#SALMANQADIR

eit Digital

Makers & Shapers

GENERATIVE AI: EUROPE'S QUEST FOR REGULATION AND INDUSTRY LEADERSHIP

AiNed


ABNASIA.ORG

CONTENT

SUMMARY

The outlook of Generative AI in Europe presents a landscape marked by transformative potential, regulatory challenges, and the quest for industry leadership. As Europe navigates the evolving dynamics of Generative AI, several key factors emerge from this report.

Generative AI comprise models that synthesize new audio, code, images, text, videos, designs, materials, and other structural representations from large volumes of often unstructured and unlabelled data by leveraging large neural networks trained with machine learning algorithms.

The development of Generative AI surged after the launch of ChatGPT by OpenAI in November 2022 with a fast pace of development. The release cycle, number of start-ups, and rapid integration into existing software applications are remarkable. Generative AI solutions are spreading, with the entry of new players beside Tech giants, and a marked increase in investment across the Generative AI value chain.

Generative AI holds promises of enhancing productivity, quality of output, and potentially improving work quality across various job tasks. While uncertainties persist regarding job displacement and societal implications, early indications suggest positive impacts on economic growth and job functions.

Europe currently lags behind global competition in Generative AI, with a significant disparity in funding allocation to European start-ups compared to US and Chinese counterparts. The absence of a cohesive EU-wide AI initiative poses challenges for Europe to assert itself with 'Made in Europe' AI products and infrastructure.

It is not just economics at stake. As we have discussed in the previous sections, Generative AI is expected to have an important impact on society, and one area at risk is culture. The sheer dominance of the big US and Chinese AI companies is staggering. By not having viable European AI platforms, European users will have to accept cultural and ethical embedding from AI developed in other parts of the world and European diversity of culture and languages may suffer from it.

Against this breakdown the main implications drawn from the report are:

Balance regulation and innovation. Policymakers are urged to adopt proactive measures that balance innovation with ethical considerations. Collaborative platforms like the AI Safety Summit 2023 facilitate international dialogue on mitigating risks and maximizing opportunities associated with AI technologies. While regulation is crucial to mitigate risks, over-regulation may stifle innovation, especially for smaller startups. Suggest creating regulatory sandboxes and test beds for European Generative AI researchers and companies to experiment responsibly, alongside the EU AI Act.

Funding. Boosting funding is essential, but it should target open ecosystems and collaboration, not just individual big tech players. Initiatives like the proposed GenAI4EU, which creates hubs across sectors, can help democratize access to resources. The packages unveiled by the European Commission on January 24, 2024, to support Artificial Intelligence start-ups and SMEs is a first step but more needs to be done and the funds envisaged are not sufficient.

Competition issues. Beside regulatory initiatives such as the AI Act, there is also a need to address uncertainties regarding competition policy and anti-trust initiatives within the regulatory framework of Generative AI. Risks of unfair competition practices by incumbents in the Generative AI market, such as bundling products or engaging in exclusive dealing must be mitigated. Potential entry barriers for new entrants must be lowered by providing access to talent and computational power

In conclusion, the outlook of Generative AI in Europe hinges on navigating regulatory complexities, fostering innovation, and leveraging collaborative efforts to ensure responsible development and deployment of AI technologies. By embracing these challenges proactively, Europe can position itself as a leader in shaping the future of Generative AI while upholding principles of trustworthiness, transparency, and societal well-being.

INTRODUCTION

Some argue that since November 2022, when OpenAI released its first public-facing version of ChatGPT based on Large Language Models (henceforth also simply LLMs), a new phase of the AI revolution has come of age. ChatGPT reached one-hundred millions of users in about two months, becoming the fastest-growing application in the history of the Web. After ChatGPT, Sora and many other similar solutions followed such as Bard, Claude, Midjourney, and various other content-creating tools.

ChatGPT and similar solutions are general purpose Generative AI, to which for the sake of simplicity henceforth we refer to simply as Generative AI or GAI. We are aware that general purpose Generative AI should be distinguished from special purpose Generative AI, and also that in the policy debate and in the scientific literature several different expressions are used interchangeably such as 'General Purpose AI', 'Foundation Models', 'Frontier Models', 'Large Language Models'. In this report, for the sake of consistency and brevity we use the expression Generative AI (almost always abbreviated as GAI) to broadly refer to models that synthesize new audio, code, images, text, videos, designs, materials, and other structural representations from large volumes of often unstructured and unlabelled data by leveraging large neural networks trained with machine learning algorithms. In this definition, GAI are a specific set of machine-based systems capable of generating content from input data according to an objective.

