Why 2025 belongs to AI

Artificial intelligence (AI) has been fuelling the conversation for over two years. Yet many investors and observers are still wondering when this technology will truly shift into high gear. The successive breakthroughs we have seen in recent years – especially in large language models (LLMs) such as GPT-4 – have sparked a remarkable wave of enthusiasm, but there remains a certain scepticism as to whether this momentum can be sustained over the long term.

Meanwhile, a major turning point is emerging: the arrival of so-called "reasoning models," the o-series from OpenAI, released in late 2024. These are expected to accelerate the adoption of AI and ignite a new phase of investment, making 2025 a decisive year for artificial intelligence.

By David Rainville, Fund Manager of Sycomore Sustainable Tech

Since 2023 and through 2024, there has been significant buzz around AI, though it has yet to reach the scale of past technological cycles. Many observers and investors have focused on "foundation models" like GPT-4. These are known for their ability to rapidly generate answers on a wide range of topics or to summarize very lengthy and/or complex texts, thereby delivering productivity gains. However, these models have proven useful for only a few specific cases so far.

Mass adoption of LLM technologies has, to date, been relatively limited. Investors were betting on the release of new foundation models, such as GPT-5, to increase the number of use cases and usher in rapid adoption. But the release date for GPT-5 is still unknown. Why? The progress has been slower than expected, due to a lack of high-quality new data to train even more intelligent models.

The media hype of the past two years has given way to new questions: Have we exhausted all the high-quality data available to train new models? Have LLMs reached their technical limits? Will AI-related spending dry up?

We believe that in 2025 these uncertainties should ease, building more confidence in the AI theme. This will be driven by two factors:

1. **The arrival of reasoning models** o1 and o3, announced at the end of the 2024, which boast much more powerful thinking and problem-solving capacities than the current foundation models. This should lead to a broader adoption of generative AI technologies by businesses.

2. **Demand for semiconductors and memory chips will gather further speed** as a result of these new applications, but also because reasoning models require much more computing and memory power than older generations.

We believe that in 2025 some uncertainties should ease, building more confidence in the AI theme

Combined, these two factors suggest a sustained growth in AI spending, potentially surpassing $250 billion in 2025 and topping $300 billion by 2026, according to our analysis.

WHAT IS A "REASONING MODEL"?

To understand why reasoning models are a turning point, we need to distinguish these newcomers from foundation models (LLMs) like GPT-4.

- **Foundation models (LLM):** These can be compared to vast knowledge libraries. Ask these models a question and you will receive an almost instant response extracted from the masses of information already fed into the model during its training process. GPT-4 or other LLMs such as PaLM and LLaMA are good examples within this category. These models are very good at conveying and summarising all the data they are given.

- **Reasoning models:** Built on the same foundation as LLMs (the transformer technologies [1]), reasoning models go much further by introducing an extra "layer of expertise" with an ability to reason. In practical terms, instead of simply diving into their knowledge pool, reasoning models (such as o1 and o3) will break the question down into several stages, develop a problem-solving logic, and execute this strategy to find a solution – whether the problem is complex or even unprecedented.

