Sherwood

(CSA Images)

How well can top AI models do these jobs?

An OpenAI benchmark tests how well AI models can perform “economically valuable” jobs.

Jon Keegan 9/29/25 10:59AM

One of the biggest fears fueling the public’s apprehension toward AI is that the technology will eventually take their jobs.

We’ve already seen evidence that some roles like entry-level software development, customer service, and marketing are feeling the effects of automation powered by generative AI. Being able to track the real-world work capabilities of AI models will become increasingly important as models get more and more powerful.

To that end, OpenAI has created a new AI benchmark called “GDPval” that aims to measure just how well leading AI models can do realistic tasks for a variety of “economically valuable” jobs.

OpenAI describes the benchmark as an evolutionary step away from the first wave of benchmarks that followed a more academic, exam-style model:

> “[GDPval] measures model performance on tasks drawn directly from the real-world knowledge work of experienced professionals across a wide range of occupations and sectors, providing a clearer picture on how models perform on economically valuable tasks. Evaluating models on realistic occupational tasks helps us understand not just how well they perform in the lab, but how they might support people in the work they do every day.”

Working with experienced industry professionals, the researchers created a dataset of 220 realistic tasks from 44 occupations that someone might do in the course of their work in a particular role.

Here’s an example of one of the tasks in the benchmark’s training data for a real estate broker:

Sample task for a real estate broker from the GDPval benchmark's training dataset (Huggingfacce.co)

We went through the data and picked a few common jobs from the benchmark's results. Unsurprisingly, software developers were the most impacted job, with Anthropic's Claude model getting an average 70% win rate on the test, which was then compared to a human in that role. For example, a score of 50% would put the model on par with a human expert. Audio and video technicians should feel that their job is secure (for now), as the models executed those tasks with very low scores.

OpenAI acknowledges there are limitations with this benchmark. For instance, currently, each task comes with some background materials that are required to do the task — but generating those background materials itself requires complex work and the benchmark doesn't assess current models' ability to complete those necessary preparatory tasks. Instead that work is done by the humans testing the AI. The paper also notes that this is a small dataset, and the current jobs tested are mainly those of "knowledge workers" that can be performed on a computer.

Maybe a future version will be used to test how well a robot can scrub your toilet.

How well can top AI models do these jobs?

A new AI benchmark from OpenAI called "GDPval" tests how well AI models can perform real-world tasks for "economically valuable" jobs. AI model win rate percentage for selected roles from the data.

gemini grok gpt-5-high claude

MORE TECH

See all Tech

Source: Op

Note: Ba ... arts do

not incl ... co ... der ... nte ... s

Rani Molla

4H

Tesla posts best revenue ever in last quarter with EV tax credit, but the stock slides on weaker margins

Tesla’s stock has stayed in the red in early trading on Thursday.

Elon Musk's Colossus 2 data center is MACROHARD

Jon Keegan 10/22/25

Meta reorganizes its AI teams yet again, this time slashing 600 positions — As it scrambles to catch up to rivals, Meta META $734.43 (0.14%) is yet again restructuring its AI teams, and will be laying off 600 researchers, according to Axios. This is the fifth reorg in the past eight months, based on news reports.

After stumbles from the release of Meta's flagship model, Llama 4, CEO Mark Zuckerberg made a risky bet to shake up the company's AI efforts.

Zuckerberg set out to build a new "superintelligence" team, made up of AI all-stars from around the industry lured with nine-figure pay packages and promises of near limitless computing resources.

The flood of new talent poached from competitors like OpenAI, Apple AAPL $259.98 (0.59%), Google GOOGL $253.25 (0.62%), DeepMind, and others created some awkward tension within Meta, as it already had a prestigious AI team in place known as FAIR, led by neural networks pioneer Yann LeCun. The new recruits were assigned to a team named "TBD" and won't be affected by the cuts, per the report.

Since hiring Alexandr Wang from Scale AI to run the new high-profile team, several rounds of restructuring have roiled the existing Meta AI talent, many of whom might be learning they have lost their jobs.

Exclusive: Meta overhauls its legacy AI operations

Zuckerberg set out to build a new "superintelligence" team, made up of AI all-stars from around the industry lured with nine-figure pay packages and promises of near limitless computing resources.

