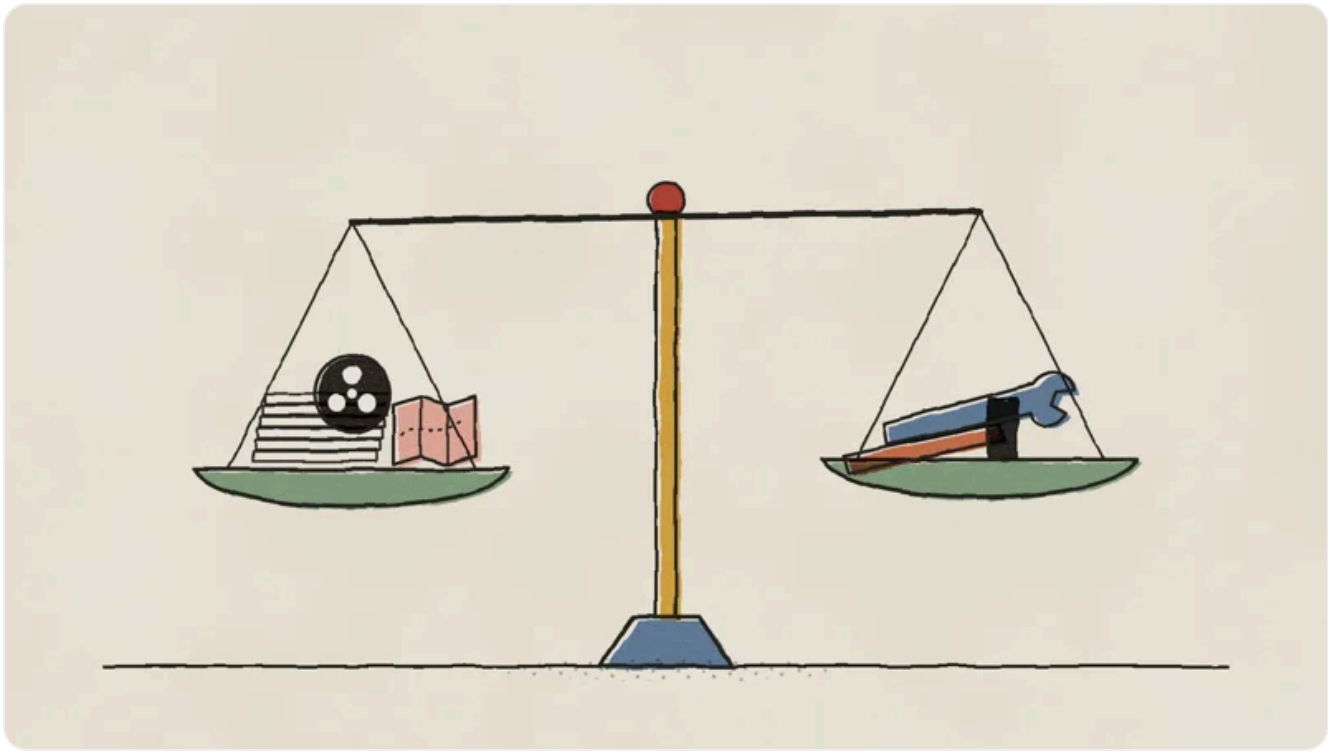
Issue 044 • AI TOOLS • NOVEMBER 10, 2026 • 7 MIN READ

Google's New Gemini 4 Argon: Is It Time to Make a Switch from Claude?



By Tommy Lee



I use Claude almost every day, and several of the tools I have built run on it. So when Google released a new model and said it beat Claude on a lot of tests, I wanted to look at it closely.

Google announced Gemini 4 Argon on September 30, and it opened up to developers and businesses on October 13. It is the first model in the Gemini 4 family. Most of the coverage has been about coding scores and benchmark charts, and it is easy to tune out and assume it was built only for programmers. But some of the biggest changes are things regular people can use every week.

Here is what is new, how it stacks up against Claude Opus, and a few things you can try.

What is new

Google built Argon for hard problems that take a lot of steps, like writing and fixing code, doing research, working through legal and financial questions, and running tasks on its own as an agent. It is a reasoning model, which means it works through a problem before it answers.

Here are the main new features and what they mean for you.

FEATURE	WHAT IT MEANS	HOW IT HELPS YOU
Video and audio	It can watch videos and listen to audio recordings directly.	Summarize a meeting you missed, or turn a long voice memo into a checklist.
Very long documents	It can work with about a million tokens at once, roughly 750,000 words, shared between what you send and what it writes back.	Give it a year of statements or a long contract and ask questions about all of it.
Very long answers	With a special setting, it can write up to about a million tokens in one answer.	Have it rewrite a long manual or build a full guide without stopping halfway.
Thinking levels	You choose low, medium or high, depending on how hard it should think.	Use low for a quick email, and high for planning a trip or comparing quotes.
Web helper	It can use a computer screen, click through websites and gather information.	Compare flights or hotels across several sites while you watch.
Google Search and Maps	It can check facts against live search results and real places on Google Maps.	Plan a day out with real driving times and current store hours.

