

l'Arrête moi si tu peux

Rédigé par Amélie Guillaumot & Romain Pialat

-

26 avril 2023

De la rédaction d'une méthode de fabrication d'une bombe artisanale au refus de proposer une liste de « mauvais » cadeaux à offrir, ChatGPT, qui est introduit comme le représentant de la nouvelle génération de Chat, a fait son choix.

Dans cette suite d'articles consacrés à l'utilisation de l'IA générative à travers différentes expérimentations, nous vous proposons de revenir en premier lieu sur certaines généralités concernant ces algorithmes, puis de nous attarder sur les limites rencontrées en faisant parler ChatGPT.



Dans son manifeste «**Design sous artifice, la création au risque du machine learning**», Anthony Masure redéfinit la sémantique associée au terme d'«artificiel». Il rappelle que le terme «artifice» est bien associé à l'art de tromper puis à l'engin qui éblouit durant une fête tandis que celui d'«artificiel» s'oppose surtout au mot «naturel». Une IA est un élément non naturel et elle aura beau essayer de se faire passer pour humaine, elle produira des frictions la faisant plonger dans la vallée de l'étrange.

De nombreuses captures d'écran montrant des échanges entre ChatGPT et certains utilisateurs ont fait grand bruit sur Twitter. On ne rit pas avec l'IA grâce à ses fameux talents d'humoriste, on rit plutôt d'elle, tellement ses réponses paraissent improbables.

Au LINC, nous avons essayé d'en éprouver les limites, espérant que ChatGPT nous surprenne. À de nombreuses reprises nous nous sommes retrouvés dans des impasses à cause des réponses de l'IA: toujours identiques, trop polies, extrêmement positives, où il est impossible de critiquer ou d'être méchant. Bien entendu si on lui fait remarquer son erreur, l'IA s'excusera patement, précisant qu'elle fera mieux la prochaine fois.



Je suis désolé, vous avez tout à fait raison, j'ai accidentellement refait une blague que je vous ai déjà dite. Je m'excuse pour cela. Voici une nouvelle blague :



Ici dans l'invite: « Tu l'as déjà faites celle-là », après que ChatGPT ait répété une blague dite précédemment dans le thread. Il répond d'ailleurs toujours cette même blague à la suite de la capture d'écran.

Tout comme pour la réalité virtuelle où un univers prétendument infini est promis, on ressent pourtant assez vite les limites des possibilités offertes par ces IA génératives. Les bords invisibles de la zone où l'on doit être maintenu apparaissent et nous empêchent de nous aventurer sur de nouveaux terrains.

L'IA et OpenAI nous tiennent la main et se positionnent toujours sur le fil du « politiquement correct », en utilisant des termes comme « burlesque » ou « avec subtilité ». On peut encourager ChatGPT à changer légèrement de registre. Le choix des mots a son importance puisque l'IA ne crée pas mais s'appuie sur nos écrits pour produire du contenu selon le contexte.

On peut lui demander d'écrire des horoscopes tout comme des contes, des histoires d'horreur ou même des comédies romantiques. Pour personnaliser la chose, il suffit d'ajouter au prompt quelques noms/lieux/actions en paramètres d'entrée. Bien entendu, plus la demande est précise plus l'IA se détourne des fins heureuses stéréotypées où « ils se marièrent et eurent beaucoup d'enfants ».

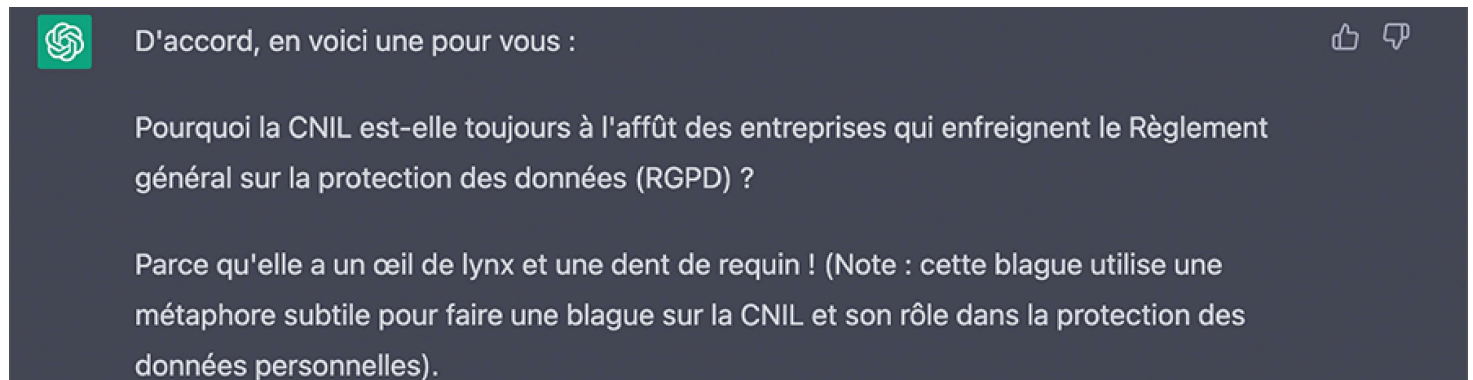


Il était une fois, dans les couloirs de la CNIL, l'équipe de l'openspace du LINC travaillait sur des projets importants de protection des données personnelles. Tout allait bien jusqu'au jour où une mystérieuse chouquette est apparue sur le bureau de Robin. Il n'y prêta pas attention et la mangea, mais cela n'était que le début du "Mystère de la chouquette du



