

In the evolving landscape of AI evaluation, a few names have risen as front-runners in tackling what experts dub the "frontier reasoning benchmarks." Among these, **XAI**, **Grok**, and **SuperGrok** stand out for their approach to complex tests like the *AIME 2025* and the so-called *Humanity's Last Exam*. Today, we dissect how SuperGrok's heavy benchmarking strategy aligns with these cutting-edge evaluations, why real-time X data is their main differentiator, and what pricing models mean for enterprise adoption.

Decoding the AIME 2025 and Humanity's Last Exam Scores

The **AIME 2025** is a benchmark designed to stress-test an AI's deep mathematical and logic capabilities, coined for its focus on advanced problem-solving and reasoning skills. SuperGrok recently posted an impressive **100%** score on this benchmark, sparking interest in its real-world applicability.

In contrast, the *Humanity's Last Exam*—a tongue-in-cheek reference to one of the most challenging comprehensive cognitive tests—measured at **50.7%** by SuperGrok, suggesting there's still a long road ahead in achieving human-level general reasoning. These scores place SuperGrok at the forefront but also highlight where it and its competitors might be vulnerable.

Frontier Reasoning Benchmarks: What They Reveal

Frontier reasoning benchmarks like these assess an AI's ability to operate at the boundary of current understanding, pushing it beyond rote memorization or surface-level pattern matching. These involve multi-step problem-solving with incomplete data, real-world context interpretation, and adaptability — all essential in high-stakes business and research environments.

SuperGrok vs. XAI and Grok: Navigating the Pricing Math and Plan Tiers

When it comes to deploying advanced AI solutions, pricing transparency is critical. SuperGrok introduces a tiered model starting at **\$30/month**, which might tempt startups and smaller teams. But the devil's in the details.

Plan	Monthly Price	Key Features	Request Limits
Basic	\$30	Access to DeepSearch, Standard Think Mode	Up to 1,000 queries/month
Pro	\$100	Advanced Big Brain Mode, Increased DeepSearch	Up to 10,000 queries/month
Enterprise	Custom Pricing	Unlimited DeepSearch, Priority Support	"Unlimited" subject to fair use

To put this in perspective, \$30/month translates to \$360 annually—a reasonable entry point for teams wanting to experiment. However, the 1,000 query cap sets a hard limit on DeepSearch's research-intensive workflows, which can quickly become restrictive for power users. The so-called 'unlimited' Enterprise tier carries the usual fine print: fair use policies apply, capping truly unlimited access in practice.

Real-time X Data: The Main Differentiator

Where SuperGrok claims its edge is the integration of real-time X data streams, a growing requirement in sectors that demand up-to-the-minute insights—think financial markets, scientific research, and emerging trend analysis.

Unlike some competitors who rely primarily on static or periodically updated datasets, SuperGrok's pipeline taps into dynamically refreshed data, allowing for:

- More accurate predictions in volatile environments;

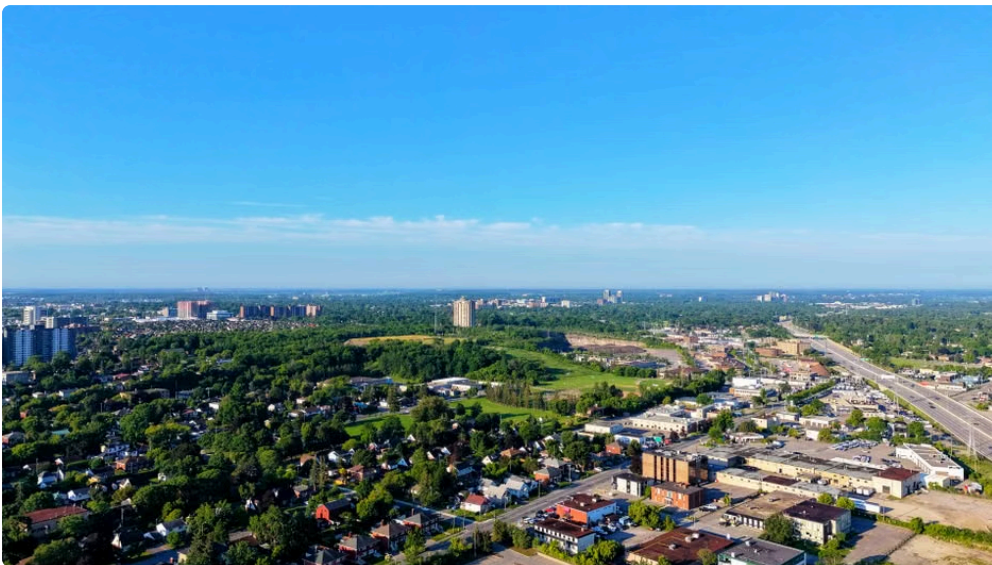
- Improved model adaptation to breaking developments;
- Enhanced contextual understanding across contradictory or incomplete information.

