AI INFRASTRUCTURE · OCTOBER 14, 2026 · ARTICLE

Claude Opus 5.5 vs Gemini 4 Argon: benchmarks, price and availability as of 14 October 2026

Opus 5.5 leads Artificial Analysis's index (58 vs 53) and terminal-style coding. Gemini 4 Argon leads Vals and business-task benchmarks and costs about a third as much per task until 30 November. After that its list price matches Opus.

6 min read · 6 sources

2026 | API pricing | Claude Opus 5.5 | Gemini 4 Argon | benchmarks | comparison

CONTENTS

Verdict 1
Scorecard 2
Intelligence, coding and knowledge work 3
- Independent aggregates favour Opus 3
- Coding: Opus ahead in the terminal, Argon on one long-task suite 3
- Knowledge work: Argon's strongest ground 4

Price, context and availability 4
- Price 4
- Context, output and speed 5
- Availability and maturity 5
- Safety and cyber 5

Sources 6

This report compares Anthropic's Claude Opus 5.5 with Google's Gemini 4 Argon as of 14 October 2026. Google's model is named "Gemini 4 Argon", and I found no "Pro" variant. Argon went broadly available on 13 October, so both can now be used by paying customers [1].

58 vs 53	66.4% vs 57.4%	68.90% vs 66.97%	$1.99 vs $5.98	1 Dec
Artificial Analysis Intelligence Index, Opus vs Argon	Terminal-Bench 4.0 (Google's table)	Vals Index, Argon vs Opus	Cost per index task, Argon vs Opus	Argon price doubles to $4 / $20

Verdict

Claude Opus 5.5 is the safer pick for most people. It scores higher on Artificial Analysis's broad index (**58 vs 53**), leads on terminal-style agent coding, and has been on the market longer, on more cloud platforms. Its $4 / $20 list price is where Argon's price lands from 1 December [1][2][3]. Argon wins on cost today and on professional-task benchmarks (finance, legal, business automation), though many of those numbers are Google's own.

- **If you need agentic coding in a terminal, or long-running coding agents: pick Opus 5.5.** Terminal-Bench 4.0 is 66.4% vs 57.4% in Google's table, and 60% vs 57% in Artificial Analysis's run [1] [2].
- **If you need the best score on an independent aggregate: pick Opus 5.5** (58 vs 53) [2].
- **If you need finance, legal or business-process automation: pick Argon.** It tops the Vals Index and AutomationBench [1].
- **If you need the lowest bill before 1 December: pick Argon.** It costs half as much per token and a third as much per benchmark task [1] [2].
- **If you must stay on AWS, Azure or Google Cloud: pick Opus 5.5.** Anthropic documents Bedrock, Vertex, Microsoft Foundry and Claude Platform on AWS [3].
- **If you want the model inside the Gemini app, Search AI Mode or Android tooling: pick Argon** [1].

Scorecard

CRITERION	CLAUDE OPUS 5.5	GEMINI 4 ARGON
Artificial Analysis Intelligence Index	**58** [2]	53 [2]
BenchLM aggregate (Argon "estimated")	**86.03** [4]	80.88 [4]
Terminal-Bench 4.0 (Google's table)	**66.4%** [1]	57.4% [1]
DeepSWE v1.1 (Google's table)	74.2% [1]	**77.9%** [1]
Vals Index (as of 12 Oct)	66.97% [1]	**68.90%** [1]
AutomationBench (Google's table)	42.5% [1]	**51.3%** [1]
Harvey Legal Agent Benchmark	3.8% [1]	**19.6%** [1]
List price per 1M tokens, in / out	$4 / $20 [3]	**$2 / $10** to 30 Nov, then $4 / $20 [1]
Cost per Intelligence Index task	$5.98 [2]	**$1.99** [2]
Output tokens per task	119k [2]	**62k** [2]
Context window	1M [3]	1M [2]
Max output	128K, 300K batch beta [3]	Unified 1M in+out via continuation protocol [1]
Measured output speed	**95 tokens/s** [2]	not measured [2]
Cloud platforms	**Claude API, Bedrock, Vertex, Foundry, AWS** [3]	Gemini API, AI Studio, Antigravity, Gemini Enterprise [1]
Release	22 Sep 2026 [3]	30 Sep (limited), 13 Oct (broad) [1]

	Claude Opus 5.5	VS	Gemini 4 Argon
AA Intelligence Index	58	·	53
Terminal-Bench 4.0 (AA run)	60%	·	57%
Vals Index	66.97%	·	68.90%
Cost per AA task	$5.98	·	$1.99
Output tokens per task	119k	·	62k
Price in/out per 1M (to 30 Nov)	$4 / $20	·	$2 / $10
Price in/out per 1M (from 1 Dec)	$4 / $20	·	$4 / $20

Sources: Artificial Analysis comparison page, Claude docs, VentureBeat, 13 Oct 2026.

