



# GROK3

A private AI platform choice for businesses

Tony Shen

02/23/2025

# Contents

Contents ..... 1

Grok3 Performance in Benchmark Tests ..... 2

Grok3 and DeepSeek Comparison at Prompt ..... 3

Apache 2.0 License and MIT License..... 5

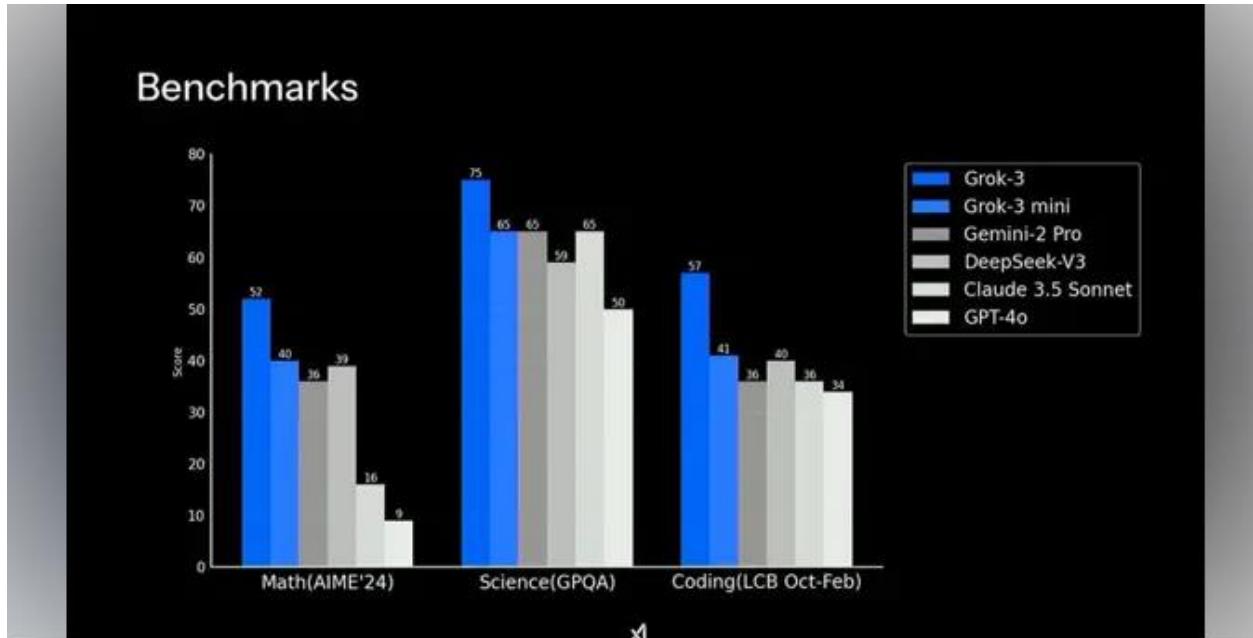
Grok3 as a powerful private AI platform choice ..... 5

Grok3 Minimum Hardware Requirements..... 6

Summary ..... 7

## Grok3 Performance in Benchmark Tests

On February 17, 2025, Elon Musk's xAI released Grok3, boasting of its superior performance over the competition, by publishing its benchmark tests in comparison with other popular models, namely, Google's Gemini-2 Pro, DeepSeek-V3, Anthropic Claude 3.5 Sonnet, and OpenAI GPT-4o [1].



