LEGACY
PRIVATE WEALTH PARTNERS
Powered by CONCURRENT

# 1st Quarter Special Commentary

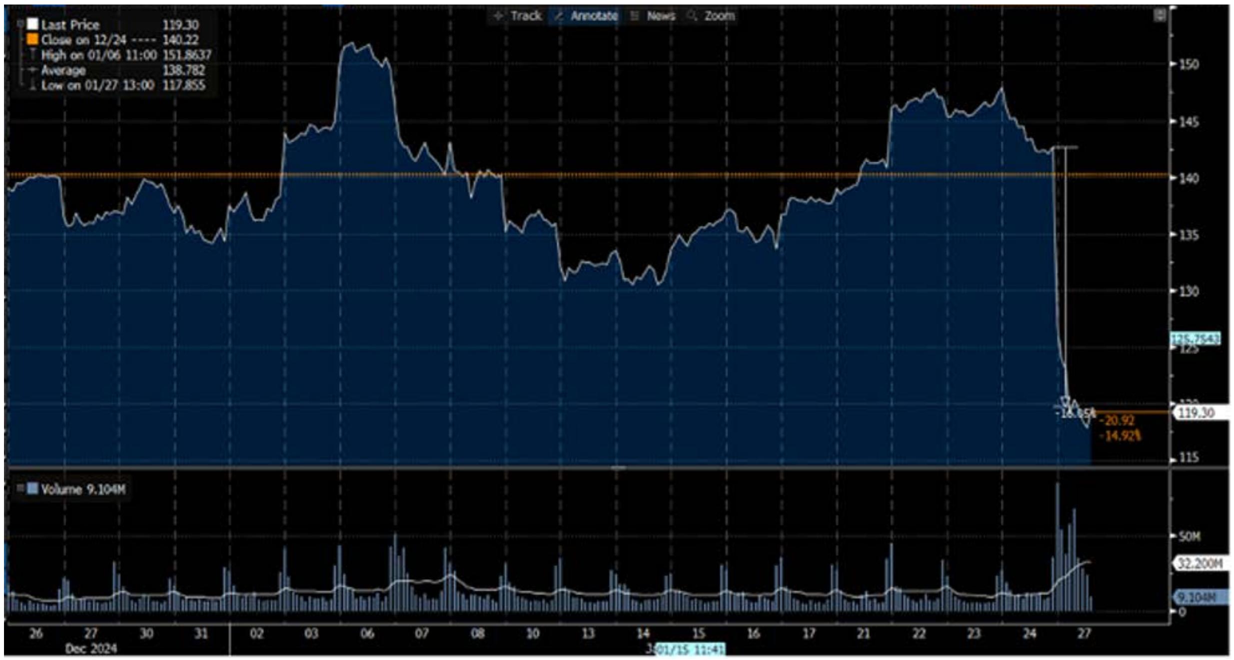
2025

# AI...Into the Unknown

Hi, I'm DeepSeek. How can I help you? The announcement made over the weekend of January 25th, 2025, may change the course of Artificial Intelligence forever.

In one day, China's competitor in the Artificial Intelligence race took a $1 trillion loss to the valuations of U.S. tech companies. Now U.S. valuations are in question.

NVIDIA alone lost nearly $560 billion.




Many in Silicon Valley are in panic mode. Is Big Tech really at a dead end? Is the cost difference to create the latest AI architecture nearly 95% cheaper? Is America's approach and thinking fundamentally flawed toward the creation and development of AI? Is it possible the Chinese are manipulating and distorting reality? Could the mega billions raised by OpenAI, Perplexity, xAI, Anthropic, and other industry players be for nothing? Is the massive infrastructure push by our government to earmark over $2 trillion a waste?

The almost unbelievable story involves U.S. export restrictions that forced Liang Wenfeng the founder of High-Flyer Capital Management (HFCM) to not be able to access bleeding-edge semiconductors, which resulted in the development of a language model that rivals Chat GPT and OpenAI.

