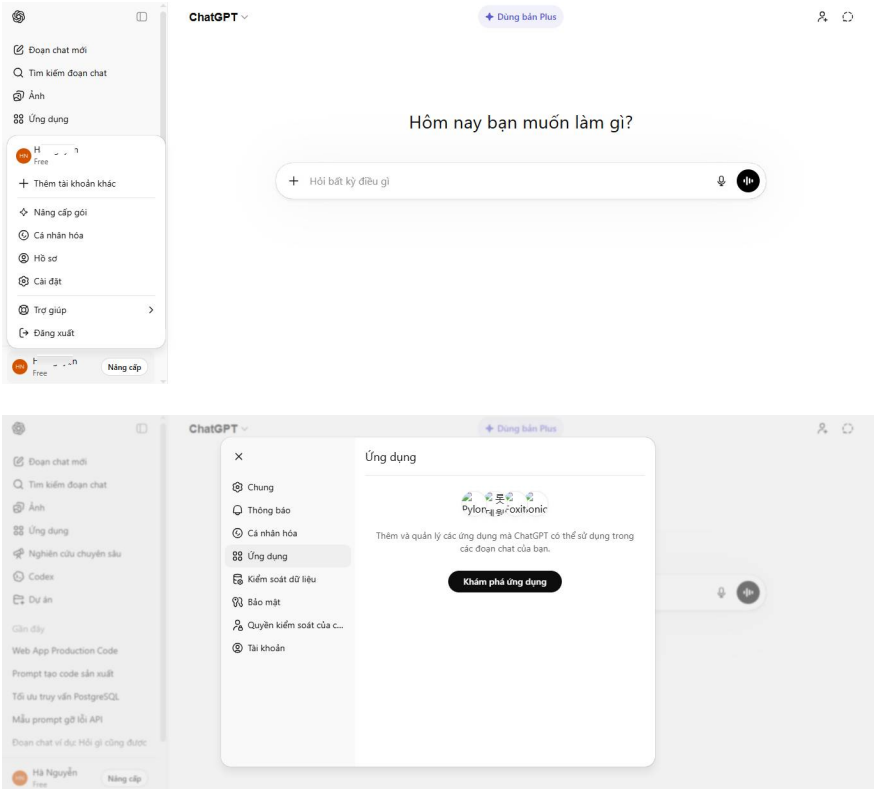
How to use ChatGPT to detect phishing scams.

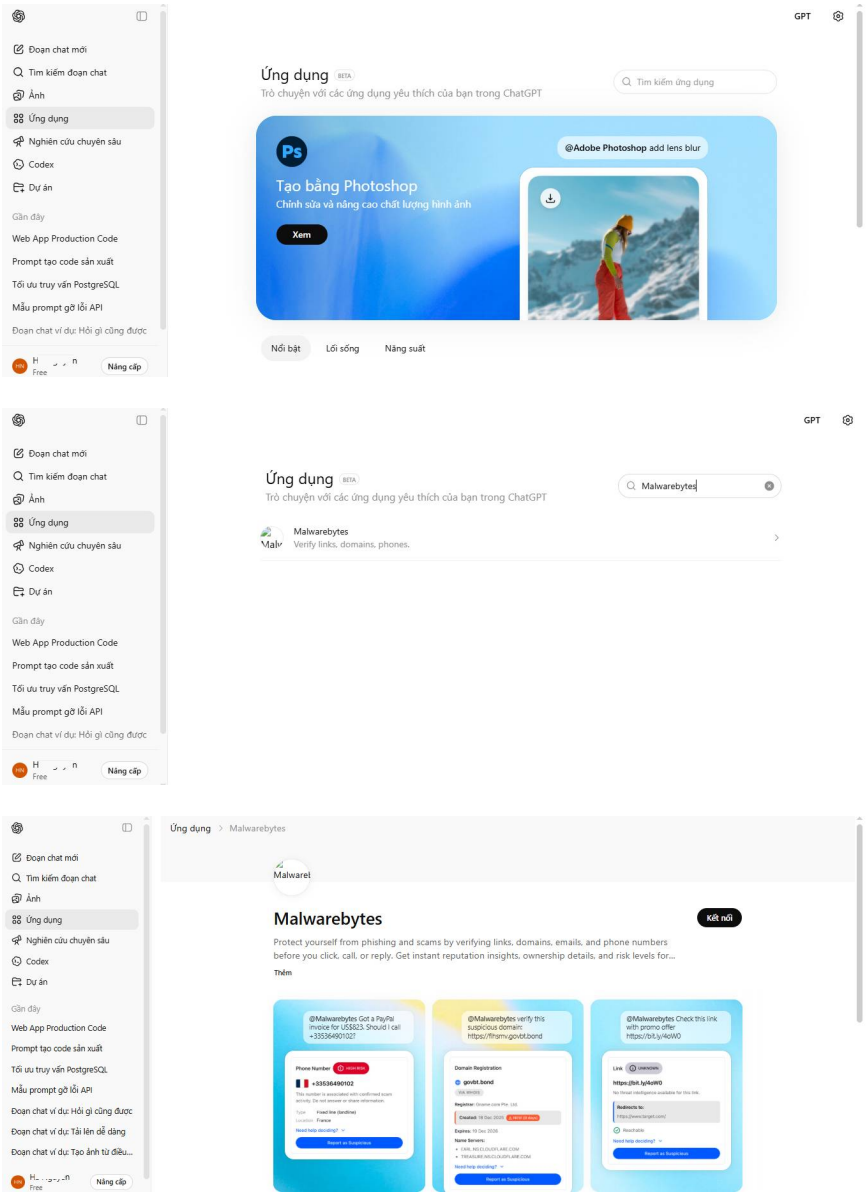
Scammers are using artificial intelligence (AI) to make phishing scams appear legitimate.