The flood of new talent poached from competitors like OpenAI, Apple AAPL $259.98 (0.59%), Google GOOGL $253.25 (0.62%), DeepMind, and others created some awkward tension within Meta, as it already had a prestigious AI team in place known as FAIR, led by neural networks pioneer Yann LeCun. The new recruits were assigned to a team named "TBD" and won't be affected by the cuts, per the report.

Since hiring Alexandr Wang from Scale AI to run the new high-profile team, several rounds of restructuring have roiled the existing Meta AI talent, many of whom might be learning they have lost their jobs.

Rani Molla 10/22/25

Applied Digital jumps after announcing $5 billion AI factory lease — Applied Digital APLD $33.19 (8.38%) was up more than 4% premarket after it announced a $5 billion, 15-year AI factory lease from a "US based investment grade hyperscaler" at its Polaris Forge 2 campus, which is expected to begin coming online next year.

On its earnings call earlier this month, the data center company's management teased the deal, saying it was "in advanced discussions with an investment-grade hyperscaler" to lease capacity at Polaris Forge 2 and "also entered negotiations with two additional hyperscalers for two new locations."

"What sets us apart isn't just the size of our pipeline — it's how fast we can deliver," Applied Digital Chairman and CEO Wes Cummins said in the press release. "The real constraint in this industry is execution, and our team continues to prove that large-scale, next-generation data centers can be designed, financed, and brought online faster and more efficiently than anyone thought possible."

LATEST STORIES

Luke Kawa

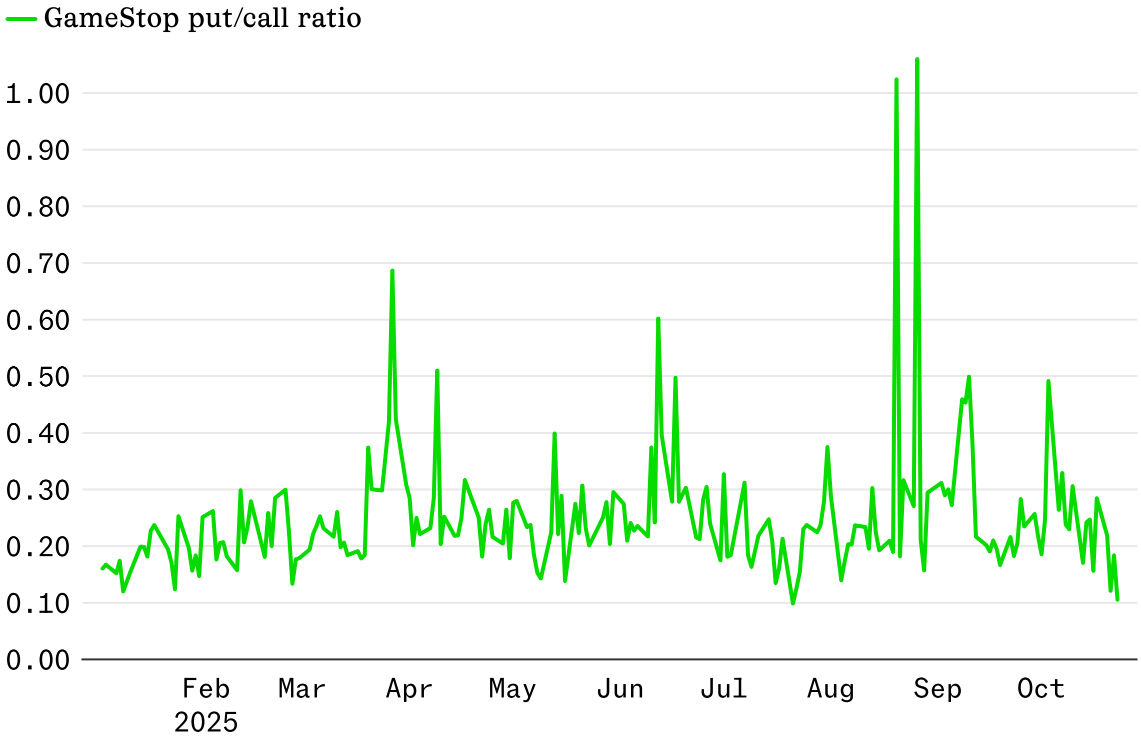
36M

GameStop surges amid bullish options flows — Shares of GameStop GME $23.66 (5.20%) are jumping on no news amid elevated options demand that's got a decidedly bullish tilt.

