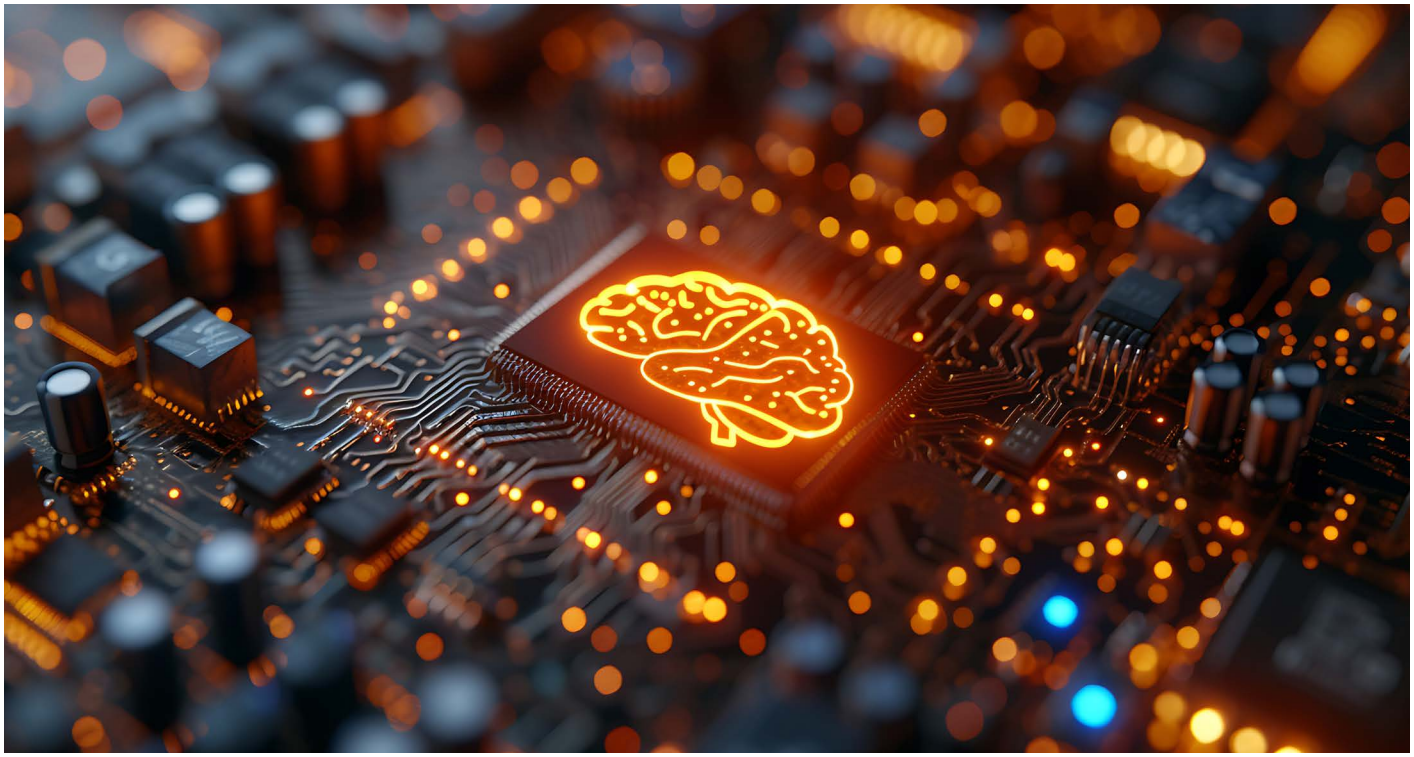
VIEW POINT

GENERATIVE AI: THE FOUNDATION FOR TRANSFORMATIVE ADVANCES IN AI ADOPTION

Abstract

Generative AI has garnered global interest, evidenced by the fact that ChatGPT-3.5 reached 100 million users in two months. AI's evolution has seen peaks and troughs, with significant growth in recent years, particularly in foundation models. Applied AI has flourished, and now integration with generative AI offers even greater potential. GenAI use cases include spoken and text customer interaction, fraud detection, pharmaceutical R&D, clinical trial success rate prediction, and marketing content generation. Software development and education are other sectors that are expected to be impacted significantly. However, safety, legality, and environmental impact concerns must be addressed. As organisations adopt generative AI as an assistive tool, transparency, governance, and sustainability are key considerations.

