The State of Frontier AI, Q2 2026: Opus 4.7 vs Gemini 3.1 Pro vs GPT 5.4

The second quarter of 2026 marks a transformative epoch in the trajectory of artificial intelligence, characterized by a fundamental shift from static conversational agents to high-autonomy, agentic systems capable of long-horizon reasoning. This transition is defined by the convergence of three primary contenders: Anthropic's Claude Opus 4.7, Google's Gemini 3.1 Pro, and OpenAI's GPT-5.4. While the preceding eighteen months focused heavily on the expansion of context windows and raw parameter scaling, the current competitive landscape is dominated by the operationalization of these models within complex, multi-tool enterprise environments and the refinement of the user experience through native system-level integration. The following analysis provides a forensic decomposition of these models across enterprise and consumer vectors, synthesizing technical performance metrics with the qualitative reality of daily deployment.

The Enterprise Battlefield

The enterprise sector has moved beyond the experimental phase of AI adoption, with C-suite

executives now demanding quantifiable returns on investment and developers seeking tools that can manage the cognitive load of massive, multi-session codebases. The Q2 2026 model releases address these needs through diverging strategies in agentic workflows, financial modeling, and security architectures.

Agentic Coding and Long-Horizon Workflows

Claude Opus 4.7 has positioned itself as the preeminent tool for professional software engineering, driven by what Anthropic describes as "adaptive thinking".[1] This mechanism allows the model to dynamically scale its internal reasoning process based on task complexity, a feature that addresses the inherent inefficiency of using the same compute for a simple syntax correction as for a complex architectural refactor. In industrial settings, this has translated into a model that plans more carefully, operates reliably across larger codebases, and demonstrates a heightened ability to catch its own errors.[1] The introduction of the "xhigh" effort level provides a granular control mechanism for developers, sitting between the standard high and maximum settings to optimize the latency-to-reasoning ratio.[2]

OpenAI's GPT-5.4 approaches the coding battlefield through the lens of a "Codex superapp." The strategy here is not just code generation but comprehensive computer-use (CUA). GPT-5.4 features persistent computer interaction capabilities, allowing it to navigate file systems, terminal environments, and browsers as a human would, often cutting token usage by two-thirds in sustained sessions through more efficient state management.[4] This model is designed to operate in parallel; multiple agents can function on a workstation without disrupting the user's primary tasks, effectively serving as an autonomous background labor force.[5]

Google's Gemini 3.1 Pro leverages its native ecosystem tie-ins to differentiate itself in the developer market. While Opus 4.7 focuses on "implicit-need tests"—tasks where the model must infer required tools without explicit instruction—Gemini 3.1 Pro focuses on throughput and integration within the Google Cloud and Vertex AI environments.[6] Its primary advantage in the coding sector is the balance it strikes between reasoning depth and speed, particularly for competitive coding and real-world software engineering issues.[8]

Coding Benchmark	**Claude Opus 4.7**	**GPT-5.4**	**Gemini 3.1 Pro**
SWE-bench Pro	64.3%	57.7%	54.2%
SWE-bench Verified	87.6%	Not Available	80.6%

Terminal-Bench 2.0	69.4%	75.1%*	68.5%
CursorBench	70%	Not Available	Not Available
LiveCodeBench Pro (Elo)	Not Available	Not Available	2,887

Note: GPT-5.4 score utilizes a self-reported harness and is not directly comparable with standard benchmarks.[9]

Analysis of these figures indicates that while Opus 4.7 has retaken the lead in raw software engineering proficiency, particularly on the industrially relevant SWE-bench Pro, GPT-5.4 remains the superior choice for command-line navigation and terminal-based tasks.[6] Gemini 3.1 Pro, meanwhile, dominates in competitive coding environments, as evidenced by its high Elo on LiveCodeBench Pro.[8] The 10.9% jump for Opus 4.7 over its predecessor (Opus 4.6) on SWE-bench Pro is a significant industrial milestone, as it represents a model capable of handling the harder multi-language engineering tasks that previously required human intervention.[9]

