



Nandan M. Nilekani
Chairman

What's next for the generative AI revolution

We are into the second year of the generative AI revolution, and some clarity is beginning to emerge from the noise and babble of the last 18 months. The initial hyperventilation of AI doomerism and the risk of human extinction by AI advances like Artificial General Intelligence (AGI) has quietened down. People have accepted that, like any other general-purpose technology be it electricity, nuclear energy, the internet or even a discovery like fire, gen AI has enormous potential for good when explored and advanced within the guardrails of responsibility. Besides, many of the doomsday prophets pleading for extensive AI regulation have revealed themselves to be just protectionists who want to limit the fruits of gen AI to a few companies and investors.

It is also clear that there won't be a scenario where we'll have 'one model to rule them all'. Every day brings new advances in large language models from a dizzying set of actors all pushing for greater innovation. These range from very large models which need massive computing infrastructure to small ones that can run locally on the phone. The real power of AI will come from configuring all the different models and tools to get the best solutions. This is not very different from previous generations of technology. What's more, the rise of powerful open-source AI models has accelerated the deployment of AI to solve tough business and societal challenges. Although there could be concentration risk in the hardware and cloud infrastructure space, as we move into actual use cases, a thousand flowers will bloom.

It is more than evident that enterprise AI will be markedly different from consumer AI. The manifestations of consumer AI will be packaged in wondrous ways to make life easier and more productive for millions of people. New ways of conducting search; agents that help plan work and leisure; intuitive interfaces that can serve up what's needed and reason with users; even speech recognition that understands the nuances of dialects and colloquialisms. Not unlike the smartphone that brought the magic of apps and touchscreen to billions, consumer AI will push the envelope of usability, convenience, and accessibility for everyone.

Enterprise AI, on the other hand, requires a root and branch surgery of the complex and multigeneration technology (both legacy and modern), that lie within firms. The AI models themselves will become commodities. The challenge will be to orchestrate the extensive data inside

the corporation, both structured and unstructured, explicit and tacit, in a way that it is consumable by AI. The quality of output needs to be managed to ensure correct and factual responses and insights with no hallucinations. Given that the leaderboard of technologies will be changing at a bewildering pace, enterprises will have to 'future proof' their AI infrastructure with suitable abstractions to be able to switch models easily and not be trapped in a technological cul de sac. As various nations come up with different ways of regulating AI, global companies will have to build their AI applications in a way that they are compliant in every country. While application can be trialled on very large models, deployment will be on narrow transformers, trained on relevant enterprise data, fully secure and efficient in their inferencing. Enterprises will need both an AI foundry for experimentation and an AI factory for scaling up. AI architecture must facilitate an approach that combines the analytical thinking of the left brain with the intuitive approach of the right brain. The constraint of resources will require a transparent way of identifying the highest value AI use cases. AI must amplify the potential of every human being in the enterprise.

Firms in the business of enabling digital transformation, like Infosys, will be in the eye of the storm. Software development will be rapidly automated and amplified. We have to seize the productivity benefits and share them with our customers. The productivity gains from automation, must lead to talent redeployment in new areas with new opportunities. We must learn from applying AI to ourselves, be it in creating an AI-first enterprise or in accelerating the massive talent amplification that's now needed. We are already doing it, at Infosys, by applying Infosys Topaz to transform all the services we offer to become AI-first and to accelerate business value using generative AI technologies. We are certain that change will have to be embraced – not resisted.

Above all, the gen AI revolution presents an unrivaled opportunity. The flux of change as the whole technological landscape is being reset will create many large openings. Packaged solutions will be reimaged, and we will see a resurgence of custom-built solutions that must be enabled with new types of AI building blocks. The puck is clearly and quickly moving to a place where the balance of advantage will be with Infosys.

Bengaluru
May 16, 2024

Nandan M. Nilekani
Chairman