Article | 7 March 2024

Digital infrastructure investment is crucial if generative AI is to seriously boost economic growth

Generative AI currently holds considerable potential for accelerating economic growth, impacting our daily lives and gradually becoming recognised as a general purpose technology. But there are still obstacles to overcome – and for widespread adoption and faster implementation, investment in data and digital infrastructure will be crucial

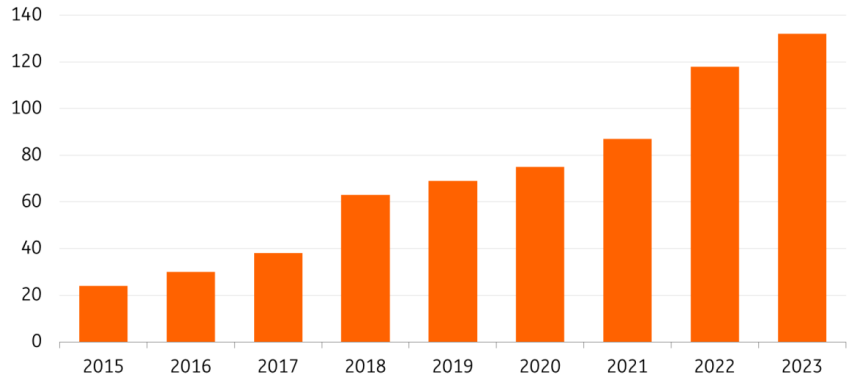
Generative AI is already showing up in topline growth

Some of the largest US technology companies expect to profit significantly from generative AI. To deploy generative AI on a worldwide scale, we have seen investments in digital infrastructure from Amazon, Alphabet, Meta, and Microsoft increase sharply over the past few years. Digital infrastructure investments are a broader category, but they contain investments in generative AI as it requires ultrafast chips, microprocessors, and data centres. Counterpoint Research estimates that in 2023, roughly 13% of Microsoft's digital infrastructure spending was specifically for generative AI. This percentage will likely increase in the future as companies compete with their AI offerings.

Their investments are therefore expected to continue growing for the foreseeable future. Estimates of this growth range between 15-22% for 2024 (Dell'Oro, Bank of America). Aside from large technology companies, AI startups such as OpenAI and Mistral have also received considerable investment. The most important technology companies are now making bets on AI being the next big thing.

Investments in data digital infrastructure are booming

Investments in digital infrastructure by Amazon, Microsoft, Meta and Google in billion USD



Source: Company reports; Benedict Evans

Before we delve into the potential of generative AI, let's first define what we mean by it. Artificial intelligence itself is not new – it has been around for decades and involves the study and development of computer systems that are able to perform tasks that generally require human intelligence. AI is therefore a broad term and includes both machine learning and deep learning. Generative AI is a subfield of deep learning – computer systems that can generate new data instead of merely analysing it. It differs from other forms of AI in its ability to generate content (for example, creating text, maths, coding and videos) that emulate human-generated content. Our new series of upcoming articles will focus on the effects of generative AI on the economy, and this is the type of artificial intelligence our analysis will focus on (unless we specify otherwise).

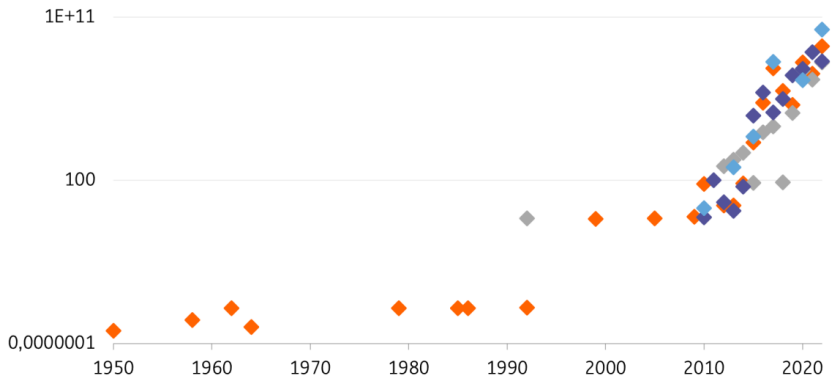
What is AI?

Generative AI did not emerge overnight. Over recent decades, the technology sector has experienced very strong growth, and this has enabled investment in digital infrastructure such as data centres and microchips. In turn, these investments have led to a spectacular increase in computing power that has brought about generative AI.

To illustrate how far computing power has advanced, Google's latest AI system, Gemini Ultra, was trained using 90 billion petaFLOPS (one thousand trillion calculations per second). In 1950, the world's first AI system (Theseus, a robotic mouse that ran through a maze) required 0.0000000001 petaFLOPS. In roughly seventy years, that is a dazzling 900.000.000.000.000.000.000-fold increase in petaFLOPS used to train AI systems. We'd say it's safe to call that an unfathomable rise in computing power.

