

# BUD ECOSYSTEM PIONEERS INNOVATIVE NEW INDIC LLM WITH AMD

Bud Ecosystem achieves faster training for India's first commercially usable open-source Indic LLM using 40% fewer GPUs on Azure with instances powered by AMD GPUs



Bud Ecosystem set out with an ambitious mission: build India's first commercially usable open-source large language model (LLM) designed for Indic languages. The company, a young AI research and development startup, believes, as Linson Joseph, Bud's CSO, explained, "that the future of AI is to make it accessible and transparent."

**"AMD Instinct™ MI300X GPUs provide us a sweet spot where we save cost, train much faster, and have the right platform to meet our requirements."**

Linson Joseph, CSO, Bud Ecosystem

That belief positioned Bud Ecosystem to develop Hex-1, a 4-billion-parameter LLM trained exclusively on AMD Instinct™ MI300X GPUs. By combining AMD Instinct accelerators and the open-source AMD ROCm™ software stack, Bud Ecosystem was able to train and optimize Hex-1 at speed, scale, and with meaningful cost efficiency. The result is an LLM model that outperforms its larger competitors in Hindi, Telugu, and other Indic languages, and a breakthrough in bringing generative AI to India's diverse population.

## BUILDING AN LLM MODEL FOR INDIA, BY INDIANS

Bud Ecosystem's leaders saw a market gap and an opportunity. High-profile LLMs such as GPT and Llama 3.2, trained primarily on the English language, have demonstrated limited performance in Indic languages, creating barriers for millions. Meanwhile, the Government of India was launching its AI Mission, with plans to build large-scale GPU infrastructure and fund models that could serve the country's needs. Bud Ecosystem's challenge was to build a competitive model for the India AI Mission in just 65 days that could meet national-scale demand and prove that Indian startups could deliver world-class AI.

As Joseph puts it, "Choosing a hardware infrastructure that takes a month to do pre-training versus one from AMD that lets me finish in 15 days would directly impact our business goals. Microsoft Azure with AMD Instinct MI300X GPUs provides us a sweet spot where we save cost, train much faster, and have the right platform to meet our requirements."

**"The exceptionally high 192 GB of HBM3 memory capacity of AMD Instinct™ MI300X lets us keep the entire model on a single device, optimize everything in place, and converge training faster."**

Ditto P S, CTO, Bud Ecosystem

## INDUSTRY

Software and Sciences

## CHALLENGES

Bud Ecosystem needed to train India's first commercially usable Indic open-source LLM in just 65 days, requiring high memory, speed, and efficiency to meet national AI mission goals.

## SOLUTION

Using Microsoft Azure instances advanced by AMD Instinct™ MI300X GPUs and AMD ROCm™ software, Bud was able to train its Hex-1 LLM with scale, flexibility, and cost efficiency.

## RESULTS

Hex1 trained 20–30% faster using 40% fewer GPUs. Compared to models like Gemma-2B, LLaMA-3.2-3B, and Sarvam-1, Hex1 delivers better performance across all supported languages on the MMLU benchmark.

## AMD TECHNOLOGY AT A GLANCE

AMD Instinct™ MI300X GPUs  
AMD ROCm™ Software

## TECHNOLOGY PARTNER



**“The portability offered by ROCm gave us confidence to migrate our code base and make full use of AMD Instinct™ MI300X accelerators.”**

Ditto P S, CTO, Bud Ecosystem

Bud Ecosystem CTO Ditto P S continues, “Our main consideration was ensuring that our model’s throughput and runtime memory requirements were fully met. High memory bandwidth, 5.3 terabytes per second on the AMD Instinct MI300X, meant we could train faster and more efficiently. The exceptionally high 192 GB of HBM3 memory capacity of AMD Instinct MI300X lets us keep the entire model on a single device, optimize everything in place, and converge training faster, giving us confidence we could deliver on schedule.”

### **ACCELERATING TRAINING WITH AMD INSTINCT MI300X GPUS AND ROCM SOFTWARE**

Bud Ecosystem selected AMD Instinct MI300X GPUs deployed via Microsoft Azure as the training platform for Hex-1. “With AMD Instinct MI300X, we don’t have to rely on complex sharding or offloading to manage memory,” says Ditto. “The large memory capacity of AMD Instinct MI300X lets us run bigger batch sizes and explore longer context lengths without compromising performance. The flexibility we have with AMD Instinct MI300X GPUs means we can iterate quickly, rather than waste time on workarounds.”

Ditto adds, “The AMD ROCm software stack is widely adopted and continues to grow. Its kernels are optimized for different algorithms and quantization, and it’s fully compatible with PyTorch. Because ROCm is open source, we can examine the kernels directly, take advantage of community contributions, and port our code directly with only minimal work. The portability offered by ROCm gave us confidence to migrate our code base and make full use of AMD Instinct MI300X accelerators.”

Joseph explains, “We approached Hex-1 with an open-source mindset—open code, open models, and sharing with the community. ROCm fits perfectly into that philosophy.”

### **WINNING WITH SUPPORT AND EARLY ACCESS FROM AMD**

For Bud Ecosystem, partnership was as important as hardware. Joseph emphasizes how AMD engagement shaped their progress, saying, “We are a startup, but the confidence AMD shows in us—aligning the right technical stakeholders, providing access to resources, and making sure we connect with their roadmap and support—plays a big role. The entire AMD team put forward the effort to bring us up to speed and give us confidence to move forward.”