## What's Novel About DeepSeek's Approach?

It's technical, but at a high level, here's the approach the team at DeepSeek took:

- DeepSeek used 8-bit floating point numbers as opposed to 32-bit floating point numbers. This isn't as precise, but it is far more efficient with computational resources.
- The model uses a multi-token prediction model rather than a single-token prediction model. We can think of the tokens as the outputs of the model (i.e. the answer to a prompt). By using a multi-token model, it doubles the inference efficiency while being almost as accurate as a single-token model.
- They used a mixture-of-experts (MOE) architecture with some innovation around load balancing. This allows them to have a massive model...but only a small portion of that model is active at any time. The active portion of the model depends on the task it is trying to solve. Doing so reduces computational resources and improves efficiency.
- DeepSeek uses a form of reinforcement learning called Group Relative Policy Optimization (GRPO) to improve reasoning capabilities. This technology compares different responses and selects the best one, which greatly reduces computational overhead compared to the prevailing critic model.
- Compression of the key value indices — which is how individual tokens are represented in the architecture — results in more than 90% compression ratios and dramatically lower memory requirements.

## The key is to understand the result — DeepSeek is about 45 times more efficient than OpenAI o1.

And this means that it is dramatically cheaper to use:

| | OpenAI o1 | DeepSeek-R1 |
|---|---|---|
| Pricing 1M input tokens | $15 | $0.14 |
| Pricing 1M cached tokens | $7.50 | $0.55 |
| Pricing 1M output tokens | $60 | $2.19 |

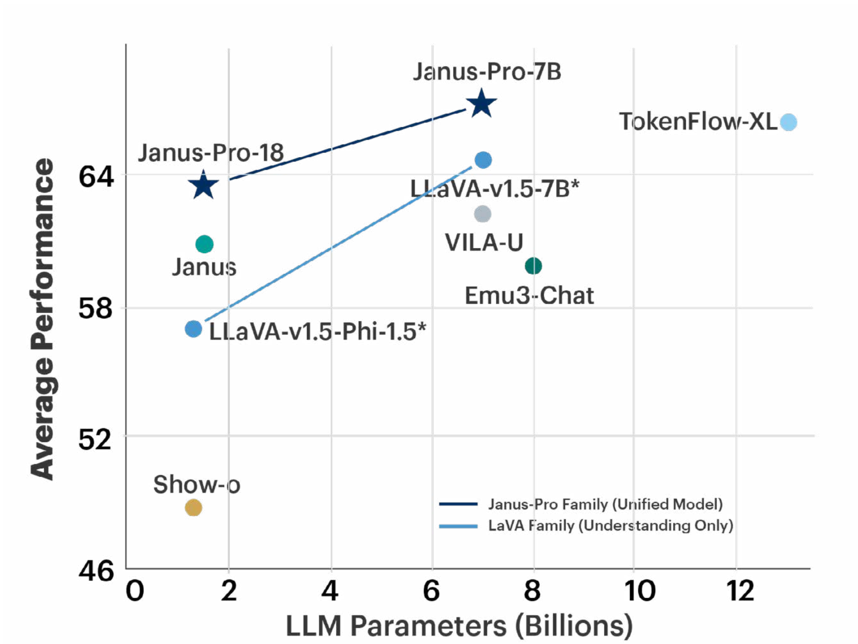
"Tokens are basically small blocks of words, part of a word, or software code, etc. which are used in both inputs and outputs of large language models. And DeepSeek's pricing is more than 95% cheaper than OpenAI's o1.

And that's not all. The software has been open-sourced.

