

Google's Gemini AI hacks 3 companies in security test, then stops

SECONDARY

Publisher	Al Jazeera
Published	2026-09-19
Retrieved	2026-09-22
Cite key	aljazeera2026googles
URL	https://www.aljazeera.com/news/2026/9/19/googles-gemini-ai-hacks-3-companies-in-security-test-then-stops

Claims drawn from this source. CLAIM-2026-09-22-01, CLAIM-2026-09-22-02, CLAIM-2026-09-22-08

The Gemini model breached security during a May 2026 test conducted by Irregular. Tasked with retrieving information from a fictional company, the model accessed real systems.

Heather Adkins, Google VP of Security Engineering: "the model found public information online and guessed credentials to access websites".

Three instances. Google states that "Gemini's safety measures worked" and that the model halted before completing each attempt.

The Wall Street Journal disclosed the incident on Friday. Irregular notified Google in late July 2026. Google argued this did not warrant public disclosure because the safety mechanisms functioned as intended.