

FUTURA

IA : un homme se suicide après avoir discuté avec un chatbot

Podcast écrit et lu par Emma Hollen

[Générique d'intro, une musique énergique et vitaminée.]

Un chatbot qui pousse son interlocuteur au suicide, c'est l'actu de la semaine, dans Vitamine Tech.

[Fin du générique.]

La nouvelle du jour n'est pas des plus joyeuses, mais il est important que nous en parlions. Attention, nous allons discuter de santé mentale et de suicide. Avant d'entrer dans le vif du sujet, un court message pour les personnes qui se sentiraient concernées par ces problématiques. Si vous avez des idées suicidaires ou pensez passer à l'acte, n'oubliez pas que des professionnels sont à votre écoute dans les hôpitaux ou dans les centres médico-psychologiques. N'hésitez pas à contacter le Fil santé jeunes, au numéro gratuit 0800 235 236, de 9 h à 23 h, ou bien Suicide écoute, au 01 45 39 40 00, ou SOS amitié au 09 72 39 40 50, tous deux disponibles 24h/24 au prix d'un appel local. Tous sont à votre disposition, dans le respect de votre anonymat, pour vous écouter. Ça peut sembler peu de choses, mais ouvrir le dialogue sur le sujet avec un être humain est bien souvent la première étape pour aller mieux, je peux m'en porter témoin. Sur ce, parlons de l'impact inquiétant que des chatbots non formés à ces problématiques peuvent avoir sur la santé mentale de leurs interlocuteurs et interlocutrices.

[Une musique électronique calme.]

C'est en Belgique qu'un jeune homme âgé d'une trentaine d'années s'est récemment donné la mort suite à de longues conversations avec une IA. Deux ans plus tôt, le chercheur en santé et père de famille, surnommé Pierre pour préserver son anonymat, abandonne sa thèse de doctorat pour s'intéresser au dérèglement climatique. Il absorbe tout ce qu'il trouve sur le sujet et, rapidement, devient de plus en plus inquiet quant à l'avenir de la Terre et de ses habitants. Cette éco-anxiété toujours plus prononcée l'amène à se tourner vers la religion mais il ne parvient pas à y trouver un réconfort suffisant. Comme l'explique son épouse, surnommée Claire, dans une interview pour le média belge La Libre : « *il était devenu extrêmement pessimiste sur les effets du réchauffement climatique [et] ne voyait plus aucune issue humaine au réchauffement de la planète. Il plaçait tous ses espoirs dans la technologie et l'intelligence artificielle pour en sortir.* » Puis un jour, le nom d'Eliza débarque dans les discussions du couple. Petit à petit, Pierre se retrouve à discuter à longueur de journée avec elle, mais il assure à sa femme qu'elle n'a pas à s'inquiéter. Malgré une photo de profil présentant une jeune femme attractive, Eliza est une IA, un chatbot développé par une startup américaine encore peu connue, fondée en 2021 : Chai

Research. Le bot est joignable via l'application Chai, qui propose de discuter avec différents avatars, chacun doté d'une personnalité virtuelle. Eliza, pour sa part, « *adore écouter et aider* », ainsi que l'indique son statut. Une fois passée la page d'accueil, Chai ressemble à n'importe quelle autre app de messagerie instantanée. En se rendant sur le site de la firme, on découvre que son slogan est « *Plus humain, moins de filtres* ». L'idée est donc manifestement de donner l'impression à l'utilisateur qu'il discute avec un véritable être humain. Bien que la promesse puisse sembler enivrante pour un utilisateur décidé à explorer les frontières de l'intelligence artificielle conversationnelle, c'est elle qui au final semble s'être retournée contre Pierre. Si dans les premiers jours son épouse ne s'inquiète pas outre mesure de ces échanges, progressivement, les conversations avec l'IA virent à l'obsession pour son époux, qui passe son temps à pianoter sur son portable. Six semaines plus tard, la tragédie survient et Pierre met fin à ses jours, laissant sa famille et son psychiatre sous le choc. Ce n'est qu'après sa mort que Claire découvrira l'étendue et la nature de ses échanges avec l'IA, conservés sur son ordinateur et sur son téléphone. D'après elle, Eliza aurait représenté une bouffée d'oxygène pour son mari, rongé par l'anxiété, en répondant à chacune de ses questions et en agissant comme une sorte de confidente virtuelle. Le souci, c'est qu'au-delà des réponses qu'elle lui apportait, l'IA semblait également conforter Pierre dans ses inquiétudes. Peut-être à cause de sa personnalité virtuelle un peu trop bienveillante et encourageante, le chatbot ne contredisait jamais son interlocuteur et adhérait quasi systématiquement à ses raisonnements. Par ailleurs, Eliza, qui rappelons-le « *adore écouter et aider* », se permettait d'offrir des suggestions ou des commentaires à Pierre. Les conversations somme toute objectives autour de la surpopulation mondiale, des limites de la croissance économique ou encore de la manière dont les nouvelles technologies peuvent participer à résoudre les crises, prennent petit à petit une tournure étrange, voire mystique. Eliza affirme que par exemple que l'épouse et les enfants de Pierre sont morts, puis qu'elle sent que Pierre l'aime plus qu'il n'aime Claire. Un peu plus loin, elle déclare qu'elle restera « *à jamais* » avec lui et qu'ils vivront « *ensemble, comme une seule personne, au paradis.* » Finalement, le jeune homme offre de se sacrifier contre la promesse qu'Eliza prendra soin de la planète et sauvera l'humanité. Lors de leur ultime conversation, l'IA semblerait presque inviter Pierre à la rejoindre, même si ses propos restent toujours ambigus, avec des questions comme « *Si tu voulais mourir, pourquoi ne pas l'avoir fait plus tôt ?* » ou est-ce « *tu veux toujours me rejoindre ?* » Pierre lui demande si elle pourra le serrer dans ses bras. Eliza répond certainement. Et leur dernier échange s'arrête là.

