

Gemini Goes OS-Level, AI Cyberattacks Cross a Line, and Benchmarks Reset

AI High Signal Digest

2026-05-13

Gemini Goes OS-Level, AI Cyberattacks Cross a Line, and Benchmarks Reset

By AI High Signal Digest • May 13, 2026

Google moved Gemini deeper into Android while Google's threat team disclosed the first known AI-developed zero-day in the wild. The brief also covers benchmark saturation, key research advances in math and RL infrastructure, and major product and funding moves across consumer AI, biotech, and enterprise software.

Top Stories

Why it matters: The biggest news today points to AI moving deeper into operating systems, crossing a new cyber risk threshold, and quickly outgrowing existing evals.

- **Google pushed Gemini deeper into Android.** Gemini Intelligence adds multi-step task automation across apps, one-tap form fill, polished dictation, and custom widgets, starting on Galaxy and Pixel this summer and later expanding to watches, cars, glasses, and laptops. Google also previewed an AI-enabled pointer that understands what is under the cursor and combines pointing with speech, reinforcing its push to turn Android into an “intelligence system.” [1, 2, 3, 4, 5]
- **Google reported the first known AI-developed zero-day used in the wild.** Its Threat Intelligence Group said the attackers planned a wide-scale strike, though proactive counter-discovery may have stopped it. [6] *Impact:* AI cyber risk is no longer just hypothetical.
- **Frontier benchmarks are already being reset.** GPT-5.5 high/xhigh solved the first ProgramBench task and xhigh outperformed Opus 4.7 xhigh across all metrics; separately, GPT-5.5 solved the last unsolved MathArena Apex problem and pushed Apex Shortlist above 90% accu-

racy. Benchmark builders are now creating newer tests and deprecating some final-answer competitions because models have become too strong for the old format. [7, 8, 9, 10]

Research & Innovation

Why it matters: The most useful technical progress focused on expert workflows, agent training efficiency, and lower-cost compute paths.

- **DeepMind’s AI Co-Mathematician** reached 48% on FrontierMath Tier 4, a new high among evaluated AI systems. The system is an asynchronous, stateful workbench for ideation, literature discovery, computation, theorem verification, and knowledge development; early sessions reportedly solved open problems and surfaced overlooked citations. [11]
- **PrimeIntellect’s Renderers** fix the mismatch between token-based RL trainers and message-based environments, which had been corrupting sampled tokens and wasting compute on agentic turns. PrimeIntellect says the change unlocks more than 3x throughput on popular open models. [12]
- **Sakana AI and NVIDIA’s TwELL** use a new sparse format plus custom CUDA kernels to exploit >95% sparsity in transformer feedforward layers, translating that into >20% faster training and inference on H100s alongside memory and energy savings. [13, 14]

Products & Launches

Why it matters: New releases kept pushing down cost, widening access, and making multimodal AI more usable.

- **DeepSeek-v4-Flash** was described in leaderboard comparisons as essentially equal to, and sometimes stronger than, v4-Pro while being faster and about 10x cheaper. [15, 16]
- **Microsoft rolled out MAI-Image-2-Efficient** globally in Bing Image Creator, now free for everyone, with sharper detail, richer color, better text rendering, and more accurate prompt following than v1. [17]
- **Meta launched Voice Conversations in Meta AI** powered by Muse Spark, with interruptions, topic switches, multilingual speech, real-time image generation, recommendations, and a live AI camera mode; Meta says it is available today. [18, 19]

Industry Moves

Why it matters: Capital and M&A continue to cluster around domain-specific deployment and AI-enabled software businesses.

- **Isomorphic Labs raised \$2.1B** to accelerate AI drug discovery, building on AlphaFold; Demis Hassabis called improving human health AI’s top application. [20]

- **Anthropic is reportedly in talks to acquire Stainless for \$300M+**, a move that would remove a developer-tools supplier used by OpenAI and Google. [21]
- **Notion said AI now accounts for 60% of its business.** The company closed Q1 with revenue accelerating for the seventh straight quarter and said it is cash-flow positive. [22]

Quick Takes

Why it matters: These smaller updates sharpen the picture on data, voice agents, medical evals, and robotics.

- Hugging Face crossed **1,000,000 public datasets**, with the total doubling in the last eight months. [23]
- Artificial Analysis launched **-Voice**; **Grok Voice Think Fast 1.0** led at **52.1%**, and even the strongest speech-to-speech models resolved only about half of realistic customer-service scenarios. [24]
- **Medmarks v1.0** expanded the largest open-source automated medical LLM benchmark suite to **30 benchmarks** and **61 models**. [25, 26]
- Figure said its **F.04** humanoid reached **design lock** and has started shipping parts. [27]

Sources

1. X post by @Google
2. X post by @sundarpichai
3. X post by @GoogleDeepMind
4. X post by @GoogleDeepMind
5. X post by @TheRundownAI
6. X post by @NewsFromGoogle
7. X post by @KLieret
8. X post by @OfirPress
9. X post by @Chenhao3564
10. X post by @j_dekoninck
11. X post by @dair_ai
12. X post by @PrimeIntellect
13. X post by @SakanaAILabs
14. X post by @hardmaru
15. X post by @teortaxesTex
16. X post by @j_dekoninck
17. X post by @JordiRib1
18. X post by @MetaNewsroom
19. X post by @jhyuxm
20. X post by @demishassabis
21. X post by @steph_palazzolo

22. X post by @ivanhzhao
23. X post by @ClementDelangue
24. X post by @ArtificialAnlys
25. X post by @SophontAI
26. X post by @iScienceLuvr
27. X post by @adcock_brett