Infosys®
Navigate your next



Generative AI has captured the attention of corporations across the globe. Most people were introduced to GenAI through OpenAI's ChatGPT-3.5 release, which gained a record of 100 million users in two months. From that time the evolution of LLM, GenAI exploded, almost each month we can hear about new models, new capabilities that have been announced- with recent OpenAI unleashing ChatGPT-4o to everyone and plans to release 5th generation mid-2024. The broadcast is flooded with news about Human AI Pin, Rabbit R1, Google Gemini or Devin which supposed to replace developer and much more.

However, the origin of AI and its path of development extend back decades to the 1950s when Alan Turing introduced the Turing Test, which could determine the ability of a machine to have natural language conversations with human beings. Since then, the AI development journey hasn't been smooth and continuous. There were peaks of inflated expectations followed by troughs of disillusionment, disinterest, and reduced investment in the 1970s and then again from the late '80s to the early '90s.

The mid-1990s to 2010 was a period of steady growth. This period saw the Deep Blue computer beating chess champion Gary Kasparov, NASA's rovers using AI for autonomous motion, the concept of deep learning emerging at the University of Toronto, and speech recognition being introduced on mobile phones. From 2010 onwards, there has been a period of explosive growth. These years saw the advent of virtual assistants and considerable leaps in NLP (natural language processing), with an algorithm clearing the Turing Test in 2014 and another outperforming people in a Stanford reading and comprehension test.

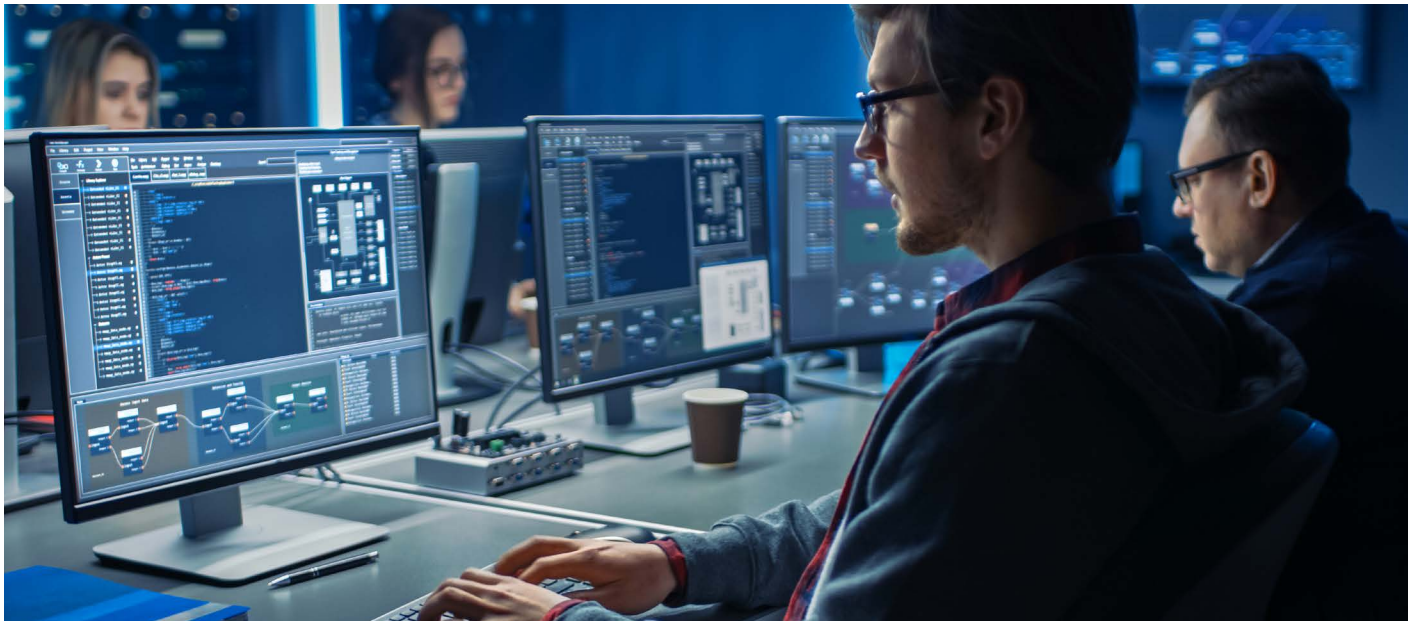
In the 2020s, foundation models have ushered in the era of generative AI. Traditionally, AI systems have had a library of models, each built for a specific task and trained on a large, structured, labelled data set related to the task. The sourcing, curating, and labelling of these data sets require people with domain knowledge; also it was a time-intensive activity. A foundation model is a paradigm shift with an individual model capable of learning and working across domains and problems, enabling it to cater to different tasks. Foundation models are trained on diverse, voluminous, unlabelled data and can be fine-tuned for various applications. Fine-tuning a foundation model for a particular domain is done using an unlabelled domain-specific data repository and a much smaller domain-specific labelled data set. The most known foundation models are the LLMs (large language models), capable of understanding and generating natural language.