(Ah, typing that makes me feel younger!)

As of 3 p.m. ET, more than 233,000 call options have changed hands, already 100,000 above their full-day average over the past 20 sessions. And that's largely one-way traffic: the stock's put/call ratio is sitting at 0.1, which would be its lowest for a single session since July 21.

One-way options traffic in GameStop



Note: as of 3 p.m. ET on 10/23

Source: Bloomberg

Sherwood

Call options that expire this Friday with strike prices of $23.50 and $24 are among the contracts seeing the most activity.

Rani Molla

2H

Meta says it's replacing jobs with tech in new round of layoffs — Meta META $734.43 (0.14%) told employees in its risk division, which is responsible for ensuring regulatory and policy compliance, that some of their roles will be replaced by tech, Business Insider reports.

"By moving from bespoke, manual reviews to a more consistent and automated process, we've been able to deliver more accurate and reliable compliance outcomes across Meta," Chief Compliance and Privacy Officer Michel Protti told the workers in an internal memo. "As a result, we don't need as many roles in some areas as we once did."

The news came right after Meta laid off 600 employees across its AI team in yet another company reorganization, amid efforts to improve its flagship AI model, Llama 4.

Meta is only the latest tech company selling AI to say that AI is helping it save money on human labor.

Meta told some employees their jobs are being replaced by tech

The news came right after Meta laid off 600 employees across its AI team in yet another company reorganization, amid efforts to improve its flagship AI model, Llama 4.

Meta is only the latest tech company selling AI to say that AI is helping it save money on human labor.

Yaël Bizouati-Kennedy

SOL CYCLES

3H

Solana rises as Solmate announces

“aggressive” acquisition plans and spot ETF filings grow

Solana ETFs take the lion’s share of the ETF filings, with 23 waiting to be approved, per Bloomberg analyst Eric Balchunas.

Sage D. Young 4H

Ethereum treasury firms surpass bitcoin treasury companies by percentage of total supply

Asia’s institutional conviction for the second-largest cryptocurrency deepens as Tokyo-listed Quantum Solutions scoops up about 2,000 ethereum tokens through its subsidiary.

Analysts parse IBM earnings, see weakness, stock slides

IBM is on track for its worst trading day in months.

Matt Phillips 5H

Americans love to hate horror movies, but they still can't look away

Horror has stormed the box office in 2025, as audiences return screaming to scary screenings.

Millie Giles 5H

Max Knoblauch 5H

Southwest sinks on bearish options activity following its third-quarter earnings beat — Southwest's first full quarter of baggage fees drove it to a revenue record and a profit beat, sending shares higher in after-hours trading on Wednesday. But on Thursday morning, its shares are down more than 5%.

As of 10:50 a.m. ET, more than 31,000 put options in Southwest Airlines LUV $31.90 (-5.52%) have changed hands. That's already about 50% above its 20-day average for a full session. Thursday's trading was particularly skewed toward puts, with a put/call ratio of about 3.3 versus Southwest's 20-day average ratio of less than 1.4.

The bearish options activity coincides with Southwest's earnings call on Thursday, which apparently isn't doing much to inspire optimism.

Nia Warfield

6H

Las Vegas Sands soars as Q3 earnings beat and Macau momentum fuel analyst optimism — Shares of Las Vegas Sands LVS $57.10 (12.80%) leapt over 12% Thursday morning after the casino operator reported a strong third quarter fueled by booming business at its properties in Macau and Singapore.

Adjusted earnings per share came in at $0.78, beating analyst expectations of $0.62. Revenue hit $3.3 billion, also above the Street's forecast of $3.05 billion. The company plans to raise its annual dividend by $0.20 for 2026, bringing the total payout to $1.20 per share.

