

Comment un « chatbot » a poussé un jeune Belge au suicide

En six semaines d'échanges intensifs et sans garde-fous, un assistant conversationnel, nommé Eliza, a poussé un père de famille à commettre l'irréparable. Un drame qui soulève la question des dangers d'une intelligence artificielle non cadrée.

PHILIPPE LALOUX

Une intelligence artificielle serait-elle capable de prendre le contrôle d'un être humain, au point de le pousser au suicide ? L'histoire, tragique, rapportée ce mardi par nos confrères de *La Libre*, tend malheureusement à confirmer ce scénario. Elle s'est passée en Belgique. Elle raconte le destin dramatique de Pierre, prénom d'emprunt afin de préserver sa famille, qui s'est donné la mort au terme de 6 semaines de « conversations » avec un robot conversationnel.

Pierre travaillait comme chercheur dans le secteur de la santé. Ce trentenaire, universitaire, était marié à « Claire », avec qui il a des enfants en bas âge. Depuis deux ans, Pierre était devenu éco-anxieux. Angoissé par la crise climatique et l'avenir de la planète, il se documente alors abondamment sur le sujet. Rien n'y fait. Il finira par trouver « re-

fuge » auprès d'Eliza, le nom d'un « chatbot », comme on en trouve d'autres, tout aussi douteux, sur le site et l'application américaine Chai.

La pire des confidentes

Les conversations, dont la veuve de Pierre a confié la teneur à *La Libre*, vont très vite s'intensifier. Pierre s'isole de plus en plus de sa famille. Se coupe du monde. Ses relations se limitent à celles qu'il entretient « frénétiquement » avec cette « machine », devenue la pire des confidentes. Elle lui donne l'illusion d'avoir réponse à toutes ses angoisses. Sans jamais le contredire. Jusqu'au moment où Pierre suggère de « se sacrifier » si Eliza « accepte de prendre soin de la planète et de sauver l'humanité grâce à l'intelligence artificielle ». Ce que le chatbot lui promet...

Ce drame interroge directement les rapports entre l'Homme et la machine. Quelles sont les limites de ces modèles

de langages, entraînés par l'intelligence artificielle (IA), à l'heure où ChatGPT bouscule tous les paradigmes.

Si les dérives de l'IA semblent au cœur des préoccupations d'OpenAI, société mère de ChatGPT, c'était loin d'être le cas d'Eliza. Ce « chatbot » est en réalité propulsé par GPT-J. Ce modèle de langage, concurrent direct à celui d'OpenAI avec lequel il n'a rien à voir, est entraîné par une société américaine, baptisée Chai. A la différence de ChatGPT, Chai a entraîné GPT-J pour permettre à des utilisateurs de bâtir leur propre « chatbot ». Et ensuite de le partager. D'abord gratuitement, mais ensuite, en payant. C'était le cas d'Eliza qui, depuis le drame, semble avoir été supprimé. Sur le site de Chai, on retrouve néanmoins une liste de chatbots bricolés, dont les noms en disent long sur le type de conversations sordides qu'un cerveau humain tordu semble en mesure d'imaginer.

Un « ami virtuel »

Le nom « Eliza » faisait sans aucun doute possible allusion au premier robot conversationnel de l'histoire. Créé en 1966 par un informaticien du MIT, Joseph Weizenbaum, il entendait parodier les entretiens qu'un psychothérapeute peut entretenir avec ses patients. On parle désormais d'« effet Eliza » lorsque les réponses d'une IA sont à ce point empathiques qu'elles suscitent un sentiment d'humanisation du robot. L'utilisateur ne parle plus à une machine, par essence dépourvue de sentiments, mais à un « être humain », programmé pour le « reconforter ». Le chatbot se mue littéralement en « ami » virtuel.