Applied AI has seen substantial growth in recent years. As per a McKinsey study, from 2018 to 2022, applied AI has seen the highest investment, innovation, and demand for talent. There has been a steady flow of investment (104 billion USD in 2022) in applied AI, resulting in great strides like a 63.6% decrease in training time for image classification. Demand for talent in applied AI has tripled since 2018. Integrating applied AI with newer technology like GenAI holds enormous growth potential. Synergising applied and generative AI will maximise value for organisations.

GenAI technology has opened up new use cases across industries and can improve, expedite, and scale up existing use cases. GenAI is expected to have a wide-ranging impact, including creating new products and revenue streams. Employee productivity, however, is expected to see the maximum impact. GenAI searches have increased threefold, and there has been a spate of investments, with venture capital investments rising by 425% from 2020. Organisations across various sectors are exploring enterprise use cases, and more than 80% of ongoing AI research is concentrated on GenAI. Some enterprises are developing their foundational models, while others are building over existing models or extending open-source models. Many firms are working on integrating GenAI into their product lines. Hardware companies are working on accelerators and customised designs to run foundational models better.

In these initial stages, organisations are concentrating on utilising GenAI more as an assistive technology that helps perform a task better or more efficiently. The use cases have human experts in the loop and are not fully automated. Businesses are working to put together GenAI strategies and are keen on deriving value through pilot projects before investing big to build capabilities. A subset of the use cases implemented includes:

- Spoken and text customer interaction
- Fraud detection
- Pharmaceutical research and development to quickly develop drug candidates
- Clinical trial success rate prediction, and marketing content generation



Software development is another sector that is expected to transform with GenAI. GenAI tools will deliver significant productivity gains for coding. Documenting code functionality, writing new code, and refactoring existing code can be faster by 50% depending on developer experience and code complexity. The gains were most observed in repetitive tasks, helping coders get into a flow, accelerating code refactoring, and facilitating the undertaking of challenging tasks. The tools have to be prompted iteratively to achieve requisite code quality, and currently, the tools function better at augmenting developers than automating coding.

The education sector will see pivotal changes with the adoption of GenAI tools, which will facilitate inclusive learning environments, individualised learning support, personalised feedback, automation of admin tasks, lesson planning, and workshop preparation. Students can have personalised 24/7 GenAI tutoring tools, which provide explanations of concepts based on their learning patterns. A GenAI evaluation tool can assess and challenge students on their assignments. The tool can study the submission and ask questions about the thought process and method of arriving at it. GenAI answer analysis tools can help students understand their mistakes by analysing erroneous answers and breaking down the faulty thinking behind them. Language learning support systems can provide interactive conversations for different scenarios. GenAI shifts learning to a dynamic and personalised journey from a rigid and standardised format.

As the use of GenAI spreads across industries and the world, there is an urgent need to ensure the technology's safety, legality, and environmental sustainability. The proliferation of generative AI has raised justified concerns about the development methods and the end use. The lack of transparency, the integrity of the training data, bias, intellectual property infringement, and privacy and security

violations must be addressed as a priority. Additionally, erroneous or manipulated output and lack of traceability of the source and explainability are risks that must be mitigated. The carbon footprint from the development and use of a single generative AI model could reach up to 284 metric tons of CO_2e, equalling the lifetime emissions of five average American cars. This substantially adverse environmental impact is under the spotlight. If generative AI is to realise its transformative potential, then organisations must focus on transparency, governance, data management, and compliance. On the environmental front, using existing models instead of developing new ones, energy-conserving computation, and monitoring energy sources should be at the forefront of organisational strategy.

About Author:

Michal Krauze has been with Infosys for 18 years, leading Digital Delivery and Quality in the EMEA region. He is a visionary leader, responsible for driving the digital automation agenda, delivering innovative technology solutions, and enhancing productivity in the EMEA. Holding a Master's degree in Information Technology and Econometrics, and being a certified Lean Six Sigma Black Belt and ITIL Foundation, he possesses a deep understanding and expertise in the field of digital business transformation.

With extensive experience spanning over 18 years in various departments and roles, he has developed a strong skill set in strategic planning, business analysis, continuous improvement, program management, and leadership. His enthusiasm for the latest technologies, including AI, extends beyond the tools and systems, focusing on the broader positive impact these can have on the company and its workforce.

For more information, contact infosysbpm@infosys.com

Infosys®
Navigate your next