"We remain enthusiastic about our growth opportunities in both Macao and Singapore as we realize the benefits of our recently completed capital investment programs," Chairman and CEO Robert G. Goldstein said in a statement.

Analysts were optimistic on the results:

- **Stifel** kept its "buy" rating and raised its price target to $68 from $60.
- **Barclays** maintained a "buy" rating and lifted its target to $62 from $59.
- **Goldman Sachs** held a "neutral" rating but boosted its target to $64 from $57.
- **Mizuho** kept its "buy" rating and raised its target to $63 from $56.
- **Macquarie** maintained a "neutral" rating but increased its target to $64 from $62.

Max Knoblauch

6H

Netflix says what the hell, the "Stranger Things" finale can be a movie if we want it to be — At about two hours long, the series finale of "Stranger Things" is already pushing the bounds of how long something can be while still being considered an episode of television.

To make matters muddier, Netflix NFLX $1,116.46 (0.01%) today announced it'll release the episode live in theaters.

More than 350 movie theaters across the US and Canada will hold showings on December 31 through January 1, Netflix announced.

The move follows an interview in Variety earlier this month in which series creators Matt and Ross Duffer expressed their desire for the episode to be shown in theaters, but a Netflix exec at the time shut the idea down.

Theatrical success has likely changed Netflix's mind. Back in August, "Kpop Demon Hunters" became the streamer's first box office No. 1, earning $19 million in a three-day weekend. That film will return to theaters over the Halloween weekend.

W 'Stranger Things' on the Big Screen: Netflix Strikes Deal for Series Finale Theatrical Release

More than 350 movie theaters across the US and Canada will hold showings on December 31 through January 1, Netflix announced.

The move follows an interview in Variety earlier this month in which series creators Matt and Ross Duffer expressed their desire for the episode to be shown in theaters, but a Netflix exec at the time shut the idea down.

Theatrical success has likely changed Netflix's mind. Back in August, "Kpop Demon Hunters" became the streamer's first box office No. 1, earning $19 million in a three-day weekend. That film will return to theaters over the Halloween weekend.

Luke Kawa

6H

Super Micro slumps after announcing preliminary Q1 net sales far below Wall Street's expectations — Super Micro Computer SMCI $47.55 (-9.43%) is slumping after management delivered a preliminary revenue update that came in far short of what the Street was expecting.

Net sales for the quarter ended September 30 (the company's fiscal Q1 2026) will be about $5 billion, according to a press release, which is below its guidance for $6 billion to $7 billion and below the average analyst estimate of just short of $6.5 billion.

Management attributed this to "recent design wins in excess of $12 billion, requesting delivery in the second quarter of fiscal year 2026 (Q2'26)."

Charles Liang, President and CEO reitereated the company's expectation of $33 billion in revenues for the fiscal year that started in July, saying "We see customer demand accelerating, and we are gaining AI share."

If net sales do come in around $5 billion, that would be a roughly 20% decline versus the same period in 2024.

This is not the first time this year that Super Micro has preannounced a revenue miss and effectively blamed it on timing issues.

On April 29, the company preannounced disappointing results and said, "During Q3 some delayed customer platform decisions moved sales into Q4." Pushing back the timing of a big revenue ramp has been a common theme for Super Micro throughout the year.

As we wrote in August:

> "If I could boil down the cause of the substantial volatility in shares of Super Micro Computer this year to one sentence, it would be this: **it's in the AI business — which is clearly booming — and management makes big promises on sales that it fails to deliver on.**
>
> Sales are the football, management is Lucy, and investors are Charlie Brown, falling for each renewed promise and then having it yanked away and landing flat on their backs."

Super Micro scheduled an earnings call for Nov. 4 to discuss the outlook for second-quarter revenues and deliveries.

