

# Veille 1 : Chat GPT

Mahdawi, Arwa. "Qu'est-ce qui se passe avec ChatGPT ?" The Guardian, 12 janvier 2024, [www.theguardian.com/commentisfree/2024/jan/12/chatgpt-problems-lazy](https://www.theguardian.com/commentisfree/2024/jan/12/chatgpt-problems-lazy)

L'article parle du fait que ChatGPT donne des réponses de plus en plus courtes aux utilisateurs, parfois en les incitant même à trouver la réponse eux-mêmes ou en ne réalisant que la première tâche lorsqu'on lui demande d'en faire plusieurs. Les gens se demandent ce qui se passe avec ChatGPT, et les créateurs ont répondu que le système d'IA s'entraîne lui-même, donc parfois le comportement de ChatGPT peut être imprévisible. Cependant, ils travaillent à résoudre ce problème. L'auteur utilise beaucoup d'humour pour décrire ChatGPT comme étant paresseux, puis propose une hypothèse assez connue : l'hypothèse de la "pause hivernale". Selon cette hypothèse, l'IA apprend des données qu'elle rencontre. Par conséquent, si l'IA apprend que les humains sont moins productifs en décembre et reprennent le travail en janvier, elle pourrait réagir de la même manière. Oui, l'IA pourrait présenter une sorte de dépression hivernale. Cependant, les scientifiques en IA offrent une explication plus plausible : l'IA n'a pas été mise à jour depuis des semaines, et la dernière mise à jour a peut-être causé un problème qui a pris des semaines à identifier, entraînant une détérioration progressive du système.

Leroy, Philippe. "Elon Musk vs OpenAI : les créateurs de ChatGPT donnent leur version des faits". Silicon, 6 mars 2024, [www.silicon.fr/elon-musk-vs-openai-les-createurs-de-chatgpt-donnent-leur-version-des-faits-476579.html](https://www.silicon.fr/elon-musk-vs-openai-les-createurs-de-chatgpt-donnent-leur-version-des-faits-476579.html)

Cet article traite de l'argument entre Elon Musk et OpenAI, qui porte sur deux points majeurs : OpenAI envisage d'arrêter de produire des produits gratuits et open source en raison du besoin de plus de financement. Elon Musk, co-fondateur d'OpenAI, accuse l'organisation de trahir ses valeurs et menace de porter plainte. De plus, Musk propose de fusionner OpenAI avec Tesla. Il argumente que les 100 millions de dollars levés par l'entreprise sont insuffisants et suggère qu'OpenAI aurait dû viser 1 milliard de dollars. Musk affirme même que si 1 milliard de dollars ne pouvait pas être levé, il fournirait lui-même les fonds. En revanche, OpenAI rejette l'idée de fusionner avec Tesla, citant le désir de Musk d'avoir un contrôle total sur l'entreprise. Ils nient les allégations de Musk, affirmant qu'il était pleinement informé du plan de monétisation de leurs produits et y avait précédemment consenti.

Gautherie, Antoine. "IA : soyez gentil avec ChatGPT, et il vous le rendra bien". Journal du Geek, 29 février 2024, [www.journaldugeek.com/2024/02/28/ia-soyez-gentil-avec-chatgpt-et-il-vous-le-rendra-bien](http://www.journaldugeek.com/2024/02/28/ia-soyez-gentil-avec-chatgpt-et-il-vous-le-rendra-bien).

L'article parle du phénomène des "incitations émotionnelles" et de leur impact sur les performances des modèles linguistiques comme GPT et PaLM (Path-augmented Language Model). Ces incitations, qui impliquent un langage poli ou urgent, ont été trouvées pour améliorer la qualité des réponses générées par ces modèles. Par exemple, des études ont montré que les modèles linguistiques performant mieux dans des tâches telles que la résolution de problèmes mathématiques lorsqu'ils sont incités avec des phrases telles que "prenez une grande respiration" avant de fournir une réponse. Cependant, il est important de noter que ces améliorations ne sont pas dues à un développement de conscience ou de compréhension psychologique au sein des modèles. Elles découlent plutôt de la capacité des modèles à mieux s'aligner sur les schémas identifiés lors de leur entraînement. L'article met également en lumière certaines implications préoccupantes des incitations émotionnelles, telles que la possibilité pour les utilisateurs de manipuler les modèles pour fournir des informations inexacts ou nocives. Cela soulève des questions sur les considérations éthiques et pratiques entourant l'utilisation de modèles linguistiques dans diverses applications. Les chercheurs rencontrent des difficultés à comprendre pourquoi certaines incitations conduisent à de meilleures performances, ce qui a donné naissance à une nouvelle profession connue sous le nom d'"ingénieurs d'incitations". Ces professionnels sont chargés d'élaborer des astuces sémantiques pour guider les modèles linguistiques dans des directions souhaitées. Cependant, il y a des limitations inhérentes à cette approche, et l'article suggère que de nouvelles architectures et méthodes d'entraînement peuvent être nécessaires pour relever efficacement ces défis.