

AI Technology Ecosystem

Opportunities and Imperatives for India

AI4India.org
March, 2023



Executive Summary

The AI4India forum organized a panel discussion on the AI Technology Ecosystem in India. Speakers from academia, startups, large organizations, and think tanks shared their perspectives on AI, the issues, technology challenges, and ways to overcome them. The panel discussion focused on Generative AI and its implications, specifically LLMs and ChatGPT, the infrastructure for training large deep learning models, and ways to strengthen the AI Technology Ecosystem in India.

Generative AI and LLMs

Generative AI is a subset of Artificial Intelligence that involves the creation of new content, such as images, text using machine learning algorithms. One of the most significant advancements in Generative AI is the development of Large Language Models or LLMs. Examples include GPT-3 from OpenAI, LLaMa from Meta, Bard from Google and many others that are capable of producing highly coherent and convincing text. ChatGPT, an application created by OpenAI based on GPT3.5 architecture, is designed to generate human-like responses to text-based prompts has shown tremendous potential in transforming the way we communicate and interact with technology.

Researchers in academia are seeking a deeper understanding of how LLMs work, their limitations, failure modes etc. LLMs are based on the transformer architecture which was introduced by Vaswani et al. in the paper “Attention is all you need” in 2017. Additionally, the concept of human alignment through Instruct GPT and Reinforcement Learning from human feedback (RLHF) are important breakthroughs that made these models powerful. Stanford Initiatives such as Alpaca Project has built an instruction-following LLaMa model and shared the datasets & code to fine-tune the model. This is helpful for researchers to experiment and gain a better understanding of how they work.

Industry is gearing up to respond to the disruption caused by LLMs. While startups are keen to build new apps with GPT, the established players are more cautious in their approach due to the reputational risk if things go wrong. Enterprises are thinking about the implications – trying to discover usecases in automation, improving decision making etc. Also looking at what it means for Legal, Finance departments, considering ways to reskill employee base and make the best of this new technology.

LLMs can also help sectors such as education, agriculture, service delivery etc. The Government of India wants to make its products and services available to every citizen in voice and in their own language. This is a part of the Bhashini mission and can be accomplished by combining Indian language models to work in conjunction with LLM to allow natural conversations in all languages. It is possible to train an IndicGPT model using data that reflects Indian culture such as videos (TV channels), audio transcripts, documents, music etc. There are challenges to overcome but it is not beyond the realm of possibility.

Training deep learning models

Training large deep learning models requires a significant amount of computational resources and infrastructure. Deep Learning models involve a large number of computations. Graphical Processing Units or GPUs are designed to perform these computations in parallel which makes them ideal for accelerating the training process. NVIDIA GPUs such as T4 and A100 are optimized for deep learning workloads. Other companies, such as AMD, Intel and Google also offer specialized hardware for deep learning. For the software stack, Pytorch and Tensorflow are very common.

Cloud platforms such as AWS, Azure and Google Cloud provide several advantages for training large deep learning models in terms of scalability, flexibility and accessibility. However GPU costs for training large deep learning models can add up and become expensive in India. There are some international providers such as Lambda labs coming up with cheaper alternatives. There are data center providers in Tier-2 cities are coming up with newer architecture and offering better performance from a cost and time perspective. However they don't have capital to host large GPU clusters.

The need to optimize costs is driving innovation in training and inference methods. Frugal approaches to train include, using Macbooks as an alternate to GPUs but this requires some hacking and hotfixes for Pytorch. One can also avoid training large models by using pre-trained deep learning model as a feature extractor and using it for a basic machine learning model which can be trained without GPUs. However these approaches can work for smaller models and narrow domains only.

Recommendations

The panel made the following recommendations to strengthen the AI Technology Ecosystem for LLMs and training large deep learning models:

1. Academic institutes in the country procure 5 or 10 GPUs for their on-premise training. These are siloed in each institute and not available as a cluster. If the money spent by each institute is combined and invested in procuring GPUs on Cloud, the combined GPUs open up a lot of possibilities to train large deep learning models. Government can ask cloud providers in the country to make lots of GPUs available which can be provided as credits to researchers
2. Availability and accessibility of data is a key challenge for both academia and industry. There is a need to unlock data with Government organizations, anonymize and make them accessible for research and building applications. This data can be used to train language models (LLMs) for India. Global LLMs can also use this data to ensure Indian context is included to a greater extent in future versions
3. Partnerships and Investments are needed to build a more vibrant AI Technology Ecosystem that can strike a balance between AI for Industry and AI for Development & Social Good. Researchers need to be encouraged to explore software and hardware architectures to make model training more efficient and indigenous. Application Developers need to be enabled to democratise sectors such as healthcare, education, agriculture & finance in India, making them more accessible and efficient for everyone, regardless of their background or socioeconomic status

Conclusion

Artificial Intelligence or AI has the potential to play a significant role in India's development and progress, and it is important for the country to invest in AI research and development to ensure that it remains competitive in the global economy. AI is also of strategic importance to India from a geopolitical perspective. India can't remain algorithm takers and data givers.