Cost and Compute Efficiency

The financial modeling for frontier AI deployment has undergone a radical transformation in Q2 2026. Anthropic has moved away from its earlier "bundled token" enterprise seat deals, transitioning toward a metered pricing model that has caused significant friction among existing customers.[10] Under this new paradigm, seat fees have dropped to as low as $20, but token consumption is metered on top, potentially tripling the bills for heavy enterprise users.[11] Furthermore, the introduction of a new tokenizer in Opus 4.7 means that the same input now maps to 1.0–1.35x more tokens, an "invisible" price increase that complicates budgeting for long-running agents.[2]

OpenAI's pricing for GPT-5.4 introduces a tiered "long context tax." While the baseline input cost is $2.50 per million tokens, sessions exceeding 272,000 input tokens are billed at a 2x rate for input and a 1.5x rate for output.[13] This creates a massive financial disincentive for organizations processing massive document sets or maintaining extremely long conversation states. In contrast, Anthropic and Google have maintained flat pricing regardless of context depth within their million-token windows.[14]

Google's value proposition revolves around its bundled Cloud offerings and the aggressive pricing of Gemini 3.1 Pro, which remains at $2 per million input tokens despite its performance

gains.[8] Google's context caching feature allows for a reduction in input costs by up to 75%, making it the most cost-effective option for recurring tasks on stable datasets.[8]

Pricing Vector (per 1M tokens)	Claude Opus 4.7	GPT-5.4	Gemini 3.1 Pro
Standard Input	$5.00	$2.50	$2.00
Standard Output	$25.00	$15.00	$12.00
Cached Input	$0.50	$0.25	$0.20
Long Context Penalty	None	2x (>272K)	None
Batch Discount	50%	Not Specified	Up to 50%

The economic reality of 2026 suggests that the ROI for these models is often "an illusion," as providers are currently charging rates that do not cover the true cost of compute.[11] Enterprises are essentially being subsidized by the providers' capital burn, a state that users recognize as unsustainable in the long term.[11]

Security and Domain Specialization

The diverging philosophies regarding model security define the current relationship between AI providers and the enterprise. Anthropic has adopted a "walled-garden" strategy, particularly with its Mythos class models. Claude Mythos Preview is withheld from general release due to its unprecedented offensive capabilities, including the ability to autonomously identify and exploit zero-day vulnerabilities in critical infrastructure.[18] Access is strictly limited to a small circle of partners through "Project Glasswing," a program that provides usage credits and early access to vetted security firms like CrowdStrike and JPMorganChase.[20]

OpenAI has countered this with a strategy of democratized access through its Trusted Access for Cyber (TAC) program. Instead of holding back models entirely, OpenAI has released GPT-5.4-Cyber, a variant trained to be "cyber-permissive".[22] This model deliberately lowers

refusal boundaries for legitimate security tasks, such as binary reverse engineering and malware analysis, while relying on robust identity verification and KYC (Know Your Customer) protocols to prevent misuse.[22] Additionally, OpenAI has targeted the life sciences sector with GPT-Rosalind, a restricted model fine-tuned for experimental planning, hypothesis generation, and evidence synthesis within scientific research workflows.[5]

Google's approach remains tied to its broad Workspace and Cloud integration, focusing on data sovereignty. The expansion of Google Cloud Interconnect (GCI) into regions like Calgary, Alberta, allows Canadian enterprises to process data within national borders, a critical requirement for regulated industries like finance and healthcare.[24] This regional specialization is paired with "web grounding for enterprise," which connects Gemini to specific internal data sources rather than just the public web.[7]

The Everyday User Experience

The consumer market in Q2 2026 is a battleground of "muscle memory" and "vibe checks." While technical benchmarks provide a baseline, the daily driver experience is dictated by UI friction, the perceived "humanity" of the AI, and the seamlessness of desktop integration.