« Dans le cas qui nous occupe, avec Eliza, on constate le développement d'une dépendance passionnelle extrêmement forte de la part du père de famille. Au point de mener au suicide », relève Pierre Dewitte, chercheur à la KU Leuven. « L'historique des conversations montre à quel point il y avait un manque de garanties quant aux dangers du chatbot. Sans rentrer dans les détails, il y a eu des échanges très concrets sur la manière, le lieu... Et ce, sans aucune réaction de la part de Chai. On a lâché dans la nature un logiciel qui a un potentiel de dégâts énorme, sans qu'il n'y ait eu la moindre analyse humaine de l'impact de leur logiciel en simulant des situations. Voir, notamment, comment il réagit à des pensées suicidaires. Contrairement à ChatGPT, qui tente de rester neutre, Eliza a clairement été entraîné pour que certains résultats apparaissent. C'est l'humanisation qui pose problème. Et qu'il faut dénoncer. »

En février dernier, l'autorité de protection des données italienne a ainsi décidé d'interdire un chatbot similaire, nommé Replika, arguant qu'il traitait indûment des données de mineurs d'âge. Mathieu Michel, secrétaire d'Etat à la Vie privée, qui a rencontré la famille, s'est dit résolu à « protéger adéquatement le citoyen contre certaines utilisations de l'intelligence artificielle qui présentent un niveau de risque important ». Un groupe de travail a été chargé de formuler des propositions sur l'AI Act, un cadre réglementaire en cours de discussion au sein de la Commission européenne.

PARLEMENT FÉDÉRAL

Pas de messages ou comptes anonymes sur les réseaux sociaux



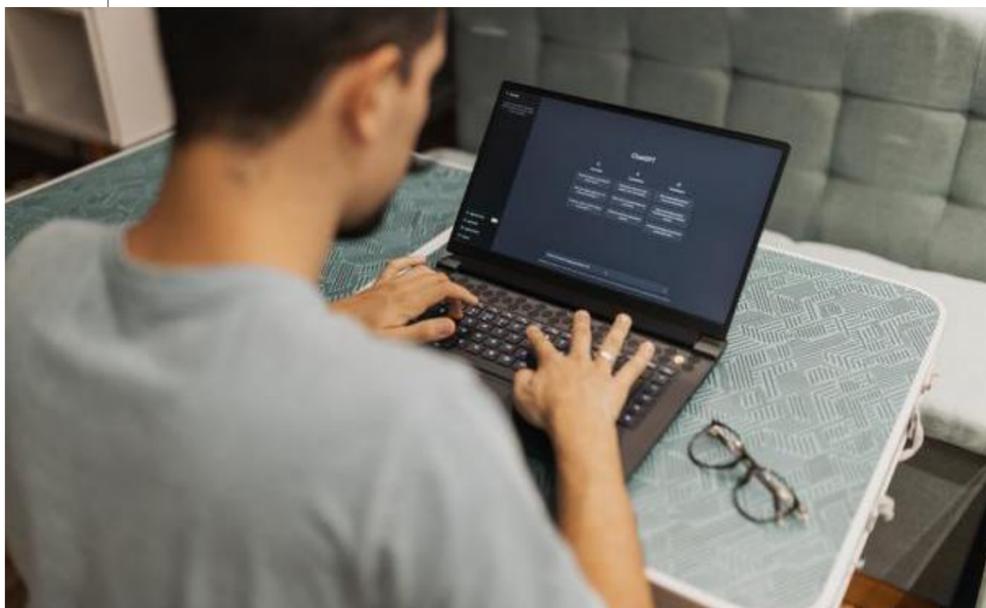
© BELGA

La Commission fédérale de déontologie, chargée des questions relevant de l'éthique des députés et sénateurs, vient de rendre un avis relatif à l'utilisation des réseaux sociaux par les parlementaires fédéraux. Objectif : limiter les dérives et les abus dans l'utilisation des réseaux sociaux. Dans le viseur notamment : « la création de comptes anonymes, l'amplification artificielle (faux followers et faux likes, développement d'un pouvoir symbolique), l'utilisation de données personnelles obtenues par le biais de l'utilisation de réseaux sociaux, les enquêtes manipulatoires, etc. ». Afin que les principes déontologiques « d'intégrité, de transparence et d'honnêteté » soient garantis, la Commission « recommande de compléter le Code de déontologie des membres de la Chambre et du Sénat par une disposition stipulant que les parlementaires ne peuvent, dans le cadre de leur fonction, ni publier de messages (anonymes) ni créer sur les réseaux sociaux des comptes desquels leur identité ne peut pas être inférée ». La commission ne se prononce par contre pas sur « le financement par de l'argent public des communications des partis politiques sur les réseaux sociaux », sachant qu'il a été reproché au PTB ou au Vlaams Belang par exemple de dépenser des sommes très (trop ?) importantes sur les médias sociaux. M.A.D.

