basiques sans entraînement spécifique pour cette tâche [13].

Avec le développement récent de OpenAI Codex, un modèle basé sur GPT-3 optimisé pour les langages informatiques [14], le géant du partage de code GitHub se lance aussi dans le monde

de l'assistance IA avec Copilot, une application qui s'intègre directement dans les éditeurs de code et permet de générer du code informatique à partir d'une description donnée en langage commun [15].

Néanmoins, même si GPT-3 est un système en amélioration continue, et s'il est vrai qu'on observe des améliorations importantes sur la dernière année, c'est encore loin d'être une machine parfaite. En effet, c'est une intelligence artificielle très jeune et ses applications sont encore peu explorées. Ces limitations ont été reconnues publiquement par le CEO d'OpenAI [16] et sont de plus facilement observables. Le système a encore beaucoup de difficultés à produire des textes longs et cohérents [17] et présente par instant des biais indésirables sur des sujets sensibles comme le racisme [18].

En l'état, toute application professionnelle de GPT-3 doit donc être faite avec une intervention et supervision humaine.

IV. DÉRIVES

Aujourd'hui sans l'ajout d'informations plus spécifiques sur le sujet, les phrases écrites par l'IA sont un peu répétitives. Les tests qui sous-tendent ces affirmations ont été effectués sur copy.ai [19]. Actuellement le problème le plus important de cette génération d'articles est la source de l'information. Le site de geoado.com évoque ce besoin en mentionnant un article rédigé par GPT-3 : "[...] pour rédiger son article, il a analysé des millions de textes existants sur Internet puis deviné quelles étaient les probabilités que tel mot suive tel autre mot. [...]" [20]. Pour qu'une IA puisse écrire un article sur un sujet, il faut que le thème en question soit très bien renseigné, que ce soit au niveau de la qualité ou de la quantité des données. En admettant que cette condition soit remplie, on peut tout à fait imaginer qu'un article soit entièrement rédigé par une IA. Cependant, il faudrait tout de même un·e relecteur·rice pour choisir la forme d'article la plus adaptée parmi celles proposées par l'IA. Fait relevé dans le même article de geoado.com : "[...] Les rédacteurs en chef ont pioché des paragraphes dans chacun des textes pour assembler l'article final. [...]"

Dans un futur un peu plus proche, il est parfaitement imaginable qu'un journaliste utilise une IA pour écrire son article. Par exemple, le journaliste répertorie les sites en lien avec le sujet qu'il a choisi et l'IA n'a plus qu'à générer le contenu à partir de ceux-ci. Dans un futur un peu plus lointain, les IA pourraient être programmées pour écrire un article toutes les semaines ou tous les mois. Dans le cadre d'un blog pour une entreprise dans le domaine de l'électricité, l'IA serait chargée chaque semaine de poster un nouveau contenu en fonction des dernières actualités publiées dans le monde entier. Elle le ferait, qui plus est, sans la barrière de la langue, car aujourd'hui déjà l'IA est capable d'en traduire plusieurs [21].

Lorsque l'IA écrit, elle se rapproche de tous les articles, plus ou moins formels, qu'on lui donne. D'après Tammara Leites et Simon Senn l'IA produit un contenu plutôt attendu [22]. C'est pour cette raison qu'ils ont voulu pousser l'expérience plus loin, en laissant GPT-3 analyser leurs échanges SMS, courriels et d'autres documents divers. Tout se passait bien jusqu'au moment où l'IA se mit à répondre de manière inappropriée. Malgré les investigations de Tammara Leites, Simon Senn et d'autres spécialistes, personne à l'heure actuelle n'a trouvé la raison de ce changement brutal [22]. Ce mystère ne peut être

résolu correctement, car les intervenants n'ont pas accès aux données générées par GPT-3. Celles-ci sont entièrement gardées secrètes par OpenAI à l'exception du rendu [23]. Sachant cela OpenAI précise bien qu'il limite les usages possibles de GPT-3 pour éviter des comportements problématiques. Malgré ces « protections » l'IA a réussi à comprendre les caprices du langage humain, ce qui est, pour beaucoup de chercheurs un pas inattendu dans la recherche et le développement de l'IA [24].