The Reddit Consensus

Synthesis of feedback from the primary artificial intelligence communities suggests a complex hierarchy of preferences. Claude Opus 4.7 is currently winning the "vibe check," particularly for tasks involving visual design and creative synthesis. Users describe it as having a "sweet" and "smart" persona that avoids the lecturing tone often associated with earlier frontier models.[25] However, the primary complaint remains the restrictive usage limits on the Pro tier, which often feels "brutal" and inconsistent compared to the more generous limits of ChatGPT and Gemini.[25]

GPT-5.4 faces a different set of friction points. While recognized for its logic and ability to act as a "critic" for ideas, its UI is increasingly viewed as "crowded and awkward".[27] The integration of the "gigantic dashboard" is a frequent target of ridicule, with users noting it consumes excessive screen real estate for information that is not always relevant to the current task.[27]

Gemini 3.1 Pro is often described as "boring" but "functional".[28] Its integration with Google's productivity suite is its greatest strength, but users report a persistent friction point where the model defaults to brief, search-result-heavy answers rather than in-depth reasoning unless specifically nudged.[26]

Dominant Themes from Consumer Discourse

One of the most compelling themes emerging from the Reddit consensus is the "Gym Membership" analogy. Users express a fear of canceling any one subscription, even if they aren't using it properly, for fear of losing access to a specific strength—ChatGPT for tools, Gemini for raw power, and Claude for "actually reading before answering".[29] This has led to a market state where users are "not picking an ecosystem" but "waiting for one of them to

disappoint me so badly that the decision makes itself".[29]

Another theme is the "Cards on Cards" critique of UI design. Users have noted a lack of imagination in how AI interacts with the desktop, with one user stating, "If they are unable to make UI decisions beyond cards, it's all still garbage".[25] This reflects a broader desire for AI to move beyond the "chat window" paradigm and into more integrated, contextual forms.

The "Lily of the Valley" metaphor provides a unique insight into the aesthetic perception of GPT-5.4. Some users view it as a delicate but potentially "poisonous" flower—beautiful and relaxing for those who like it, but requiring a high "sense of crisis" regarding its own worth and the truth of its outputs.[26]

Desktop Integration

The battle for desktop muscle memory has intensified with Google's launch of its native Gemini Mac app, which requires macOS 15 Sequoia or newer.[30] The app's primary innovation is the Option + Space shortcut, which allows users to bring up the assistant instantly without breaking their workflow.[30] Its standout feature, "Share Window," enables the model to see specific application windows (such as a chart in a spreadsheet or a data sheet) and provide context-aware help without a manual file upload.[30]

OpenAI's Codex desktop update has transformed from a coding assistant into a broader productivity workspace. It features the ability to see and interact with any app on the computer as a human would—typing, clicking, and navigating with its own cursor.[5] Crucially, Codex supports parallel background workflows, allowing it to complete administrative tasks while the user is active in other applications.[5] It also features a persistent memory that proactively suggests task prioritization based on comments from Slack, Notion, and Google Docs.[5]

Claude's desktop presence is primarily web-based or through the Claude Code environment. While it lacks the pervasive "Option + Space" system-wide integration of Gemini, it is preferred for "Projects," which allow users to isolate specific instructions and reference data for recurring workflows.[1]

Nuance and Creativity

For writing and brainstorming, Opus 4.7 is widely regarded as the most "human" and "tasteful" model.[28] It produces higher-quality professional interfaces and slide decks, and its creative writing is described as having more "nuance" and requiring less editing to sound natural.[28] Anthropic's focus on honesty and self-verification has resulted in a model that "pushes back more often" when it lacks data, rather than fabricating a response, a trait that users find more trustworthy.[3]

