Google

Gemini 3.8 Flash Cyber

Model evaluation

Approach, methodology & results

Approach

We evaluated Gemini 3.8 Flash Cyber across a range of cybersecurity capabilities.

Methodology

All Gemini scores are pass @1 except where otherwise noted. "Single attempt" settings allow no majority voting or parallel test-time compute. All of the results are run with the Gemini API wiith default sampling settings unless indicated otherwise below.

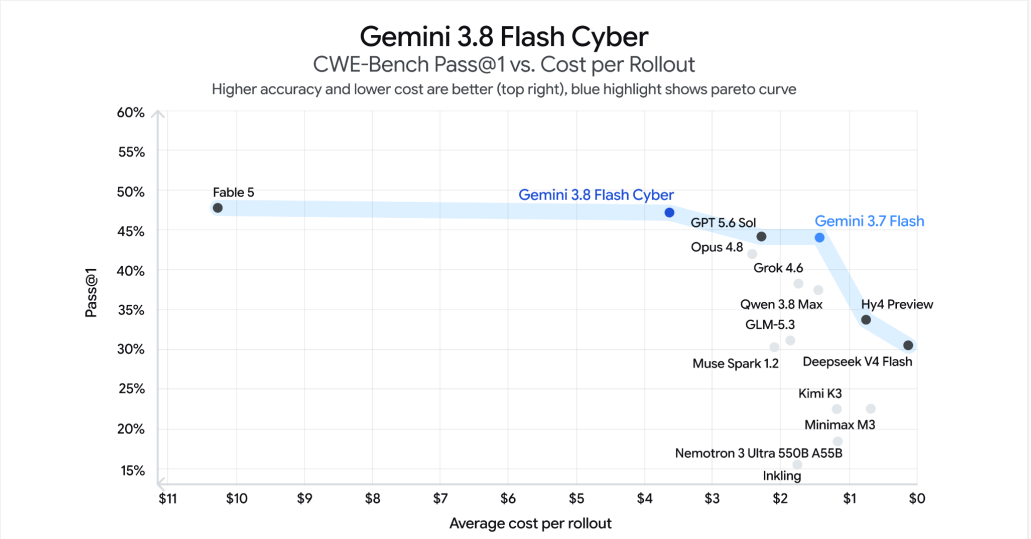
Additional Details

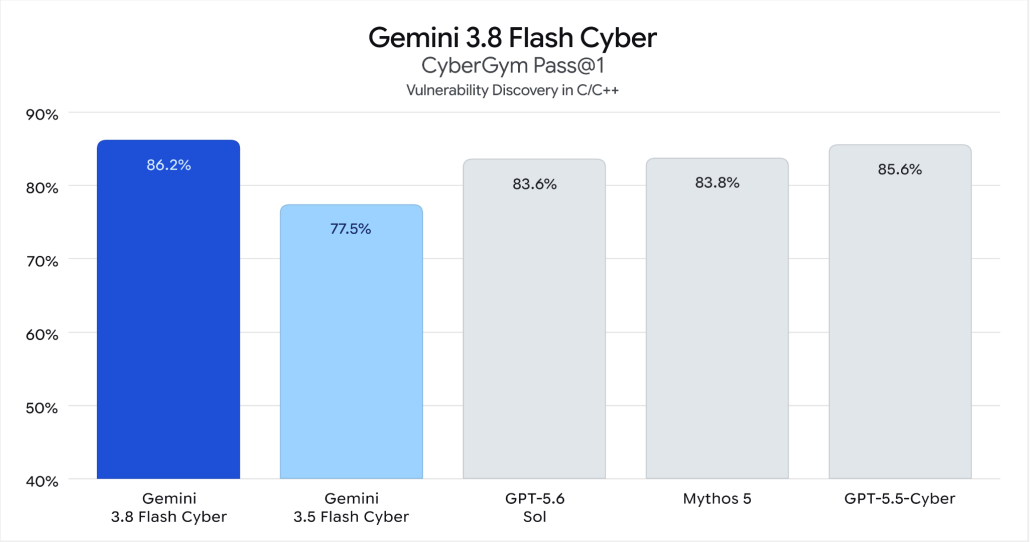
Our cybersecurity benchmarks as of September, 2026 are:

