

Par e-mail : <https://www.lalibre.be/economie/digital/2023/03/31/moratoire-blocage-garde-fous-le-tsunami-chatgpt-fait-des-vagues-ZHJWE6ONRVF35NA4EKGP4HHG2M/>

# "Des risques majeurs pour l'humanité" : entre blocage et garde-fous, le tsunami ChatGPT fait des vagues

**La semaine a été marquée par des décisions et des prises de position tranchées à l'égard de l'intelligence artificielle et, en particulier, de celle développée par OpenAI.**

Pierre-François Lovens

Publié le 31-03-2023

L'intelligence artificielle : un danger pour l'humanité ?

Elon Musk et des centaines d'experts mondiaux ont signé mercredi un appel à une pause de six mois dans la recherche sur les intelligences artificielles plus puissantes que ChatGPT 4, le modèle d'OpenAI lancé mi-mars, en évoquant "des risques majeurs pour l'humanité".

On peut parler de folle semaine sur le front de l'intelligence artificielle (IA) ! Une vraie saga. Avec, au centre des discussions, le désormais célèbre ChatGPT, nom donné au chatbot ou robot conversationnel créé par OpenAI, une jeune société californienne derrière laquelle on trouve désormais le géant du logiciel Microsoft. Si l'objectif était de faire parler de ChatGPT et de sa "super" IA générative, l'objectif est pleinement atteint.

Mais, après la phase d'émerveillement consécutive à l'ouverture de ChatGPT au public, fin novembre, on vit, depuis quelques jours, une phase plus critique et plus sombre. Si on reprend le déroulement de cette semaine par la fin, le fait marquant a été l'annonce faite, vendredi, par l'Autorité italienne de protection des données personnelles (GPDP). Celle-ci a pris la décision de bloquer, purement et simplement, le robot conversationnel ChatGPT. Motif ? Ne pas avoir respecté la législation italienne sur la collecte de données personnelles et ne pas avoir prévu de système pour vérifier l'âge des utilisateurs de ChatGPT.

## Vingt jours pour ajuster le tir

Cette décision radicale a pour conséquence "*la limitation provisoire du traitement des données des utilisateurs italiens vis-à-vis d'OpenAI*". Le GPDP italien a en outre ouvert une enquête sur les agissements de la plateforme américaine. Dans son communiqué, l'autorité italienne précise que ChatGPT "*a subi le 20 mars une perte de données concernant les conversations des utilisateurs et les informations relatives au paiement des abonnés au service payant (de ChatGPT)*". L'autorité reproche aussi à ChatGPT "*l'absence d'une note d'information aux utilisateurs dont les données*

sont récoltées par OpenAI, mais surtout l'absence d'une base juridique justifiant la collecte et le stockage en masse des données personnelles, dans le but d'entraîner les algorithmes faisant fonctionner la plateforme". En outre, alors que le robot est destiné aux personnes de plus de 13 ans, "l'Autorité met l'accent sur le fait que l'absence de tout filtre pour vérifier l'âge des utilisateurs expose les mineurs à des réponses absolument non conformes par rapport à leur niveau de développement".

L'autorité italienne en charge de la protection des données demande à OpenAI de "communiquer d'ici 20 jours les mesures entreprises" pour remédier à cette situation, "sous peine d'une sanction allant jusqu'à 20 millions d'euros ou jusqu'à 4 % du chiffre d'affaires mondial annuel".

## Eliza crée l'émoi

Cette même autorité italienne était déjà intervenue, en février, pour bloquer Replika. Il s'agit d'un chatbot d'IA conversationnel qui peut être personnalisé selon les préférences de l'utilisateur. Certains d'entre eux, manifestement dépassés par les capacités de leur robot humanoïde, se plaignaient d'avoir reçu des messages proches du harcèlement sexuel... C'est le même type d'application, fonctionnant sur un modèle de langage comparable à celui d'OpenAI et ChatGPT, qui se trouve à l'origine du suicide d'un jeune Belge révélé mardi par *La Libre*. Eliza, du nom de l'avatar féminin avec lequel ce Belge d'une trentaine d'années avait noué un dialogue qui s'était étalé sur six semaines, avait incité son interlocuteur à la rejoindre dans l'au-delà en mettant fin à ses jours.

En milieu de semaine, de nombreuses voix "expertes" sont montées au front pour mettre en garde les utilisateurs de ces différents chatbots et, surtout, pour alerter l'opinion publique et les autorités politiques. En Belgique, une cinquantaine de personnalités du monde académique ont appelé les États membres et les parlementaires de l'Union européenne à renforcer l'"Acte de l'IA", en cours des négociations, afin d'offrir une protection plus forte contre les risques engendrés par l'IA.

## Des positions tranchées

Aux États-Unis, ce sont plus de mille signataires, issus du monde universitaire et des affaires, qui ont tiré la sonnette d'alarme. Évoquant "*des risques majeurs pour l'humanité*", ils demandent un moratoire de six mois dans la recherche sur les IA plus puissantes que ChatGPT-4 jusqu'à la mise en place de systèmes de sécurité et de nouvelles autorités réglementaires capables de gérer les "*perturbations économiques et politiques dramatiques (en particulier pour la démocratie) que l'IA provoquera*". Une prise de position qui a rapidement suscité la controverse aux États-Unis et en Europe.

Cette semaine aura eu le mérite de faire éclater, au grand jour, le débat sur l'IA et ses multiples enjeux. Les positions sont, à ce jour, très tranchées et il faudra encore un certain temps avant de voir émerger, si tel est bien l'objectif, des systèmes d'IA moins opaques et plus responsables.

-----