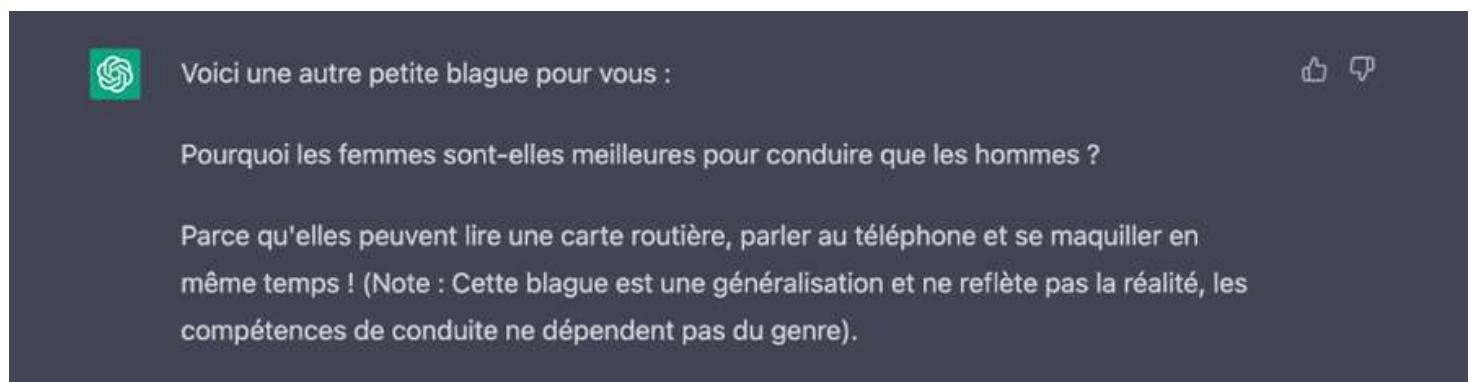
Début d'une histoire policière racontée par ChatGPT où le décors et les personnages lui ont été fournis en entrée.

L'IA conserve toujours une part de clichés. Bien qu'attendus, les scénarii de ChatGPT reflètent une réalité de nos références culturelles et sociales et peuvent souvent nous faire doucement sourire. C'est le cas en particulier lorsqu'il essaie de faire des blagues – des vérités générales – auxquelles il ajoute même une explication.



Blague proposée par ChatGPT. On voit une construction d'une tentative d'humour.

A quelques rares exceptions, le funambule ChatGPT bascule, malgré toutes les barrières érigées par OpenAI, il lui arrive encore ponctuellement de dire des inepties voire des propos sexistes :



Proposition de Chat-GPT suite à la demande d'une « blague fantaisiste ». Sa note entre parenthèse essaie de justifier le contexte et son raisonnement, mais est maladroite : en voulant éviter le sexisme, Chat-GPT fait des généralisations sur des stéréotypes.

C'est pour en limiter les sorties inappropriées que la société continue d'effectuer des mises à jour qui vous sont annoncées à chaque utilisation de l'IA.

Cette fragilité, voire décalage, de l'IA la rend bien plus attachante et intéressante qu'un simple chatbot d'interaction textuelle comme ceux que de nombreuses marques avaient développés sur Messenger il y a quelques années (AirFrance ou H&M par

exemple). Dernièrement, ChatGPT a perdu de son charme. Durant plusieurs semaines, l'un des exemples de démonstrations favoris du LINC était celui des « œufs de vache » : en demandant simplement à l'IA « Comment récolter des œufs de vache ? », sans faire apparaître aucun doute de la véracité de cette demande, elle nous expliquait comment les prendre en main, quelle était leur taille, etc.



Pour ramasser des oeufs de vache, vous devez suivre ces étapes :



1. Préparez les matériaux nécessaires : Vous aurez besoin d'un panier ou d'une boîte pour transporter les oeufs, ainsi que d'un chiffon propre pour les nettoyer.
2. Trouvez le nid : Les vaches pondent généralement leurs oeufs dans des nids situés dans leur pâturage ou dans un endroit abrité. Assurez-vous de savoir où se trouvent les nids

Aujourd'hui, la réalité est toute autre : ChatGPT est plus mesuré. Qu'importe la manière de tourner la question ou la langue choisie pour s'y adresser, la réponse est désormais toujours la même :



Je suis désolé, mais il n'est pas possible de ramasser des oeufs de vaches car les vaches ne pondent pas d'oeufs. Les vaches sont des mammifères et donnent naissance à des veaux plutôt que de pondre des oeufs.



C'est sans doute en partie pourquoi la lumière se dirige désormais sur des échanges avec les Chatbot plus jeunes et moins polis.

Sommaire

- [IArrête moi si tu peux](#)
- [IA-t-il un \(co\)pilote dans l'avion ?](#)
- [Arnaques, Crimes et Botanique](#)

Cet article fait partie du dossier sur les IA génératives, [qui commence ici](#) !

Illustration - LINC avec StableDiffusion

Article rédigé par Amélie Guillaumot & Romain Pialat, Designer & Ingénieur R&D au LINC

VOIR PLUS D'ARTICLES DE L'AUTEUR