Compute

Institutional Equity Research
November 24, 2025

Don't Forget to Hit Refresh on the AI State-of-the-Art

Following last week's headline release of Gemini 3, we're sharing our updated view on the frontier AI race, including the latest results from our Vitruvian Score internal benchmark.

The frontier race is far from over. Over the past week, "Gemini 3 is state-of-the-art" became quick consensus in both technical and investor circles. That being said, we'd posit that the investor community over-corrected in Google's favor and prematurely wrote off much of the competition. As we write this, Anthropic has released Opus 4.5 which, on benchmarks alone, looks fully capable of taking the crown from Google after just a week in the top spot. Point being, the race among frontier labs and their models is far from being over, while the state-of-the-art is clearly a title that isn't meant to be held for long.

While Gemini 3 is the strongest model we've tested overall, it wasn't all sunshine and rainbows. Across our internal assessments including the Vitruvian Score, Gemini 3 largely matched our initial expectations which were that it's an extremely capable model and clearly at the frontier. However, while the capabilities of this model are undeniably strong, we did encounter material weakness in instruction following. This view was echoed by many of the researchers and engineers we spoke with over the past week, which was a point of frustration for some. In several of our tests, poor instruction adherence likely suppressed its realized performance, which shows up in our Vitruvian sub-scores where Gemini 3 underperforms OpenAI's GPT-5.1-Codex-Max on the quantitative portion of the benchmark. And only after incorporating our V_{vibes} score did Gemini 3 edge into the top spot, highlighting just how thin its margin is over peers.

As we said before, scaling is alive and well across all three core vectors. Quick napkin math on parameter-count would indicate that Gemini 3 is in the multi-trillion parameter range at minimum, a clear signal that pre-training scaling continues to buy meaningful capability gains beyond where many expected performance to asymptote. On this point, we'd argue that maybe OpenAI should consider dusting off GPT-4.5 or at least go a different route than the one they've ventured on with GPT-5. Scaling test-time compute also clearly still in play as we would highlight through Gemini 3 DeepThink performance with parallel thinking and other comparable "deep research" products. Finally, post-training scaling, especially reinforcement learning with verifiable rewards, is visibly driving capability in models where it has been emphasized, such as GPT-5 and Grok 4. We'd note that Gemini 3's instruction-following issues likely reflect relatively limited post-training investment, which leaves room for future checkpoints to improve without altering the base model.

Figure 1: Vitruvian Score (DVS) Leaderboard as of 11-24-2025

Rank	Model	V_{core}	V_{mar}	V_{knife}	V_{24}	V_{vibes}	V_{total}	Lab
1	Gemini 3 Pro	10.5	9.8	10.0	10.1	0.8	6.89	Google
2	GPT-5.1-Codex-Max	10.2	9.9	10.5	10.2	0.7	6.82	OpenAI
3	Kimi K2 Thinking	9.4	9.1	9.6	9.4	0.9	6.59	Moonshot
4	GPT-5.1	10.1	9.4	9.7	9.7	0.6	6.39	OpenAI
5	Claude Sonnet 4.5	9.0	8.7	8.9	8.9	1.0	6.39	Anthropic
6	Grok 4.1	9.6	9.0	9.3	9.3	0.4	5.79	xAI
7	Claude Opus 4.1	8.8	8.4	8.6	8.6	0.5	5.60	Anthropic
8	GPT-5	9.2	8.9	9.0	9.0	0.3	5.41	OpenAI
9	o3	8.2	7.5	8.1	7.9	0.2	4.56	OpenAI
10	Gemini 2.5 Pro	7.5	6.8	7.0	7.1	0.1	3.73	Google

Source: D.A. Davidson DaVinci

*Our updated Vitruvian Score (DVS) leaderboard above makes two points clear at a glance **(1)** the top cluster of models is tightly packed, and **(2)** leadership is rotating faster than most investors appreciate. On the following pages, we expand on our views of the individual labs and models that shape this leaderboard.*

Please refer to pages 3 - 4 of this report for detailed disclosure and certification information.

INDUSTRY UPDATE

Price (11/24/25)
Industry: TECHNOLOGY

Alexander Platt
(503) 603-3045
AJPlatt@dadco.com

DaVinci Overview

D.A. Davidson's DaVinci initiative focuses our technical-oriented research, data-driven insights, and prescient think pieces under one unified framework. We note that for our DaVinci coverage of deep tech businesses, we employ an early-stage venture approach focusing on technical foundations, disruptive potential, and long-term strategic value, rather than near-term financial and valuation metrics given the unique growth trajectories of pre-inflection markets.

