

Why India Doesn't Have Its Own LLM Yet?

This report explains why India hasn't made a big language model (LLM) by February 2025. It looks at problems like not enough money, missing data, weak tech setups, and a lack of experts. It also talks about what's happening now, like the IndiaAI Mission and projects from companies like Krutrim and Project Indus. The report shows how India plans to build its own LLM by the end of 2025 and what it could mean for the country.

Key Points

- India has not developed a foundational Large Language Model (LLM) by February 2025 due to high costs, data scarcity, and infrastructure challenges.
- The government and private sector are working on it, with a target completion by the end of 2025.

Challenges and Current Status

India faces several hurdles in developing a foundational LLM, which is a model trained from scratch to understand and generate text across multiple languages and contexts. These include:

- **High Costs and Funding Issues:** Developing LLMs requires significant investment, and Indian AI start-ups have seen reduced funding, with only \$166 million raised in 2024 compared to \$518.2 million in 2022 ([Why India Needs Its Own LLM: Insights into the Challenges and Opportunities](#)).
- **Data Scarcity:** Quality data in Indian languages is limited, with less than 1% of global digital content in these languages, complicated by linguistic diversity and oral traditions ([India's Large Language Language: Challenges and Opportunities](#)).
- **Infrastructure Constraints:** Until recently, India lacked sufficient GPUs, though the IndiaAI Mission plans to supply 18,693 GPUs ([India to build home-grown foundational model by 2025](#)).
- **Talent Shortage:** There is a need for more AI experts, with many skilled professionals working abroad or for international firms.

Future Prospects

Surprisingly, despite these challenges, India is making rapid progress. The IndiaAI Mission, with a \$1.2 billion budget, aims to have a domestic LLM ready by December 2025, focusing on Indian languages and contexts. Private efforts, like Krutrim by Ola's founder and Project Indus by Tech Mahindra, are also advancing, showing a collaborative push towards self-reliance in AI ([India's LLM race is heating up! Here's a look at who's building what](#)).

Comprehensive Analysis of India's LLM Development

India's journey towards developing its own Large Language Model (LLM) is a critical component of its broader ambition to become a global leader in artificial intelligence (AI). As of February 2025, India has not yet developed a single foundational LLM, which is a model trained from scratch to handle a wide range of natural language processing tasks across multiple languages and contexts. This section provides a detailed examination of the reasons behind this delay, the current efforts underway, and the prospects for future development, ensuring a thorough understanding of the landscape.

Background and Definition

LLMs, such as OpenAI's ChatGPT or Google's LaMDA, are advanced AI models capable of understanding and generating human-like text, typically trained using unsupervised learning on vast datasets. A foundational LLM is one trained from scratch, as opposed to fine-tuned versions based on pre-existing models. India's linguistic diversity, with over 22 scheduled languages and numerous dialects, necessitates such models to cater to local needs, including education, healthcare, and governance.

Reasons for the Current State

Several factors have contributed to India's inability to develop a foundational LLM by February 2025:

1. Financial Constraints and High Costs:

- Developing LLMs requires significant computational resources, such as thousands of GPUs, and large datasets, making it a costly endeavor. The text highlights that hosting and training LLMs is expensive, with only a few global companies able to afford it ([4 major challenges we must overcome to make LLM mainstream](#)).
- Indian AI start-ups have faced funding challenges, raising only \$166 million in 2024, a sharp decline from \$518.2 million in 2022, limiting their capacity to invest in such projects ([Why India Needs Its Own LLM: Insights into the Challenges and Opportunities](#)).

2. Data Scarcity and Linguistic Diversity:

- India's rich linguistic landscape, with many languages having limited digital content, poses a significant challenge. Less than 1% of global digital content is in Indian languages, and the predominantly oral nature of some languages complicates data collection ([India's Large Language Language: Challenges and Opportunities](#)).
- The need to handle code-switching, where speakers blend languages, and varied scripts further complicates model training, requiring advanced processing capabilities ([How Indian LLMs are gaining momentum](#)).

3. Infrastructure Limitations:

- Until the IndiaAI Mission, India lacked sufficient high-performance computing infrastructure. The mission plans to supply 18,693 GPUs, but this is part of a broader

AI ecosystem, and whether it's sufficient for training a foundational LLM remains uncertain ([India to build home-grown foundational model by 2025](#)).

- The computational demands of LLMs, often requiring thousands of GPUs for training, have historically been a bottleneck, with India relying on foreign hardware.

4. Talent Shortage:

- While India has a large IT workforce of 5.4 million, with over 200,000 professionals having AI/ML skills, the specialized expertise required for LLM development is limited. Many talented individuals work for international firms or abroad, reducing domestic capacity ([The India LLM Project — my thoughts](#)).

5. Regulatory and Ethical Considerations:

- Data privacy concerns and the need for ethical AI development, especially given India's diverse cultural context, add layers of complexity. Initiatives like Bhasha Daan aim to crowdsource data, but scaling this remains challenging ([India's Large Language Language: Challenges and Opportunities](#)).

Current Efforts and Progress

Despite these challenges, India is making significant strides towards developing its own LLM, with both government and private sector initiatives:

1. Government Initiatives – IndiaAI Mission:

- Launched with a budget of INR103.7 billion (approximately US\$1.2 billion), the IndiaAI Mission aims to build a domestic LLM within 10 months from February 2025, targeting completion by December 2025 ([India to build home-grown foundational model by 2025](#)).
- The mission includes seven pillars: IndiaAI Compute, Innovation Center, Dataset Platform, Application Development, FutureSkills, Startup Financing, and Safe & Trusted AI, with a focus on indigenous models trained on Indian datasets ([Call for proposals for building India's foundational AI models](#)).
- It plans to establish a computing capacity of over 10,000 GPUs and develop models with over 100 billion parameters, focusing on sectors like healthcare, agriculture, and governance ([IndiaAI Mission](#)).

