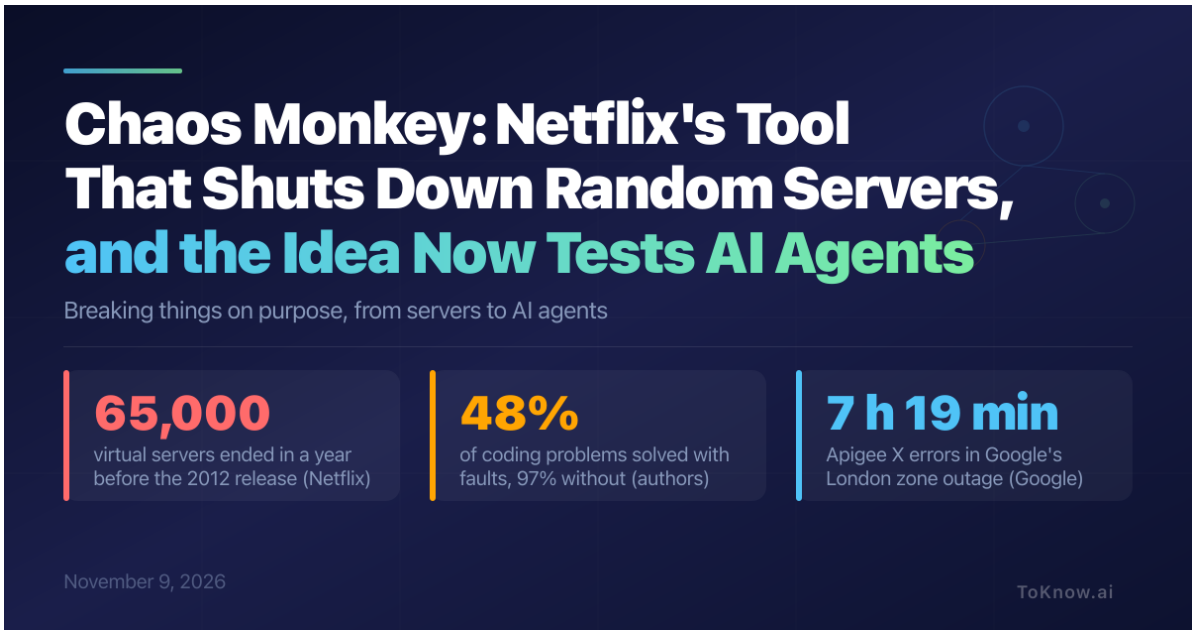
Chaos Monkey: Netflix's Tool That Shuts Down Random Servers, and the Idea Now Tests AI Agents

Kabui, Charles

2026-11-09

Read at ToKnow.ai



Chaos Monkey is a tool that shuts down random servers in a live service on purpose, to test whether it survives. It is for any team running a website or app in the cloud. Netflix described it in 2010 and released the code in July 2012. It ends virtual servers, not hardware, by default

on weekdays from 9am to 3pm, with engineers on hand. In the year before the release it ended over 65,000 of them in Netflix's live and test systems, and Netflix wrote that "most of the time nobody notices."

Researchers in 2026 ran the same test on AI agents, programs that use an AI model and tools to finish tasks. The AgentChaos study fed five agent programs damaged model replies (missing, cut off or corrupted), and all did worse. With Claude Sonnet 4.5, the coding agent MapCoder solved 48% of problems with faults against 97% without. With AutoGen, 65.71% of failed tasks ended in a plausible but wrong answer with no warning. ReliabilityBench, a one-author test, saw Gemini 2.0 Flash agents fall from 96.9% to 84.0% under reworded requests and faults.

Chaos Gorilla, another Netflix tool, simulates losing a whole Amazon availability zone (a group of data centres in one region). On October 20, 2026, a power failure took one of three zones in Google Cloud's London region offline. Google reports its Apigee X service, built to survive losing a zone, had put two of three data copies there and returned errors for 7 hours 19 minutes.

Read More: OpenSRE, an AI agent that works out why a system broke, tested by killing Postgres, Redis and the backend one at a time

Sources:

- 5 Lessons We've Learned Using AWS (Netflix Tech Blog, December 2010)
- Chaos Monkey Released Into The Wild (Netflix Tech Blog, July 30, 2012, archived copy)
- AgentChaos: Chaos Engineering for Agent Systems via Programmatic Fault Injection (arXiv)
- ReliabilityBench: Evaluating LLM Agent Reliability Under Production-Like Stress Conditions (arXiv)
- Incident report: multiple services in europe-west2 (Google Cloud Service Health)

Disclaimer: *For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation.* ***Read more:*** */terms-of-service*