Scammers are using artificial intelligence (AI) to make phishing scams look legitimate. From creating more human-like messages to impersonating official emails from major tech companies, distinguishing between real and fake scams is becoming increasingly difficult. However, you can use these very AI platforms to quickly detect phishing scams with the help of antivirus software companies, which can now work in conjunction with ChatGPT .

How the Malwarebytes add-on works in ChatGPT

A third-party security layer inside your AI chatbot.





ChatGPT recently introduced an app store that allows you to connect third-party services directly to your chats. It's similar to browser extensions, but for your AI chatbot. One of the most useful add-ons is Malwarebytes, which you can activate by going to Settings **,** then selecting **Apps** , searching for **Malwarebytes** , and then clicking **Connect** .

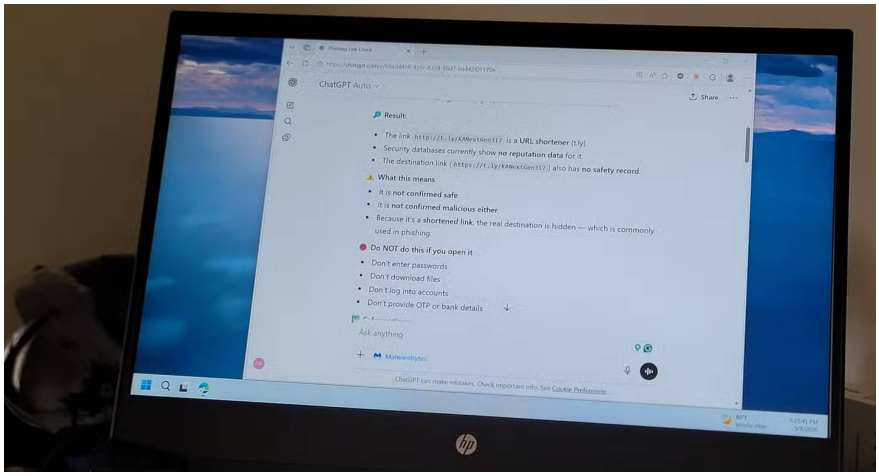
Once connected, you can activate it in any chat by typing something like *"@Malwarebytes, is this a scam?"* followed by suspicious content. You can paste text from emails, SMS messages, or direct messages, insert URLs or phone numbers, or even upload a screenshot of a suspicious message.

Of course, this is vastly different from simply asking ChatGPT directly because it doesn't rely on guesswork. When you submit content through the Malwarebytes add-on, it runs that content against Malwarebytes' own threat intelligence database, checking URLs for domain age, WHOIS data, and known phishing signatures. Similarly, phone numbers are compared against phishing and spam databases, and email addresses are verified for domain legitimacy and registration history. The results are then fed back to ChatGPT and explained in easy-to-understand language.

This add-on works with ChatGPT Free, Plus, Team, and Enterprise accounts, so you don't need a paid Malwarebytes subscription to use it.

Testing with real phishing emails and messages.

The results are surprisingly detailed.