2. Private Sector Efforts:

- Several start-ups are making progress, including:
 - **Sarvam AI:** Raised \$41 million in 2023, focusing on LLMs for multiple Indian languages, with Sarvam-1 being a two-billion parameter model optimized for local needs ([DeepSeek's LLM success triggers big debate: Is India's hesitation a strategic mistake?](#)).
 - **Krutrim:** Led by Ola's founder Bhavish Aggarwal, developing multilingual models trained on 2 trillion tokens, with Krutrim Pro slated for advanced problem-solving ([India's LLM race is heating up! Here's a look at who's building what](#)).

- **Tech Mahindra's Project Indus:** Launched in 2024, focusing on Hindi and 37 dialects, aiming for release by December 2024 or January 2025, with plans for domain-specific applications ([tech mahindra hindi llm: Tech Mahindra eyes Dec-Jan release of Hindi LLM under 'Project Indus'](#)).
- Other notable models include OpenHathi by Sarvam AI, the first Hindi LLM, and Tamil-LLAMA, both building on existing frameworks like LLaMA ([7 Top LLMs from India to Watch Out for in 2024](#)).

3. Collaborative Efforts:

- The IndiaAI Mission invites proposals from startups, researchers, and entrepreneurs, fostering public-private partnerships. Collaborators include IBM, Microsoft, NVIDIA, and academic institutes like IIT Madras ([INDIAai | About Us](#)).
- Initiatives like Bhashini and Bhasha Daan are working on digitizing Indian languages and crowdsourcing data to enhance LLM development ([India's Large Language Language: Challenges and Opportunities](#)).

Prospects for the Near Future

Given the momentum, India is poised to develop its own LLM in the near future, likely by the end of 2025. Key factors supporting this include:

- The government's aggressive timeline and substantial budget, with IT Minister Ashwini Vaishnaw announcing a world-class AI model by year-end ([India to develop its own AI model like ChatGPT and DeepSeek in 10 months: Ashwini Vaishnaw](#)).
- Increasing private sector investment and innovation, with models like Krutrim and Project Indus showing promise.
- The potential to leverage open-source models, such as DeepSeek R1, as a starting point, democratizing access to AI development ([The India LLM Project — my thoughts](#)).

However, success depends on overcoming remaining challenges, such as scaling infrastructure, ensuring data quality, and attracting talent. The debate among industry leaders, with figures like Nandan Nilekani suggesting focusing on applications rather than models, highlights differing views, but the government's commitment suggests a strong push forward ([New Tech Debate: Should India Build its Own LLM?](#)).

Comparative Analysis

To contextualize, global leaders like the US and China invest heavily in AI, with the US's Stargate Project allocating \$500 billion. India's \$1.2 billion, while significant, is smaller, but its focus on indigenous models tailored to local needs could provide a competitive edge, especially in sectors like agriculture and governance ([Why India Needs Its Own LLM: Insights into the Challenges and Opportunities](#)).

Conclusion

India's path to developing a foundational LLM is marked by significant challenges but also promising developments. The combination of government support through the IndiaAI Mission and private sector innovation positions India to achieve this goal by the end of 2025, enhancing its AI self-reliance and digital inclusivity.

Key Citations

- [Top 10 LLM That Are Built In India](#)
- [India and its Own Large Language Model \(LLM\) — Do We Need One?](#)
- [Top Large Language Model \(LLM\) Companies in India](#)
- [Does India need its own large language model \(LLM\).](#)
- [India's LLM race is heating up! Here's a look at who's building what](#)
- [India to build home-grown foundational model by 2025](#)
- [The India LLM Project — my thoughts](#)
- [7 Top LLMs from India to Watch Out for in 2024](#)
- [New Tech Debate: Should India Build its Own LLM?](#)
- [tech mahindra hindi llm: Tech Mahindra eyes Dec-Jan release of Hindi LLM under 'Project Indus'](#)
- [The Future of Large Language Models \(LLMs\): Strategy, Opportunities and Challenges](#)
- [Why India Needs Its Own LLM: Insights into the Challenges and Opportunities](#)
- [India's Large Language Language: Challenges and Opportunities](#)
- [4 major challenges we must overcome to make LLM mainstream](#)
- [Building LLMs: A strategic necessity](#)
- [7 Top LLMs from India to Watch Out for in 2024](#)
- [How Indian LLMs are gaining momentum](#)
- [India braces up for AI challenge, plans own LLM foundational model to rival ChatGPT, DeepSeek R1](#)
- [DeepSeek's LLM success triggers big debate: Is India's hesitation a strategic mistake?](#)
- [Call for proposals for building India's foundational AI models](#)
- [INDIAai | About Us](#)
- [IndiaAI Mission](#)
- [INDIAai | Pillars](#)
- [Cabinet Approves Ambitious IndiaAI Mission to Strengthen the AI Innovation Ecosystem](#)
- [IndiaAI mission: Call for Proposals to Build Foundational AI Models](#)
- [IndiaAI Mission](#)
- [Cabinet Approves Over Rs 10,300 Crore for IndiaAI Mission, will Empower AI Startups and Expand Compute Infrastructure Access](#)
- [Website of IndiaAI Mission| National Portal of India](#)
- [INDIAai | National Missions](#)
- [India to develop its own AI model like ChatGPT and DeepSeek in 10 months: Ashwini Vaishnaw](#)

- [With DeepSeek, are India's foundational AI model dreams closer to reality?](#)
- [Tech Mahindra launches Project Indus Large Language Model \(LLM\)](#)
- [Tech Mahindra Launches Project Indus Large Language Model \(LLM\)](#)