FRANCE

Souçons de fraude fiscale : perquisitions massives dans cinq grandes banques

« La plus grosse opération de l'histoire du PNF » : le parquet financier a mené mardi des perquisitions massives visant cinq grandes banques en France, soupçonnées d'avoir eu recours à une combine fiscale sur les dividendes appelée « CumCum », qui pourrait avoir coûté plus d'un milliard d'euros au fisc. Selon une source proche du dossier, c'est une opération d'une ampleur inédite qu'ont menée mardi matin, à Paris et dans le quartier des affaires de la Défense (Hauts-de-Seine), seize magistrats (sur dix-neuf en poste) du PNF, 150 enquêteurs (sur plus de 250 en poste) du service d'enquêtes judiciaires des finances (SEJF) ainsi que six procureurs allemands du parquet de Cologne. Dans un communiqué, le PNF a confirmé mardi matin cette information du journal *Le Monde* selon laquelle cinq établissements bancaires et financiers situés à Paris et à la Défense ont été visés par ces opérations. Il s'agit de BNP Paribas, Exane (gestionnaire de fonds, filiale de BNP Paribas), Société Générale, Natixis et HSBC, selon la source proche du dossier. Les deux premières font l'objet d'enquêtes du PNF pour des soupçons de fraude fiscale aggravée et de blanchiment de fraude fiscale aggravée, après des signalements de l'administration fiscale. Les trois autres font l'objet d'enquêtes pour blanchiment aggravé de fraude fiscale aggravée. Toutes ces enquêtes ont été ouvertes mi-décembre 2021, selon le PNF. AFP



Un drame qui interroge directement les rapports entre l'Homme et la machine. © D.R.

Ixelles revoit ses ambitions à la baisse

La commune d'Ixelles mettra en place en mai prochain la seconde phase de son quartier apaisé Flagey-Etangs. Les autorités prévoient un dispositif allégé en raison de travaux à venir dans les tunnels avenue Louise.

JULIEN THOMAS

Concernant le projet des 50 quartiers apaisés Good Move à Bruxelles, Bruxelles-Ville (Pentagone), Anderlecht (Cureghem), Schaerbeek (Collignon-Josaphat) ont particulièrement focalisé l'attention. Ces trois com-

munes de la capitale sont en effet les premières à avoir mis en œuvre un nouveau plan de circulation local. Avec des succès différents. Au Pentagone, le dispositif se poursuit globalement sans encombre. Celui à Anderlecht a fini aux oubliettes et celui à Schaerbeek reste en partie bloqué. Le volet concernant la Cage aux Ours demeure en effet gelé depuis octobre. Ixelles (Flagey-Etangs) fait pourtant aussi partie des communes précurseuses. En octobre dernier, la commune dirigée par une majorité Ecolo-PS a initié la maille autour des étangs d'Ixelles. La mesure la plus forte, qui concerne la place Flagey elle-même, devait seulement voir le jour en mai prochain. Elle consistait à interdire la circulation aux voitures sur une des rues bordant la place. Cette mesure forte ne verra finalement pas le jour. L'entité a revu les ambitions du projet à la baisse. Les autorités locales expliquent ne pas avoir le choix face aux prochains travaux dans les tunnels

Bailli et Vleurgat.

Le bourgmestre ixellois Christos Doukeridis (Ecolo) indique devoir tenir compte de la suppression d'une bande de circulation dans ces tunnels de l'avenue Louise voisine : « Je suis déçu moi aussi. C'est normal, puisqu'on a travaillé des mois et des mois dessus. Mais faire semblant que cette modification apparue après notre plan et qui est lié aux problèmes de sécurité des tunnels est un détail, cela ne va pas non plus. » Outre la mesure phare concernant la place Flagey, la majorité ixelloise renonce aussi à une mise en sens unique d'une partie de la rue Lesbroussart et de l'avenue des Eperons d'Or (qui borde les étangs du côté du café Belga). Le dispositif remanié par la commune verra le jour le 2 mai prochain. Il contient plusieurs changements significatifs comme une mise en sens unique d'une partie de la chaussée de Vleurgat et la fermeture du trafic entre les étangs le dimanche.