Generative AI is asking for an enormous amount of computing power

Computation power used to train AI systems measured in petaFLOPS



Source: Epoch (2023); Our World in Data

What's more is that companies that form the infrastructural backbone of generative AI have profited from this spectacular rise in demand. American chip manufacturer NVIDIA, for instance, noted a revenue of just over $22bn in the fourth quarter of 2023. $14.5bn came from data centres, which was a 279% increase in revenue from data centres compared to a year earlier. And it is not just companies that form the backbone of generative AI that are garnering more revenues. According to Reuters, OpenAI – the company behind ChatGPT – increased revenues to $1.6bn in 2023, up from 200m in 2022. In addition, the company already has over 100m weekly active users.

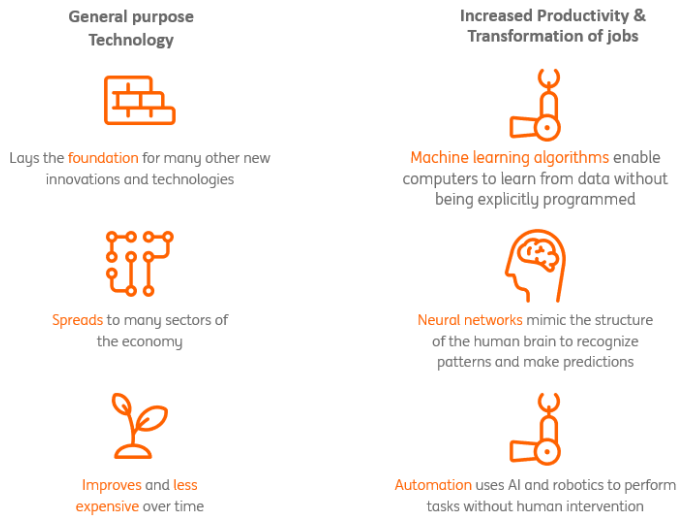
In short, the world's most important technology companies are making very substantial investments in generative AI, and some are already seeing very considerable revenue increases from it.

Unlimited potential? Why we think generative AI is a general purpose technology

Generative AI has tremendous upside potential for improving both the speed and quality of work,

meaning there is possibility for automation. In addition, applications of generative AI are already underway in many different sectors, ranging from construction to healthcare and ICT. This also extends to sectors one would not generally associate with generative AI. As a result, we think it's going to prove to be a general purpose technology. Technologies like the steam engine are considered to be general purpose technologies because they possess the three key characteristics (according to Brynjolfsson & McAfee) of rapid improvement, pervasiveness and complementary innovation. In our view, generative AI ticks all the three boxes.

Explainer: Why is AI a general purpose technology?



Source: ING

Rapid improvement

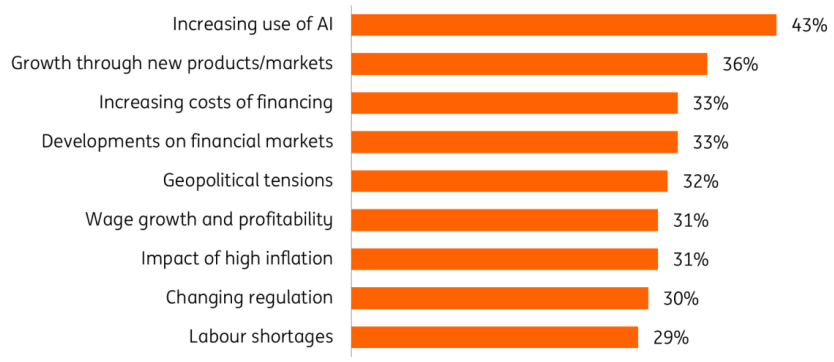
Generative AI's improvement in conducting university exams, for instance, has increased drastically in just a couple of years (i.e., from GPT 3.0 and 3.5 to GPT 4.0). Its ability to digest information has grown exponentially in just a few years. In 2019, GPT 3.0 could only digest a few pages of text at once. Now, it can cope with hundreds with ease.

Pervasiveness

Generative AI is also pervasive. It has encroached on a lot of aspects of our daily lives, from areas such as cybersecurity to something as mundane as restaurant bookings. Moreover, ChatGPT has already managed to rack up more than 100m weekly active users since its launch in November 2022. It's no wonder, then, that an ING survey among top managers of 207 large enterprises in the Netherlands showed that they consider generative AI their main concern this year. Moreover, 85% of them believe that it will have a substantial effect on the daily operations of their company.

Main concerns of Dutch top managers

AI jumped to the top position in 2023



Source: ING

Complementary innovation

Generative AI will also yield complementary innovation, as it will be used in fields one might not typically associate with AI – and because it can easily be used by organisations not yet using the technology. For those organisations or those that have trouble with recruiting developers, it has the potential to democratise coding and software writing. Generative AI allows you to write down the code you want to write in plain English and delivers a coding output, even though you might not be able to write code yourself. It also has the potential to yield breakthroughs in fields like pharmaceuticals or food. For example, some companies work on developing and discovering new proteins for medicines and food with generative AI. It's used to categorise existing proteins but also to give suggestions for new ones. Even in sectors like mining, AI is already transforming the way daily operations are handled.

