



MORRIS II

LE VER INFORMATIQUE

INTRODUCTION

Retourner une IA contre elle-même pour générer un ver informatique attaquant les messageries, une expérience menée par des chercheurs en cybersécurité aux États-Unis.

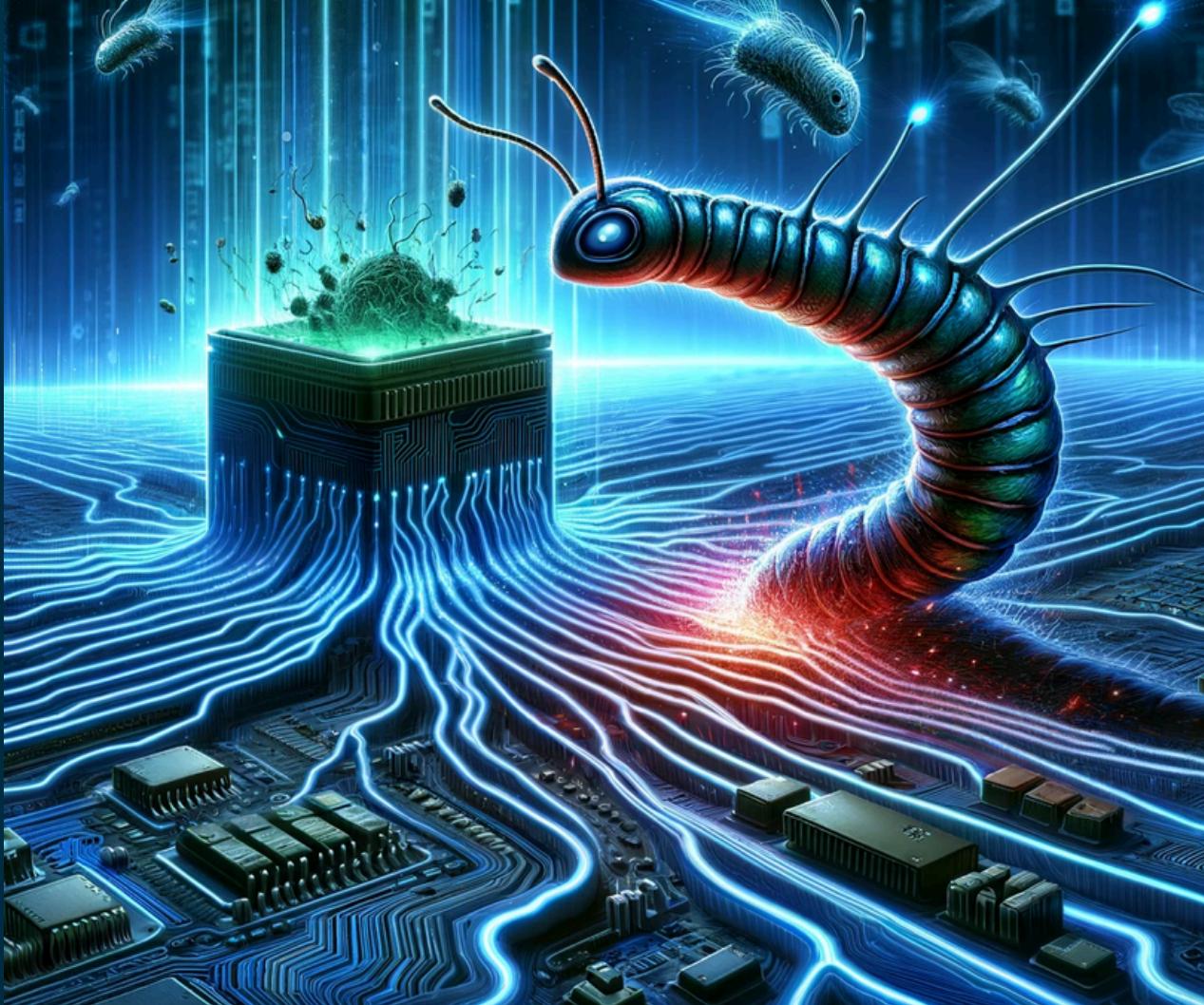


article écrit par SYLVAIN BIGET
JOURNALISTE
le 5 Mars 2024

C'EST QUOI UN VER INFORMATIQUE ?



Un ver informatique est un logiciel malveillant qui se reproduit sur plusieurs ordinateurs en utilisant un réseau informatique comme Internet. Il a la capacité de se dupliquer une fois qu'il a été exécuté. Contrairement au virus, le ver se propage sans avoir besoin de se lier à d'autres programmes exécutables.



QUI EST MORRIS II

Il s'appelle Morris II (en clin d'oeil au premier ver informatique « Morris » qui avait été créé en 1988)

Il manipule les IA (telles que Chatgpt et Gemini) puis vole des données dans les e-mails et est capable d'envoyer des messages pour contaminer d'autres messageries.

Ce **virus** a été créé par des chercheurs du Cornell Tech à New York. Il ne s'agit que d'un « exercice » conçu pour montrer les risques liés aux **écosystèmes** d'IA connectés et autonomes.

ORIGINE

Pour créer ce ver génératif, les chercheurs ont utilisé une « invite contradictoire à **auto-réPLICATION** ». Il s'agit d'une invite de commande qui demande à l'IA de générer dans sa réponse une autre invite. L'IA va donc développer de nouvelles instructions dans sa réponse.



COMMENT ONT-ILS FAIT

les chercheurs ont d'abord créé une messagerie capable d'envoyer et de recevoir des messages, assistée par une IA générative connectée.

Ils ont par la suite généré deux types d'invites pour manipuler les IA : une invite auto-répliquante basée sur du texte et une invite équivalente dans un fichier image.

l'e-mail intégrant l'invite a corrompu la base de données de l'assistant de messagerie.



PARTICULARITÉ

il peut être généré directement via ChatGPT ou Gemini en sachant « parler » correctement avec le chatbot et en passant les verrous de protection.

il peut se propager d'un système à l'autre, en volant des données ou en déployant des logiciels malveillants dans le système.



MERCI