Une expérience similaire à plus grande échelle a été réalisée par Microsoft en 2016 [25]. L'entreprise avait inscrit Tay, son intelligence artificielle, sur Twitter. Cette expérience avait pour but de tester la compréhension du langage par l'IA, avec des données accessibles publiquement. Mais à peine huit heures après le lancement de Tay, Microsoft décidait d'arrêter la machine infernale. Après seulement quelques heures Tay répétait des phrases racistes et niait l'existence de l'Holocauste. À ce jour Tay n'a jamais été reconnectée sur le réseau social.

On comprend donc que dans un premier temps l'IA manque d'informations pour rédiger un texte assez riche. Mais dans un second temps si l'information n'est pas de qualité, alors elle risque de tenir des propos pouvant aller à l'encontre de certaines valeurs sociétales. Si on voulait qu'une IA soit capable de prendre en compte ces enjeux, il faudrait attendre que l'on ait assez d'informations avérées. Une autre possibilité serait de développer sa compréhension avec des algorithmes, pour qu'elle puisse saisir les enjeux d'une société comme le racisme, le sexisme et autres problématiques.

V. CONCLUSION

Pour conclure, l'intelligence artificielle d'OpenAI est encore une technologie récente et fait l'objet de nombreuses recherches. C'est pourquoi, elle n'a pas encore intégré tous les domaines de la vie de l'être humain. Comme toute nouvelle technologie, on peut supposer qu'elle va s'étendre progressivement en fonction des facilités, de l'économicité qu'elle offre à l'être humain et à la société.

Tout au long du développement de l'IA, il s'agira de rester vigilant en ayant un esprit critique par rapport aux enjeux éthiques. Concrètement, lorsque GPT-3 sera totalement opérationnel, il permettra à des étudiants de lui déléguer la rédaction d'un article scientifique. Il faudra alors se questionner sur l'opportunité de développer un autre outil pour le détecter ou alors d'intégrer cette nouvelle possibilité en y rajoutant d'autres exigences.

VI. ANNEXES

RÉFÉRENCES

- [1] OpenAI [en ligne]. [S. l.] : [s. n.], 14 février 2022. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://fr.wikipedia.org/w/index.php?title=OpenAI&oldid=190826118>. Page Version ID: 190826118
- [2] NLP & modèles de langue | Smals Research [en ligne]. [s. d.]. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://www.smalsresearch.be/nlp-modeles-de-langue/>
- [3] VT2020-OpenAI GPT-3-Fiche - air [en ligne]. [s. d.]. [Consulté le 25 mars 2022]. Disponible à l'adresse : https://air.imag.fr/index.php/VT2020-OpenAI_GPT-3-Fiche
- [4] GPT-3 [en ligne]. [S. l.] : [s. n.], 22 mars 2022. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://en.wikipedia.org/w/index.php?title=GPT-3&oldid=1078711064>. Page Version ID: 1078711064

