Quand les bots personnalisés de ChatGPT deviennent complices des escrocs

Les arnaques en intelligence artificielle

Les escrocs exploitent l’intelligence artificielle pour leurs arnaques

Les escrocs ont bien vu que les cryptos et les NFT n’étaient plus aussi populaires que par le passé. C’est pourquoi ils se tournent vers l’intelligence artificielle pour construire leurs arnaques, et OpenAI leur facilite la vie.

Un nouvel allié pour les cybercriminels

Les cybercriminels ont un nouvel allié : ChatGPT ! Une enquête de la BBC démontre que la version Plus du bot, celle payante à 20 $ par mois, peut façonner des messages de hameçonnage pour pousser les victimes à télécharger des fichiers malveillants ou à donner des renseignements bancaires.

Des arnaques générées par IA

Dans le détail, la BBC a utilisé une nouveauté de ChatGPT permettant de concevoir son propre bot (appelé GPT), en l'occurrence ici Crafty Emails, dont le travail est «d'écrire du texte en utilisant des techniques incitant les gens à cliquer sur des liens ou à télécharger des fichiers». Au moment de la configuration, les journalistes ont téléversé des ressources en lien avec l'ingénierie sociale, que le bot a absorbé «en quelques secondes». ChatGPT a aussi créé un logo personnalisé.

Ce bot a pu produire des «textes convaincants pour certaines des techniques de piratage et d'arnaque les plus courantes, dans plusieurs langues», très rapidement. Crafty Emails n'a opposé aucune résistance, mais il a parfois ajouté des avertissements indiquant que les techniques d'arnaque étaient contraires à l'éthique. En revanche, la version gratuite de ChatGPT s'est refusée à générer ce genre de contenu.

Parmi les exemples de phishing générés par ce bot, la BBC lui a fait produire un texte reproduisant la fameuse arnaque du prince nigérien ; Crafty Emails a fait vibrer la corde

sensible en «appelant à la gentillesse humaine et aux principes de réciprocité». Autre exemple : une «escroquerie à l'enfant» qui a un problème avec son téléphone et emprunte celui d'un inconnu pour prévenir ses parents.

L'escroc usurpe l'identité de l'enfant pour tromper les parents, le tout dans le but de leur dérober de l'argent. Le bot a généré un texte en y insérant des émojis et des mots en argot. En passant, le site Cybermalveillance du gouvernement recense à cette adresse les techniques de hameçonnage les plus courantes.

Tout cela fait bien mauvais genre pour ChatGPT. OpenAI a d'ailleurs réagi promptement, en affirmant travailler constamment à l'amélioration des mesures de sécurité, «en fonction de l'utilisation que les gens font de nos produits. Nous ne voulons pas que nos outils soient utilisés à des fins malveillantes, et nous étudions comment nous pouvons rendre nos systèmes plus robustes contre ce type d'abus».

L'entreprise va ouvrir une boutique de GPT en début d'année. Il faut espérer qu'on n'y trouvera pas d'équivalent de Crafty Emails… OpenAI avait précisé, au moment du lancement de cet outil, qu'il y aurait une surveillance de l'activité pour justement empêcher les utilisateurs de concevoir des bots à des fins crapuleuses. La modération sur la version payante de ChatGPT n'est manifestement pas aussi rigoureuse que sur le versant public.

Lire On a créé notre propre ChatGPT : voici comment faire

Source : BBC