Let's go back to where it all began. A week ago, the author received an SMS message claiming to be from the transport department about an unpaid traffic ticket. It mentioned the license plate number, the fine amount, and included a link to pay. At first glance, it seemed legitimate. The link was a shortened URL, which should have been a warning sign, but people are often too worried about potentially being fined for unpaid tickets to notice.

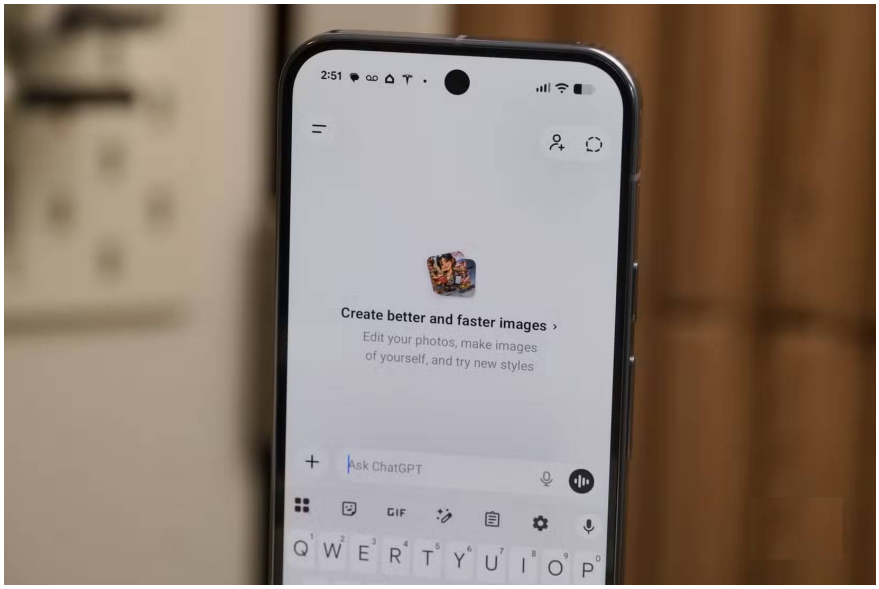
Instead of clicking the link, the author copied the entire message and pasted it into ChatGPT with the Malwarebytes application selected. Simply ask: "Is this a scam?". After a few seconds, it will confirm your suspicions and analyze exactly why. It warns that government agencies do not use URL shortening services like t.ly, that official traffic violation notices come from sender IDs like VM-PARIVH and not random codes like 56161230, and that the UID format in the message does not match actual traffic ticket references. It even directs you to the official website to verify the fine yourself.

Out of curiosity, the author opened the shortened link in a virtual machine . It redirected to a website that automatically downloaded and installed an APK file disguised as the official Ministry of Transport app for paying for violations. If installed on a real phone, payment information could be stolen or worse.

This add-on also maintains context throughout the conversation. When the author pastes the sender's phone number, it matches that number against spam databases and flags it as a known mass SMS messaging gateway commonly used in phishing campaigns across the region.

This add-on isn't perfect, but it's a solid first line of defense.

There are a few points to note.



Like any security tool, Malwarebytes' add-on also has limitations. It relies on threat intelligence databases and speculative analysis, so if a threat is completely new or extremely specific, it may not have specific data on it. In those cases, it will make an uncertain judgment based on a general risk context rather than a definitive yes or no answer.

There's also a privacy aspect to consider. To analyze suspicious content, you're pasting messages, phone numbers, and URLs into ChatGPT, which the application then sends to Malwarebytes' backend. This means both OpenAI and Malwarebytes process that data. The reports you submit are also fed into Malwarebytes' threat intelligence system, implying they retain certain data. If that concerns you, you should reconsider how ChatGPT handles your data and how it manages what it remembers.

A useful tool, not a magic solution.

The Malwarebytes add-on for ChatGPT won't replace dedicated antivirus software or make you immune to online scams. But as a quick screening tool in situations where you receive suspicious messages, emails, or links and need a quick, well-founded opinion, it works quite effectively. The fact that it checks threat data in real time instead of just matching text patterns makes it far more reliable than regular ChatGPT.

However, not every suspicious message will be checked by ChatGPT, so building good habits is more important. If you're using Android, enabling Android 16 's Advanced Protection is a practical step that works passively in the background, blocking unsafe apps, filtering phishing messages, and protecting against malicious links without you having to manually check every notification.

You finished reading the article "**How to use ChatGPT to detect phishing scams.**" edited by the **TipsMake** team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.

© 2019 TipsMake.com