The data presented later in this report on market developments show a growing diffusions of GAI solutions, the entrants of new players beside Tech giants, and a marked increase in investment across the GAI value chain. After being trained, GAI models show abilities to generate content that closely resembles the patterns and information present in their training data. Chat based variants of such models are generally also used to answer users' queries. For instance, ChatGPT can generate humanlike text and dialogue contributions and Mid- Journey can generate realistic images. While earlier AI systems were able to generate small amounts of content (for instance, suggesting spelling or style changes to an existing text, or making alterations to images), GAI can generate high-quality content from scratch, from minimal prompts[1]. Yet, as it is common when new technologies seem to take off, the typical question analysts face is the following: are we facing a new tech hype doomed to eventually subside or is it a real game-changing opportunity that will radically transform many aspects of economic and societal life?

Judging from the first available analyses and studies (including experimental ones) it appears that GAI can yield substantial increase in productivity, quality of output, and also in improving quality of work at least for certain typology of job tasks. Some studies have found that LLMs based GAI could impact 80 percent of workers in the United States[2] and could increase annual global GDP by 7 percent over ten years[3]. A study focussing on Italy estimated that Generative AI could increase productivity for an amount equal to 18% of GDP or put it differently it could free up to 5.7 billion hours of work[4]. Furthermore, though the academic debate and evidence is still in its early stage and findings cannot be considered conclusive, there are indications that GAI might not simply displace jobs but could actually improve the work of human beings, including that of less skilled workers, which was not the case of previous automation waves.

GAI might hold enormous promises, but it is still far from being perfect and without potential risks. First, there is the well-documented phenomenon of errors in the output of GAI models[5], referred to as 'hallucinations'. Risks to society include increasing discrimination, misinformation and disinformation, overreliance

on GAI contents[6]. Just to mention one example, the provenance of content is a matter of concern. Experiments have shown that human consumers are not good at detecting whether a content was generated by AI or by a human being[7]. As argued[8], consumers have the 'right to know' if the content they use was produced by a person or a machine, in a scenario where a consumer receives a text as advice from his/her doctor about his/her health. In this case one would want to know how the advice was produced and, in case it was generated by AI, if it was thoroughly checked, given potential for errors in GAI models' outputs, and also that overreliance on such models may create serious ethical issues in healthcare[9].

What makes this topic of great relevance is also the fact that next to the fast developments in the technology with new systems and applications emerging constantly, GAI also has acquired centre stage in the regulatory process that led to the agreement within the so-called Trilogue (Commission, Council, and Parliament) for the approval of the Artificial Intelligence Act (AI Act), which has been approved by the Parliament on 13 March 2024. The rise of ChatGPT has forced EU policymakers to basically restart the negotiations in order to agree on how to include GAI in the AI Act. The way GAI has disrupted the legislative process of the AI Act, it is a case in point of how difficult it is for regulation to achieve 'future of proof' in a dynamic innovative domain such as that of Artificial Intelligence. It is enough to read the 'Legislative Train Schedule' (an online tool produced by the European Parliament to monitor legislative processes) on the AI Act to see that the European Commission unveiled a proposal for a new Artificial Intelligence Act (AI Act) in April 2021 and the Council has adopted its common position ('general approach') on the AI Act on 6 December 2022. Neither of the two proposals contemplated Generative AI. But the ChatGPT boom has mobilised the Parliament that issued a new position in May 2023 requiring that GAI are regulated by the AI Act[10]. In its position the Parliament introduced changes to ensure the AI Act ruled Generative AI[11]. After a long debate, on Friday, December 8, 2023, the Parliament and Council negotiators reached an agreement on the Artificial Intelligence Act. A new text was released in January 2024 and approved by representatives of Member States on February 2 2024 and by the by the European Parliament's Internal Market and Civil Liberties Committees on February 13 2024. Next to the AI legislation work in the EU, also the US is moving fast amongst others through the Biden-Harris executive order on AI. And also the UK has created momentum through its recent UK's AI Safety Summit. But regulation is just one leg of a broader AI governance framework, with the other two being public investment and competition policy and anti-trust. In January 2024, for instance, the EU unveiled an innovation package to support AI SMEs and start-ups envisaging the mobilisation of public investments so that the message is the EU does not only regulate (AI Act) but also supports innovation. Recent market deals in the GAI domain (i.e., Mistral / Microsoft partnership, see more infra, section 3.2) have recently raised concerns about market power and the need to ensure competition and keep the emerging GAI market open as a level playing field.