FEATURE	WHAT IT MEANS	HOW IT HELPS YOU
Fewer made up answers	Artificial Analysis measured a hallucination rate of about 15%, the lowest of the top scoring models they tested.	You can trust it a little more, but you still need to check it.
Reading web pages	Paste a link to an article or web page, and it reads it for you.	Ask questions about a long article, or compare two pages without copying and pasting.
Searching your files	It can search through documents you have uploaded to find the answer.	Find one decision in years of meeting minutes or newsletters.
Charts and graphs	It is strong at reading charts. Google reports 71.6% on a chart test, compared to 66.3% for Claude Opus 5.5.	Upload a chart from a report and ask what it actually shows.
Harder to trick	It is built to resist hidden instructions on web pages or in documents that try to fool it.	Safer to let it browse websites or read files for you.

There are a few things it can’t do yet. It only answers in text, so it can’t make pictures or talk back to you out loud. And what it knows on its own stops in March 2026, unless it looks something up with Google Search.

Google launched it at an introductory price for developers of $2 per million tokens going in and $10 per million coming out. On December 1 that doubles to $4 and $20. At launch, Artificial Analysis scored it even with OpenAI’s top model on their overall score, at about 60% of the cost while the introductory price lasts. Google also trained it to find and fix security holes in software, and it gave a small group of security partners early access before anyone else.

How it can help in everyday life

Here are a few ways regular people like you and me could use it.

- **Catch up on a meeting you missed.** Upload the recording of a school board meeting, a church leadership update or a work call, and ask for a short summary of what was decided and anything you need to follow up on.
- **Turn a voice memo into a plan.** If you think out loud in the car, record it on your phone. Argon can listen to the recording and turn it into a to do list or a first draft of an email.
- **Get through a pile of paperwork.** Give it a year of bank statements and photos of your receipts, and ask it to match them up and show you which charges are missing a receipt.
- **Understand the fine print.** Drop in an insurance policy, a lease or a contract and ask what is actually covered, what the fees are, and what deadlines you need to watch.
- **Plan a trip or a day out.** Because it is connected to Google Search and Google Maps, it can check driving times, store hours and reviews while it plans your day.

How it compares to Claude Opus

Google published a long table of test results comparing Argon with OpenAI's GPT 6 Astra and Anthropic's Claude Opus 5.5 and Claude Fable 5.1. Argon came out on top in most of them.

