PRIMUS PARTNERS®

Quote by Devroop Dhar, Co-founder & CEO, Primus Partners

Published in CSO
September 22, 2025

AI-powered phishing scams now use fake captcha pages to evade detection

Attackers are exploiting low-code AI platforms such as Vercel, Netlify, and Lovable to rapidly build phishing sites that look legitimate, trick users, and bypass automated security tools.

Read on: https://www.csoonline.com/article/4060817/ai-powered-phishing-scams-now-use-fake-captcha-pages-to-evade-detection.html

Article Content- In an attempt to evade security tools, cybercriminals are now leveraging AI to craft sophisticated phishing campaigns using fake captcha pages. The pages appear legitimate to users, effectively bypassing security filters and capturing sensitive information.

Identified by Trend Micro, these AI-generated captcha pages are designed to mimic the appearance and functionality of genuine verification systems. The fake captcha pages have been hosted on such platforms since January, and there has been a renewed spike in these types of phishing campaigns in August.

Minimal coding, maximum impact

Platforms such as Lovable, Netlify, and Vercel that are designed to simplify development and lower barriers to entry to build and host applications are now being exploited by attackers.

“On Lovable, attackers can use vibe coding to generate a fake captcha or phishing page, while Netlify and Vercel make it simple to integrate AI coding assistants in the CI/CD pipeline to churn out fake captcha pages,” said Trend Micro.

Other than ease of deployment requiring minimal technical skills, free hosting lowers the cost of launching phishing operations. Also, with domains ending *.vercel.app or *.netlify.app, attackers also inherit credibility from the platform’s reputation, which the attackers can leverage.

“Unlike traditional phishing pages, the AI-generated ones are a step up in speed and scale rather than using some new technical trick,” said Devroop Dhar, MD and co-founder at Primus Partners. “They can iterate and create brand-looking pages very quickly. Phishing sites used to take time to create, but now can be generated and cloned across many domains in minutes. That increases the volume of attacks and the chance that an employee will see a convincing fake.”

Dhar added that it also drops the skill lever way down as attackers grab a template, tweak a few things, and are suddenly able to create a phishing kit that looks professional.

Trend Micro has identified 52 malicious sites on Vercel.app, compared with 43 on Lovable.app and 3 on Netlify.app. Lovable has been the primary target for such abuse, but Vercel is currently hosting even more fake CAPTCHA pages.

The attack playbook

The phishing campaigns follow a familiar playbook at the outset. Victims typically receive spam emails that carry urgent, action-oriented messages such as “Password Reset Required” or “USPS Change of Address Notification”.

Clicking on the embedded link doesn’t take the user directly to a credential-stealing site but instead loads what appears to be a harmless captcha verification page. This actively engages the victim, making them feel they are completing a legitimate security check, which lowers their suspicion and makes it less likely they will recognize the page as fraudulent.

Secondly, the automated scanners crawling the page encounter only a captcha, not the underlying credential-harvesting form, reducing the likelihood of the scam being flagged, noted Trend Micro.

Once the captcha is completed, the victim is redirected to the actual phishing page, where their credentials and other sensitive data can be stolen, such as Microsoft 365 credentials.

Strengthening defenses

Enterprises are rethinking defenses as AI-driven phishing campaigns push past legacy filters. Passkeys and phishing-resistant MFA are gaining traction, particularly in financial services and tech. But to combat the growing threat of AI-driven phishing attacks, organizations must adopt a multi-layered security approach.

“The most effective strategies now blend behavioural detection with platform accountability. Tools must be able to simulate clicks and follow redirects, and hosting providers must build safeguards that prevent abuse,” said Sanchit Vir Gogia, CEO and chief analyst at Greyhound Research.

Yet detection alone is not enough. The ultimate resilience lies in reducing the value of stolen credentials altogether through phishing-resistant authentication. Gogia added that organizations must modernise training from checkbox exercises to realistic immersion. That includes phishing simulations with CAPTCHA fronts, policies that block newly registered domains, and strict governance of identity logins. The goal is not to prevent every click, but to shorten the time from incident to containment.

“You need to be aware if the page suddenly redirects to a login form or starts pulling data from untrustworthy domains. Those patterns are harder to hide for attackers. One should also keep an eye on outbound traffic. Stolen data leaving the network is often the first sign,” added Dhar.

User awareness remains the frontline. Training employees to spot suspicious CAPTCHA challenges, verify URLs before interacting, rely on password managers that won’t autofill on fake pages, and promptly report anomalies remains critical.