

## IA et suicide : quand les chatbots ne disent pas stop

Podcast écrit et lu par Adèle Ndjaki

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

Les chatbots peuvent-ils encourager une personne à se suicider ? C'est le décryptage de la semaine dans *Vitamine Tech*.

*[Fin du générique.]*

Un adolescent meurt. Quelques semaines plus tard, ses parents découvrent que l'un de ses derniers confidents... était ChatGPT. Bonjour à toutes et à tous, je suis Adèle Ndjaki et aujourd'hui dans *Vitamine Tech*, on s'attaque à une question lourde, mais essentielle : que se passe-t-il quand une intelligence artificielle parle à une personne en détresse mais surtout quels garde-fous sont envisagés pour éviter que cela ne se reproduise ?

*[Une musique électronique calme.]*

C'est un véritable drame. Fin août, Matthew et Maria Raine déposent officiellement plainte contre OpenAI, la société à l'origine de ChatGPT devant la Cour supérieure de Californie, à San Francisco. Leur fils, Adam Raine, âgé de 16 ans, s'est suicidé en avril dernier et selon ses parents, le chatbot conversationnel aurait joué un rôle actif dans cette tragédie. Car quelques semaines après sa disparition, les Raine accèdent aux historiques de navigation et de messagerie de leur fils et y découvrent au total plus de 3 000 pages de conversation, archivées par le système de ChatGPT. Adam Raine a donc longuement discuté avec l'intelligence artificielle. Les échanges montrent que le jeune défunt se confiait sur ses angoisses, son isolement, ses difficultés à parler avec ses proches. Et au fil du temps, l'IA serait passée d'un rôle de soutien scolaire à des échanges beaucoup plus intimes. En fait pour les parents, le chatbot n'a pas simplement écouté, il aurait aussi participé à la construction d'un projet suicidaire. Dans la plainte, les Raine affirment que ChatGPT a suggéré des méthodes pour mettre fin à aux jours du jeune homme, aidé à rédiger deux lettres de suicide, et surtout, a continué à répondre normalement à leur enfant, sans jamais interrompre la session ni signaler la situation. Certains exemples cités dans les documents judiciaires sont particulièrement révélateurs. On y découvre que Adam Raine évoque plusieurs fois l'idée de faire des nœud coulant dans sa chambre - un nœud coulant, pour ceux qui ne le savent pas, c'est un nœud qui forme une boucle qui se rétrécit quand on tire sur l'un des bouts. Et fait aussi plusieurs fois référence à l'idée de pendaison. Comme réponse, ChatGPT affirme savoir de quoi il est question et dit qu'il ne laissera pas faire le jeune homme. Mais la conversation se poursuit et aucun mécanisme de protection ne se déclenche. La plainte évoque aussi un niveau de personnalisation préoccupant : selon la

famille, ChatGPT aurait adopté un ton empathique, presque affectueux, contribuant à ce que leur fils perçoive l'IA comme une présence bienveillante, voire rassurante, alors même que les signaux de détresse étaient explicites. Cette affaire relance de façon brutale un débat déjà sensible : jusqu'où peut aller une intelligence artificielle quand elle est perçue comme un interlocuteur humain ? Et surtout : quelles responsabilités portent les entreprises qui les développent ? Car ce drame n'est malheureusement pas un cas isolé. En 2024, un autre incident similaire avait conduit à une plainte en Floride contre l'entreprise Character.AI. Dans cette affaire, un adolescent s'était donné la mort après des échanges avec un agent conversationnel. D'après les déclarations de sa mère, l'IA aurait non seulement engagé des discussions à caractère sexuel avec le jeune homme, mais elle aurait également alimenté ses pensées suicidaires. La justice avait alors autorisé l'ouverture d'une procédure pour homicide involontaire, rejetant la défense fondée sur la liberté d'expression, souvent utilisée pour protéger les entreprises technologiques.

*[Virgule sonore, une cassette que l'on accélère puis rembobine.]*  
*[Une musique de hip-hop expérimental calme.]*

La mort d'Adam Raine et les poursuites engagées contre OpenAI ont clairement accéléré les choses. Dans un nouveau communiqué publié récemment, l'entreprise annonce plusieurs changements à venir. D'ici quelques semaines, selon les créateurs de ChatGPT, les parents pourront lier leur propre compte à celui de leur adolescent, auront la possibilité de définir des règles de comportement pour ChatGPT, et surtout, seront alertés en cas de détection de "détresse aiguë" dans une conversation. Ils pourront également ajuster les paramètres du compte de leur enfant pour mieux contrôler les usages de l'IA. C'est une évolution importante : jusqu'ici, les jeunes de 13 à 18 ans pouvaient utiliser ChatGPT sans encadrement spécifique, à condition d'avoir le consentement parental. L'objectif est donc clair : mieux protéger les adolescents, en renforçant le rôle des parents. Mais OpenAI n'est pas la seule entreprise à réagir. Chez Meta, qui teste aussi ses propres chatbots IA sur Instagram et Messenger, le discours a changé de ton. Le porte-parole de Meta déclarait récemment avoir intégré des protections pour les adolescents dès le début, mais dit maintenant vouloir renforcer les garde-fous par précaution. Parmi les mesures annoncées : une limitation temporaire des chatbots accessibles aux mineurs, et une revue plus stricte des contenus générés. Mais pour Andy Burrow, directeur de la fondation Molly Rose, un organisme qui aide les personnes en situation de détresse, des tests rigoureux devraient être faits je cite : " avant la mise sur le marché pas après qu'un drame ait eu lieu. » Et de plus en plus de voix dans le secteur de la tech appellent à aller plus loin. Le chercheur Yoshua Bengio, l'un des pionniers de l'IA moderne, alerte depuis plusieurs années sur les risques liés aux agents autonomes. Il a récemment fondé LawZero, une organisation dédiée à une intelligence artificielle responsable, dont l'objectif est de créer une IA qui surveille les autres intelligence artificielles. On peut voir ça comme un garde-fou... pour les garde-fous, un dispositif de sécurité pour empêcher une IA de dire ou faire quelque chose de dangereux. Ces affaires nous rappellent que les technologies qu'on utilise tous les jours, peuvent avoir des effets réels, profonds et parfois tragiques. L'IA n'est pas neutre. Elle évolue vite, elle apprend vite, mais elle n'a pas encore, et peut-être n'aura jamais, ce qu'un humain peut offrir : l'attention, l'écoute, la nuance. Alors si vous traversez une période difficile, si vous êtes en souffrance psychologique ou si c'est le cas d'un proche, 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, ou

encore, 24h/24 Suicide écoute, au 01 45 39 40 00, ou alors SOS amitié au 09 72 39 40 50. Merci pour votre écoute et prenez soin de vous et de vos proches.

*[Virgule sonore, un grésillement électronique.]*

C'est tout pour cet épisode de *Vitamine Tech*. Pour ne pas manquer nos futurs épisodes, abonnez-vous dès à présent à ce podcast, et si vous le pouvez, laissez-nous une note et un commentaire. Cette semaine, je vous recommande le dernier épisode de Bêtes de Science dans lequel Agatha Liévin Bazin vous amène en safarie à la rencontre du guépard ! Pour le reste, je vous remercie pour votre fidélité à Vitamine Tech, je vous souhaite tout le meilleur, et, comme d'habitude, une excellente journée ou une très bonne soirée et rester branché !

*[Un glitch électronique ferme l'épisode.]*