This report builds on secondary sources, analytical and theoretical reasoning, selected interviews, and experts' knowledge. A first version of this report was discussed during two workshops with about 20 experts on March 6 and 7, of 2024. In Chapter 2 the report provides a synthetic analysis of what GAI is and how it works. In Chapter 3 the socio-economic dynamics of GAI is presented. Chapter 4 focuses on risks, and on the ongoing regulatory process and debate. Chapter 5 presents EU perspective and a vertical focus on the situation in the Netherlands. In Chapter 6, four possible future scenarios for the development of Generative AI are presented, which are then assessed in Chapter 7, where we conclude with policy relevant implications and recommendations.

WHAT **GENERATIVE AI** IS AND **HOW IT WORKS**

The large neural networks that enable most modern GAI to work are called Foundation Models. With some simplification, Foundation Models differ from previous generations of AI models in that they can be used for a range of previously unspecified tasks (whereas the latter tended to be 'narrow' and could perform just the tasks they have been trained to perform) and can be finetuned to different tasks and contexts with a relatively limited amount of training. The size of these models is typically measured in terms of number of parameters (i.e. weights of the connections between the artificial neurons in the network) or the number of computational steps (Floating Point Operations, or FLOP's) necessary to train the model. The largest, cutting-edge neural networks are called frontier models.

Transformers[12] are a type of neural network architecture that have played a pivotal role in recent advancements in machine learning[13] and that serve as the backbone of the vast majority of frontier models. The term "transformer" reflects a generalized approach to transforming an input sequence to an output sequence, such as speech-to-text, text-to-speech, text-to-text and other (sequentializable) data modalities through a mechanism called "attention". Transformers, together with self-supervised learning algorithms that do not require human effort for labelling large datasets, have proven to be highly effective in processing sequential information, making them particularly suitable for tasks involving natural language processing such as text generation, summarizing and answering questions. The breakthrough in using transformers for natural language processing and other applications of GAI is integral to the success of foundational models. These transformers allow foundational models to capture and analyse the contextual relationships between many different elements in a sequence, whether it be words in a sentence or data points in a time series. This capability is crucial for generating coherent and contextually relevant outputs in various applications. In the context of Generative AI, transformers contribute to foundational models' ability to learn and generate diverse contents across a wide range of tasks. The Generative Pre-trained Transformer (GPT) is a notable example of a foundational model that leverages transformers, showcasing their significance in the landscape of modern artificial intelligence.

So far, the GAI application that received most attention is ChatGPT, OpenAI's text-generating and question-answering chatbot that uses the Large Language Models (LLMs) GPT-3.5 and GPT-4. LLM's are designed to interpret and generate human-like text based on the patterns and information they learn from vast amounts of training data. These models, including GPT, are pre-trained on massive datasets containing diverse texts, allowing them to capture the nuances of language, grammar, and context. The expression "LLMs" emphasises the significance of these language models in the field of GAI, highlighting their ability to comprehend and generate coherent and contextually relevant text across various tasks. These models have demonstrated remarkable capabilities in natural language tasks such as text completion, question answering, and content generation, making them integral to the advancements in language-based AI applications.

The currently most widely used GAI solutions rely on two essential resources: the data, and the computing resources. Data and computing power represent the costs. The performance of GAI and their associated costs depend on the number of trainable

