Article | 5 December 2024

AI Monthly: The world of AI from A to Z

As another year comes to an end, we reflect on how Generative AI has become more embedded in our daily lives and what we can expect from 2025. Here are some of our main takeaways, key terms, and observations

AI Index Report

If you want to know which country is leading the way, the number of AI patents, how much is being invested, what current AI models cost and more, then you need to look at the "AI Index Annual Report" from the Stanford Institute for Human-Centered Artificial Intelligence (HAI). Look out for the release of the 2025 AI Index Report here, probably in April 2025.

Boom

Back in May, a global survey by McKinsey reported that 65% of respondents' organisations regularly use GenAI, nearly double the percentage from a previous survey just 10 months ago. The share of organisations that have adopted AI in at least one business function increased from 55% to 72% with the biggest increase in adoption found in professional services (HR, legal services, management consulting, tax preparation). Meanwhile, 50% responded that their organisations have adopted AI in two or more business functions.

Computing power

Increased investment in AI has led to an immense increase in IT power. In 2023, the Americas had

a live supply of 17.4 GigaWatt (GW) from 8.1 GW in 2018. This compares to 11.1 GW in APAC and 8.8 GW in EMEA from just 4.6 GW in 2018. As investments in the US are higher than elsewhere in the world, this gap is likely to widen in the coming years.

Domains

The AI boom can create surprising macroeconomic side effects, as an IMF paper has pointed out. Anguilla, a small island located in the Caribbean, generated just over 20% of the government's total revenue, resulting in approximately USD$32 million for 2023, by selling ".ai" domains. In the years before, revenue was hovering around 5%. For a two-year domain registration, Anguilla charges USD$140, while expired .ai domains are auctioned, with one domain being sold for USD$13,000 in 2023.

Ethical AI

Ethical considerations in AI development have gained prominence with a focus on ensuring that AI systems are designed and used in ways that are fair, transparent, and beneficial to society. Organisations are establishing guidelines and frameworks to ensure AI is developed and used responsibly, addressing issues like fairness and accountability.

Fiscal policies

Fiscal policy could play a role in curbing the potential harmful effects of Generative AI, as acknowledged in an interesting research paper from the IMF. These harmful effects include labour disruptions and increased inequality, both of which are unwanted for policymakers. Governments should not institute taxes on AI but look at corporate tax incentives that encourage a rapid displacement of labour. This would decelerate possible displacement of labour. Moreover, governments should consider raising general taxes on capital income to offset rising inequality.

GPU (Graphics Processing Unit)

Specialised electronic circuits, or GPUs, are designed to accelerate the processing of images and videos, ensuring smooth visuals in videos and video games. GPUs are also used in non-graphical tasks such as neural network training, scientific simulations and cryptocurrency mining due to their high computational power. They excel at parallel processing, making them ideal for tasks that require handling multiple operations simultaneously. The leading GPU manufacturers are NVIDIA, AMD, and Intel, all based in the US.

Human in the Loop (HITL)

HITL is a concept of incorporating human feedback into the training and refinement of AI models to improve their accuracy and reliability, i.e. human expertise is combined with machine learning in order to augment human capabilities rather than replace them.

Investment Gap

The AI sector is full of high-flying ideas and potential companies, but their success is uncertain. This uncertainty makes the sector reliant on venture capital. Sequoia, a top venture fund, has highlighted a growing gap between AI sales expectations and actual sales growth, now projected to reach $600 billion. High initial investment costs are nothing new and are part and parcel of

revolutionary inventions. However, the problem with developments in AI is that today's state-of-the-art technology can quickly become obsolete, meaning that investments may never pay off for some.

Job Transformation

We've looked closer into how AI might transform labour markets this year. And we believe that AI will fundamentally transform the job market, impacting workers of all skill levels. While some jobs will become obsolete, AI is not expected to lead to mass unemployment. Instead, it will create new roles and enhance existing ones, particularly in high-skilled occupations. Advanced economies will be more affected due to their higher share of high-exposure jobs. Workers will need to adapt to changing skill demands, with digital competencies becoming increasingly important. The transition will be gradual, with AI complementing rather than completely replacing human labour.

Knowledge Distillation

This is a technique in machine learning where a smaller model (the "student") learns to mimic a larger, more complex model (the "teacher"). This way, the smaller model can perform almost as well as the big one but is far more efficient and easier to run on devices with limited power, like smartphones or embedded systems. This technique is super helpful for cutting down the energy use of Generative AI models, which can be pretty power-hungry.

Large Language Models (LLMs)

Large Language Models (LLMs) are advanced AI systems designed to understand and generate human language. They revolutionise how we interact with technology. The competition to become the leading Generative AI company, meanwhile, has intensified greatly this year. The available cutting-edge LLMs have required billions of dollars in investment to get to this level of sophistication. Grok recently raised $6 billion based on a company valuation of $18 billion and is looking to raise an additional $5 billion, according to the Financial Times. Rival OpenAI is in talks to raise billions of dollars after a $13 billion commitment from Microsoft while Anthropic has attracted about $8 billion, including significant funding from Amazon and Google, according to news reports.

Multi-Token AI

The pace of advancements in AI technology is incredible. Another milestone was met this year by the company Meta, which released pre-trained models with a multi-token approach. But what makes them so revolutionary? Traditional LLMs learn tokens sequentially, meaning they process one unit of text, such as a word or symbol, at a time. In contrast, a multi-token model can directly understand and predict entire sentences. 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.

