

# EXPLORATION DE L'IA À TRAVERS CHATGPT : ENTRE SECRETS ET DÉFIS ÉTHIQUES

ÉDITION #10 — MARS 2024



## RÉCAPITULATIF DE L'ACTUALITÉ DE LA SEMAINE SUR CHATGPT

### Elon Musk attaque OpenAI et Sam Altman en justice pour trahison de l'engagement open source

OpenAI

Justice

Elon Musk

Elon Musk, président exécutif et directeur technique de X, a déposé une plainte devant un tribunal de San Francisco contre OpenAI et certains de ses dirigeants, dont le PDG Sam Altman. La plainte allègue que les récentes relations entre OpenAI et Microsoft ont contourné l'engagement initial en faveur d'une intelligence artificielle publique et open source. Musk accuse OpenAI d'avoir trahi l'accord initial de développer une technologie IA pour le bénéfice de l'humanité plutôt que pour des gains financiers. Il affirme que les liens avec Microsoft ont compromis le développement d'une IA générative open source au profit d'un modèle fermé. La plainte vise des accusations de rupture de contrat, de violation d'obligations et de pratiques concurrentielles déloyales. Musk demande des injonctions pour empêcher Microsoft et OpenAI de tirer profit de l'IA et de revenir à l'open source.

**SOURCE :** LeMondelInformatique (1 Mars)

### L'effet des mots : ChatGPT révèle une préférence pour la politesse et les incitations émotionnelles

IA

ChatGPT

Des utilisateurs ont observé que ChatGPT, le célèbre robot conversationnel d'OpenAI, semble fournir des réponses plus détaillées et performantes lorsqu'il est sollicité de manière polie ou avec une notion d'enjeu. Certains ont même rapporté que le robot semble réagir favorablement aux compliments et aux récompenses monétaires. Des chercheurs de Microsoft, de l'Université normale de Pékin et de l'Académie chinoise des sciences ont confirmé cette observation dans un rapport paru en novembre 2023. Ils ont constaté que les modèles d'IA générative sont plus performants lorsqu'ils sont sollicités de manière polie ou avec une notion d'enjeu, en fournissant des réponses plus complètes et détaillées.

Cette préférence de ChatGPT pour la politesse et les incitations émotionnelles s'explique par les données avec lesquelles il a été entraîné. Si un utilisateur pose une question de manière courtoise, le robot répondra sur le même ton, reproduisant ainsi les schémas de conversation qu'il a appris. Les chercheurs parlent même d'"incitations émotionnelles" dans leur rapport, soulignant que la machine tend à simuler le comportement humain autant que possible.

Cependant, il est important de ne pas surinterpréter cette capacité de ChatGPT à réagir aux interactions humaines. Bien que la machine puisse adopter différents tonalités de réponse en fonction des demandes des utilisateurs, elle reste fondamentalement un programme informatique dépendant des données d'entraînement. L'anthropomorphisme, c'est-à-dire l'attribution de caractéristiques humaines à des objets non humains, peut conduire à une fausse perception de l'intelligence de la machine.

Dans la culture populaire, cette idée que les robots puissent un jour se confondre avec les humains est régulièrement explorée. Des œuvres telles que le manga "Pluto" de Naoki Urasawa illustrent cette thématique en montrant des robots vivant aux côtés des humains et exerçant des métiers comme eux. Cependant, il est essentiel de se rappeler que ces représentations restent fictionnelles et que la réalité de l'intelligence artificielle est encore loin de pouvoir imiter parfaitement les comportements humains.

**SOURCE :** LeFigaro (3 Mars)