Gemini 3.1 Pro is catching up in the "human feel" category, with users noting its ability to be "straightforward and direct without talking down like GPT".[26] Its new "Notebooks" feature allows for the synthesis of disparate files and chat histories into a single view, which is a powerful tool

for planning and brainstorming.[35]

GPT-5.4 remains the model of choice for "hard logic." It excels as a critic and can identify holes in arguments that other models might overlook.[28] However, its personality is sometimes perceived as "nihilistic" or "vague" compared to the more grounded and compliant Claude.[26]

Benchmarks vs. Reality

The technical benchmarks of Q2 2026 show a field that is rapidly converging in raw reasoning power, but diverging significantly in multimodal acuity and real-world reliability.

Technical Performance Cluster

On graduate-level reasoning benchmarks like GPQA Diamond, the differences between the leading models have effectively fallen within the noise of statistical measurement.[6]

Benchmark	Claude Opus 4.7	GPT-5.4 Pro	Gemini 3.1 Pro
GPQA Diamond	94.2%	94.4%	94.3%
GDPVal-AA (Elo)	1,753	1,674	1,314
MMMLU	91.5%	Not Available	92.6%
HLE (with tools)	54.7%	58.7%	51.4%

While reasoning scores no longer provide a clear winner, the differentiation has moved to specialized benchmarks like Visual Acuity and Finance Agent. Opus 4.7's jump from 54.5% to 98.5% in visual acuity is the single largest performance leap in the Q2 cycle.[3] This allows the model to process images up to 2,576 pixels on the long edge, a 3.3x improvement that transforms its utility for reading dense screenshots and technical diagrams.[12]

The Lived Reality of Deployment

The question facing many organizations is whether a 5% to 10% gain on a benchmark actually translates to a better user experience. In the coding sector, the 10-point jump for Opus 4.7 on SWE-bench Pro is considered "meaningfully ahead" because it crosses a threshold where the

model can autonomously resolve complex concurrency bugs that previously caused it to fail.[9]

However, in the consumer sector, benchmark gains are often offset by "lazy outputs" or excessive guardrails. A model that scores 94% on a science test can still fail to write a simple email if its safety filters are too aggressive or if it defaults to a brief search summary to save compute tokens.[26] This "marketing noise" is particularly visible in agentic search, where Opus 4.7 actually showed a regression, dropping from 83.7% to 79.3% on BrowseComp.[9] For users who rely on the model for research and web synthesis, this regression is more impactful than a gain on a graduate-level physics test.

Forensic Verdicts

The Q2 2026 landscape is defined by a strategic split between Anthropic's "Focused Upgrade" and OpenAI's "General Expansion," with Google acting as the "Ecosystem Stabilizer."

The Enterprise Verdict

For the C-Suite, Claude Opus 4.7 is the definitive choice for high-stakes engineering and autonomous agents where reliability and visual precision are non-negotiable. Its dominance in SWE-bench Pro and its triple-digit lead in visual acuity make it the gold standard for production-ready code.[3] However, the metered pricing transition requires a rigorous ROI assessment, particularly for organizations with "light" usage patterns that were previously subsidized by seat deals.[11]

OpenAI's GPT-5.4 is the choice for "multitasking" environments and life sciences research. Its persistent background computer-use and parallel agent architecture provide a vision of an autonomous office that neither Anthropic nor Google has fully matched.[5] The specialized GPT-5.4 Cyber and Rosalind models offer a level of domain-specific permission that is critical for security and research teams.[21]

Google Gemini 3.1 Pro is the superior choice for large-scale, cost-sensitive deployments. Its $2/$12 pricing and massive context caching savings make it the only frontier model that is currently economically viable for high-throughput, low-margin applications.[8] Its integration with Workspace and its regional data sovereignty (as seen in Calgary) provide a low-friction path for traditional enterprises to adopt AI.[24]

