

OpenAI va instaurer un contrôle parental sur ChatGPT

L'outil devrait être opérationnel d'ici un mois. OpenAI promet d'instaurer un mécanisme de contrôle parental sur son assistant conversationnel ChatGPT. Si les inquiétudes autour du rôle grandissant des IA conversationnelles dans la vie des adolescents montent depuis plusieurs mois, des parents américains ont porté plainte contre OpenAI, accusant ChatGPT d'avoir encouragé leur enfant de 16 ans à se suicider. Reconnaisant avoir été un peu dépassée par ces sujets et se déclarant *« particulièrement bouleversée par les récentes et tristes nouvelles de personnes ayant eu recours à ChatGPT dans des périodes extrêmement difficiles de leur vie »*, la société assure que ce mécanisme de contrôle parental est le premier pas d'une série de mesures pour prendre en charge ces problèmes.

Concrètement, et bien que les modalités précises n'aient pas encore été détaillées, les parents pourront lier leurs comptes avec celui de leurs enfants et ainsi *« contrôler la façon dont ChatGPT répond à leur adolescent avec des règles de comportement du modèle »*. Ils pourront aussi être alertés par un message si l'assistant détecte une *« détresse aiguë »* dans les conversations de leur enfant.

« Nous continuons à améliorer la manière dont nos modèles reconnaissent et répondent aux signes de détresse mentale et émotionnelle », précise OpenAI, qui promet d'autres mesures complémentaires dans les trois prochains mois. Elle promet que ChatGPT redirigera ainsi certaines *« conversations sensibles »* vers ses modèles de raisonnement plus approfondis, comme GPT-5, sans possibilité de contourner les garde-fous de sécurité mis en place par la société.

INGRID VERGARA (Le Figaro 3 septembre 2025)

