GRANT'S
INTEREST RATE OBSERVER®
Vol. 44, No. 13b 233 Broadway, New York, New York 10279 • www.grantspub.com JULY 3, 2026

AI outsmarts itself

Evan Lorenz writes:

"The AI revolution has only just begun," Masayoshi Son declared two Wednesdays ago. "For an industry that is still at such an early stage, calling it a bubble is an insult."

The founder of SoftBank Group Corp. is half right. Artificial intelligence is indeed a groundbreaking technology, but for that very reason it coexists with an epic stock-market bubble. The twain coincided in the railway buildout of the late 19th century, in the electrification boom of the 1920s and during the dot-com levitation at the turn of the millennium. It's the way of the world.

In preview, we find that the tech-heavy S&P 500 is capitalized as if the American frontier model-makers were fated to win in the marketplace and to charge monopoly-like rents in the bargain. They will not. How the bubble might pop and which companies may benefit in its aftermath are the topics before the house.

In the first phase of the AI rollout, 2020–25, betting on the likes of OpenAI and Anthropic looked like a sure thing, notes James Grogan, CEO of Universitas AI, a company that designs software for colleges and research groups. The cost to train leading-edge models rose exponentially, with OpenAI's GPT-4.5 model, released in 2025, coming in at around $500 million versus just $43,000 for GPT-2, which debuted in 2019.

Indeed, you did get what you paid for, as the most expensive models excelled. In April 2023, the top-of-the-line product outperformed the 90th percentile second-best by a whopping 26%, according to the Epoch Capabilities Index, which combines multiple AI benchmarks into a single score.

This formed a "self-reinforcing economic loop," Grogan points out—we quote from his March 9 paper, "The End of the Foundation Model Era: Open-Weight Models, Sovereign AI, and Inference as Infrastructure," which was published on arXiv. "In each cycle," he writes,

> the hyperscaler [i.e., large tech companies such as Microsoft Corp.] commits capital to the foundation model company. The foundation model company spends that capital on cloud compute—predominantly from the same hyperscaler's infrastructure.…The hyperscalers earned returns on both sides of each transaction: equity appreciation in the foundation model company and high-margin revenue on the infrastructure the company was required to purchase.…High pre-training costs, far from being an unavoidable technical constraint, were the mechanism that made the structure work.

The dollars at stake are immense. Equity gains from Anthropic and OpenAI generated no fewer than 12 percentage points of the 28% leap in first-quarter S&P 500 earnings, according to calculations by Baolian Wang, an associate professor at the University of Florida. OpenAI and Anthropic account for $1.05 trillion's worth of backlog across Microsoft, Oracle Corp., Alphabet, Inc. and Amazon.com, Inc. AI-related spending is projected to make up 2.1% of U.S. GDP this year, approximately the same as the railroads' contribution between 1850 and 1859 (*Grant's*, June 5).

In phase two of the AI revolution, that formerly self-reinforcing loop is unraveling. Grogan dates this from January 2025, when DeepSeek released its R1 model, which offered comparable performance to frontier AI products but cost only around $6 million to train. Overnight, the high training-cost market moat vanished.

Using the Epoch Index as a gauge, the leading AI product is now just 6% more capable than the one at the 90th percentile, an "almost invisible" difference, Paul Kedrosky, a general partner at SK Ventures and a research fellow at the Massachusetts Institute of Technology, writes in a June 23 blog post. In 1941, Kedrosky reminds us, Ted Williams became the last man to hit over .400 in a Major League season. This is not because today's players are less talented, but because "the gap between average and elite had closed." All players, not forgetting the pitchers, are better at their jobs.

At the start of the AI boom, more data meant better models, but it's not so simple today. Software engineers have already ransacked libraries and the world wide web. Now they're turning to companies like Outlier AI to engage chatbots in high-quality conversations with human interlocutors. Well, at least, that is what they intend. Last week *New Scientist* reported that the lowly paid conversationalists are using AI chatbots to hold up the human side of the conversation, too. "If these companies want quality data, then they should offer quality contracts," one worker griped. Performance degrades when AI outputs become AI inputs.

It appears that tech companies have overestimated what ordinary consum-

ers—not the data-scientist early movers—are willing pay for the chatbot experience. "Software engineers have this unusual property that they'll write short prompts that generate huge amounts of code," Kedrosky tells me. "I call that an expansive use of LLMs. That early user became representative, but no one uses tokens the way software engineers do."

