Article | 1 August 2024

AI Monthly: The world of the unknown

This month, AI investments get a reality check while progress in technology keeps hopes high. But regulators never sleep, and the European Central Bank is now looking at the impact of AI on inflation and monetary policy. Our focus this month shifts to just how interconnected the AI world and the major issues surrounding it are

It's widely recognised that the impact of AI could soon affect the labour market and productivity levels, but it could also reach further to influence inflation, monetary policy transmission and the natural rate of interest

Reality check: The 500 billion-dollar AI investment gap

The Artificial Intelligence (AI) sector is characterised by high-flying ideas and many possible up-and-coming companies. But since it isn't clear whether those companies will be successful – at best, turning the established corporate landscape upside down, or at worst failing and destroying the capital invested – the AI landscape relies heavily on venture capital. So, it's no wonder that ears prick up when one of the world's biggest venture capitalists begins to wonder whether all the

investments in AI are justified and if they'll ever be rewarded with a decent return. Sequoia, ranked number one in all global venture funds by Dealroom, has renewed its analysis from September 2023 on the gap between AI sales expectations and AI actual sales growth, concluding that the **125 billion dollar gap** is now set to become a 500 billion dollar gap.

The problem with developments in AI is that today's state of the art technology quickly becomes obsolete

---

The calculation goes as follows: for every dollar spent on a Graphics Processing Unit (GPU), roughly one dollar needs to be spent on energy costs to run the GPU in a data centre, meaning that your expenses amount to two dollars per GPU. Now, the end user of the GPU also needs to earn a margin, assumed at 50% of the final price, which brings the total costs to four dollars. That is, for each year of GPU CapEx, a four dollar revenue needs to be generated to pay back the upfront capital investment. So if, for example, Nvidia sells 50 billion dollars in run-rate GPU revenue, running a data centre costs 100 billion dollars. Adding the end users' margin of 50% implies that for the GPU investment, a 200 billion dollar revenue is needed to pay back the upfront capital investment, estimated as of September 2023. Even if the big tech companies and other AI-related companies were able to generate a part of these expenses, we're currently some way off from closing the upfront investment gap – resulting in a 500 billion dollar gap, according to David Cahn from Sequoia.

High initial investment costs are nothing new and are part and parcel of revolutionary inventions. The problem with developments in AI is that today's state of the art technology quickly becomes obsolete, meaning that investments may never pay off for some.

Progress in AI models: The ever-changing world of AI

Indeed, the pace of advancements in AI technology is incredible. Another milestone was recently met by the company Meta, which released pre-trained models with a multi-token approach. But what makes them so revolutionary?

Currently available large language models (LLM) learn tokens one at a time, i.e., one unit of text that is used to represent a word or symbol such as dots or spaces. A multi-token model, on the other hand, can directly understand and predict an entire sentence. A good example to illustrate this is an orchestra. Each instrument plays beautifully on its own (current standard LLM), but what is really fascinating is the interplay of all the instruments together (a multi-token LLM). But of course, it's not just limited to sentences. Complex formulas can also be obtained in seconds, for example. And what are the benefits? **Faster** and **more efficient** results that require **less computing power**. The models currently available can predict four tokens simultaneously, performing some 15% better than comparable LLMs and generating output three times faster.

The models of Meta are available as open source; they are available to researchers and developers free of charge, so that they're able to research and further develop the technology behind the model. This is a good way to innovate, but brings us back to the beginning of our billion-dollar investment gap.

Regulators never sleep

But AI isn't left to run alone. Well, to some extent, it is. Microsoft, which has invested 13 billion dollars in OpenAI, has given up its [observer seat at OpenAI](#) in hopes of easing concerns surrounding the extent of its control over the startup. The US Federal Trade Commission (FTC) and the Department of Justice are currently conducting an antitrust investigation into big tech companies and their dealings with AI companies.

There's also a question of whether we move towards a higher or lower concentration in the future, scenarios raised by [Erik Brynjolfsson and Gabriel Unger](#). Will only some large firms and their business partners be able to develop proprietary AI, stem the upfront investment costs and the costs to run AI models? Or will we live in an open-source AI model world of for-profit companies, nonprofits, academics, and individual coders, where everyone profits?

AI's effect on monetary policy

And lastly, AI will not only affect the labour market or productivity, but also monetary policy by affecting inflation, monetary policy transmission and the natural rate of interest as Piero Cipollone, member of the Executive Board of the European Central Bank, explained [in a speech](#):

1 Inflation

Downward pressure

- Reduced risk of **labour shortages** and downward pressure on **unit labour cost growth** if the net effect of AI is substituting labour and increasing productivity
- Decline in **energy prices** through enhanced grid management and more efficient energy consumption on the supply side, better tools for price comparison on the demand side

Upward pressure

- Increase in **energy prices** due to higher global demand for energy due to the computing power required
- **Discriminatory pricing** by facilitating the real-time analysis of consumer demand and price elasticities

2 Monetary policy transmission

- If AI leads to an increase in intermediation outside the banking sector, monetary policy transmission could be faster as non-banks react more strongly to monetary policy measures affecting longer-term interest rates, such as asset purchases. They have a higher credit, liquidity and duration risk compared to the banking sector, which affects people's marginal propensity to consume and their access to credit.

3 The natural rate of interest

Downward pressure

- Labour displacement and rising income inequality leads to an increase in **precautionary savings** and a subsequent boost to the supply of **loanable funds**

Upward pressure

- Productivity and output boost lead to higher demand for **capital investment** and expansion of **production capabilities**

However, it is not yet clear to the European Central Bank which effects will dominate. In general, much is still unclear in the world of AI. But that is also what makes it so fascinating. To conclude in the language of the ECB, we must remain vigilant and will report more on the world of AI in our September issue.

Author

Inga Fechner
Senior Economist, Germany, Global Trade
inga.fechner@ing.de