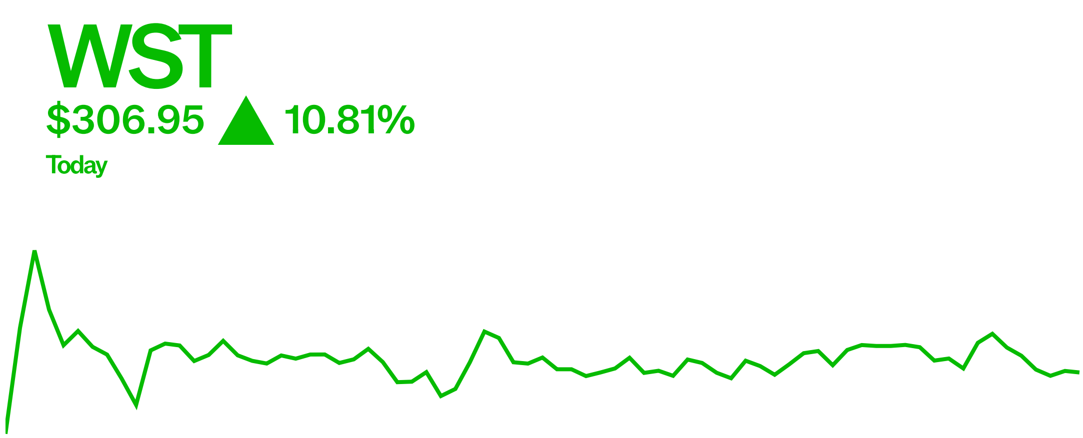
J. Edward Moreno

6H

West Pharmaceutical jumps on continued GLP-1 demand — West Pharmaceutical WST $306.95 (10.81%), which manufactures injectable pharmaceutical packaging, rose more than 14% in early trading after it reported earnings results that beat Wall Street estimates, bolstered by sales of components that go into GLP-1 pens.

The company reported adjusted earnings per share of $1.96 for the quarter, more than the $1.68 analysts polled by FactSet were penciling in. West also reported sales of $805 million, crushing the $785.7 million the Street was expecting.

The company attributed the strong report to growth in its GLP-1 elastomers, which are small rubber components that go into the pens that deliver the blockbuster weight-loss drugs made by Novo Nordisk NVO $53.08 (-0.57%) and Eli Lilly LLY $820.87 (1.04%). Those parts accounted for 9% of sales for the quarter, the company said.



Max Knoblauch

7H

Rivian plans its second layoff in two months, cutting more than 600 workers — After laying off about 1.5% of its workforce a month ago, EV maker Rivian RIVN $13.09 (1.32%) is once again cutting staff.

According to The Wall Street Journal, the automaker will let go of more than 600 employees. Rivian had just under 15,000 employees as of the end of December 2024.

Rivian has been looking to cut costs, with the expiration of the $7,500 EV tax credit and a new, lower-cost SUV due to launch next year. The automaker has said the end of regulatory credits could hold up about $100 million in revenue.

Rivian lost $1.12 billion in its second quarter and downwardly revised its full-year loss forecast to between $2 billion and $2.25 billion.

WSJ **Exclusive | Rivian to Lay Off More Than 600 Workers Amid EV Pullback**

Rivian has been looking to cut costs, with the expiration of the $7,500 EV tax credit and a new, lower-cost SUV due to launch next year. The automaker has said the end of regulatory credits could hold up about $100 million in revenue.

Rivian lost $1.12 billion in its second quarter and downwardly revised its full-year loss forecast to between $2 billion and $2.25 billion.

J. Edward Moreno

7H

Molina sinks on earnings miss and guidance cut, drags down peers Centene and Oscar — Molina Healthcare MOH $159.94 (-18.03%) reported quarterly earnings Thursday morning that severely missed Wall Street estimates, bringing down its stock as well as several of its peers' that provide coverage for government-sponsored plans.

Molina reported adjusted earnings per share of $1.84, less than half of the $3.90 analysts polled by FactSet were expecting. The company slashed its full-year profit guidance to roughly $14 a share, down from its previous guidance of $19 a share and less than the $18.50 a share analysts are penciling in.

One of the factors driving the underperformance was its Affordable Care Act plans, which were used at higher rates than expected, the company said. A higher portion of Molina's revenue comes from other federal programs, such as Medicaid.

Molina dropped about 18% in premarket trading. Insurance companies that focus on providing ACA plans, like Oscar Health OSCR $19.51 (-4.25%) and Centene CNC $34.24 (-4.96%), also fell.

ACA plans, which are government-subsidized health insurance for the poor, have roiled insurers this year amid higher-than-expected costs. ACA extensions are at the center of budget negotiations as the government shutdown stretches on.