— Benchmark tests comparing Grok 3 and Grok 3 mini to the competition. xAI

### Grok Key Highlights [2]

Highlight	Description
Open Source:	Fully open under Apache 2.0 license
Performance:	73% on MMLU and 63.2% on Human Eval benchmarks
Context Memory:	Supports conversations up to 8,000 words.
Customization:	Fine-tune with proprietary data for tailored models
Pricing:	Premium+ access costs \$16/month or \$168/year

### Grok Comparison with ChatGPT, Claude, and Anthropic [2]

Feature	Grok	ChatGPT	Claude	Anthropic
Open Source	Yes	No	No	No
Context Memory	8,000 words	4,096 tokens	4,096 tokens	4,096 tokens
Parameters	314B	175B	10B	20B
Real-time Data	Yes	Limited	Limited	Limited
Pricing	\$16/month	Free-\$20	Varied	Varied

## Grok3 and DeepSeek Comparison at Prompt

Not all Grok3's competitors are open-sourced, namely ChatGPT, Gemini, and Claude are all close sourced. DeepSeek, on the other hand, is fully open sourced, like Grok. Let's take Grok3 and DeepSeek to a quick prompt test by asking them each a same question "[How Russia gained the upper hand in Ukraine?](#)", and compare their responses.

**Grok3** responded:

Russia has gained the upper hand in the ongoing conflict with Ukraine through a combination of military, strategic, and external factors, particularly evident in developments up to February 22, 2025. While the war, which escalated with Russia's full-scale invasion on February 24, 2022, has seen periods of stalemate and Ukrainian resilience, Russia's recent battlefield momentum stems from several key advantages.

First, Russia has leveraged its superior manpower and industrial capacity. With a military of approximately 1.3 million personnel compared to Ukraine's 900,000, Russia has sustained high casualty rates—estimated at 30-50,000 troops per month—while continuing to replenish its forces through mobilization and external support. Ukraine, facing troop shortages and low morale, has struggled to match this, despite efforts to recruit an additional 160,000 soldiers. Russia's ability to produce or procure vast quantities of ammunition and equipment, firing an estimated 10,000 rounds daily compared to Ukraine's rationed 2,000, has allowed it to maintain relentless pressure, particularly in eastern Ukraine, such as Donetsk.

Second, territorial gains have bolstered Russia's position. In 2024 alone, Russia captured over 4,000 square kilometers, including 2,656 square kilometers by December, with November marking Ukraine's worst month for territorial losses since September 2022 (1,202 square kilometers, roughly the size of New York City). Key victories, like the capture of Avdiivka in February 2024 and advances toward Pokrovsk, have been facilitated by fortified defenses, extensive minefields, and Ukraine's inability to break through due to insufficient manpower and ammunition. Russia now controls about 20% of Ukraine's territory, including Crimea and parts of four eastern regions it illegally annexed in 2022—Donetsk, Luhansk, Kherson, and Zaporizhzhia.

Third, Russia has capitalized on Ukraine's dependence on Western aid, which has wavered. The U.S., providing over \$118 billion since 2022, has not secured new congressional funding under the Trump administration, threatening to halt the flow of critical weapons like HIMARS and Patriot systems. Europe's defense industry lacks the capacity to fill this gap fully, leaving Ukraine vulnerable. Delays and limits on advanced weaponry earlier in the conflict allowed Russia to build fortifications and store ammunition beyond Ukraine's reach, tilting the ground war in Moscow's favor.

External alliances have also played a role. North Korea has supplied 10,000 troops and ballistic missiles, while Iran provides drones and production materials, and China offers geospatial intelligence, machine tools, and economic support via discounted oil purchases. These partnerships have offset Western sanctions, helping Russia sustain its war effort despite isolation from Europe and the U.S.

However, Russia's gains are not absolute. Its naval dominance in the Black Sea has been challenged by Ukrainian drones and missiles, and its territorial advances have come at a steep cost, with losses deemed unsustainable long-term by some analysts. Ukraine's resilience, including its Kursk incursion in

2024, and Western backing have prevented a total collapse. Still, as of February 2025, Russia's strategy of attrition, mass troop deployments, and exploitation of Ukraine's logistical and political vulnerabilities have given it a clear edge in the ground war, shifting the conflict's trajectory toward a prolonged struggle favoring Moscow's endurance.