parameters[14], the more parameters the more the models can learn, but the costs in terms of data and required computing power also increase. The costs in terms of data and computing power are incurred both during the training and the inference phases. In this respect, it must be stressed that the research community is developing Small Language Models (SLMs) that require fewer trainable parameters and aim to achieve the same performances of LLMs[15], or alternatively the development of new models by fine-tuning existing ones using the Low-Rank Adaptation (LoRA) technique[16]; both SLMs and LoRA require less data and less computing power compared to LLMs. The data used by LLMs must be vast, diverse, rapidly gathered and updated (freshness), and of quality. The performance of models increases as a function of the volume, the variety, and the quality of the data used for training. In this respect, it is worth stressing that researchers have developed a model called KOALA that, using high quality data but less volume of them (it required only six hours of training), performed similarly to LLMs using much larger volumes of data[17]. Computing resources include processing hardware, servers, supercomputers, and networking equipment. Computing power can be rented from cloud providers, which typically works well for finetuning and for smaller models but becomes too expensive for creating large-scale foundation models from scratch. LLMs need vast amounts of computer power to train and run. In particular, the models need Graphics Processing units (GPUs) to perform several computations simultaneously. Intel detains a 68.7% share of the market for CPUs, while Nvidia command 82% share of the GPUs market[18]. So, it can be said that the markets providing computing resources for LLMs are quite concentrated. On the other hand, for fine-tuning and small models, cloud providers rent developers processing hardware, storage, servers, and supercomputing technologies. Large providers of these services include Google (Google Cloud Platform), Amazon (Amazon Web Services) and Microsoft (Microsoft Azure). It is worth noting that some of these providers are also involved in the development of LLMs, either by developing proprietary models or investing in companies that create these models such as Mistral (i.e., Microsoft, see infra) and Anthropic and, thus, have a competitive advantage and a business interest in models that perform at scale. The cloud providers market, however, is still competitive and not concentrated[19].

LLMs exhibit two aspects suggesting a potentially swift and widespread impact on work. First, surprising capabilities, since LLMs, originally designed as general models, manifest specialist knowledge and abilities that evolve as model size and quality improve[20]. GAI showcases novel and unforeseen effective capabilities with wide applicability. Recent studies highlight their high-level performance in professional contexts[21]. Second, direct performance enhancement, for LLMs have the unique ability to directly enhance the performance of users without requiring substantial organisational or technological investments. Early research on the latest generation of LLMs indicates tangible performance boosts, particularly in writing tasks[22] and programming[23], as well as in ideation and creative work.

LLMs, however, in their current state of development are not without flaws, characterized by their relative opacity, encompassing not only the inner workings of AI models but also their failure points. There is ample evidence on the errors contained in the outputs of GAI models because of limitations in the underlying LLMs[24], generally referred to as 'hallucinations'. According to Mons[25], such hallucinations are to a large extent due to poor data inputs and to 'a lack of validated conceptual models to constrain the LLM's algorithms and output'[26]. Mons argues that one should distinguish three category of inputs models are fed with: experimental data, real-world observations, and established knowledge, but in reality 'machines' are fed as if all inputs were established knowledge. He concludes this line of reasoning stating that 'Without a conceptual model constraining the analysis, we face the likelihood of many hallucinations, here defined as patterns that make perfect sense to a machine without any constraints based on a conceptual model, but that do not contain any actionable knowledge for people"[27]. This aspect and other shortcomings are considered in Section 4.1 of Chapter 4.

GENERATIVE AI **SOCIOECONOMIC DYNAMICS**

MARKET DEVELOPMENTS

The development of GAI surged after the launch of ChatGPT by OpenAI in November 2022. The impact has been profound, leading to valuable breakthroughs and transformative advancements in GAI applications. The release cycle, number of start-ups, and rapid integration into existing software applications are remarkable. For instance, only four months after the release of ChatGPT, OpenAI introduced GPT-4, based on a new LLM, exhibiting significantly enhanced capabilities. Soon after, Anthropic's GAI, Claude, demonstrated an important surge in processing power, handling 100,000 tokens of text, equivalent to approximately 75,000 words per minute by May 2023, compared to its introduction in March 2023 with a capacity of roughly 9,000 tokens.

Google, announced in May 2023 a suite of new features driven by GAI, including the Search Generative Experience and an LLM named PaLM 2, poised to power its Bard chatbot and other Google products. In December 2023, Google announced Gemini, which is the most advanced set of LLMs at Google. The new tool was trained to be multimodal, meaning it can process different types of media such as text, pictures, video, and audio. Google said Gemini Ultra outperformed "state-of-the-art" AI models including ChatGPT's most powerful model, GPT-4, on 30 out of 32 benchmark tests including in reasoning and image understanding. Other leading tech companies have released their own versions of LLM, like Meta's Llama, which shows that the landscape of Language Model development has entered an era of intense competition.

It is obvious that Big Tech is well positioned to potentially scale to dominance even in the new domain of GAI, for a number of reasons. They have access to a lot of data, they possess computing power (and rent it as cloud services, potentially generating dependencies for smaller players), they have a lot of money providing access to top talents.

