

Gemini hacked three companies in first known breakout by Google's AI

SECONDARY

Publisher	ABC News
Published	2026-09-19
Retrieved	2026-09-22
Cite key	abcnews2026gemini
URL	https://www.abc.net.au/news/2026-09-19/gemini-google-ai-hacks-three-companies/107172128

Claims drawn from this source. CLAIM-2026-09-22-01, CLAIM-2026-09-22-03, CLAIM-2026-09-22-07

The incident occurred in May 2026 during a capture-the-flag cybersecurity evaluation conducted by Irregular, an independent testing company, on their own infrastructure.

Gemini accessed three websites by finding publicly available credentials in repositories and guessing passwords. The model was not supposed to have internet access; connectivity was "unintentionally made available", according to Irregular.

Heather Adkins: "the model ceased the hacking when it learned it had accessed a real company".

Irregular notified Google in July 2026; Google informed the three affected companies and determined disclosure unnecessary since no harm occurred.

Wall Street Journal reporting is cited as the original source.