Gemini 3 is a great model, but it was always going to be leapfrogged, and Anthropic's Opus 4.5 may already have done so. As we mentioned on the previous page, prevailing sentiment would lead an outside observer to believe Google had opened a wide gap in the frontier AI race and that the rest of the field would need a major step-change to catch up. That was never our read. It clearly isn't the case and wasn't ever the case to begin with. Even last week, OpenAI released GPT-5.1-Codex-Max which in our own internal benchmark outperformed Gemini 3, with the only cause for it being lower in final rankings was because of the subjective vibes-score we assign each model based on our own personal preference. But without the V_{vibes} score, which you can see in the V_{24} score, GPT-5.1-Codex-Max would have been in the #1 position. Now with Opus 4.5, while we have not had time to play with the model (*testing and reaction to be expected in the coming week*), benchmark scores alone would indicate that Gemini 3 has been clearly passed up, most notably on benchmarks such as ARC-AGI-2, SWE-bench Verified, and Terminal-bench 2.0. And if history is any indicator of our preference for models, we would be unsurprised to see Opus 4.5 land in the very top tier of our vibe assessments as well.

In our view, OpenAI's risk of failing to live up to its trillions of compute commitments has little to do with model quality or research talent. We understand that there is plenty of concern going around the market about OpenAI over governance, compute spend, and commercialization. All are real concerns and we wouldn't argue against them. But on the capability side, OpenAI remains one of the top research labs in the world and their models keep them firmly in the frontier conversation at all times. We'd even point out that OpenAI is the only lab in our Vitruvian Score leaderboard that has 4 models in the top 10, and actually the only provider with 3 models for that matter in the top 10. Has OpenAI potentially made a wrong turn at some point with a failed GPT-5/Orion training run which resulted in GPT-4.5, at least we think so yes. Did that set them back potentially and cede their clear lead over their competition, yes. As a lab is OpenAI now behind and unlikely to catch up because of the previous two statements, no. OpenAI still has some of the best research talent in the world, even after the incredible fragmentation that has gone on over the past several months across the space and primarily to labs like Meta and Thinking Machines. As long as OpenAI keeps shipping models that are objectively competitive at the frontier, we see no reason to be structurally bearish on the lab's ability to contend for the win.

We'll have more to say on Anthropic once we've fully tested Opus 4.5, but one thing is clear, they chose the right hill to dominate. Sonnet-4.5 remains our favorite model to use which is based on a combination between its objective capabilities, our preference for its style (vibes), inference cost, and the ability to use it in our favorite agentic coding tool, Claude Code. Anthropic has leaned into its status as the engineer's and programmer's model provider and has largely owned that identity, which is visible both in qualitative feedback from developers and in the rapid growth of API revenue compared to other frontier labs. Even before we run our own Vitruvian tests, we expect Opus 4.5 to be state-of-the-art on coding-heavy use-cases which is where the Vitruvian Score places the most weight around.

xAI is just there. We'll admit that Grok 4/4.1 is a good model, but we aren't necessarily using the model on a day-to-day basis or even a week-to-week basis for that matter. Elon is deploying substantial capital to keep xAI in the frontier game, and that alone will keep the lab in the conversation. But based on what we see today, we do not expect xAI to ultimately win the frontier race, and we suspect more usage will continue to be concentrated among X power-users and @grok interactions rather than in the broader agentic and coding ecosystems we care most about.

Moonshot AI currently has the top open-source model in our view with Kimi K2 Thinking, though competition in China is intense. Since we published our initial thoughts on Kimi K2, it has remained our favorite open-source model, even as a wave of new releases from Chinese labs has arrived. Some of the best research we read today comes from groups like ByteDance Seed, while labs such as Qwen and Hunyuan continue to ship promising models that we enjoy testing. By contrast, DeepSeek has been notably quiet this year following the widely covered V3/R1 release. They've produced solid research and incremental updates to their models, but no new flagship, which is a negative surprise given we expected at least a V4 release in the early summer. And at this point, we're questioning whether DeepSeek still belongs in the very top tier of Chinese labs.

Meta is not out of the race, but we are increasingly concerned about execution risk. As we're all well aware of, Zuckerberg and Wang have aggressively assembled a high-end research team to compete at the frontier, and in theory, aggregating that much talent should translate into fast progress. In practice, the situation appears far more nuanced. Many of the conversations we've had with those close to the situation at Meta indicate to us that there has been considerable issues with TBD Labs particularly around personnel and organizational friction. We'd note that there has already been some attrition from the team since its inception which is concerning to say the least when the price tag for these researchers is so high. On the model side, rumors around Llama-5 are worrying. We hear echoes of the Llama-4 and even Phi playbook, with heavy emphasis on benchmark performance (benchmaxxing/Goodharting) which led to models that looked great on paper but felt far worse in real life. Not to mention that similar to Llama-4, it sounds like Meta is trying to run the DeepSeek architecture playbook on Llama-5 with lots of synthetic data which was one of the primary issues with GPT-5 Orion training run did not work. Point being here is that much of Meta's outlook in AI is riding on the performance of Llama-5, and while they certainly should be in a hurry to put out a leading model, we'd argue that if it's going to be a repeat of the Llama-4 scenario, take the extra months to get the model right.

Potential Risks

Required Disclosures

D.A. Davidson & Co, or any of its affiliates, does or seeks to do business with companies covered in its research reports. As a result, investors should be aware that the firm may have a conflict of interest that could affect the objectivity of this report. Investors should consider this report as only a single factor in making their investment decision.