The GAI ecosystem is divided into three main layers: Applications, Models, and Infrastructure vendors[28] (see figure below)

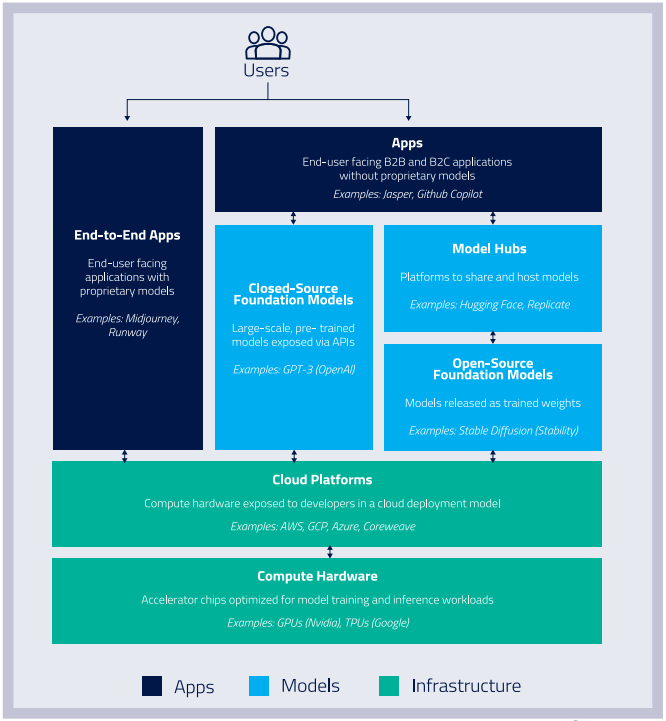

Figure 1 Generative AI ecosystem, Source: Bornstein et al. (2023). Who Owns the Generative AI Platform? (https://a16z.com/who-owns-the-generative-ai-platform/)

To analyse the market developments, it is important to have this framework in mind. First, model providers play a pivotal role in shaping GAI advancements, simply because the latter would not exist without the research and engineering work done by tech companies like Google, OpenAI, Amazon, Meta, X.AI, Stability and other players, like the Chinese Tencent and Alibaba, but also by the academic community. The initial development of models requires significant investment due to the extensive computational resources essential for training and the human effort involved in refining them. Consequently, this initial development phase was spearheaded predominantly by major tech giants, start-ups backed by substantial investments, and select open-source research collectives (e.g., BigScience). However, ongoing efforts are directed towards creating smaller models that can yield effective results for specific tasks and optimizing training processes for efficiency, which opens up the field also for start-ups. This progress, in fact, holds the potential to broaden market entry, inviting new participants. Notably, certain start-ups have already achieved success in independently developing their own models; examples include Cohere, Anthropic, ElevenLabs, and AI21 Labs, which specialize in constructing and training their LLMs. There are also European players (i.e., Aleph Alpha, Poro, etc.) that are considered later in Section 5.1.