Those heavy users are generating more lines of code, but not necessarily more revenue. According to a May paper by two MIT researchers, Mert Demirer and Liyuan Yang, and Leon Musolff, assistant professor at the Wharton School of the University of Pennsylvania, programmers using autonomous agents generate 180% more coding activity than the luddites who continue to write code by hand. While this has led to a surge in the number of software apps at major marketplaces like Apple's App Store, total app usage has not followed suit.

Leading-edge AI devices don't come cheap. By Kedrosky's calculations, OpenAI's GPT-5.5 and Anthropic's Claude Opus each charge $5 for a million tokens, compared with just $0.019 for DeepSeek. As companies blow through their AI budgets, they are beginning to ration AI use (see the prior issue of *Grant's*) or are adopting the cheap Chinese upstarts. Lindy AI, which makes virtual executive assistants, switched to DeepSeek from Claude, and Coinbase Global, Inc. has a goal to move 80% of its workloads to low-priced alternatives by the end of the year. Even Microsoft, which owns a 27% stake in OpenAI, has produced a suite of low-cost AI models and is considering adding DeepSeek to its product lineup.

"We're hitting this inflection point where AI is becoming material to the cost structure," Justice Kwak, the head of agentic AI at Accenture plc, accidently told the world (the press having picked up his comments at an in-house meeting). "Spend is becoming very unpredictable, and leadership, especially at the CFO, COO and CIO level, are still asking the question of whether they're getting value from what we're spending…in the context of AI. It's really not a niche problem. It is a problem that every enterprise will face if they are bullish on AI, if they haven't already."

Not surprisingly, the cheap models are starting to win favor. "Data from OpenRouter," The Information reported on June 23,

Mr. Market miscalculates



source: The Bloomberg

a service that compares AI models so developers know which ones should power their apps or handle different tasks within them, show that the share of all tokens processed on open-source models—primarily DeepSeek and MiniMax from China—climbed to roughly 65% in the second week of June, up from 34% during the same week in January and 26% the year prior.

The collapse in the performance gap between best and second-best is only one reason why AI will not be a winner-takes-all market. A second reason is that users can switch between models with little or no operational disruption. As if on cue two weeks ago, *The Wall Street Journal* reported that both Anthropic and OpenAI are weighing drastic price cuts. Monopoly-like rents are a dream.

The growing efficiency of AI software constitutes another sand trap, Kedrosky adds. This includes quantization (i.e., cutting off numbers to the right of the decimal point in complex calculations), speculative decoding (when models guess what you'll ask next) and prefetching (gathering data you need before you know you need it). Chinese open-weight products allow users to train AI on their own data sets, which can enable the cheap models to outperform the closed-weight American ones.

"For individuals, probably close to 70–75% of AI will migrate to edge devices over the next 3–4 years," instead of in the massive data centers that hyperscalers are borrowing to build, Kedrosky tells me. "That's really just memory-constrained right now," a problem that new and improved iterations of consumer electronics devices may solve.

Advances in silicon, too, are making AI cheaper to deliver. As it is, training models require expensive graphics cards from the likes of Nvidia Corp., which earns a fat 75% gross margin on its wares. Inference, i.e., using the already trained devices, can run on application-specific integrated circuits (ASICs), which are custom-designed to process data cheaply. Last week, OpenAI announced that, in partnership with Broadcom, Inc., it will launch its own ASIC named Jalapeño that "will deliver performance per watt substantially better than the current state-of-the-art." The AI-maker joins firms like Amazon, Meta Platforms, Inc. and Microsoft, which have produced their own ASICs.

The shift in computational demand to inference from training is a threat to the ultimate AI picks-and-shovels business model. Thus, it's no mystery why Nvidia last December paid $20 billion to license Groq, Inc.'s ASIC designs and to acquire several of Groq's senior executives.

. . .

The dot-com boom of a quarter-century ago may provide a road map for today. In 1995, cutting-edge ethernet speeds reached 100 megabits per second. In 1997, WorldCom, Inc. estimated that internet traffic was doubling

every 100 days. Extrapolating that rapidly growing demand and limited throughput, telecommunication firms went on a fiber-optics building jag.

WorldCom was correct that internet usage was doubling, but it compounded about once every year instead of every three months. Meanwhile, advances in dense wavelength-division multiplexing, which allows fiber-optic cables to carry many signals simultaneously, allowed telecoms to deliver ethernet speeds of 10 gigabits per second, a hundred-fold jump in capacity, by 2002 (*Grant's*, June 6, 2025).

