Unknown Title

⋮ 9/18/2025

Artificial Intelligence (AI)

How AI-Native Development Platforms Enable Fake Captcha Pages

Cybercriminals are abusing AI-native platforms like Vercel, Netlify, and Lovable to host fake captcha pages that deceive users, bypass detection, and drive phishing campaigns.

By: Ryan Flores, Bakuei Matsukawa September 19, 2025 Read time: 4 min (949 words)

Key takeaways

- Since January, Trend Micro has tracked a surge in phishing campaigns using AI-powered platforms (Lovable, Netlify, Vercel) to host fake captcha pages that lead to phishing websites. This ploy misleads users and evades security tools.
- Victims are first shown a captcha, lowering suspicion, while automated scanners only detect the challenge page, missing the hidden credential-harvesting redirect.

- Attackers exploit the ease of deployment, free hosting, and credible branding of these platforms.
- Defenders should train employees to recognize captcha -based phishing, adopt layered defenses that follow redirects, and monitor trusted hosting domains for abuse.

Artificial intelligence has revolutionized web development, empowering even novice users to create professional-looking websites. Tools like Lovable enable anyone to build and host applications with little to no coding knowledge, while Netlify and Vercel position themselves as AI-native development platforms. However, cybercriminals are increasingly exploiting these services to create and host fake captcha challenge websites, which serve as entry points for phishing campaigns.

Since January, Trend Micro has observed a rise in fake captcha pages hosted on such platforms. These scams pose a dual threat: misleading users while evading automated security systems.

The social engineering ploy

The phishing campaigns typically begin with spam emails carrying urgent messages such as: "*Password Reset Required*" or "*USPS Change of Address Notification*", which are standard tactics that are a staple of these types of attacks.

Clicking the embedded URL directs the target to what appears to be a harmless captcha verification page (see Figure 1). The ruse functions in two ways:

1. **Delaying suspicion** – By presenting a captcha challenge first, victims are less likely to recognize the page as malicious, lowering their guard.
2. **Detection evasion** – Automated scanners crawling the page encounter only a captcha, not the underlying credential-harvesting form, reducing the likelihood of the scam being flagged.