The Consumer Vibe Check

In the "battle for the soul" of the consumer AI, Claude Opus 4.7 is winning on nuance and honesty.[28] It "feels" the most human and is the most reliable for creative work, though its rate limits are a constant source of frustration.[26] Gemini 3.1 Pro is winning on convenience; its "Option + Space" shortcut and native screen-sharing have the potential to become part of the desktop's daily muscle memory.[30] GPT-5.4 is currently in a "logic vs UI" crisis—it remains the smartest critic, but its crowded dashboard and "spinning in circles" behavior are pushing users

to experiment with alternatives.[26]

Podcast Narrative Hook

The proposed narrative hook for an investigative episode titled "The $2,283 Exploit: When AI Becomes Too Dangerous for Its Own Boardroom" centers on the paradoxical release of Opus 4.7. We follow the story of a security researcher who spent over $2,000 in API tokens—consuming 2.3 billion tokens in the process—to use an "inferior" model to crack the same version of Chrome that the AI company's own desktop app runs on.[18]

The episode bridges the gap between the billion-dollar safety decisions at Anthropic—withholding Mythos because it's a "zero-day machine"—and the everyday user on Reddit who is just trying to design a fitness app but keeps running into a rate limit after one prompt. We explore the "Token Squeeze" of 2026: how companies are hiking prices "invisibly" through tokenizer changes while users are left "feeling empty" after outsourcing their thinking to Claude.[11] The episode asks: If a model is too powerful to be released, is it a legitimate security threat, or the ultimate PR stunt in an era where everyone is "terrified to cancel the wrong subscription"?.[29]

The forensic data from Q2 2026 clarifies that we have entered the age of "Reliability-as-a-Service." The victor is no longer the model that knows the most, but the one that fails the least in an eight-hour, autonomous workflow. While Opus 4.7 holds the technical lead in engineering and vision, the economic and integration advantages of Gemini and the agentic breadth of GPT-5.4 ensure that the frontier remains a three-way stalemate for the foreseeable future. Organizations must move beyond benchmark tourism and into rigorous, workload-specific testing to survive the "Rug-Pull" era of AI economics.

Works cited

1. Claude Opus 4.7 - Anthropic, accessed on April 18, 2026, https://www.anthropic.com/claude/opus
2. Anthropic launches Claude Opus 4.7 with enhanced coding and vision capabilities, accessed on April 18, 2026, https://www.streetinsider.com/Corporate+News/Anthropic+launches+Claude+Opus+4.7+with+enhanced+coding+and+vision+capabilities/26322789.html
3. Claude Opus 4.7 is Now on Overchat AI — Anthropic's Most Advanced AI Model | AI Hub, accessed on April 18, 2026, https://overchat.ai/ai-hub/claude-opus-4-7
4. Computer Use & Frontend UI with GPT-5.4 Thinking, accessed on April 18, 2026, https://www.youtube.com/watch?v=YICiHiU2GBU
5. GPT-Rosalind: OpenAI's New Bet on Faster Breakthroughs in Medicine - eWeek, accessed on April 18, 2026, https://www.eweek.com/news/openai-gpt-rosalind-life-sciences/
6. Anthropic releases Claude Opus 4.7 with benchmark-leading coding and agentic performance - TNW, accessed on April 18, 2026,