**DeepSeek** responded:

Russia's perceived upper hand in the Ukraine conflict stems from a combination of military, geopolitical, economic, and strategic factors, though the situation remains dynamic and contested. Key elements include:

1. **Military Adaptation and Attrition:**
  - Superior Resources: Russia's larger military, artillery, and ammunition reserves enable prolonged attrition warfare, grinding down Ukrainian forces.
  - Territorial Gains: Control of strategic areas like Mariupol (securing a land bridge to Crimea) and incremental advances in Donbas bolster Russia's position.
  - Mobilization: The 2022 partial mobilization and expanded defense production mitigated manpower and equipment shortages.
2. **Geopolitical Alliances and Economic Resilience:**
  - International Support: Assistance from Iran (drones), North Korea (artillery), and China (economic partnerships) offsets Western sanctions.
  - Energy Leverage: Revenue from oil/gas exports to India, China, and others funds the war effort, while Europe's prior energy dependence constrained sanctions.
  - Sanction Evasion: Russia circumvented restrictions via third countries and shadow fleets, maintaining economic stability.
3. **Strategic Warfare Tactics:**
  - Cyber and Information Operations: Disinformation campaigns sow division in Ukraine and abroad, undermining Western solidarity.
  - Nuclear Posturing: Threats of tactical nuclear use deter direct NATO intervention and limit escalation of Western aid.
4. **Exploitation of Internal Divisions:**
  - Ethnic Narratives: Russia leverages historical ties to Russian-speaking populations in eastern Ukraine to justify territorial claims.
  - Political Fatigue: Prolonged conflict risks draining Ukrainian morale and Western support, particularly amid debates over aid in the U.S. and EU.
5. **Territorial Consolidation:**
  - Annexation of Donetsk, Luhansk, Kherson, and Zaporizhzhia (2022) creates political facts on the ground, complicating future negotiations.
6. **Defensive Fortifications:**
  - Extensive trenches, minefields, and air defenses in occupied territories slow Ukrainian counteroffensives (e.g., the 2023 push toward Tokmak).

**Challenges and Uncertainties:**

- Ukraine's Western-backed military modernization (e.g., F-16s, ATACMS) and asymmetric tactics (e.g., drone strikes on Crimea) could shift momentum.
- Global geopolitical shifts, such as U.S. election outcomes or EU energy diversification, may alter the conflict's trajectory.

While Russia holds advantages in resources and strategic patience, Ukraine's resilience and international backing ensure no decisive victory for either side is imminent. The conflict remains a protracted stalemate with episodic gains.

#### Grok3 and DeepSeek Comparison at Prompt

Response	Grok3	DeepSeek
Response is structured	Yes	Yes
Total words in response	504	346
Supporting quantitative data or facts	11 data points	0

Furthermore, in the simple prompt test described above, Grok3 response was instant; DeepSeek was slower than Grok3

## Apache 2.0 License and MIT License

Though both Grok3 and DeepSeek are fully open sourced, they use different open-source licenses. Grok opens its code under Apache 2 License whereas DeepSeek opens its code under MIT License.

#### Apache 2.0 License and MIT License Comparison

Highlight	Description	Apache 2.0	MIT
Permissiveness	Allow users to freely use, modify, distribute, and sublicense the software	Yes	Yes
Attribution	Requires that any modifications to the original code are documented, and notices must be included in the distributed software.	Yes	Yes
License Notice	A copy of the license must be included with any distribution of the software.	Yes	Not explicitly required
Contribution	Contributions made to the software are automatically licensed back to the original licensor under the same license	Yes	Not explicitly stated
Documentation of Changes	Changes made to the original code must be documented	Yes	Not required
Patent Grant	It includes an explicit patent grant	Yes	No
Complexity	License verbiage is complexed	Yes	No

The key difference between the Apache 2.0 and MIT licenses can be summarized as follows:

**Apache 2.0 License:** Requires more thorough documentation of changes and has stricter compliance requirements. This includes an explicit patent grant and the obligation to provide a copy of the license in any distribution of the software.

**MIT License:** More relaxed and easier to comply with. It is simpler and shorter, only requiring the inclusion of the original copyright and license notice in any distribution.