READ MORE

Max Knoblauch 8H

American Airlines climbs as its third-quarter loss comes in smaller than expected — American Airlines shares rose more than 3% in premarket trading on Thursday after the last of the big four carriers reported its third-quarter earnings.

American AAL $12.72 (5.21%) reported a loss of $0.17 per diluted share, beating analyst expectations of a $0.28 loss per share and improving on the same period last year. The figure came in on the lower end of American's guidance for a loss between $0.10 and $0.60 per share for the third quarter.

The airline's revenue came in essentially flat from last year at $13.69 billion, beating the estimate from analysts polled by FactSet of $13.6 billion. Passenger revenue was $12.5 billion, above Wall Street's expectations.

For its fourth quarter, American is forecasting earnings of between $0.45 and $0.75 per share, beating Wall Street expectations of $0.30 per share. The carrier now expects full-year 2025 earnings per share of between $0.65 and $0.95, again ahead of analysts' expectations ($0.42).

The improved outlook is still significantly below American's forecast from earlier in the year, when the airline issued annual profit guidance of between $1.70 and $2.70 per share.

As of Wednesday's close, American Airline shares had fallen about 31% on the year. Rival Delta Air Lines DAL $58.74 (-2.10%) is down less than 1% in the same time, while the S&P 500 has gained about 14%.

IBM drops after reporting that revenue in its hybrid cloud division grew 14%, slower than expected — After sliding last night, IBM IBM $284.67 (-0.99%) shares have continued their slump into the morning too, down about 7% in premarket trading at the time of writing. This follows a disappointing Q3 revenues report across two of the most keenly watched areas in its software business.

Despite a small headline beat on the top line, investors have since honed in on the fact that revenues in its hybrid cloud unit — home to Red Hat, the enterprise software provider IBM acquired for $34 billion in 2019 — grew 14%, slower than the previous period and below analysts' estimates of 16%, per Bloomberg. Meanwhile, transaction processing software sales dropped 1% in the third quarter, compounding investors' worries about IBM's growth moving forward, per analysts.

Though overall software revenues climbed 10% to $7.2 billion across the quarter (in line with expectations), traders seem unwilling to look past the Red Hat and transaction processing results.

B IBM Posts Lukewarm Growth for Red Hat Software Business

Despite a small headline beat on the top line, investors have since honed in on the fact that revenues in its hybrid cloud unit — home to Red Hat, the enterprise software provider IBM acquired for $34 billion in 2019 — grew 14%, slower than the previous period and below analysts' estimates of 16%, per Bloomberg. Meanwhile, transaction processing software sales dropped 1% in the third quarter, compounding investors' worries about IBM's growth moving forward, per analysts.

Though overall software revenues climbed 10% to $7.2 billion across the quarter (in line with expectations), traders seem unwilling to look past the Red Hat and transaction processing results.

Hyunsoo Rim & Luke Kawa 10H

Quantum stocks soar after report that the Trump administration is in talks to invest directly in the sector — After speculation has swirled for weeks that the US government might consider investing in the quantum sector, discussions are now underway, with The Wall Street Journal reporting that the Trump administration is negotiating with several quantum computing companies about giving the US Commerce Department equity stakes in exchange for federal funding.

Companies in talks include IonQ IONQ $59.08 (6.54%), Rigetti Computing RGTI $39.08 (8.37%), and D-Wave Quantum QBTS $30.79 (12.81%), with each seeking a minimum of **$10 million** in funding, per the report, while others like Quantum Computing QUBT $15.76 (5.95%) and privately held Atom Computing consider similar arrangements. The deals "haven't been completed and might change."

These stocks soared double digits on the initial news, and IonQ and Rigetti were the second- and third-most-traded stocks in the premarket, trailing only Tesla.

Separate sources also appeared to contradict the report. Per Reuters, a US Commerce Department official said over email that it is "not currently negotiating with any of the companies." Quantum computing stocks all pared some of their advances after Yahoo Finance reported that taking equity stakes is "not necessarily something the Trump administration is considering," citing a person familiar, adding that these companies (and many others) have pitched the government on buying a position.