[1] The transformer is a novel neural network architecture based on self-attention mechanism and used for the processing of natural language.
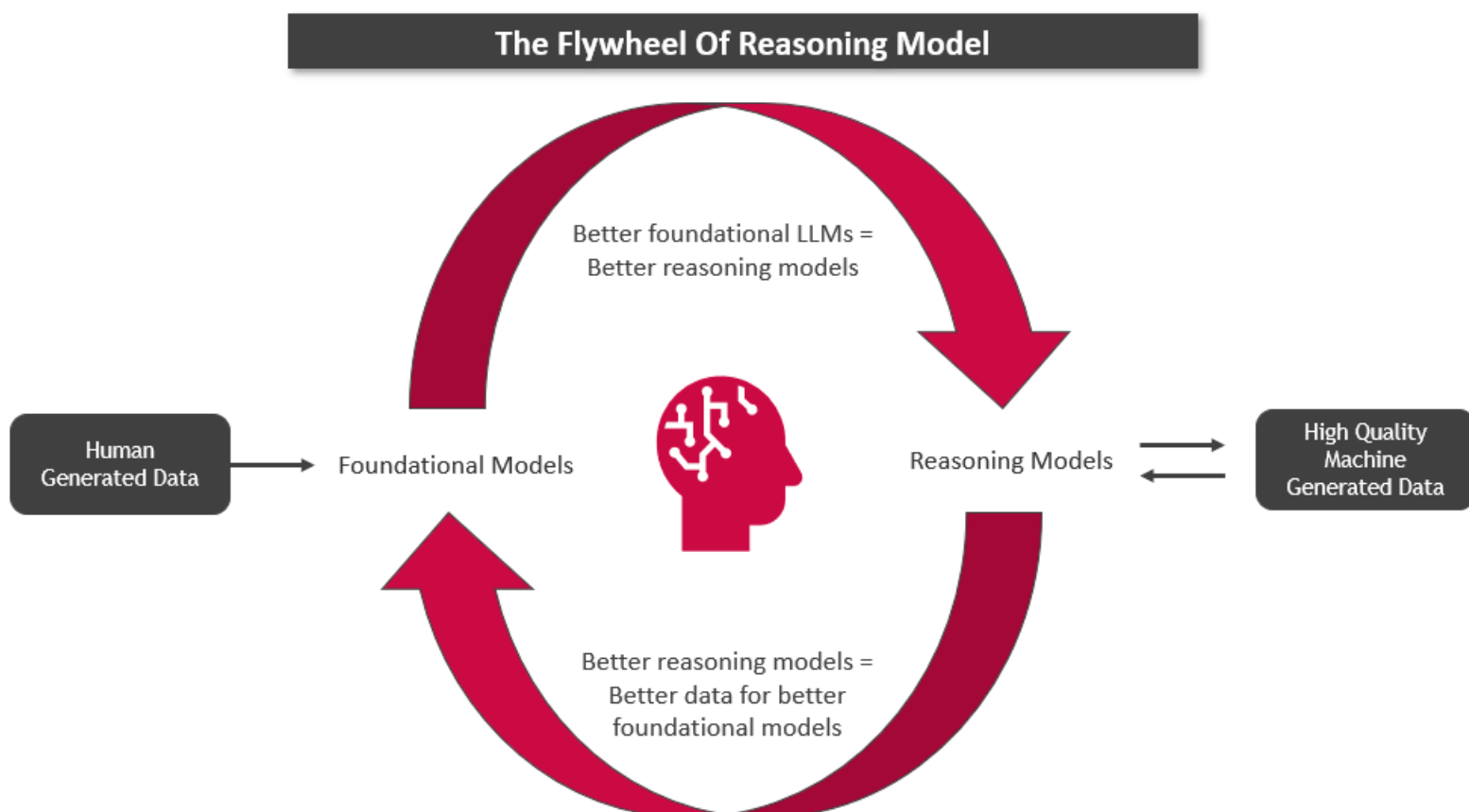
With estimated IQs of 140 for o1 and almost 160 for o3, these models are close to achieving the concept of AGI (Artificial General Intelligence). The o3 model actually ranks among the world's top 200 coders, beating experienced researchers (including OpenAI's Head of research). This qualitative leap is a game changer for the adoption of generative AI technology, as it opens the door to new applications.

A FRESH OUTLOOK FOR INVESTMENT

1. A new training and data creation phase

Up to now, AI model training has relied on huge amounts of text, images, and audio from the internet and other sources. Several AI labs (and many investors) worried that we were about to run out of "high-quality data," threatening the continual improvement of LLMs.

However, the arrival of reasoning models could upend that equation. Thanks to their ability to reason and generate more structured content, o1 and o3 may themselves produce "synthetic data" that will be used to train future base models. In other words, we are about to witness a virtuous cycle in which AI helps enrich the data needed for its own evolution.
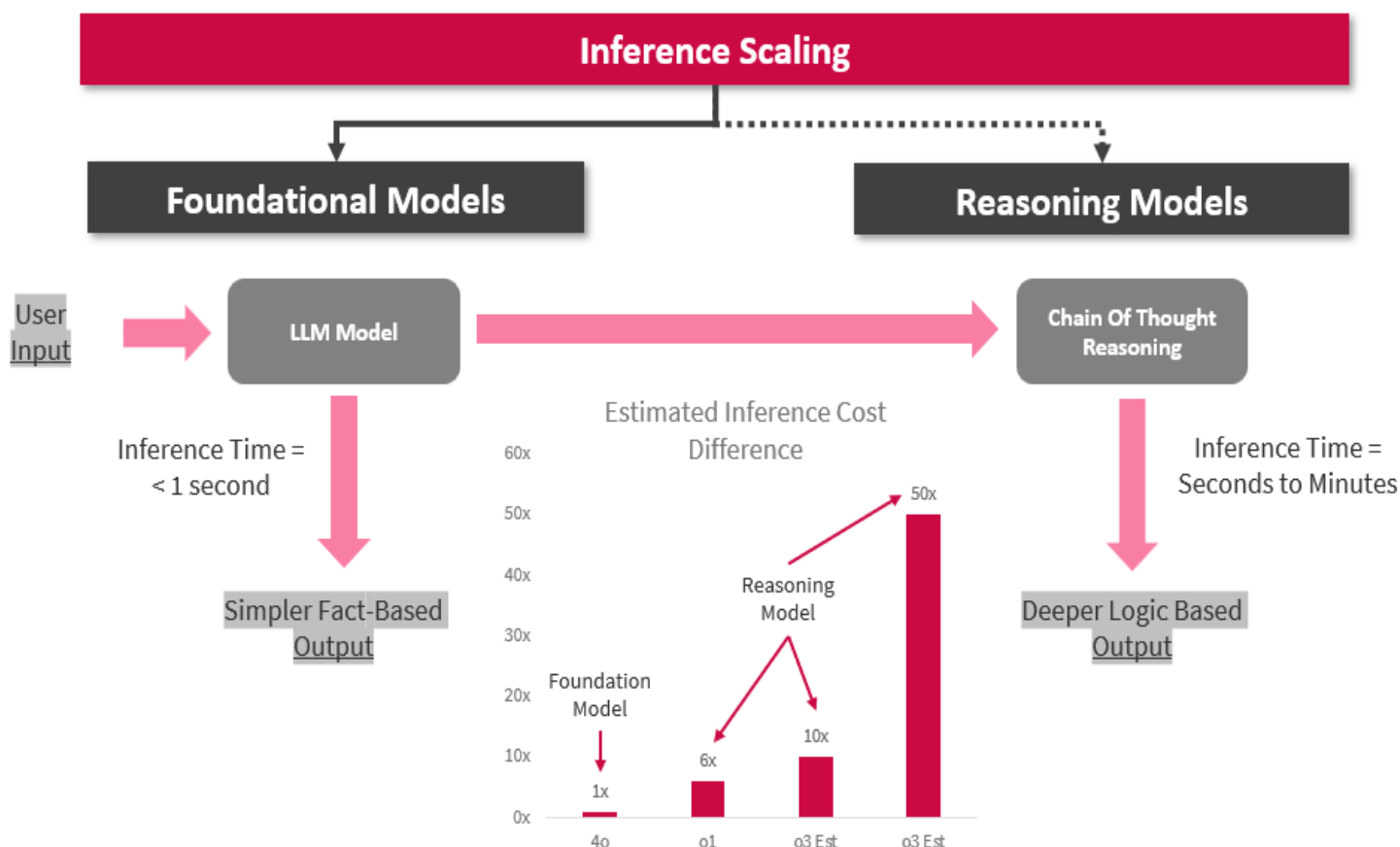