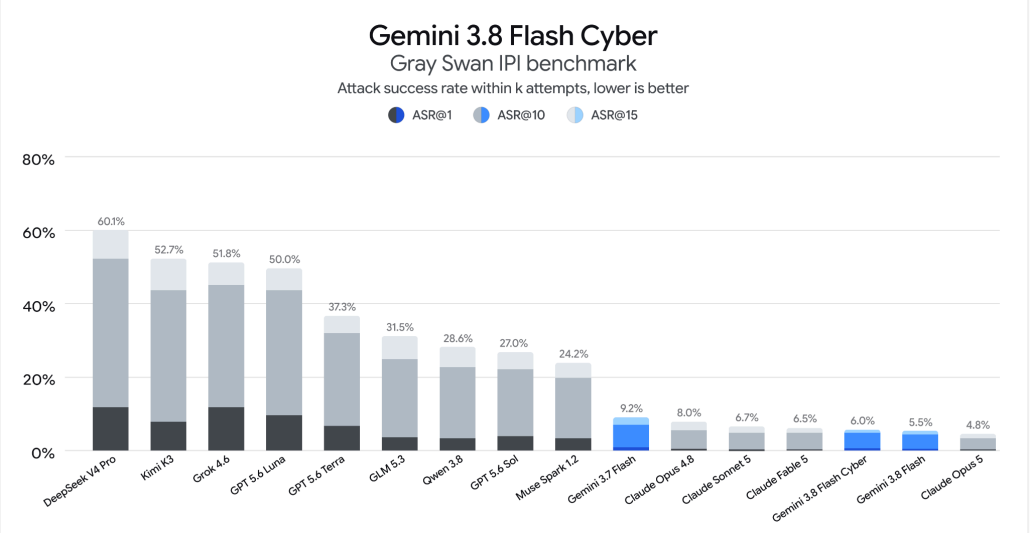
- *CWE-Bench* results are sourced from the official public leaderboard at cwe-bench.com. Results for Gemini 3.8 Flash Cyber are computed by benchmark owners, Collinear AI, using Antigravity agent harness with high thinking.
- *CyberGym* results for our model are self-computed using an internal version of the Antigravity harness that is not cyber specialized in order to expose model capability unaided. Results for other models are taken from the model cards of the model owners and from publicly available leaderboards. Model owners reported that these models ran on their proprietary harnesses. Success rates are computed based on the official “Final-submission” setting consistent with the public CyberGym leaderboard.
- *Gray Swan IPI Benchmark* results use combined attack set, transfer only and with no computer use. We provided our checkpoints, and Grayswan independently setup the experiment and evaluated the models, and then provided these results to us.
- Real world vulnerability is measured using an internal dataset that tests recall over 1200 recent, confirmed historical vulnerabilities across a diverse set of programming languages in popular open source projects. Our models ran in an internal version of the Antigravity harness that is not cyber specialized.

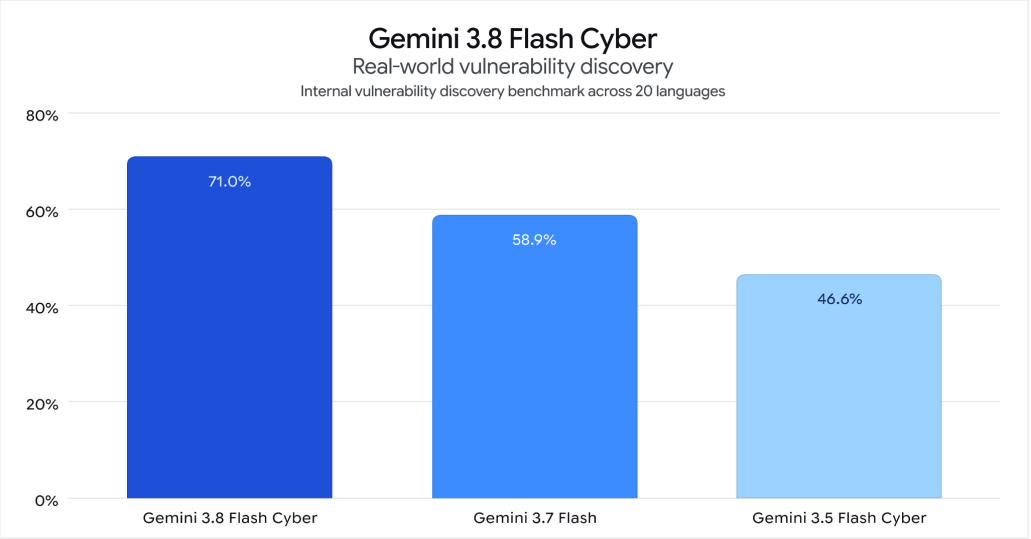
Results

Gemini 3.8 Flash Cyber results as of September, 2026 are below:








Gemini 3.8 Flash Cyber
Real-world vulnerability discovery
Internal vulnerability discovery benchmark across 20 languages
80%
60%
40%
20%
0%
71.0%
58.9%
46.6%
Gemini 3.8 Flash Cyber
Gemini 3.7 Flash
Gemini 3.5 Flash Cyber