Ditto adds, “Early access to machines with AMD Instinct MI300X in AMD labs was invaluable. We used those systems to validate our stack, run initial optimizations, and see how the cards handled specific kernels. The AMD support team also set up calls to walk us through how the ROCm stack works, what optimizations were available, and how to integrate them with our software. That work gave us the confidence to then scale to larger clusters on Azure.”

### **GAINING SPEED, EFFICIENCY, AND COST SAVINGS**

“The higher memory and bandwidth of AMD Instinct MI300X shortened our training cycles by roughly 20 to 30% compared to previous platforms we had studied,” says Ditto. “The resource efficiency of AMD Instinct MI300X meant we required about 40% fewer GPUs than we would have using other platforms, which plays a big role in controlling our costs. And, overall performance of AMD Instinct MI300X was much better than anything else we tried.”

**“The higher memory and bandwidth of AMD Instinct™ MI300X shortened our training cycles by roughly 20 to 30%.”**

Ditto P S, CTO, Bud Ecosystem

Joseph emphasizes the benefits to Bud customers. “Our 4-billion parameter model running on AMD Instinct MI300X GPUs provides greater accuracy and will help customers reduce deployment costs for Indic language applications. For example, in inference, they can save significantly while also getting better accuracy. That’s forward-looking cost savings we can deliver right now.”



*Bud trained Hex-1 up to 30% faster using AMD Instinct™ MI300X GPUs.*

The combination of Hex-1 with AMD Instinct MI300X proved its strength in Bud's benchmark testing. Bud's 4B-parameter model outperformed larger global models such as ChatGPT and Llama 3.2. For example, in Hindi, Hex-1 achieved an MMLU score of 46.73 versus 3744 for Llama-3.2-3B, 3740 for Gemma-2B, and 43.80 for Sarvam. In Telugu, Hex-1 scored 46.85 compared to 33.15 for LLaMA-3.2-3B, 35.13 for Gemma-2B, and 43.00 for Sarvam-1. MMLU (Massive Multitask Language Understanding) measures knowledge and reasoning across many subjects, making it a strong indicator of real-world capability. Hex-1 also achieved a BLEU score of 0.52 in Hindi, compared to 0.47 for LLaMA-3.2-3B, 0.46 for Gemma-2B, 0.49 for Sarvam-1, and 0.45 for GPT-3.5. In Telugu, Hex-1 scored 0.51, compared to 0.45 for LLaMA-3.2-3B, 0.44 for Gemma-2B, 0.48 for Sarvam-1, and 0.43 for GPT-3.5. BLEU (Bilingual Evaluation Understudy) measures how closely machine-generated text matches human output in terms of both accuracy and fluency.

With the success of Hex-1, Bud Ecosystem is looking to expand the applications of its Indic-language models and contribute to India's growing AI ecosystem. The success of this first model validates Bud's open-source approach and highlights what startups can achieve with the right collaboration and infrastructure. Hex-1 stands as proof of progress toward a more inclusive future for language technology.



*Access to AMD hardware and technical expertise helped Bud meet its aggressive timetable.*



#### EXPLORE [AMD INSTINCT GPU](#)

Sign up to receive [AMD data center communications](#).

#### ABOUT BUD ECOSYSTEM

Bud Ecosystem is on a mission to democratize access to generative AI, making it practical, affordable, profitable, and scalable for everyone. To achieve this, Bud is reengineering the fundamentals of GenAI systems, from runtime environments to model architectures to agent frameworks. Bud Ecosystem makes GenAI portable, scalable, and independent of specialized hardware. For more visit [budecosystem.com](https://budecosystem.com).

#### ABOUT AMD

For more than 50 years AMD has driven innovation in high-performance computing, graphics, and visualization technologies. Billions of people, leading Fortune 500 businesses, and cutting-edge scientific research institutions around the world rely on AMD technology daily to improve how they live, work and play. AMD employees are focused on building leadership high-performance and adaptive products that push the boundaries of what is possible. For more information about how AMD is enabling today and inspiring tomorrow, visit the AMD (NASDAQ: AMD) [website](#), [blog](#), [LinkedIn](#), and [X](#) pages.

#### DISCLAIMERS

All performance and cost savings claims are provided by Bud Ecosystem and have not been independently verified by AMD. Performance and cost benefits are impacted by a variety of variables. Results herein are specific to Bud Ecosystem and may not be typical. GD-181

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. The information contained herein is subject to change and may be rendered inaccurate for many reasons, including but not limited to product and roadmap changes, component and motherboard version changes, new model and/or product releases, product differences between differing manufacturers, software changes, BIOS flashes, firmware upgrades, or the like. Any computer system has risks of security vulnerabilities that cannot be completely prevented or mitigated. AMD assumes no obligation to update or otherwise correct or revise this information. However, AMD reserves the right to revise this information and to make changes from time to time to the content hereof without obligation of AMD to notify any person of such revisions or changes. GD-18.

#### COPYRIGHT NOTICE

© 2025 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, AMD Instinct, ROCm, and combinations thereof are trademarks of Advanced Micro Devices, Inc. Other product names contained herein are for identification purposes only and may be trademarks of their respective owners. Certain AMD technologies may require third-party enablement or activation. Supported features may vary by operating system. Please confirm with the system manufacturer for specific features. No technology or product can be completely secure.