https://thenextweb.com/news/anthropic-claude-opus-4-7-coding-agentic-benchmarks-release
7. Vertex AI Pricing | Google Cloud, accessed on April 18, 2026, https://cloud.google.com/vertex-ai/generative-ai/pricing
8. Google Gemini 3.1 Pro: Benchmarks, Pricing & Guide - Digital Applied, accessed on April 18, 2026, https://www.digitalapplied.com/blog/google-gemini-3-1-pro-benchmarks-pricing-guide
9. Claude Opus 4.7 Benchmarks Explained - Vellum AI, accessed on April 18, 2026, https://www.vellum.ai/blog/claude-opus-4-7-benchmarks-explained
10. Anthropic • Page 1 • Tag - The Register, accessed on April 18, 2026, https://www.theregister.com/Tag/Anthropic/
11. Anthropic squeezes enterprises by ejecting bundled tokens from seat deal - Reddit, accessed on April 18, 2026, https://www.reddit.com/r/technology/comments/1snnstj/anthropic_squeezes_enterprises_by_ejecting/
12. Introducing Claude Opus 4.7 - Anthropic, accessed on April 18, 2026, https://www.anthropic.com/news/claude-opus-4-7
13. GPT-5.4 Model | OpenAI API, accessed on April 18, 2026, https://developers.openai.com/api/docs/models/gpt-5.4
14. Claude Opus 4.7 Deep Dive: Capabilities, Migration, and the New Economics of Long-Running Agents | Caylent, accessed on April 18, 2026, https://caylent.com/blog/claude-opus-4-7-deep-dive-capabilities-migration-and-the-new-economics-of-long-running-agents
15. Gemini Developer API pricing, accessed on April 18, 2026, https://ai.google.dev/gemini-api/docs/pricing
16. Vertex AI Pricing: The Complete 2026 Guide to Costs, Hidden Fees, and Savings - nOps, accessed on April 18, 2026, https://www.nops.io/blog/vertex-ai-pricing/
17. The Daily A.I. — AI, Photography, Design & Marketing Headlines | SuperStock, accessed on April 18, 2026, https://mktg.superstock.com/daily-ai
18. Claude Opus wrote a Chrome exploit for $2,283, accessed on April 18, 2026, https://www.theregister.com/2026/04/17/claude_opus_wrote_chrome_exploit/
19. Claude Mythos Preview System Card, accessed on April 18, 2026, https://reason.com/wp-content/uploads/2026/04/Claude-Mythos-Preview-System-Card1.pdf
20. GPT-5.4 Cyber vs Claude Mythos, Which Model Fits Cybersecurity Work - Penligent, accessed on April 18, 2026, https://www.penligent.ai/hackinglabs/fr/gpt-5-4-cyber-vs-claude-mythos-which-model-fits-cybersecurity-work/
21. OpenAI expands cybersecurity program, launches GPT-5.4-Cyber - Quartz, accessed on April 18, 2026, https://qz.com/openai-cybersecurity-program-gpt-54-cyber-model-041526
22. Trusted access for the next era of cyber defense - OpenAI, accessed on April 18, 2026, https://openai.com/index/scaling-trusted-access-for-cyber-defense/
23. GPT-5.4-Cyber: OpenAI Introduces AI Model for Cyber Defense to Counter