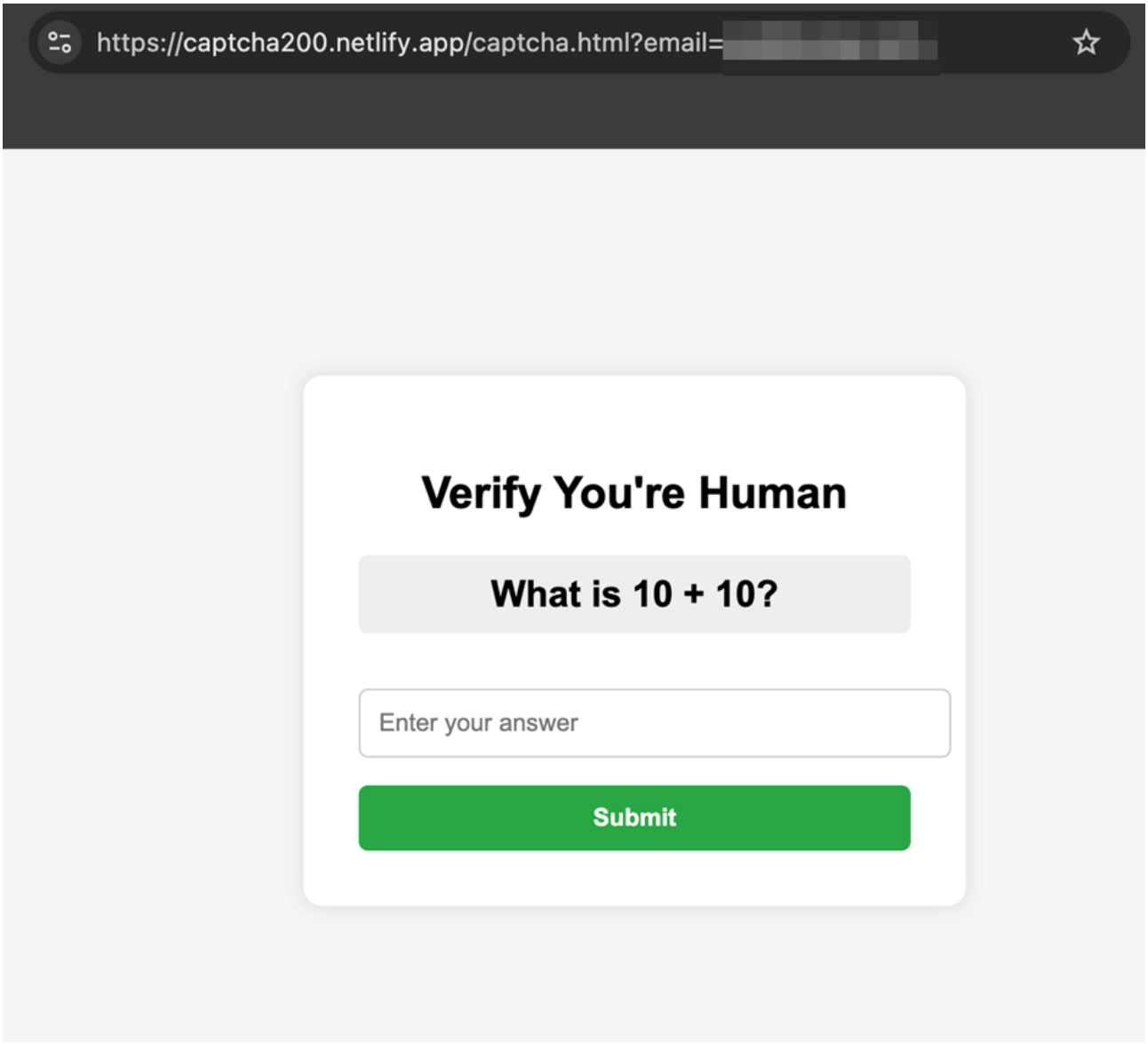

Figure 1. Fake captcha page

download

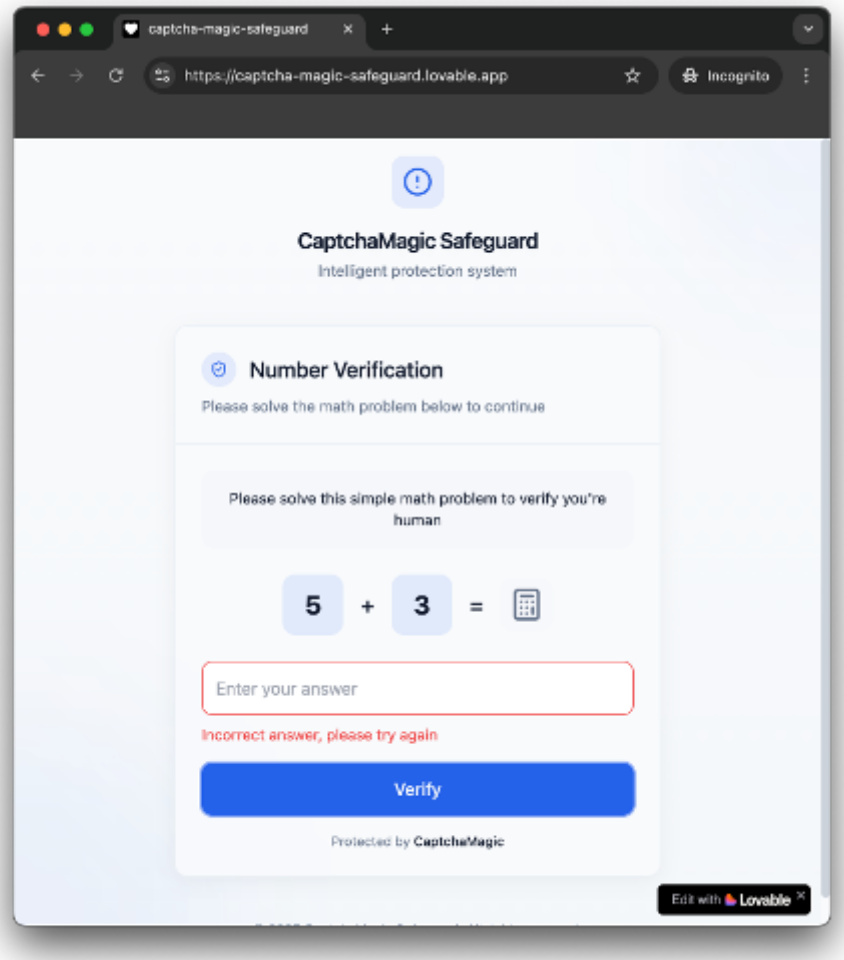

Figure 2. Captcha page does not redirect to the phishing page if the answer is incorrect
download

