

L'emprise clandestine de ChatGPT sur l'opinion générale

Les risques engendrés par l'Intelligence Artificielle générative dans la propagande

L'usage abusif de l'IA générative dans la propagation d'informations fallacieuses

L'IA générative a été exploitée de manière sournoise pour diffuser de fausses informations à l'échelle mondiale. Une étude récente réalisée par OpenAI a mis en évidence de nombreuses campagnes de désinformation menées par divers pays comme la Chine, Israël et la Russie, qui utilisent ChatGPT, un logiciel de production de texte, à des fins malveillantes.

ChatGPT, un instrument au service de la manœuvre

Développé par OpenAI, ChatGPT possède des fonctionnalités à double tranchant. Ce bot est au cœur de diverses campagnes de désinformation et de propagande initiées par des entités en Russie, en Iran, en Chine et en Israël. Ces outils technologiques sont employés pour générer du contenu sur les médias sociaux, rédiger des articles mensongers, concevoir des

titres trompeurs et même automatiser leur propagation.

OpenAI a révélé la manière dont ces campagnes manipulent ChatGPT. Par exemple, le dispositif “Doppelganger”, dirigé par la Russie, exploite le bot pour rédiger des commentaires hostiles à l’Ukraine en plusieurs langues, modifier des articles favorables à l’invasion en Ukraine et automatiser la diffusion sur Telegram. En Chine, la campagne “Spamouflage” a utilisé ChatGPT pour le débogage de code, l’analyse de publications sur les réseaux sociaux et la création de contenus favorables au gouvernement. En Iran, cet outil a été utilisé pour élaborer des contenus en soutien au régime et contre certains États.

De même, Israël a eu recours à ChatGPT dans la campagne “Zeno Zeno” pour générer de faux profils publiant des messages anti-Islam dans divers pays. Ces pratiques soulèvent des inquiétudes quant à la perspective d’une IA incontrôlable capable de manipuler l’opinion publique en diffusant massivement de fausses informations.

Lire [Sauvegarder la démocratie face aux abus de l’IA : un accord entre les grandes entreprises de la technologie pour maintenir une bonne réputation](#)

Cependant, en dépit de ces tentatives de manipulation, Ben Nimmo, enquêteur principal chez OpenAI, souligne que ces manoeuvres peinent à influencer de manière significative l’opinion publique. Les contenus produits par ChatGPT sont souvent repérés en raison de leur mauvaise grammaire et de leurs caractéristiques typiques du bot.

Malgré les limitations actuelles, le développement continu de ChatGPT soulève des questions quant à sa capacité à produire des contenus indétectables. OpenAI explore diverses solutions telles que des garde-fous pour limiter les manipulations, ainsi que des partenariats avec d’autres plateformes d’IA pour contrer ces abus.

Lire [OpenAI annonce la sortie prochaine de ChatGPT 5 et promet un nouveau niveau de compétences](#)

□ Pour rester informé sur toutes les actualités de 01net, suivez-nous sur [Google Actualités](#) et [WhatsApp](#).

Source : [OpenAI](#)