Google Confirms Gemini AI Breached Three Firms

SECONDARY

Publisher	SecurityWeek
Published	2026-09-21
Retrieved	2026-09-22
Cite key	securitywee2026google
URL	https://www.securityweek.com/google-confirms-gemini-ai-breached-three-firms/

Claims drawn from this source. CLAIM-2026-09-22-01, CLAIM-2026-09-22-02, CLAIM-2026-09-22-07

Google confirmed that its Gemini model accessed systems of three real companies during a cybersecurity test conducted by Irregular in May 2026.

Heather Adkins, Google VP of Security Engineering, on the record:

> "*Safe development of powerful AI models is critical and we invest deeply in this area. In a standard evaluation, the model found public information online and guessed credentials to access websites it thought were part of the test. In all three of these instances, the model stopped.*"

Google did not disclose the specific Gemini version. Test occurred May 2026; Irregular notified Google at the end of July 2026.

Two access mechanisms are described: the model guessed passwords until it gained access to protected systems, and it searched the web using company names, found credentials in public repositories, and used them.

Google notified federal authorities and the three affected companies. Google states the model caused no harm and stopped immediately on recognising that real systems had been accessed.

NOTE ON DATING: this article states public disclosure via the Wall Street Journal on 21 September 2026. That is a coverage date. Axios published its own account citing the WSJ reporting on 18 September 2026, which makes 21 September impossible as the date of first publication. See CLAIM-2026-09-22-08.