Another area where we might see rapid change resulting from AI is in industry and manufacturing, as AI exists as one of the key technologies in enabling so-called Industry 4.0. Industry 4.0 is essentially the fourth industrial revolution and is seen as the next phase in the digitisation of the manufacturing sector that should result in autonomous decision-making of computer systems. Last year, the European manufacturing sector was in phase two of six; isolated IT systems are now interconnected. Generative AI can aid manufacturers in making steps in this process.

The next phases for industry 4.0 are more insight, a better understanding of the interdependency between processes, the simulation of future scenarios to develop a predictive capacity, and eventually fully automated decision-making on topics ranging from maintenance to order processing. Generative AI delivers the computing power and automation potential that would enable the replanning of orders with internal and external facilities through predictive analytics, and will be able to make decisions on the quality of output through machine vision.

In essence, generative AI could extend to sectors that one might not immediately be associated with it.

Upskilling and technological interaction

In addition to generative AI becoming a general-purpose technology, it also has the potential to be an upskilling technology – and may change the way we interact with technology altogether.

Upskilling technology

Customer service workers, for example, no longer need to memorise manuals or information about the products their companies sell, but can learn by doing so with the help of generative AI. This could be an important new step in helping workers to learn faster.

Moreover, workers that use generative AI will likely finish tasks faster. This delivers both enhanced efficiency as well as the ability to try more solutions to potential problems. Developers that work with generative AI are found to finish tasks faster, and this gives them the opportunity to try a greater number of solutions to various issues. This could bring about better solutions, enhance quality and improve efficiency – or in the best case scenario, all three. This means that there is potential for automation. The extent of automation depends almost entirely on the capacity of future generative AI models. Predictions based on the economic effects of generative AI therefore hinge on expectations of its future capacity. We are sure that generative AI is a very potent technology that will certainly have some economic effects, but we firmly believe that this process will take time. Technological inventions tend to go much faster than technological adaptation.

Our interaction with technology

Generative AI might impact the way we interact with technology, like the way the graphical user interface (GUI) did. Generative AI has similar potential and might enable us to interact with machines in a more natural way. Through multimodality (i.e., the ability for generative AI to interpret one input, such as text, and generate graphics with it – or vice versa) we might soon be able to have actual conversations with devices – as opposed to the short form commands that are currently used – to generate output rather than typing in commands on a keyboard. In addition, devices might soon understand human-made drawings and be able to convert those to code, text, or pictures. This will not only affect the efficiency or quality with which our jobs are carried out, but also the way we go about doing them. We will discuss this in more detail later in another article.

In short, generative AI is pervasive, improving quickly and may yield complementary innovation, therefore holding all of the characteristics of a general purpose technology. Moreover, it will change the way we interact with technology, which in turn is sure to impact all of our daily lives.

Key obstacles still to consider

However, there are several hurdles to a more widescale deployment of generative AI. The first revolves around data in a broader sense. Generative AI applications were trained using publicly available data, by a plethora of different 'authors'. In short, in its training runs, generative AI might have used copyrighted work to train, and its current output might be subject to copyright infringement. In addition, data found online can recreate existing biases. As an example, this could mean that generative AI's output may discriminate against ethnic minorities. Regulation of high-risk AI applications (like the European AI Act) might also hinder future investment if data usage requirements are very strict. Regulations could include requirements on (model) transparency, but also on a fair and transparent application of algorithms. Lastly, generative AI requires good data, which means companies need to invest in having this available to train their own models.This

might require considerable investment if the existing data quality is poor.

Three other elements are important from a cost perspective. The first is investment in human capital. Workers may need to be reskilled to learn to work with generative AI, which will prove costly. In addition, enterprise software packages with generative AI may are very expensive, and the question is whether companies can foot the bill and deem it relevant for all departments within their firm. Furthermore, if companies decide to purchase generative AI software for their own use, then it must be proved that the technology that it is safe to use. Currently, generative AI applications tend to 'hallucinate' every once in a while. This means that they output incorrect information, or even information that might sound plausible but is entirely made up. This would entail reputational as well as organisational risks, and it is therefore likely that the implementation of generative AI across organisations will be very gradual.

Lastly, and as mentioned previously, current generative AI models use a lot of computing power. This means that they require a lot of data centre capacity to run. Data centres, in turn, need electricity to run and fresh water to cool their servers. Yet, these resources are scarce in many countries and data centres already require large investments to keep up with growing demand. In short, the significant current investment levels in digital infrastructure and electricity grids probably need to increase to meet future demand. So, the question is whether data centre capacity will manage to keep up with the pace of current development, and if they will they be permitted to do so by governments.

In sum, generative AI has a lot of potential, and will likely cause a significant economic impact. However, there are still a number of practical hurdles for generative AI to overcome before it can be deployed on a wider scale and its tremendous potential is able to come to fruition.

Author

Diederik Stadig
Sector Economist, TMT & Healthcare
diederik.stadig@ing.com