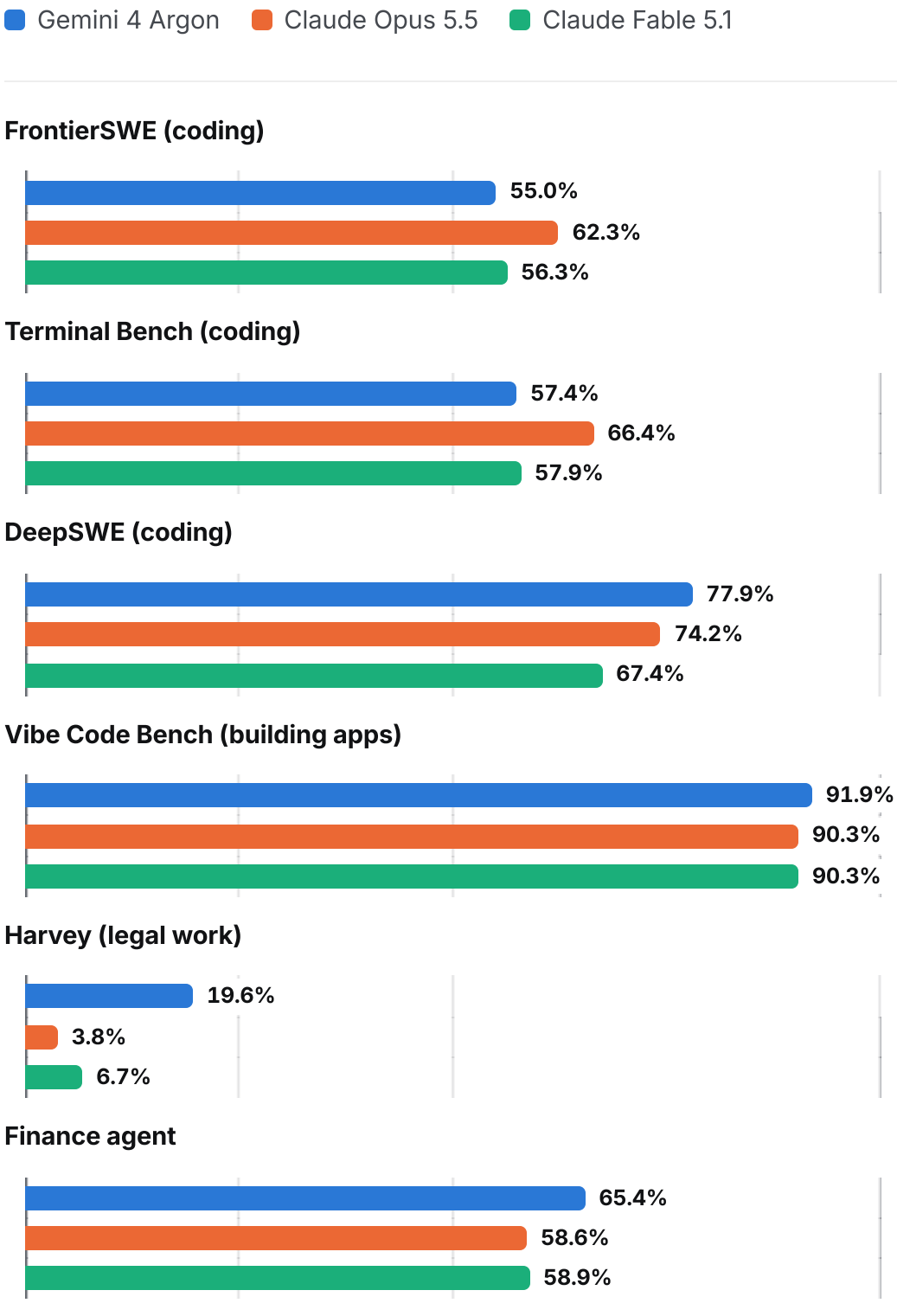
Some of the gaps are big. On a test of legal work from Harvey, Argon scored 19.6% and Opus 5.5 scored 3.8%. On a finance agent test, it was 65.4% to 58.6%. On a long context test, Argon scored 84.2% and Opus scored 66.8%.

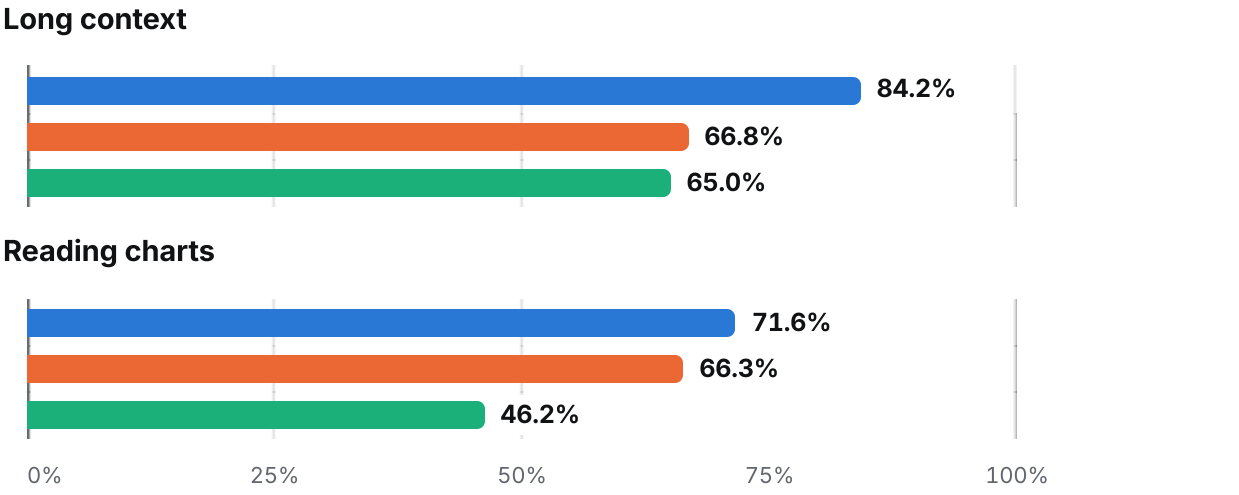
But Claude is still better at coding. That is what I use it for every day, and it is where Claude held its ground in Google's own table. Opus 5.5 scored higher on a coding test called FrontierSWE, 62.3% to 55.0%. It also scored higher on Terminal Bench, which tests how well a model can do real work on a computer through the command line, 66.4% to 57.4%. Argon scored a little higher on two other coding tests, and OpenAI's model actually had the top score on FrontierSWE. So it is close, but for building software Claude is still my first choice.

