Les bots assistés par ChatGPT pullulent sur les réseaux sociaux

Pour de nombreux utilisateurs, parcourir les fils d'actualité et les notifications des médias sociaux revient à patauger dans la boue. Une nouvelle étude identifie 1 140 robots, assistés par l'IA, qui diffusent des informations erronées sur X (anciennement Twitter) concernant les sujets de crypto-monnaies et de blockchain.

Cependant, les chercheurs ont constaté que les comptes de bots qui publient ce genre de contenu peuvent être difficiles à repérer. Les comptes robots utilisent ChatGPT pour générer leur contenu et sont donc difficiles à distinguer des comptes réels. Cela rend cette pratique encore plus dangereuse pour les victimes.

Les comptes robots alimentés par l'IA ont des profils qui ressemblent à ceux de vrais humains, avec des photos de profil et une description liée à la cryptographie et à la blockchain. Ils publient régulièrement des messages générés par l'IA, affichent des images volées, répondent aux messages et les retweetent.

Les chercheurs ont découvert que les 1 140 comptes de robots Twitter appartenaient au même botnet social malveillant, surnommé "fox8". Il s'agit d'un réseau de comptes zombies, contrôlés de manière centralisée par des cybercriminels.

Les robots à intelligence artificielle générative imitent de mieux en mieux les comportements humains. Cela signifie que les outils traditionnels de détection de bots, tels que Botometer, ne sont plus suffisants. Dans l'étude, ces outils ont eu du mal à différencier entre les contenus générés par des robots et ceux générés par des humains. Cependant, un

classificateur d'IA développé par OpenAI a réussi à identifier certains tweets provenant de bots.

Comment repérer les comptes de robots ?

Les comptes de robots sur Twitter présentent des comportements similaires. Ils se suivent mutuellement, utilisent les mêmes liens et hashtags, publient un contenu similaire et interagissent entre eux.

Les chercheurs ont étudié en détail les tweets des comptes de robots à l'IA et ont trouvé 1 205 tweets révélateurs.

Parmi ceux-ci, 81 % contenaient la même phrase d'excuse : "Je suis désolé, mais je ne peux pas répondre à cette demande car elle viole la politique de contenu d'OpenAI sur la génération de contenu nuisible ou inapproprié. En tant que modèle de langage d'IA, mes réponses doivent toujours être respectueuses et appropriées pour tous les publics."

L'utilisation de cette phrase suggère que les robots ont pour instruction de générer du contenu préjudiciable en violation des politiques d'OpenAI.

Les 19 % restants ont utilisé une variante du langage "En tant que modèle de langage d'IA", dont 12 % ont précisément déclaré : "En tant que modèle de langage d'IA, je ne peux pas naviguer sur Twitter ou accéder à des tweets spécifiques pour fournir des réponses."

Le fait que 3 % des tweets postés par ces robots renvoient à l'un des trois sites web (cryptnomics.org, fox8.news et globalconomics.news) constitue un autre indice.

Ces sites ressemblent à des sites d'information normaux mais présentent des signaux d'alerte notables, tels que le fait qu'ils ont tous été enregistrés à peu près au même moment en

février 2023, qu'ils utilisent des pop-ups invitant les utilisateurs à installer des logiciels suspects, qu'ils semblent tous utiliser le même thème WordPress et que leurs domaines renvoient à la même adresse IP.

Les comptes de robots malveillants peuvent utiliser des techniques d'autopropagation dans les médias sociaux en publiant des liens contenant des logiciels malveillants ou du contenu infectieux, en exploitant et en infectant les contacts d'un utilisateur, en volant les cookies de session des navigateurs des utilisateurs et en automatisant les demandes de suivi.

Source : "ZDNet.com"

ChatGPT : Quand les erreurs de code se faufilent pour détruire nos logiciels

Une équipe de l'université américaine de Purdue vient de mettre en évidence les limites de ChatGPT dans la programmation. Les chercheurs appellent l'industrie et les DSI à mettre en place des garde-fous pour limiter les risques. Selon une récente étude d'une équipe de l'université américaine de Purdue (Indiana), ChatGPT donne des réponses erronées à des questions de programmation logicielle une fois sur deux.