The result: a glut of capacity. "In late 2002, consulting firm TeleGeography estimated that only 10% of the long-haul fibers installed in Europe and North America carried any signals and that only 10% of the wavelengths in those fibers were lit," a 2016 post mortem in *Optics & Photonics News* concluded. "Soon, creditors were trying to unload dark-fiber networks for pennies on the dollar."

If AI follows fiber optic's precedent, which firms would fail and which would succeed? As to the former, "companies that are using a lot of debt to finance data centers predicated on predictable earnings and stable token prices are in trouble," says Kedrosky, naming Oracle as the top example. (For the bear case on Oracle, see the Jan. 16 issue of *Grant's*.)

"Foundation model companies now face the trajectory that early utilities, railroad companies and first-generation telecommunications providers followed," Grogan notes in his aforementioned essay. "They built transformative infrastructure, proved the concept at enormous cost and discovered that the infrastructure converged toward commodity pricing while value accumulated in the hands of those who used it."

In other words, the winners exist outside the AI industry, "in the exact same way that most of the winners in rural electrification were outside of the electrical industry," Kedrosky adds. "It was Westinghouse and companies that were building appliances that could benefit from ubiquitous and cheap electricity to sell you a fridge." Today, the beneficiaries include banks, which spend enormous sums on technology and compliance that can now be automated with AI chatbots, and companies that utilize commoditized AI models to solve their customers' problems. Today's disfavored software companies are near the head of that line.

. . .

Wolters Kluwer N.V. (WKL on Euronext Amsterdam) is another prospective winner, says Drew Estes, portfolio manager of Banyan Capital Management, Inc., which owns the stock. Netherlands-based Wolters is a broadly diversified provider of professional information, software and services. Its customers operate in health care (providing 26% of last year's revenues) and tax and accounting (27%), with the balance in financial and corporate compliance (20%), law and regulation (16%) and corporate performance and environmental matters (10%). North America furnished 63% of last year's top line; subscriptions and other forms of recurring revenue contribute 83% of sales.

"Their product is pretty dominant in health-care systems in the United States," Estes says. "It's called UpToDate.... They've infused their product with AI, and it's really caught on. It's a trusted solution for practitioners. In our estimation, the big health systems are going to be very slow to move from a trusted provider to an upstart, and I think AI benefits them and makes their product better."

The market seems to agree. Over half of the Wolters legacy customers signed on to secure an AI-enhanced version of UpToDate within six months of the product's unveiling in October 2025.

Aren't OpenAI and Anthropic the wave of the future? A quartet of Wolters bulls at Deutsche Bank A.G. took issue with that narrative in a report dated April 10. It's not so easy to displace business-to-business vendors that boast "proprietary data/content/IP and embedded relationships," the analysts observe. Wolters deploys proprietary data and expert networks in each business vertical to design AI programs that don't hallucinate. For any in health care and tax compliance especially, non-hallucinating sources of information are just what the doctor ordered.

Since Feb. 12, 2025, Wolters Klewer's share price has slumped by 68.8% in euro terms to value the company at 10.1 times the estimate for 2026 earnings. In the five years before the selloff, the Dutch software company traded at an average of 30.9 times trailing earnings. Using a 7.6% weighted average cost of capital and running a discounted cash flow analysis, the Deutsche Bank team finds that investors are implicitly expecting a terminal decline in WK's revenues. Last year, for the record, adjusting for acquisitions and currencies, revenues grew by 6%, and management is guiding for a little bit better growth in 2026.

Insiders, at least, are buying, having picked up a net 10,700 shares over the past 12 months for a net outlay of €1.2 million. As of March 31, Wolters reported €3.8 billion in net debt, equivalent to 1.9 trailing Ebitda. Of the 16 analysts on the case, 12 say buy and none says sell. (The issues of *Grant's* dated Feb. 27 and April 24, 2026 mention a few other software companies that may just survive the age of AI.)

"What is underway is a market correction," Grogan's paper concludes, "not of AI's underlying capabilities, which are real and consequential, but of the assumptions about where AI value accrues, which proved to be badly wrong."

•

Grant's® and Grant's Interest Rate Observer® are registered trademarks of Grant's Financial Publishing, Inc.
PLEASE do not post this on any website, forward it to anyone else, or make copies (print or electronic) for anyone else.
Copyright ©2026 Grant's Financial Publishing Inc. All rights reserved.