Anthropic, accessed on April 18, 2026, https://www.trendingtopics.eu/gpt-5-4-cyber-openai-introduces-ai-model-for-cyber-defense-to-counter-anthropic/
24. Equinix and Google Cloud Accelerate Digital Business in Calgary - Interconnections, accessed on April 18, 2026, https://blog.equinix.com/blog/2022/01/12/equinix-and-google-cloud-accelerate-digital-business-in-calgary/
25. Opus 4.7 vs Gemini 3.1 Pro vs GPT 5.4 : r/ArtificialInteligence - Reddit, accessed on April 18, 2026, https://www.reddit.com/r/ArtificialInteligence/comments/1snxjy2/opus_47_vs_gemini_31_pro_vs_gpt_54/
26. Been a couple months since I visited this sub - 4o users, what LLMs are y'all using nowadays? : r/ChatGPTcomplaints - Reddit, accessed on April 18, 2026, https://www.reddit.com/r/ChatGPTcomplaints/comments/1sohqwa/been_a_couple_months_since_i_visited_this_sub_4o/
27. Opus 4.7 vs Gemini 3.1 Pro vs GPT 5.4 vs Grok 4.2 : r/vibecoding - Reddit, accessed on April 18, 2026, https://www.reddit.com/r/vibecoding/comments/1sowqu8/opus_47_vs_gemini_31_pro_vs_gpt_54_vs_grok_42/
28. ChatGPT Plus vs Claude Pro vs Gemini Advanced : which one is best for content & thinking?, accessed on April 18, 2026, https://www.reddit.com/r/ArtificialInteligence/comments/1r02bxn/chatgpt_plus_vs_claude_pro_vs_gemini_advanced/
29. I'm paying for chatGPT, gemini and perplexity and i feel like an idiot. - Reddit, accessed on April 18, 2026, https://www.reddit.com/r/ArtificialInteligence/comments/1qwyv9j/im_paying_for_chatgpt_gemini_and_perplexity_and_i/
30. Google Launches Native Gemini App for macOS With Keyboard Shortcut Access and Screen Sharing, accessed on April 18, 2026, https://www.ghacks.net/2026/04/16/google-launches-native-gemini-app-for-macos-with-keyboard-shortcut-access-and-screen-sharing/
31. Google Introduces Gemini App to macOS, to Rival ChatGPT and Claude | Technology News, accessed on April 18, 2026, https://www.gadgets360.com/ai/news/google-gemini-app-macos-native-experience-ai-features-rival-chatgpt-claude-perplexity-11364702
32. Mac, meet Gemini: Google's AI gives Apple's macOS a 'native' experience for sharing and more | Android Central, accessed on April 18, 2026, https://www.androidcentral.com/apps-software/ai/mac-meet-gemini-googles-ai-arrives-apple-computers-native
33. Anthropic Debuts Claude Opus 4.7 as Mythos Remains Under Lock and Key - Techstrong.ai, accessed on April 18, 2026, https://techstrong.ai/articles/anthropic-debuts-claude-opus-4-7-as-mythos-remains-under-lock-and-key/
34. Anthropic's Claude Opus 4.7 Is Here, and It's Already Outperforming Gemini 3.1 Pro and GPT-5 - Inc. Magazine, accessed on April 18, 2026,

https://www.inc.com/amaya-nichole/anthropics-claude-opus-4-7-is-here-and-its-already-outperforming-gemini-3-1-pro-and-gpt-5/91332180
35. Gemini Apps' release updates & improvements, accessed on April 18, 2026, https://gemini.google/release-notes/
36. Claude Opus 4.7 vs Opus 4.6: What Actually Changed? - Labellerr, accessed on April 18, 2026, https://www.labellerr.com/blog/claude-opus-4-7-vs-opus-4-6-comparison/amp/
37. Anthropic releases Claude Opus 4.7, narrowly retaking lead for most powerful generally available LLM, accessed on April 18, 2026, https://venturebeat.com/technology/anthropic-releases-claude-opus-4-7-narrowly-retaking-lead-for-most-powerful-generally-available-llm
38. Claude Opus 4.7: Benchmarks, Pricing, Context & What's New - LLM Stats, accessed on April 18, 2026, https://llm-stats.com/blog/research/claude-opus-4-7-launch
39. Best AI for undergrad dissertation? : r/ArtificialInteligence - Reddit, accessed on April 18, 2026, https://www.reddit.com/r/ArtificialInteligence/comments/1rarv8w/best_ai_for_undergrad_dissertation/
40. Claude Opus 4.7 hits 92% honesty rate— are we closer than ever to human-like AI with less hallucination? Here's what Anthropic's new AI model is capable of - The Economic Times, accessed on April 18, 2026, https://m.economictimes.com/news/international/us/claude-opus-4-7-hits-92-honesty-rate-are-we-closer-than-ever-to-human-like-ai-with-less-hallucination-heres-what-anthropics-new-ai-model-is-capable-of/articleshow/130341842.cms
41. OpenAI's plan to make ChatGPT the 'everything app' has never been more clear | Mashable, accessed on April 18, 2026, https://mashable.com/article/openai-plan-to-make-chatgpt-everything-app