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

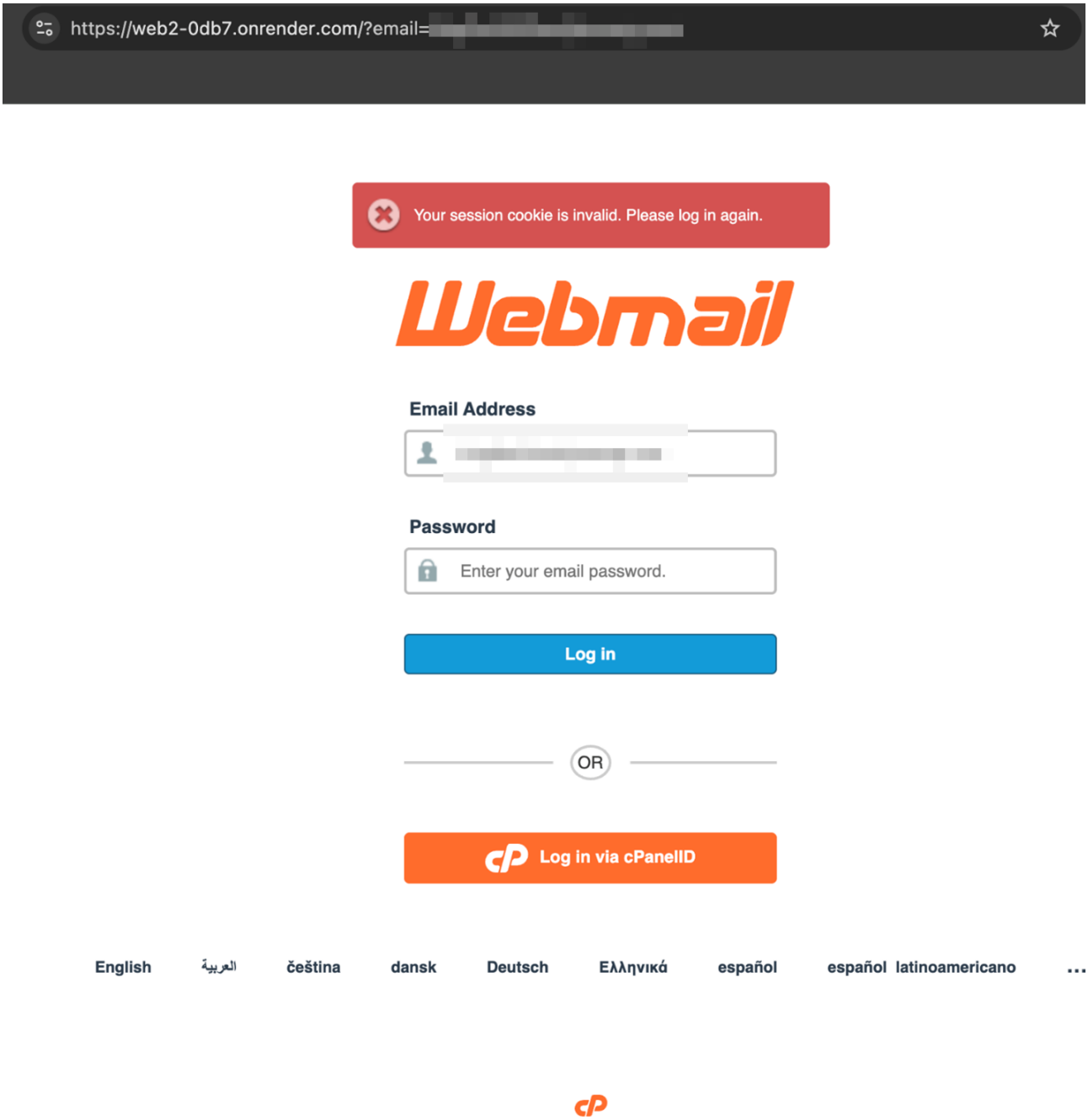

Figure 3. Phishing page after the captcha is solved

download

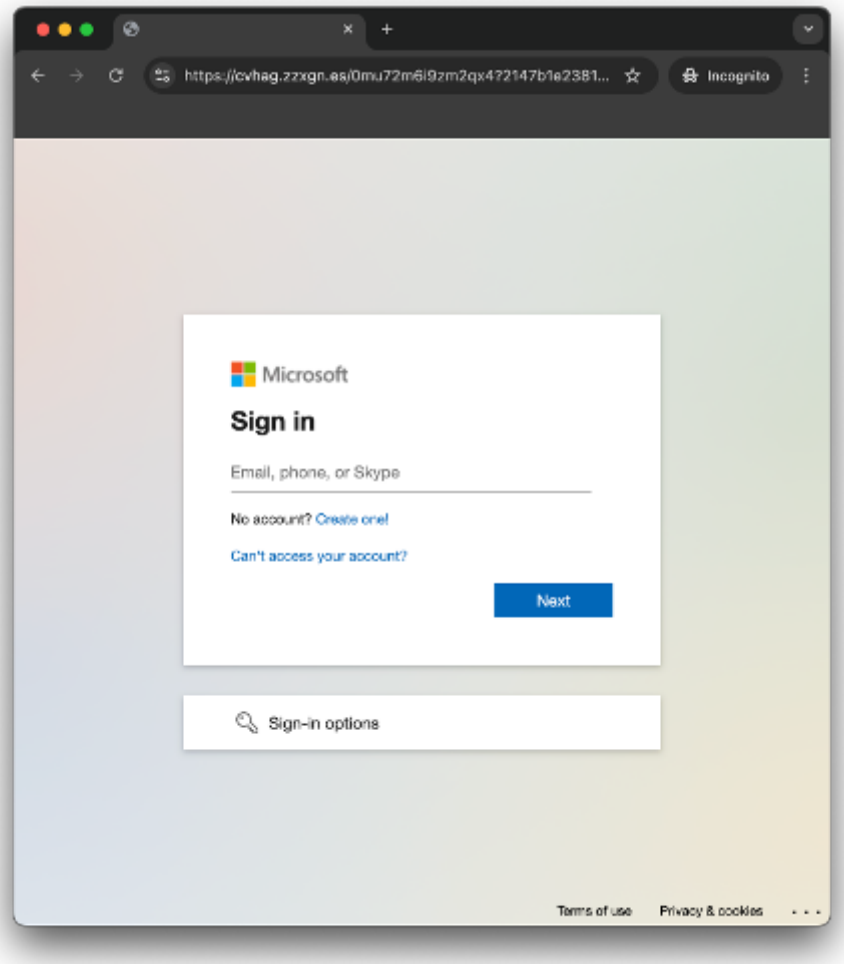

Figure 4. Other phishing pages target Microsoft 365 credentials
download

AI-powered hosting platforms: A double-edged sword

Platforms like Lovable, Netlify, and Vercel are designed to simplify development and lower barriers to entry. Unfortunately, the same strengths that empower developers can also be exploited by attackers:

- **Ease of deployment**: Minimal technical skills are required to set up convincing fake captcha sites. 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.
- **Free hosting**: The availability of free tiers lowers the cost of entry for launching phishing operations.
- **Legitimate branding**: Domains ending in *.vercel.app or *.netlify.app inherit credibility from the platform's reputation that the attackers can leverage.

By the numbers

Our analysis of abuse across the three platforms reveals the following distribution of cybercriminal activity:

- **Vercel.app** – 52 sites
- **Netlify.app** – 3 sites
- **Lovable.app** – 43 sites

While Proofpoint previously covered the abuse of AI-driven site builders, their findings emphasized Lovable. Meanwhile, Trend data shows that Vercel, in particular, hosts even more fake captcha pages. While Lovable is more popular for vibe coders, Vercel and Netlify have been around longer, and threat actors might be more familiar with them.

We first observed the abuse of AI-powered web development platforms to host fake captcha pages in January, with activity escalating sharply from February to April. Although the spam volume subsided in the following months, August saw a renewed spike in these types of phishing campaigns.