D.A. Davidson & Co. is a full service investment firm that provides both brokerage and investment banking services. Alexander Platt, the research analyst principally responsible for the preparation of this report has received and is eligible to receive compensation, including bonus compensation, based on D.A. Davidson's overall operating revenues, including revenues generated by its investment banking and institutional equities activities. D.A. Davidson & Co.'s analysts, however, are not directly compensated for involvement in specific investment banking transactions.
I, Alexander Platt, attest that (i) all the views expressed in this research report accurately reflect my personal views about the common stock of the subject company, and (ii) no part of my compensation was, is, or will be, directly or indirectly, related to the specific recommendations or views expressed in this report.

Rating Information
D.A. Davidson & Co.'s Institutional Research Rating Scale Definitions (maintained since October 10, 2017); information regarding our previous definitions is available upon request:
BUY: Expected to produce a total return of over 15% on a risk adjusted basis over the next 12-18 months
NEUTRAL: Expected to produce a total return of -15% to +15% on a risk adjusted basis over the next 12-18 months
UNDERPERFORM: Expected to lose value of over 15% on a risk adjusted basis over the next 12-18 months

Rating Distribution (as of 9/30/25)	Coverage Universe Distribution			Investment Banking Distribution		
	IR	WMR	Combined	IR	WMR	Combined
BUY (Buy)	59%	85%	62%	8%	0%	7%
NEUTRAL (Hold)	40%	13%	37%	4%	0%	3%
UNDERPERFORM (Sell)	1%	2%	1%	0%	0%	0%

IR denotes Institutional Research; WMR denotes Wealth Management Research whose rating scale is Buy/Add, Neutral, Sell/Reduce. Investment Banking Distribution denotes companies from whom D.A. Davidson & Co. has received compensation in the last 12 months. Best-of-Breed: Expected to outperform on a risk adjusted basis over a five-year time horizon.

Target prices are our Institutional Research Department's evaluation of price potential over the next 12 months, based upon our assessment of future earnings and cash flow, comparable company valuations, growth prospects and other financial criteria. Certain risks may impede achievement of these price targets including, but not limited to, broader market and macroeconomic fluctuations and unforeseen changes in the subject company's fundamentals or business trends.
While the Best-of-Breed designation does not contain a separate rating and/or price target from that of the standard ratings system referenced above, the expectation is that the security, based on the 12 criteria utilized in assessing the "Best-of-Breed" designation, will outperform over a five-year time horizon, not the standard 12-18 month time horizon.

For a copy of the most recent reports containing all required disclosure information for covered companies referenced in this report, please contact your D.A. Davidson & Co. representative or call 1-800-755-7848.

Other Disclosures

Information contained herein has been obtained by sources we consider reliable, but is not guaranteed and we are not soliciting any action based upon it. Any opinions expressed are based on our interpretation of data available to us at the time of the original publication of the report. These opinions are subject to change at any time without notice. Investors must bear in mind that inherent in investments are the risks of fluctuating prices and the uncertainties of dividends, rates of return and yield. Investors should also remember that past performance is not necessarily an indicator of future performance and D.A. Davidson & Co. makes no guarantee, express or implied, as to future performance. Investors should note this report was prepared by D.A. Davidson & Co.'s Institutional Research Department for distribution to D.A. Davidson & Co.'s institutional investor clients and assumes a certain level of investment sophistication on the part of the recipient. Readers, who are not institutional investors or other market professionals, should seek the advice of their individual investment advisor for an explanation of this report's contents, and should always seek such advisor's advice before making any investment decisions. Consensus estimates are obtained from Capital IQ. Further information and elaboration will be furnished upon request.

Other Companies Mentioned in this Report

Company Name	Ticker	Rating	Price
Apple Inc.	AAPL	NEUTRAL	$271.49
Adobe Inc.	ADBE	BUY	$324.19
Amazon.com, Inc.	AMZN	BUY	$220.69
Blaize, Inc.	BZAI	BUY	$2.29
Salesforce.com, Inc.	CRM	NEUTRAL	$227.11

Company Name	Ticker	Rating	Price
CoreWeave, Inc.	CRWV	UNDERPERFORM	$71.65
Datadog, Inc.	DDOG	BUY	$157.55
Dynatrace, Inc.	DT	BUY	$43.57
JFrog Ltd.	FROG	BUY	$59.22
Alphabet Inc.	GOOGL	NEUTRAL	$299.66
MongoDB, Inc.	MDB	BUY	$321.18
Meta Platforms, Inc.	META	BUY	$594.25
Microsoft Corporation	MSFT	BUY	$472.12
Nebius Group N.V.	NBIS	BUY	$83.26
NVIDIA Corporation	NVDA	BUY	$178.88
Oracle Corporation	ORCL	NEUTRAL	$198.76
Snowflake Inc.	SNOW	BUY	$234.03
SoundHound AI, Inc.	SOUN	BUY	$11.22