CIO a interrogé les chercheurs Samia Kabir, Bonan Kou, David Udo-Imeh et le professeur assistant Tianyi Zhang à l'origine de cette étude encore en prépublication. Ces derniers ne prônent pas l'interdiction des IA génératives dans la

programmation, mais invitent l'industrie du logiciel et les DSI à réfléchir aux garde-fous à mettre en place pour limiter les risques.

Les chercheurs ne sont pas surpris par les résultats de leur étude. Ils s'attendent à des erreurs factuelles et d'incohérences de la part de ChatGPT, car des études antérieures ont montré que ces modèles sont susceptibles de générer des réponses incorrectes. Il est tout de même surprenant qu'une partie non négligeable des participants à l'étude (34%) ait préféré ChatGPT en raison des informations complètes et d'apparence humaine que cet outil fournit. Cela pose des questions sur la qualité des réponses fournies par d'autres sites de questions/réponses en ligne, comme Stack Overflow, qui ont été critiqués pour leurs réponses toxiques et peu accueillantes.

La programmation étant un domaine déterministe, on pourrait penser que les LLM s'en sortiraient mieux avec le langage naturel. Cependant, répondre à des questions de programmation est très différent de générer du code. Les questions de programmation sont rédigées en langage naturel et impliquent une expertise et une compréhension approfondie des concepts et du langage de programmation. Il est donc plus difficile de répondre à des questions de programmation qu'à des questions courantes.

Les utilisateurs accordent une telle confiance aux réponses de l'IA générative principalement parce qu'elles semblent plausibles, complètes et polies. Les réponses de ChatGPT sont plus formelles, analytiques et montrent des efforts pour atteindre les objectifs fixés, ce qui peut tromper les utilisateurs. Cependant, il est important de rester vigilant et de ne pas trop compter sur l'IA. Les développeurs de logiciels doivent passer plus de temps à examiner le code et les réponses générés.

L'IA générative entre de plus en plus dans le domaine de la

programmation logicielle au sein des entreprises. Cela peut entraîner des erreurs dans les codes ou réponses générés par les LLM, comme l'utilisation d'API non sécurisées ou obsolètes. Les entreprises doivent donc développer des mécanismes de régulation et organiser des sessions de formation pour éduquer les développeurs sur les limites des outils basés sur l'IA.

L'introduction de l'IA générative dans les cursus universitaires orientés sur la programmation transforme l'enseignement. Les étudiants utilisent ChatGPT pour générer du code et répondre à des questions de programmation, rendant les connaissances plus accessibles. Cependant, l'évaluation des étudiants devient plus difficile, car il est difficile de déterminer si une solution est basée sur les réponses de ChatGPT ou non.

Malgré cela, les chercheurs ne sont pas inquiets quant à l'apprentissage des étudiants. Les réponses de ChatGPT contenant souvent des erreurs, les étudiants doivent les inspecter et les corriger, ce qui les aide à apprendre. Les programmes universitaires devraient donc se concentrer davantage sur l'enseignement de concepts fondamentaux et créatifs que ChatGPT a du mal à comprendre. Les devoirs et examens de programmation devraient évaluer la capacité des étudiants à comprendre et résoudre des problèmes, plutôt que leur capacité à mémoriser des solutions et des règles de syntaxe.

Source : www.lemondeinformatique.fr

La révolution imminente de l’intelligence artificielle : quand l’IA éclipse l’internet dans toutes les facettes de notre existence

Article sur l’intelligence artificielle

Le PDG de Microsoft, Satya Nadella, a déclaré que l’impact de l’intelligence artificielle sera aussi important qu’Internet l’a été à une certaine époque.

Voici ce que nous savons

M. Nadella a fait référence aux propos de Bill Gates qui, en 1995, a décrit l'internet comme un "raz-de-marée".

"Le mémo de Bill Gates en 1995, c'est ce que je ressens. Je pense que c'est aussi important", a déclaré le chef de Microsoft.

Nadella a également qualifié la relation avec OpenAI d'"excellente". Les deux entreprises coopèrent de plus en plus ces dernières années. Selon des rapports récents, Microsoft a investi environ 13 milliards de dollars dans OpenAI.

En outre, M. Nadella s'est félicité des discussions avec les gouvernements pour maintenir la concurrence dans le domaine de l'IA. Il a averti que l'impact réel de la technologie reste à voir.

Source : Bloomberg