GOOGLE'S NUMBERS

Google's test results: Argon vs. Claude

Higher is better. These are Google's numbers from its own announcement.





Source: Google DeepMind, Gemini 4 Argon model card.

Here is how I would sum it up for everyday use.

EVERYDAY TASK	GEMINI 4 ARGON	CLAUDE OPUS 5.5	EDGE
Coding	Higher on two coding tests, DeepSWE and Vibe Code Bench.	Higher on FrontierSWE (62.3% vs 55.0%) and Terminal Bench (66.4% vs 57.4%).	Claude
Video and audio files	Watches and listens directly.	Reads text, images and PDFs, but not video or audio files.	Argon
Long documents	84.2% on Google's long context test.	66.8% on the same test.	Argon
Legal and finance work	Higher scores on Google's legal and finance tests.	Lower scores on those tests.	Argon
Local plans and errands	Built in Google Maps and Google Search.	Can search the web, but has no built in Google Maps tool.	Argon

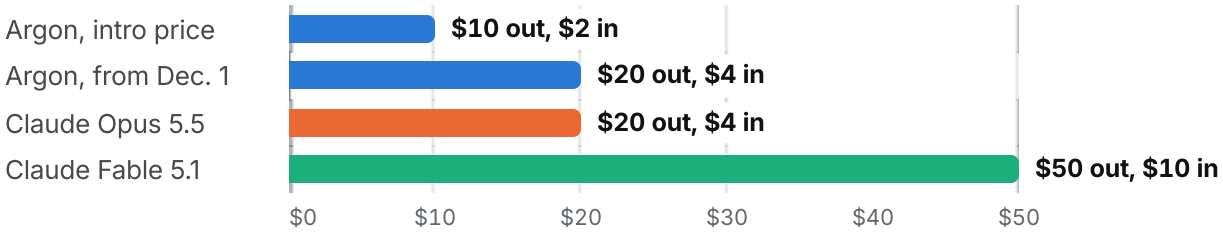
What about Claude Fable 5.1?

Google put Claude Fable 5.1 in its table too, and when Google ran these tests it was Anthropic's top model and the most expensive. Then yesterday, November 9, Anthropic released Claude Fable 5.2. It costs $10 per million tokens in and $50 out. Opus 5.5 is $4 and $20. Argon is $2 and $10 through November 30, then $4 and $20 starting December 1.

PRICE

What it costs developers

Dollars per million tokens the model writes back. Lower is cheaper.



Sources: Google, Anthropic, Anthropic docs.

Even so, Fable beat Argon on only 2 of the 17 tests, and both by about a point. One of those was Terminal Bench, and both Anthropic's own page and Artificial Analysis put Argon ahead there, so I wouldn't count it. Argon won the rest, some by a lot, like a legal test from Harvey (19.6% to 6.7%) and a long context test (84.2% to 65.0%).

Artificial Analysis had Fable on top overall at Argon's launch: Fable 66, Opus 58, Argon 53. On their newer version from October 27, Fable and Argon are about tied at 49 and 50, and Opus is ahead at 59. They also found Fable got more questions right on their knowledge test than any model they had measured, 67% to Argon's 50%. But when Fable doesn't know something, it tends to guess instead of saying "I don't know." Its hallucination rate was around 73%, compared to about 15% for Argon.

So Fable 5.1 was the most powerful Claude and the priciest when Google tested it, but on Google's tests it doesn't beat Argon in the everyday areas this post covers. For everyday use and coding, Opus 5.5 is the better Claude to compare against, and it beat Fable on coding in Google's own table.

Five things you can try

You don’t need special commands. Just talk to it the way you would talk to a helpful friend. Here are five you can copy.

- **A meeting you missed:** “Watch this video for me. Give me the five most important takeaways, list any dates or deadlines, and tell me what I need to follow up on.”
- **Receipts:** “Match these receipts to the charges on my statement. Put them in a table by date, and list any charges that are still missing a receipt.”
- **Fine print:** “Read this whole document. Explain what I am actually agreeing to in simple bullet points, and point out any cancellation penalties or automatic renewal dates.”
- **A voice memo:** “Listen to this recording and sort my thoughts into three lists: things to do today, projects for this week, and questions I need to ask other people.”
- **A weekend plan:** “We have six people visiting for the weekend. Plan two relaxed days with good casual places to eat, keep the drives under 20 minutes, and put it in a simple table.” For something like this, turn the thinking level up to high.

Tommy

Issue 044 · AI Tools · November 10, 2026



ABOUT THE AUTHOR

Tommy Lee is the founder and president of Resource Global, a network that teaches, mentors, and connects emerging Christian marketplace leaders in cities around the world. He writes about faith, leadership, work, and AI at tommyplee.ai.

Read more issues at tommyplee.ai