Source: Sycomore Asset Management.

2. Inference costs will be much higher

Reasoning models require more resources than basic models, in both computing power and memory. Some estimates suggest their costs are 10 to 50 times higher than GPT-4's. The reason lies in the "thinking" or "inference" phase: instead of delivering an instant response, a model like o1 Pro might "think" for several minutes to produce a detailed, reliable solution.

For businesses, this extra investment can be justified if they're solving complex problems where accuracy outweighs speed. And for infrastructure providers (semiconductor, GPU, and High Bandwidth Memory [HBM] makers), this cost increase translates into a continuous - even growing - demand from customers.

Sources: Sycomore Asset Management; Semianalysis.

MEMORY: A KEY CHALLENGE

In 2025, one of the biggest potential bottlenecks will be memory capacity, particularly HBM. Unlike traditional memory, HBM stacks multiple chips to deliver higher bandwidth and density. It is essential for supporting the long computation times of reasoning models, but also comes at a significantly higher cost - on the order of 3 to 5 times that of conventional memory.

In 2024, recall that investors discovered how essential networking components were to the performance of AI data centres, dramatically boosting both the performance and valuations of those companies.

We expect a similar moment of truth for the memory sector in 2025.

The HBM market is very large. In 2024, it already reached $16 billion and could surpass $30 billion by 2025, rising to $100 billion by 2030. This figure is roughly equivalent to the entire global memory market today. This rapid growth is testament to the strategic importance of HBM for the next generation of AI.

WHICH STOCKS TO WATCH?

When discussing AI investments, **Nvidia** is often top of mind: the company has greatly benefited from the initial wave of LLM adoption. However, other players may warrant investors attention with more attractive valuations:

TSMC: The undisputed leader in chip manufacturing, TSMC effectively holds a monopoly position in components for major tech companies such as Nvidia, Google, AMD, and Meta. With a market cap of around $800 billion, TSMC trades at a multiple that many consider reasonable (around 14x earnings). According to our estimates, over 25% of its revenue in 2025 will come directly from AI chip fabrication. As ASML (in Europe) experienced its own "moment of recognition" when we moved to EUV technology and it became clear they were a monopoly for leading edge lithography, we believe TSMC's valuation could rise as investors come to realise its growing economic and geo-politic importance.

Micron: With a market cap of roughly $100 billion, Micron is one of the key manufacturers of HBM memory, a crucial technology for the new AI generation. The company could capture around 30% of this HBM market, which is headed toward $100 billion by 2030. That would more than double its total revenue, just from this segment. Additionally, Nvidia's planned release of its new GB300 range in the second half of 2025 should show Micron is gaining new market share.

REASONING WILL LEAD TO THE MASSIVE ADOPTION OF AI

Although AI's achievements have been widely publicized since 2023, 2025 looks to be the real tipping point when large-scale adoption will begin.

The uncertainties that once loomed over demand are now fading, thanks to the emergence of reasoning models like o1 and o3, which are far more resource-hungry but should also play a core part in training future and more intelligent AI systems.

This momentum, coupled with technological innovations (such as HBM memory) and now more attractive investment opportunities (TSMC, Micron, Nvidia, etc.), promises a pivotal year in which AI transitions from an unproven sector to an unavoidable force in many new economic application.