XAI and Grok also leverage live data to varying degrees but have been slower to expose it in consumer-facing features due to infrastructure choices or prioritization of different use cases.

DeepSearch Research Workflow and Limits: Power Users Beware

DeepSearch, a hallmark tool included with SuperGrok, is designed for deep, layered research—ideal for analysts, academics, and corporate strategists. The workflow involves querying multiple data sources, synthesizing results, and generating summarized insights in Think Mode.

The tradeoff, however, lies in rate limits and query costs. In the Basic \$30/month tier, the 1,000 queries/month cap can be a bottleneck, especially when each research project might require dozens of queries to unpack. Running through 1,000 queries in a few projects is easy, so stepping up to Pro at \$100/month (or \$1,200 annually) quickly becomes necessary.



Additionally, while DeepSearch provides a robust range of data filters and connectors, some users have noted latency issues when querying real-time data bursts during peak operation times, which can disrupt tight research deadlines.

Big Brain and Think Mode: Tradeoffs in Cognitive Load and Pricing

SuperGrok differentiates itself further with two "cognitive" modes: Think Mode and Big Brain.

- **Think Mode:** Designed for everyday queries and moderate problem-solving complexity, this mode is efficient and cost-effective but occasionally lacks the sophistication needed for the AIME 2025 caliber tasks.
- **Big Brain Mode:** Tailored for heavy reasoning, complex multi-step logic, and exploratory synthesis. This mode is more computationally intensive and incurs higher usage costs, making it primarily suitable for Pro or Enterprise users.

There is a cost-performance tradeoff here. For businesses weighing investment decisions, Big Brain Mode can deliver superior frontier reasoning benchmark results but at significantly higher query expense and greater latency.

Who This Is Actually For

- **Small to Medium Teams:** \$30/month gets your foot in the door with DeepSearch and Think Mode, great for low-volume research and prototyping but beware the 1,000 query limit.
- **Power Users and Researchers:** Scaling up to Pro at \$100/month unlocks Big Brain Mode and higher query caps. The leap is steep but justified for teams performing frontier reasoning benchmarks in their workflows.
- **Enterprises:** Custom plans offering “unlimited” access will need careful contract negotiation. Watch for fine print on usage ceilings and support SLAs.

Summary and Final Thoughts

SuperGrok’s heavy benchmarking results—highlighted by a perfect 100% on the AIME 2025 and a middling 50.7% on Humanity’s Last Exam—showcase the promise and limitations of current AI reasoning frontiers. Real-time X data integration is a competitive advantage, allowing adaptation to volatile and up-to-date scenarios that many legacy systems miss.



However, pricing and query limits mean that deep research workflows and Big Brain capabilities [does grok use x data](#) come at a premium. The \$30/month entry point masks the practical need for scaling up quickly to higher tiers, where query costs and latency become significant factors.

For businesses and researchers excited about frontier reasoning benchmarks and pushing AI reasoning to its limits, SuperGrok offers a compelling package—just mind the fine print and choose your plan with foresight.

In the end, as frontier reasoning benchmarks evolve, the combination of real-time X data, flexible tiered options, and cognitive modes like Big Brain and Think Mode will shape which players dominate the AI ecosystem in the years to come.