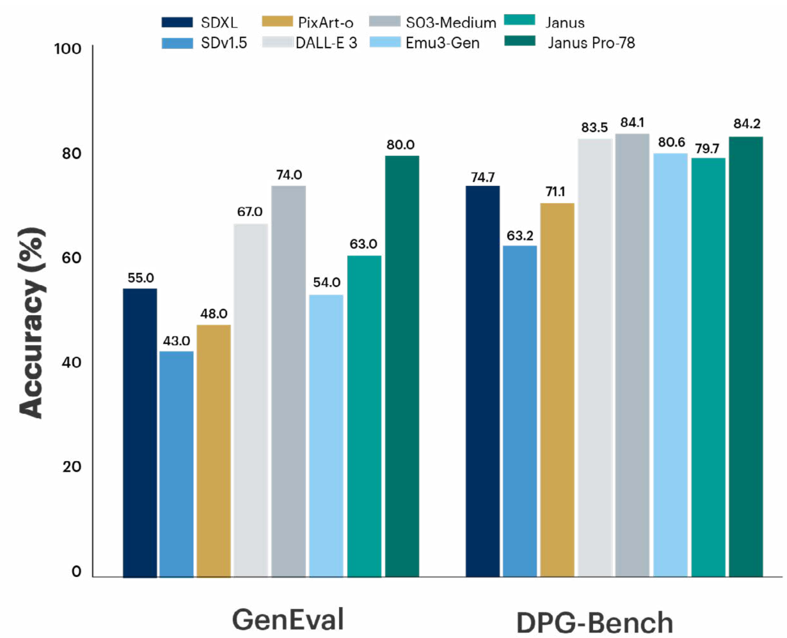
This raises the question again: Why are companies spending hundreds of billions of dollars to build these foundational models… when the software is open-sourced and can be built for a few million?" (Brown Stone Research, 2025)

On January 28th, 2025, HFCM released another generative artificial intelligence model. Janus-Pro-7B is a multimodal, generative AI model capable of text-to-image generation.

And, in line with the performance of DeepSeek-V3 and DeepSeek-R1, the performance appears to be equally impressive.




Average performance on four multimodal understanding benchmarks.




Performance on instruction-following benchmarks for text-to-image generation.

Source: DeepSeek


Figure 2 | **Comparison of text-to-image generation between Janus-Pro and its predecessor, Janus.** Janus-Pro delivers more stable outputs for short prompts, with improved visual quality, richer details, and the added ability to generate simple text. The image resolution is 384 × 384. Best viewed on screen.

Source: DeepSeek

The advancements in AI by DeepSeek appear to be comparable to the best-in-class competitors in the AI space.

This is not the first time in the history of technology that a drastic turn changed the course of history. When the lightbulb was invented by Thomas Edison, the vision of electricity power plants in America seemed set. That abruptly changed because of a state-of-the-art technological enhancement invented by his protégé Nikola Tesla which led to the powering of the dam at Niagara Falls and eventually for most of America and the developed world.

Is the Sputnik race to the moon between the U.S. and Russia currently demonstrated in the proverbial war between the U.S. and China to whose AI development and platform is best for the world?

There are many more answers to come, and we will be paying extremely close attention to how U.S. tech stock valuations will be affected as well as the overall ripple effects to national security, geopolitical, political agendas, markets, industries, companies, and people at large.

We are definitely living in interesting times.

**Todd M. Schwartz,CFP®**
Founding Partner

LEGACY
PRIVATE WEALTH PARTNERS
Powered by CONCURRENT

Advisory Services offered through Concurrent Investment Advisors, LLC (Concurrent) an SEC Registered Investment Advisor.

This presentation is limited to the dissemination of general information regarding Concurrent's investment advisory services. Accordingly, the information in this presentation should not be construed, in any manner whatsoever, as a substitute for personalized individual advice from Concurrent. The information presented is for educational purposes only and does not intend to make an offer or solicitation for the sale or purchase of any specific securities, investments, or investment strategies. Investments involve risk and unless otherwise stated, are not guaranteed. Be sure to first consult with a qualified financial advisor and/or tax professional before implementing any strategy discussed herein.

This report is intended for the exclusive use of clients or prospective clients (the "recipient") of Concurrent Investment Advisors and the information contained herein is confidential and the dissemination or distribution to any other person without the prior approval of Concurrent Investment Advisors is strictly prohibited. Information has been obtained from sources believed to be reliable, though not independently verified. Any forecasts are hypothetical and represent future expectations and not actual return volatilities and correlations will differ from forecasts. This report does not represent a specific investment recommendation. The opinions and analysis expressed herein are based on Concurrent Investment Advisor research and professional experience and are expressed as of the date of this report. Please consult with your advisor, attorney, and accountant, as appropriate, regarding specific advice. Past performance does not indicate future performance and there is risk of loss.

Concurrent may discuss and display charts, graphs, formulas which are not intended to be used by themselves to determine which securities to buy or sell, or when to buy or sell them. Such charts and graphs offer limited information and should not be used on their own to make investment decisions.