Nvidia

Why mention one single company? Because Nvidia is special in the world of AI, having revolutionised computing with its GPUs. Founded in 1993, Nvidia initially focused on graphics for

gaming but expanded into AI, data centres, and autonomous vehicles. Their GPUs are now crucial for AI applications, offering unmatched performance and efficiency. Its success can also be seen from the strong increase in its revenues, something that is expected to continue.

Open-source

Open-source means that the source code, datasets, and model parameters of AI technologies are freely accessible for anyone to use, study, modify, and share thus making AI models transparent, customisable, innovative, and cost-efficient. Open-source models fall into the ethical AI category. An example of an open-source model is Meta's LLaMa (Large Language Model Meta AI), in contrast to closed-source models such as ChatGPT.

Power

The International Energy Agency (IEA) estimates that data centres and data transmission networks, the powerhouses behind the AI infrastructure, are each responsible for 1-1.5% of global electricity use. By 2026, power demand from data centres could reach 1,000 terawatt-hours (TWh), double the levels of 2022 and equivalent to Germany's annual power demand (512 TWh).

Quantum Computing

Quantum computing might bring AI to the next level as quantum computers can process complex calculations much faster than classical ones. The Quantum Insider projects that quantum computing will contribute $1 trillion in value creation by 2035, starting to deliver value to commercial end users as of next year. The US, the UK, Germany, France, China, and Japan are best positioned to benefit from the value generated by quantum computing according to the report.

Regulation

The European Union has finalised the AI Act, which sets comprehensive legal frameworks for AI applications, categorising them by risk levels. Ahead of its legal deadline in two years' time, more than 100 companies signed the EU AI Pact in September. This means they are making voluntary commitments to start implementing the requirements of the upcoming AI Act.

Sectors and AI

Wondering how AI could impact sectors across Europe? We've taken a look at the details, and we're inclined to think that smaller sectors stand to benefit the most in the near term. Larger sectors may not be ready to implement AI just yet – and for some, perhaps not at all. We think productivity gains will be limited to around one percentage point per year. Smaller, more digitalised sectors in Europe, such as Information and Communications Technology (ICT), business services, and financial services, are set to gain the most from AI in the near term. These sectors can quickly implement AI to enhance productivity. Larger sectors
like manufacturing and transportation will benefit more slowly due to lower digitalisation levels. Healthcare and real estate also have significant potential but face challenges like data availability and privacy issues.

Trade and AI

AI not only plays a role for individuals and companies but will also influence global trade. A WTO

report highlights how AI intersects with international trade and which benefits AI holds such as enhancing trade efficiency by optimising supply chains, improving trade logistics, and facilitating better market analysis. AI technologies could help reduce trade costs, increase transparency, and streamline customs procedures.

US dominance

The US is not only home to the best LLMs, the country's investment volumes are larger than anywhere else, and the same goes for the availability of computing power. This means that the US will likely dominate the AI market for the foreseeable future.

Virtual agents

Virtual agents, also known as intelligent virtual agents (IVAs), are AI-driven tools designed to enhance customer experience through personalised interactions. They are becoming more advanced and widespread, assisting with tasks ranging from customer service to personal organisation. Well-known examples are Alexa, Siri, Cortana or chatbots used on websites.

WormGPT

There are always two sides to a coin! While we've discussed the numerous opportunities AI brings, it's important to remember that the underground world doesn't sleep either. WormGPT is a malicious AI tool designed to assist cybercriminals. Unlike typical AI models that adhere to ethical guidelines, WormGPT is specifically created to generate harmful content, such as phishing emails and malware. It has been promoted on the dark web as a tool for automating cyberattacks, making it easier for criminals to carry out their activities, with successors called FraudGPT, Escape GPT, Evil GPT and many more.

XAI (Explainable Artificial Intelligence)

Refers to a set of processes and methods that make the results and workings of AI models understandable to humans. Unlike traditional AI, which can be a "black box" with decisions that are hard to interpret, XAI ensures that every decision made during the machine learning process can be tracked and explained.

Youth and AI

Some 1.2 billion young people aged 15-24 face a skills mismatch with labour market demands, with GenAI expected to automate many entry-level jobs, impacting youth employment prospects. The report "Time to Act: Preparing Youth for Work in an AI-Powered World" therefore stresses the urgent need for coordinated efforts to equip young people with GenAI skills, bridge the digital divide, and focus on uniquely human skills like empathy and critical thinking. Free and accessible GenAI training, along with inclusive policies, should be available to benefit all young people, especially in low- and middle-income countries so they can fully leverage AI advancements.

Zettabytes (ZB)

A zettabyte is 10^{21} or 1,000,000,000,000,000,000,000 bytes and they are crucial in AI because they represent the massive scale of data generated and processed globally. Data is essential for training AI models, especially large-scale models like GPT and other foundation models. As data

enhances the models' accuracy and capabilities, it is no wonder that data grew exponentially over the last decade. While in 2010 the volume of data/information created, captured, copied, and consumed worldwide amounted to 2 ZB, it increased to 123 ZB in 2023, according to Statista. That means that we would need 123 billion iPhones with the largest storage capacity of 1 terabyte (TB) each to store our current global data volume. By 2025, global data volume could even surpass 200 zettabytes, underscoring the importance of efficient data management and processing in AI.

While recent geopolitical events have tempered some of the hype around AI, the future holds significant promise. The impact of President-elect Trump on the AI landscape, the evolution of the world of work, and the potential returns on AI investments are all key areas to watch. Additionally, the arrival of the first five-gigawatt data centre is highly anticipated. Stay tuned as we navigate these exciting and uncertain times in our AI monthly, resuming in February 2025.

Author

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