What if generative AI is reaching its limits?

The ability of generative AI ('GenAI') to generate plausible text, images, music and computer code in response to human prompts is impressive. GenAI promises huge productivity gains in many domains, and large amounts of financial and political capital are staked on its success. Nevertheless, the current generation of models exhibit well-publicised weaknesses that might not simply disappear by using more data and processing power or smarter training. This paper looks at those limitations, and the lower-profile alternative approaches to AI that could overcome them and even provide the EU with a competitive advantage.

Since the unveiling of ChatGPT in 2022, we have witnessed what many commentators consider to be the **most rapid adoption** of new technology in recent times in virtually all segments of society. One cannot fail to be impressed by the speed at which it can produce outputs that some say can be 'PhD-level'.

At the same time, there is growing awareness that these algorithms often produce plausible but demonstrably false outputs ('**hallucinations**') or outputs that are **inconsistent with logic, maths** or 'how the real world works'; they regurgitate problematic **biases** present in their training data and **cannot provide a reliable explanation** of how the outputs were derived. 'GenAI-optimists' believe that accuracy and tractability will emerge through **scaling up** the data and processing power available and adding external (human or automated) processes to fine-tune the models. 'GenAI-pessimists', on the other hand, point to the fact that finding solutions to the above problems is slowing down, strongly suggesting that the approach is reaching its limits. In addition, the **very high energy and resource consumption** of the infrastructure necessary for training and running these 'brute force', data-driven applications is putting pressure on already stretched supplies. Given the **several trillions** of private and public capital invested, now and over the next few years, in the **geopolitically strategic** race for an 'all-purpose' artificial general intelligence (AGI), it is worth asking where GenAI's limits lie and what other approaches might be explored to put Europe at the forefront of AI innovation.

Potential impacts and developments

GenAI models, in essence, generate a multi-dimensional **map of statistical relationships** in the training data between '**tokens**' – digital representations of linguistic, graphical, musical or other data. The models can then be prompted to produce new combinations of text, images, music, etc. that are consistent with those relationships. In contrast to traditional statistical regression analysis, these models contain **billions of parameters** and make almost no assumptions about the form of the relationships between tokens. This gives GenAI its enormous advantage in finding the subtle, multivariate patterns that allow it to output plausible text, code, music, images, etc. in response to prompts.

The flip side of this underlying design, however, is that GenAI (like traditional statistical regression) is prone to **mistaking 'noise'** in the training sample for meaningful patterns and producing **bizarre results** when it is applied to data outside the training set. In addition, pure **GenAI trades transparency of 'reasoning' for power**: a printout of the billion-parameter map of relationships provides no explanation. This is a problem for the required 'human in the loop': how can one have confidence in the result if no one knows exactly how it was reached?

Clearly, these **design weaknesses** limit mission-critical uses of AI. Scaling GenAI does not appear to eliminate these problems, all the while adding additional problems linked to resource usage. Ideally, we would make the most of the powerful pattern detection that GenAI offers, and overcome some of the limitations of a purely data-driven approach. This requires boosting research into **'hybrid' approaches** which, like the human brain, leverage GenAI's pattern extraction and matching abilities using built-in,

Author: David Kemp, Scientific Foresight Unit (STOA)
PE 774.701 – November 2025

explicit models of efficient reasoning strategies and the real world, as well as strategies for maintaining and developing those models. These approaches provide a **reality check** to improve the reliability of GenAI's probabilistic outputs. They are also **more efficient** (algorithmically and energetically) as they directly encode readily available, fundamental knowledge rather than requiring everything to be extracted by brute force from the data. Finally, the use of explicit representations and strategies allows their reasoning to be inspected.

Neuro-symbolic approaches, such as IBM's neuro-symbolic concept learner, combine GenAI with explicit rules for symbolic reasoning (for logic, abstraction and generalisation) to enhance reasoning and explainability. **Embodied AI** – such as Meta's Habitat – involves training agents in virtual or physical environments where they are designed to learn through **perception, action and feedback**, promoting causal learning and the development of sensorimotor intelligence. **Cognitive architectures**, such as Soar, ACT-R and OpenCog, include explicit models of human cognitive processes, integrating perception, memory, learning, planning and reasoning in a modular way. This enables continuity of learning, goal-directed behaviour and long-term memory. **World model learning** approaches such as DeepMind's MuZero, Ha & Schmidhuber's world model agents and DreamerV3 focus on training agents to derive **compact, predictive models of their environment** to support causal reasoning, generalisation and efficient planning. Finally, it must be underlined that, regardless of the underlying technology, the pursuit of artificial *general* intelligence is not necessarily the most efficient route to useful applications. Artificial *specific* intelligence (AI approaches focused on a specific domain, such as the Nobel prize-winning, protein-folding algorithm, AlphaFold2) gives more reliable and transparent results by combining the subtle pattern detection at which GenAI excels with explicitly encoded, domain-specific knowledge.

Anticipatory policymaking

The reliability issues mentioned above call into question GenAI's ability to deliver the 'trustworthy and human centric AI … pivotal for economic growth and … [preserve] the fundamental rights and principles that underpin our societies', as promised in the AI continent action plan. Most AI initiatives from the European Commission have so far concentrated on the implementation of GenAI rather than on research and development of alternative and complementary AI approaches. The €700 million flagship GenAI4EU programme, for example, states its aim as being to 'integrate generative Artificial Intelligence (AI) in Europe's strategic sectors, and keep their competitive edge'. Consequently, most of the calls for projects focus on applying GenAI in particular sectors and providing the data and computing power it needs. The pursuit of artificial general intelligence has attracted an enormous amount of political and economic interest, potentially to the **detriment of equally interesting and possibly more efficient and effective alternatives**, including those mentioned above.

To best serve the EU's goals of competitivity and innovation in AI, EU policy and funding could be targeted more directly to support the whole value chain of **alternative and complementary** approaches to GenAI. It is also important to actively further domain-specific AI (**artificial specific intelligence**) applications alongside GenAI. This can be achieved through proactively promoting such projects for **existing funding programmes** and by **policy guidance** in the next multiannual financial framework and the Competitiveness Fund for Digital Leadership.