Per JPMorgan's Arun Jain, retail traders "are actively participating in the sharp rebound of quantum stocks," with net purchases of about $136 million in these four stocks through 11 a.m. ET.

D-Wave Quantum is leading the rally in the cohort, and that makes some fundamental sense: this news would constitute a bigger shift in how the government feels about this particular company relative to its peers.

D-Wave CEO Dr. Alan Baratz had previously expressed feeling left out in the cold by the US government because its most prominent quantum computing technology utilizes annealing models, while its peers use gate-base models. Back in May, he told us he "couldn't even get a foot in the door" with the US government, calling its focus on gate-based models "profoundly disappointing." Now, if these reports are realized, the government won't just have its foot in the door; it'll have a seat at D-Wave's table.

READ MORE

WSJ **Exclusive | Trump Administration in Talks to Take Equity Stakes in Quantum-Computing Firms**

Companies in talks include IonQ IONQ $59.08 (6.54%), Rigetti Computing RGTI $39.08 (8.37%), and D-Wave Quantum QBTS $30.79 (12.81%), with each seeking a minimum of **$10 million** in funding, per the report, while others like Quantum Computing QUBT $15.76 (5.95%) and privately held Atom Computing consider similar arrangements. The deals "haven't been completed and might change."

These stocks soared double digits on the initial news, and IonQ and Rigetti were the second- and third-most-traded stocks in the premarket, trailing only Tesla.

Separate sources also appeared to contradict the report. Per Reuters, a US Commerce Department official said over email that it is "not currently negotiating with any of the companies." Quantum computing stocks all pared some of their advances after Yahoo Finance reported that taking equity stakes is "not necessarily something the Trump administration is considering," citing a person familiar, adding that these companies (and many others) have pitched the government on buying a position.

Per JPMorgan's Arun Jain, retail traders "are actively participating in the sharp rebound of quantum stocks," with net purchases of about $136 million in these four stocks through 11 a.m. ET.

D-Wave Quantum is leading the rally in the cohort, and that makes some fundamental sense: this news would constitute a bigger shift in how the government feels about this particular company relative to its peers.

D-Wave CEO Dr. Alan Baratz had previously expressed feeling left out in the cold by the US government because its most prominent quantum computing technology utilizes annealing models, while its peers use gate-base models. Back in May, he told us he "couldn't even get a foot in the door" with the US government, calling its focus on gate-based models "profoundly disappointing." Now, if these reports are realized, the government won't just have its foot in the door; it'll have a seat at D-Wave's table.

READ MORE

Beyond Meat craters in premarket trading on heavy volumes — Beyond Meat BYND $2.87 (-19.97%)'s rally has seemingly gone from well done to overcooked in a hurry.

Shares of the plant-based meat company are cratering, down nearly 20% in premarket trading on heavy volume.

As of 5:05 a.m. ET, the stock has traded more than $60 million in the premarket. That's the fifth-highest of all stocks listed on US exchanges, trailing only the pure-play quantum computing companies (after a report that the US government may take stakes in these companies) as well as Tesla, which reported earnings after the close on Wednesday.

It's an about-face for Beyond, which has seen the premarket session be a source of particular strength over the past few days. In fact, the stock's recent peak of $8.86, which marked a rally of 1,672% from its all-time low in less than four sessions, occurred in the premarket session on Wednesday. It's down more than 60% since that time.

What's different about the Beyond Meat move compared to the so-called short squeezes we've seen in the past is that the cost of borrowing shares has gone down, not up, during its period of intense gains. However, that's what we would expect given its shares outstanding jumped from ~76 million to close to 400 million in light of its debt-for-equity swap, which temporarily gave those creditors control over the company.

So, while its previous parabolic move may have been extremely painful for existing shorts, it's also making it arguably more attractive for any doubters to put on fresh bets that the surge in the stock is half-baked.

Southwest's first full quarter charging for checked bags drives it to record Q3 revenue

Southwest became the third major airline to report its third-quarter earnings when it dropped its results after the bell Wednesday.

Max Knoblauch 23H

IBM slides despite earnings beat

Here's how Big Blue did.

Matt Phillips 23H

Load more