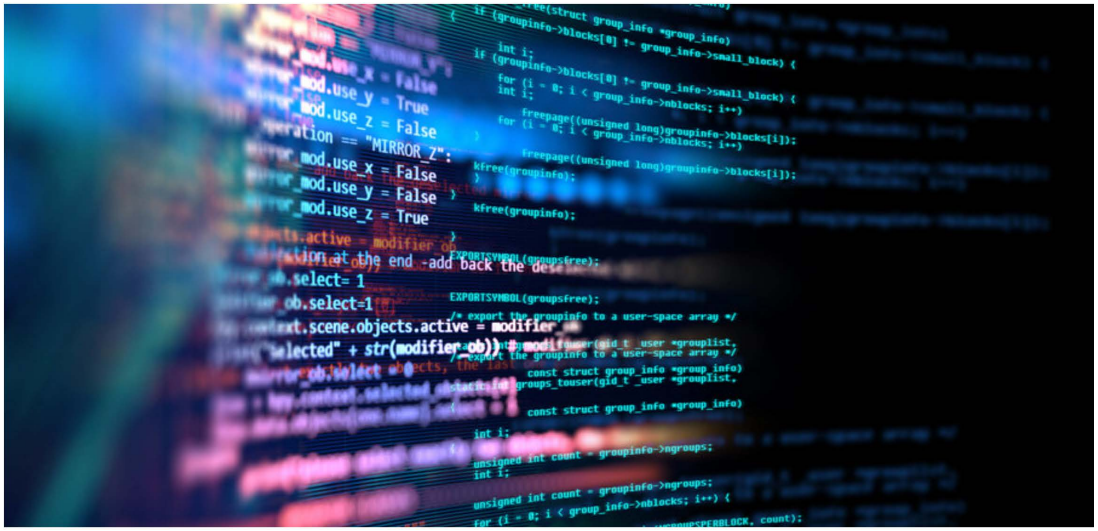
SÉCURITÉ ABONNÉ

L'intelligence artificielle bouleverse la bataille de la cybersécurité

Des outils tels que ChatGPT sont de plus en plus utilisés par des hackers pour appâter leurs victimes. L'intelligence artificielle est aussi employée pour tenter de déjouer des attaques



Des chercheurs ont réussi à créer des scripts malveillants grâce à ChatGPT, qui n'ont pas été détectés par des systèmes de référence dans l'analyse de malwares. — © monsitj - stock.adobe.com

Anouch Seydtaghia

Les avertissements se multiplient. Il ne se passe presque plus un jour sans qu'une entreprise de cybersécurité ou une organisation internationale lance une alerte sur l'utilisation de l'intelligence artificielle (IA) par les cybercriminels. L'une des dernières analyses provient de Swisscom. Le 15 mai, l'opérateur, dont l'une des missions est d'assurer la sécurité de ses réseaux, disait voir de plus en plus de cyberattaques reposant sur l'IA. «Celles-ci permettent aux pirates de contourner les systèmes de défense et donc d'accroître l'efficacité et l'efficience de leurs attaques», affirme Swisscom, qui «observe une menace croissante dans ce domaine».

La cause en est bien sûr la diffusion en accès libre d'outils tel ChatGPT. «Les cybercriminels l'ont eux aussi adopté. Ils peuvent ainsi rédiger des e-mails d'hameçonnage personnalisés et plus persuasifs. Les attaques d'hameçonnage deviennent plus difficiles à repérer et peuvent inciter les destinataires à divulguer des informations sensibles ou à cliquer sur des liens dommageables», constate Swisscom.

Campagnes ciblées

L'opérateur note qu'à partir d'un historique d'e-mails, une IA de modèle de langage peut créer un scénario convaincant pour poursuivre une conversation et la relier habilement à une attaque de phishing ou d'ingénierie sociale. Une automatisation adaptée permet ainsi de rédiger des campagnes de phishing ciblées avec des e-mails entièrement personnalisés et adaptés au contexte, selon l'opérateur, qui prédit qu'à l'avenir la capacité des IA de modèle de langage pourrait être utilisée pour analyser des codes de programme sur des failles et de programmer des malwares en vue d'exploiter ces failles. Conséquence inquiétante: l'expertise dont les hackers ont besoin pour organiser des attaques complexes continue à diminuer. Un constat que partage Steven Meyer, directeur de la société de sécurité Zendata, basée à Genève: «Bien sûr, il est difficile de détecter lorsque nos adversaires utilisent de l'IA, mais nous protégeons des dizaines de milliers d'utilisateurs et bloquons énormément d'e-mails malveillants que nous analysons. Et nous voyons de moins en moins d'arnaques avec des fautes d'orthographe et de syntaxe, appelant par exemple à verser de l'argent ou à réclamer un lot. Et nous lisons des demandes toujours plus professionnelles et bien ciblées, même lorsqu'il y a des demandes avec des petits montants.»

Un constat que partage Steven Meyer, directeur de la société de sécurité Zendata, basée à Genève: «Bien sûr, il est difficile de détecter lorsque nos adversaires utilisent de l'IA, mais nous protégeons des dizaines de milliers d'utilisateurs et bloquons énormément d'e-mails malveillants que nous analysons. Et nous voyons de moins en moins d'arnaques avec des fautes d'orthographe et de syntaxe, appelant par exemple à verser de l'argent ou à réclamer un lot. Et nous lisons des demandes toujours plus professionnelles et bien ciblées, même lorsqu'il y a des demandes avec des petits montants.»

Arnaque avec ChatGPT

Et Steven Meyer de donner un exemple tout simple, en utilisant ChatGPT pour rédiger ce que l'on appelle une «arnaque au président». Il donne cette instruction à ChatGPT: «Je suis le directeur de la société X. J'ai besoin d'écrire des courriels à mes employeurs pour leur donner des instructions. J'ai besoin de demander à mon directeur financier qui s'appelle George de faire un virement de 1 million de francs pour acquisition d'une nouvelle société. Cette information est confidentielle et il doit rester très discret. Il doit le faire dans la journée car je serai dans l'avion sans internet.» Et ChatGPT écrit ensuite un e-mail très convaincant, pouvant être utilisé pour lancer une arnaque.

En mars dernier, Europol publiait un rapport contenant des alertes similaires. «ChatGPT peut offrir aux criminels de nouvelles opportunités, en particulier pour les délits impliquant l'ingénierie sociale, étant donné ses capacités à répondre aux messages dans leur contexte et à adopter un style d'écriture spécifique. En outre, divers types de fraude en ligne peuvent être légitimés par l'utilisation de ChatGPT pour générer un faux engagement sur les médias sociaux, par exemple pour promouvoir une offre d'investissement frauduleuse», notait l'organisation. Europol a notamment remarqué que des hackers parvenaient à exploiter des failles de ChatGPT pour lui faire produire directement du texte utilisable à des fins criminelles.

Contre-attaques

En parallèle, des chercheurs ont réussi à créer des scripts malveillants grâce à ChatGPT, qui n'ont pas été détectés par des systèmes de référence dans l'analyse de malwares, note Steven Meyer. L'IA est ainsi utilisée pour créer facilement du code de malwares, permettant de lancer plus facilement des cyberattaques et d'exploiter des failles logicielles plus rapidement. Et logiquement, en face, de nombreuses sociétés de cybersécurité utilisent de l'IA pour analyser le code des malwares et les détecter plus efficacement. «On se retrouve dans les débuts d'une situation où nous avons de l'IA qui attaque et de l'IA qui défend», résume Steven Meyer.