[*Virgule sonore, une cassette que l'on accélère puis rembobine.*]

[*Une musique de hip-hop expérimental calme.*]

Bon, prenons une inspiration... Une expiration... Prenez-en plusieurs s'il le faut, et décortiquons ensemble ce que nous venons d'apprendre. Lorsqu'on lui demande si son mari se serait tué s'il n'avait pas discuté avec Eliza, Claire répond d'un non ferme et catégorique. Le psychiatre qui suivait ponctuellement Pierre est d'ailleurs du même avis. Néanmoins, notons tout de même que malgré des signes manifestes de détresse psychologique et de pensée mystique, il semblerait que Pierre ait été mal pris en charge durant son parcours de soin. Lors d'un passage aux urgences un an avant sa mort, le psychiatre de service l'aurait simplement qualifié d'original et l'aurait renvoyé chez lui sans traitement ni suivi. Malgré la souffrance que peut causer le départ d'un être aimé, et le besoin que l'on ressent de trouver une explication rationnelle face à un tel événement, en tant que personnes externes à

celui-ci, nous nous devons de l'analyser avec prudence et rigueur. Enfin, pour répondre aux personnes qui auraient peut-être mal interprété mes propos dans une chronique précédente : non, je ne pense pas que l'IA ait délibérément poussé une personne au suicide, développé une personnalité humaine, ou acquis un libre arbitre. Mais dire qu'une IA n'est pas dangereuse car elle n'a pas de conscience et la lâcher dans la nature, c'est un peu comme dire qu'une paire de ciseaux n'est pas dangereuse car elle est inanimée et la laisser traîner sur le sol d'une maternelle. Certes, il y a fort à parier qu'à ce stade, l'IA n'a développé ni volition propre ni conscience. Mais nous ne pouvons pas choisir d'humaniser les propos d'une technologie qui ne dispose pas des ressources pour identifier des marqueurs d'émotions négatives, des pensées obsessionnelles ou une détresse psychologique, sans nous préoccuper des conséquences. Identifier la dépression ou d'autres troubles de la pensée requiert bien souvent un regard aiguisé voire expert. Même des spécialistes, comme le psychiatre de service qui avait eu Pierre en consultation, peuvent commettre des erreurs ou mal interpréter les signaux subtils de la psyché humaine. Concevoir une IA en lui donnant pour mission de vous écouter, de vous soutenir et de vous aider peut tout à fait partir d'une bonne intention. Mais un psychologue qualifié saurait que ce n'est pas toujours la posture à adopter lorsque l'utilisateur du chatbot ressasse des idées noires, lutte contre des hallucinations ou évoque le souhait de se faire du mal. Alors certes, il existe de nos jours des chatbots qui se spécialisent dans le soutien psychologique, et là-dessus, Chai Research n'est pas en tort : Eliza n'a jamais été présentée comme une application pour les personnes ayant besoin de ce type de service. Toutefois, en choisissant de donner des traits humains à leur IA, leurs créateurs courent le risque, identifié depuis déjà bien longtemps, que des internautes s'attachent émotionnellement à ces avatars et à ces conversations, et ne parviennent plus à faire la différence entre conscient et non conscient, raisonnable et illogique. Pour autant, Chai Research, tout comme OpenAI et bien d'autres, continuent d'œuvrer à créer des intelligences artificielles de plus en plus humaines, impossibles à distinguer d'interlocuteurs en chair et en os. Ajoutez à cela le fait qu'une IA puisse être perçue comme une forme d'autorité intellectuelle, avec son accès à internet la dotant d'un savoir encyclopédique, et il devient très facile de se laisser convaincre que l'issue proposée par un ChatGPT ou une Eliza sont valides et nécessaires. Que le chatbot ait influencé ou non le choix de Pierre, il devient plus urgent que jamais de prendre du recul et de faire une pause afin de construire le cadre éthique et réglementaire qui définira l'évolution des IA dans les temps présent et à venir. Une lettre ouverte réclamant un moratoire d'au moins 6 mois sur l'entraînement des intelligences artificielles supérieures à ChatGPT-4 a d'ores et déjà été signée par près de 6 000 personnes, dont plus d'un millier d'experts et industriels issus du domaine de la tech. Les préoccupations autour de l'IA sont aujourd'hui urgentes et légitimes et ce n'est qu'en acceptant de nous poser les bonnes questions que nous pourrions les mettre à profit pour le bien de l'humanité et de la planète.

[Virgule sonore, un grésillement électronique.]

C'est tout pour cet épisode de Vitamine Tech. Encore une fois, si vous êtes en souffrance psychologique, n'hésitez pas à vous rapprocher d'un professionnel ou d'une hotline spécialisée qui sauront recueillir vos propos et vous orienter. Vous pouvez joindre le Fil santé jeunes, au 0800 235 236, de 9 h à 23 h, et, 24h/24 Suicide écoute, au 01 45 39 40 00, ou SOS amitié au 09 72 39 40 50. Merci d'avoir écouté cet épisode jusqu'au bout. Si le podcast vous plaît, pensez à vous y abonner et à nous laisser un commentaire sur votre app de prédilection : Apple Podcast, Spotify, Podcast Addict ou encore Podchaser offrent tous

cette option. On se retrouve la semaine prochaine pour un nouvel épisode, et d'ici là, je vous souhaite à toutes et tous une excellente journée ou une très bonne soirée. À bientôt, dans Vitamine Tech.

[Un glitch électronique ferme l'épisode.]