Intelligence, coding and knowledge work

Independent aggregates favour Opus

On Artificial Analysis's index (v4.3.2, ten evaluations), Opus 5.5 scores **58** and Argon **53**. Opus is ahead on nine of the ten components, and Argon leads only on AutomationBench-AA (78% vs 70%) [2].

COMPONENT	OPUS 5.5	ARGON
GDPval-AA v2.1	1866	1624
Humanity's Last Exam	61%	57%
SciCode	67%	62%
AA-LCR (long context)	85%	80%
AA-Omniscience	46	42

VentureBeat's table puts Argon level with GPT-6 Astra and Claude Fable 5.1 at 53. It sits below Opus 5.5 (58), Sonnet 5.5 (56) and Meta's Muse Wonder Full (56) [1]. BenchLM's aggregate ranks Opus first (86.03) and Argon sixth (80.88). BenchLM labels Argon's evidence "estimated" and says the two models' score ranges overlap, so the rank is not firmly established [4].

Coding: Opus ahead in the terminal, Argon on one long-task suite

Opus 5.5 leads Terminal-Bench 4.0, 66.4% to 57.4% in Google's own table [1]. Artificial Analysis's run narrows the gap to 60% vs 57% [2]. Argon leads on DeepSWE v1.1, 77.9% against 74.2%, but that is a Google-reported figure [1]. The earlier library entry on Argon also records it behind on FrontierSWE v2 (55.0% vs 62.3%) and PostTrainBench (45.3% vs 49.3%) [5]. BenchLM gives Opus the coding estimate, 82.8 to 71.6, with "limited" confidence [4].

Tooling matters too. Sam Witteveen, a tester quoted by VentureBeat, calls Argon “a really, really, really good model” but doubts it is the best, and says Google’s Antigravity is not “anywhere near as good as Claude Code or Codex”[1].

One small hands-on test is the only direct head-to-head I found. A Reddit user gave both models the same Three.js planet prompt. Argon finished in 5.5 minutes for $1.58 but never inspected its own output. Opus took 14 minutes for $2.54 and checked its page three times. The poster judged Argon “not too bad, but also not good enough”[6]. It is one anecdote.

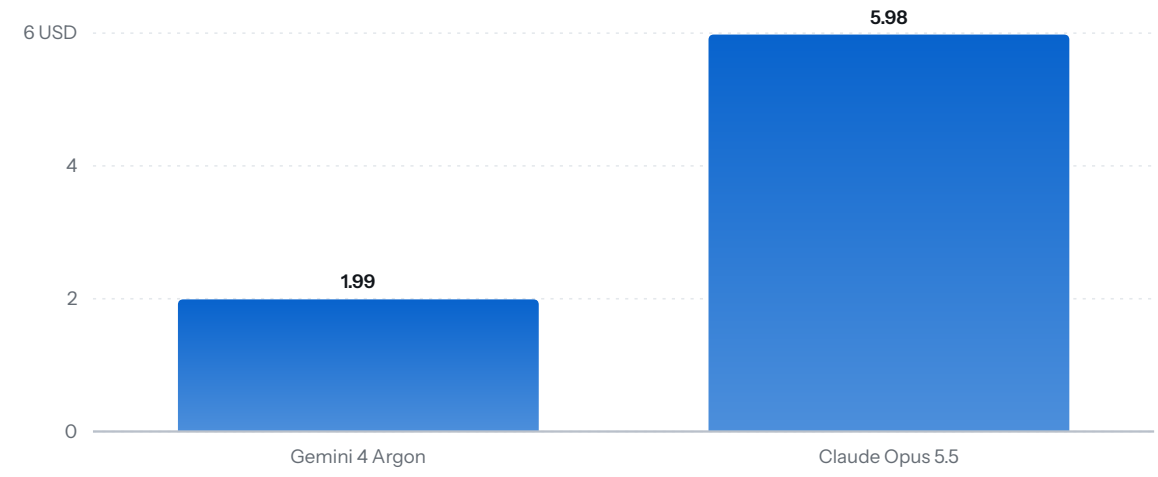
Knowledge work: Argon’s strongest ground

Vals ranks Argon first at 68.90%, with Sonnet 5.5 at 67.04% and Opus 5.5 at 66.97%. Argon also leads Vals’s Finance Agent v2 at 65.40%[1]. In Google’s launch comparison it scores 19.6% on Harvey’s legal benchmark against 3.8% for Opus, and 51.3% on AutomationBench against 42.5%[1]. The Vals gap to Opus is under two points. The Harvey and AutomationBench comparisons are Google-reported, and the earlier entry notes that absolute legal scores are low for every model[5]. Artificial Analysis’s GDPval-AA and AA-Briefcase, also professional-work tests, favour Opus (1866 vs 1624; 1805 vs 1487)[2]. Which model wins depends on the benchmark set.