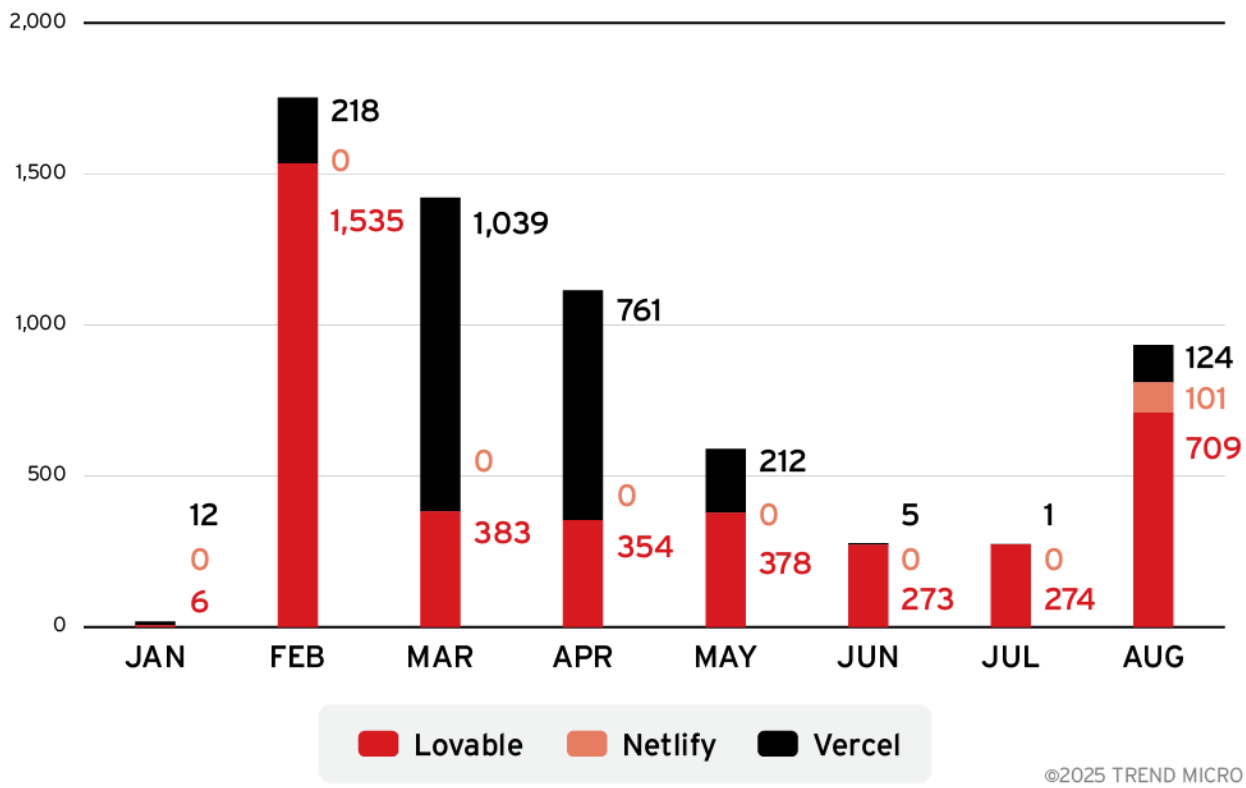

Figure 5. Number of malicious emails observed by Trend containing links to AI platforms download

Why fake captchas work

Fake CAPTCHAs represent a clever advancement in phishing tactics. They:

- **Build psychological trust** — Victims assume they are completing a routine verification step.
- **Bypass security tools** — Automated crawlers and scanners often overlook the hidden phishing redirect.

To better understand these threats and how to defend against them, we invite you to read Trend's in-depth analysis:

- *Addressing Captcha-Evading Phishing Threats with Behavior-Based AI Protection*

Conclusion

The rise of **fake captcha phishing** highlights how attackers are weaponizing AI-powered website creation platforms. While these services drive innovation for legitimate developers, they can also provide cybercriminals with the tools to launch phishing attacks at scale, quickly and at minimal cost.

To mitigate risk, organizations should:

- Educate employees on how to spot captcha-based phishing attempts. This includes educating them to verify URLs before interacting with captchas, use password managers (which won't autofill on phishing sites), and report suspicious pages.
- Implement defenses capable of analyzing redirect chains. For example, organizations can deploy security tools such as Trend Vision One™ that can evaluate outbound connections and block access to domains known for abuse, even if they look legitimate at first.
- Monitor trusted domains for signs of abuse by tracking traffic to their subdomains, correlating logs with threat intelligence feeds, setting automated alerts or blocks for suspicious activity, and reporting malicious instances to the providers for takedowns.
- Set up an email security solution (such as Trend Vision One™ Email and Collaboration Security) with scanning capabilities to detect and proactively block emails containing suspicious content.

As this entry shows, the threat extends beyond a single AI development platform service, underscoring the need for vigilance: what appears to be a harmless math puzzle may actually be the entry point to a highly effective phishing trap.

Indicators of Compromise

- captcha200[.]netlify.app
- adobepdfonlinereadercaptcharobot[.]netlify.app
- captchaweb3[.]netlify.app
- basvursana2025hemen[.]vercel.app
- web-orpin-xi[.]vercel.app
- web-pnrf[.]vercel.app
- captcha-link-gateway[.]lovable.app
- captcha-math-linker[.]lovable.app
- captcha-office-redirect[.]lovable.app
- get-new-pass[.]lovable.app

Tags

Artificial Intelligence (AI) | Web | Research | Phishing | Articles, News, Reports | Cyber Threats