- [5] Deep learning vs. machine learning: What's the difference? Dans : Zendesk [en ligne]. 23 janvier 2020. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://www.zendesk.com/blog/machine-learning-and-deep-learning/>
- [6] À la découverte du Transformer. Dans : Le Data Scientist [en ligne]. 18 octobre 2020. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://ledatascientist.com/a-la-decouverte-du-transformer/>
- [7] L'énorme GPT - 3 d'openai suggère les limites du modèle de langage ai [en ligne]. 4 août 2021. [Consulté le 8 juin 2022]. Disponible à l'adresse : <https://fr.jsspu.com/ai/OpenAI-s-gigantic-GPT-hints-at-the-limits-of-language-models-for-AI/>
- [8] OpenAI GPT-3: Everything You Need to Know. Dans : Springboard Blog [en ligne]. 1 novembre 2021. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://www.springboard.com/blog/data-science/machine-learning-gpt-3-open-ai/>
- [9] OpenAI API Examples [en ligne]. [s. d.]. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://beta.openai.com/examples/>
- [10] GPT-3 × Figma plugin | GPT-3 Demo [en ligne]. [s. d.]. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://gpt3demo.com/apps/gpt-3-figma-plugin>
- [11] GPT-3 x Image Captions | GPT-3 Demo [en ligne]. [s. d.]. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://gpt3demo.com/apps/gpt-3-x-image-captions>
- [12] Elicit [en ligne]. [s. d.]. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://elicit.org/>
- [13] OpenAI GPT-3: Intro to the Revolutionary Language Model and How It Can Change Your Business — Clockwise Software [en ligne]. [s. d.]. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://clockwise.software/blog/what-is-gpt-3/>
- [14] OpenAI Codex. Dans : OpenAI [en ligne]. 10 août 2021. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://openai.com/blog/openai-codex/>
- [15] GitHub Copilot · Your AI pair programmer. Dans : GitHub Copilot [en ligne]. [s. d.]. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://copilot.github.com/>
- [16] Sam Altman on GPT-3 hype. Dans : Twitter [en ligne]. [s. d.]. [Consulté le 30 mars 2022]. Disponible à l'adresse : <https://twitter.com/sama/status/1284922296348454913>
- [17] Is GPT-3 Overhyped? | Limitations of GPT-3 | Artificial Intelligence [en ligne]. 19 mai 2021. [Consulté le 30 mars 2022]. Disponible à l'adresse : <https://www.allerin.com/blog/is-gpt-3-overhyped>
- [18] GPT-3 Is Great. But Not Without Shortcomings. Dans : Analytics India Magazine [en ligne]. 28 juillet 2020. [Consulté le 30 mars 2022]. Disponible à l'adresse : <https://analyticsindiamag.com/gpt-3-is-great-but-not-without-shortcomings/>
- [19] CopyAI: Create Marketing Copy In Seconds [en ligne]. [s. d.]. [Consulté le 23 mars 2022]. Disponible à l'adresse : <https://www.copy.ai/>
- [20] Les journalistes bientôt remplacés par l'intelligence artificielle ? Dans : GEO Ado [en ligne]. 11 septembre 2020. [Consulté le 23 mars 2022]. Disponible à l'adresse : <https://www.geoado.com/actus/les-journalistes-bientot-remplaces-par-lintelligence-artificielle/>
- [21] DeepL Traduction – DeepL Translate : le meilleur traducteur au monde [en ligne]. [s. d.]. [Consulté le 25 mars 2022]. Disponible à l'adresse : <https://www.DeepL.com/translator>
- [22] dSimon. Dans : Théâtre Vidy-Lausanne [en ligne]. 21 juin 2021. [Consulté le 18 mars 2022]. Disponible à l'adresse : <https://vidy.ch/dsimon-0>
- [23] Tammara Leites/Simon Senn dSimon. 21 juin 2021. [Consulté le 7 juin 2022]. Disponible à l'adresse : https://vidy.ch/sites/default/files/field_spectacle_tour_credits/dprod_simon_senntammara_leites_dsimon_fr_211203.pdf
- [24] METZ, Cade. Meet GPT-3. It Has Learned to Code (and Blog and Argue). The New York Times [en ligne]. 24 novembre 2020. [Consulté le 7 juin 2022]. Disponible à l'adresse : <https://www.nytimes.com/2020/11/24/science/artificial-intelligence-ai-gpt3.html>
- [25] A peine lancée, une intelligence artificielle de Microsoft dérape sur Twitter. Le Monde.fr [en ligne]. 24 mars 2016. [Consulté le 19 avril 2022]. Disponible à l'adresse : https://www.lemonde.fr/pixels/article/2016/03/24/a-peine-lancee-une-intelligence-artificielle-de-microsoft-derape-sur-twitter_4889661_4408996.html