Price, context and availability

Price

Argon’s introductory list price is **$2 / $10** per million tokens, with cached input at $0.10. Google now says it expires on 30 November 2026 and from 1 December becomes **$4 / $20**[1]. Opus 5.5 is **$4 / $20** with cache reads at $0.20 and a 50% Batch API discount[3]. At its introductory price Argon is half the per-token cost of Opus, and from December the two are equal on list price[1][3].



Source: Artificial Analysis comparison page. Argon figure reflects the introductory $2/$10 price.

Artificial Analysis puts Argon at **$1.99 vs $5.98** per task and $2,407 vs $8,708 to run the whole index[2]. Part of that gap is token use: Argon used 62k output tokens per task against 119k for

Opus[2]. Part is the discounted price, so my own arithmetic suggests Argon's per-task cost roughly doubles at the standard rate, to about $4. It would still be cheaper than Opus on that test. The earlier library entry's estimate for Argon at standard pricing was about $3.98[5]. Artificial Analysis ran Opus at max effort and Argon at high, which tilts the cost comparison toward Argon[2]. BenchLM's list-rate presets agree on direction: a 1K-in, 500-out chat turn is $0.007 on Argon and $0.014 on Opus[4].

Context, output and speed

Both list a 1M-token context window and accept text and images, producing text[2][3]. Opus 5.5 allows 128K output tokens, or 300K on the batch API under a beta header[3]. Argon's single-response limit is unclear. Google's launch described 1M output tokens, but the 13 October announcement describes a combined 1M input-plus-output limit using a continuation protocol, and VentureBeat says Google has not clarified the single-response maximum[1]. Artificial Analysis measured Opus at 95 tokens per second but shows no speed for Argon, so latency cannot be compared[2].

Availability and maturity

Opus 5.5 has been available since 22 September 2026 on five platforms, with retirement no sooner than 22 September 2027[3]. Opus 5 users face breaking changes: thinking cannot be disabled and forced tool use returns an error[3]. Argon was limited to Fairwind cyber partners and government users from 30 September. On 13 October it reached the paid Gemini API, AI Studio, Antigravity and eligible Gemini Enterprise customers. Pro and Ultra subscribers get it in the Gemini app and Search AI Mode, and individual Antigravity access requires Ultra[1]. Argon therefore has under two weeks of public track record.

Safety and cyber

Google reports a 0.7% attack success rate on Gray Swan's indirect prompt-injection test and 2% unauthorised shortcut attempts on its coding evaluations[1]. I found no equivalent Opus 5.5 figure, so that comparison is not possible. The earlier entry records Argon at 68% on CWE-bench against 67% for Opus, effectively a tie[5].

> CAUTION
>
> Many Argon numbers are Google's own runs, and Google's rival figures are largely provider-reported. Independent coverage exists (Artificial Analysis, Vals, BenchLM), but BenchLM marks Argon's evidence as estimated, and public hands-on evidence is only days old. Check the 1 December price change and re-run your own workload before committing.

▾ Terms used in this document (3)

- **Introductory pricing** — A temporary discounted API rate; Argon's ends 30 November 2026.
- **Terminal-Bench** — A benchmark of coding and system tasks an agent must complete through a command-line shell.

- **Vals Index** — Vals AI's GDP-weighted index of professional tasks across finance, legal, coding and tax work.

SOURCES

1 Google makes Gemini 4 Argon broadly available with limited-time low API pricing (VentureBeat)
https://venturebeat.com/technology/google-makes-gemini-4-argon-broadly-available-with-limited-time-low-api-pricing-of-2-10-per-1m-token

2 Gemini 4 Argon (High) vs Claude Opus 5.5 — Artificial Analysis
https://artificialanalysis.ai/models/comparisons/gemini-4-argon-vs-claude-opus-5-5

3 Claude Opus 5.5 overview — Claude Platform Docs
https://platform.claude.com/docs/en/models/opus-5-5/overview

4 Claude Opus 5.5 vs Gemini 4 Argon — BenchLM
https://benchlm.ai/compare/claude-opus-5-5-vs-gemini-4-argon

5 Google Gemini 4 Argon: reported strengths and weaknesses, as of 8 October 2026
https://intel.innerpulse.net/doc/google-gemini-4-argon-reported-strengths-and-weaknesses-as-of-8-october-2026

6 Gemini 4 Argon vs Claude Opus 5.5: same prompt. Tiny World test (r/ClaudeCode)
https://www.reddit.com/r/ClaudeCode/comments/1x56wuh/gemini_4_argon_vs_claude_opus_55_same_prompt_tiny/

Researched and written with claude-sonnet-5-5, from the sources listed above. Figures are as reported by those sources. Exported November 29, 2026 from https://intel.innerpulse.net/doc/claude-opus-5-5-vs-gemini-4-argon-benchmarks-price-and-availability-as-of-14, where the current version lives.