## Grok3 as a powerful private AI platform choice

You can download the Grok3 model from GitHub, install it, and run it locally on-premises, creating your own powerful private AI platform with Grok3. This setup ensures that your data, both input and output, remains fully protected within your premises, without any risk of being lost, stolen, captured, or used without your knowledge and authorization. Since your private AI system operates entirely within your premises, it can function without an internet connection, unlike AI models hosted on public clouds or cloud-based SaaS subscriptions, where achieving such total protection is challenging, if not impossible.

With a fully protected private AI platform, your organization will reap significant benefits. You can enhance day-to-day internal email communications among staff and management across all departments with Grok3. Let Grok3 serve as a personal assistant to each employee, improving their writing before emails are sent, thereby minimizing miscommunication. The same applies to your internal documentation systems, which require easy-to-read, easy-to-understand, and accurate written texts.

Your CRM system will also benefit from Grok3, providing customers with more comprehensive answers to their concerns and issues. Your internal IT Help Desk will see improvements as Grok3 empowers agents with the knowledge they need to resolve tickets, eliminating the silos of tribal knowledge that often hinder teamwork.

Business intelligence will be significantly enhanced by feeding the model with all your business data, records, documents, and other materials, including trade secrets, PII, and other sensitive information. This can be done without security and compliance concerns, as all data flows within your premises. The more input you provide to Grok3, the more AI benefits it can deliver. Applying this local AI to customer relationship management, marketing, sales, pricing, cost controls, budgeting, planning, and R&D will exponentially increase the benefits.

Since Grok3 is an open-source model, you can modify and further develop it to integrate it with your ERP, HR, accounting, finance, inventory control, manufacturing systems, and more to maximize AI benefits. When the next release from xAI becomes available, you can easily update your private AI platform with the new model with minimal downtime.

## Grok3 minimum hardware requirements

Due to its advanced transformer architecture and computational demands, Grok3 based private AI platform requires robust hardware to run. Here are the minimum hardware requirements:

Component	Description	Apache 2.0
Computing Power	CPUs	Intel Xeon Scalable processors or AMD equivalent
	GPUs	8-16 high-end GPUs such as NVIDIA A100 or H100
Memory	System RAM	512GB – 1TB DDR5
	GPU VRAM	80GB per GPU
Storage	Primary storage	4TB+ PCIe Gen 4 NVMe SSD for fast model access
	Secondary storage	10TB+ high-speed HDD or SSD storage for datasets and logs
Networking	Cluster(s) on premises	100GbE+ Ethernet or InfiniBand for cluster GPU-equipped work nodes networking
	Remote scale-out networking to public clouds	High speed dedicated private links between on-prem and clouds such as ExpressRoute for Azure or Direct Connect for AWS

## Summary

This short article briefly discusses the benefits of using open-source LLM models like Grok3 to create a private AI platform, effectively addressing data protection concerns associated with AI models hosted on public clouds or AI SaaS subscriptions on the Internet. Running Grok3 locally ensures that all user data and data flows involving the AI model stay within the premises, providing total control over data protection and maximizing the use of AI and its benefits.

Building a private AI infrastructure to run Grok3 requires significant investments in computing, storage, and networking hardware due to its advanced architecture and large size. Additional considerations include datacenter floor space, racks, wiring, and cooling, power supply, etc., which this short article omitted on purpose to keep the discussion simple. However, with careful planning, preparation, and execution, businesses can successfully build their own private AI platforms to run leading open-source LLM models like Grok3 and reap substantial benefits.

Other open-source LLM models are available for businesses to build similar private AI platforms that require fewer hardware resources. Examples include DeepSeek and Meta Llama3.3 [3].

## References

1. [Musk's xAI Launches Grok 3: Here's What You Need to Know](#)
2. [Grok is now officially open-source](#)
3. [2025 DeepSeek Test Run. Computer Configuration | by Tony Shen | Feb, 2025 | Medium](#)