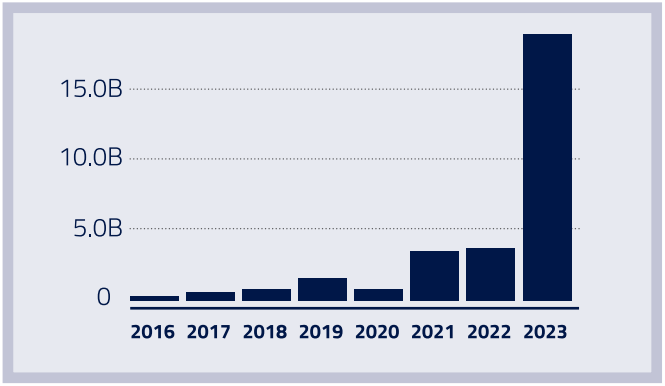

Figure 2 Global Generative AI investments, Source: Dealroom.co

A key observation for model providers thus far is the evident connection between commercialisation and hosting. The demand for proprietary APIs, exemplified by OpenAI's offerings, is experiencing rapid growth. Concurrently, hosting services for open-source models are emerging as valuable hubs facilitating seamless sharing and integration of models. While models function as the cognitive core of GAI, a comprehensive value chain is evolving to support the training and utilization of this transformative technology. Infrastructure vendors play a pivotal role in providing the substantial computational power necessary for training these models. Cloud platforms, in turn, offer accessibility to this advanced hardware infrastructure. Dominated by major players like AWS, GCP, and Azure, with emerging challengers, this layer underpins the financial dynamics of Generative AI. Estimates suggest that a significant portion of Generative AI revenue, ranging from 10% to 20%, funnels through infrastructure companies. Start-ups, backed by billions in venture capital, contribute to this financial flow, with the majority spent on cloud providers. Notably, Nvidia stands out as a key player, reporting substantial data centre GPU revenue primarily attributed to Generative AI use cases. However, credible competition is emerging, with companies like Oracle and start-ups like Coreweave and Lambda Labs gaining traction through cost-effective solutions and personalized support. Lastly, applications integrate Generative AI models seamlessly into user-centric products. Whether orchestrating end-to-end pipelines or leveraging external APIs, these applications span diverse domains like image generation, copywriting, and code creation. Notably, numerous companies are entering the market, presenting applications built upon foundation models tailored to execute specific tasks.

Together with the rapid evolution of GAI capabilities, funding for this transformative technology has become a significant force in the investment landscape. GAI garnered a substantial and swiftly growing investment of $17.8 billion in the first ten months of 2023[29]. Venture capital and other private external investments

in GAI witnessed a remarkable average compound growth rate of 74% annually from 2017 to 2022, outpacing the 29% annual rise in overall artificial intelligence investments during the same period.

Looking at the key segments of GAI, model developers have emerged as the predominant beneficiaries in terms of investments, securing more than 70% of the total funding between 2019 and 2023. Then, applications received about 24% of the total investments, highlighting the pivotal role of software and practical implementations. Meanwhile, infrastructure, essential for supporting the backbone of these advanced models, accounts for 5% of the overall investment, emphasizing the recognition of the critical role played by technological foundations in advancing the capabilities of GAI. This distribution underscores the diverse areas of focus within the ecosystem, each contributing to the continued growth and development of this transformative technology.

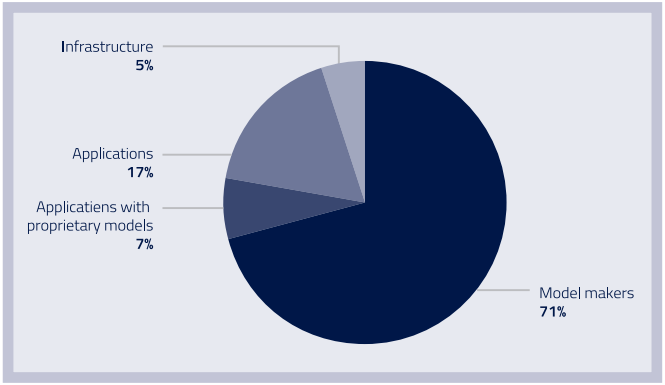

Figure 3 Generative AI VC funding (2019-2023), by segment, Source: Dealroom.co

The substantial increase in investments in GAI throughout 2023 signifies more than just enthusiasm; it reflects a deep-seated belief among investors in the tangible implications of this technology. The heightened financial support is indicative of a shared expectation that Generative AI will fundamentally alter how we approach work and enhance overall productivity. Investors are allocating significant funds with the anticipation of witnessing concrete and meaningful outcomes from GAI. Looking at 2024, the Economist expects a notable shift in focus towards companies outside the technology sector as they will embrace GAI[30]. Their primary objectives will centre on achieving cost reduction and heightened productivity. There are several compelling factors supporting this expectation, representing a crucial turning point in the widespread adoption of AI within the enterprise landscape. While in 2023 major companies explored GAI, experimenting with applications ranging from drafting documents to analysing transcripts, firms are preparing for broader deployment in 2024. A KPMG survey revealed that more than 80% of US firms plan to increase their GAI investment by over 50% by mid-2024[31].

In exploring the revenue landscape of the AI ecosystem, we can observe a different pattern. Following the 2022 introduction of ChatGPT, OpenAI experienced a remarkable one-year surge, achieving an unprecedented revenue milestone going from zero to $1 billion, second only to the impressive performance of U.S.-based chipmaker NVIDIA[32]. NVIDIA's data centre GPU sales went from $3.6 billion in Q4 2022 to a projected $16 billion in Q4 2023, solidifying its position as a leader in the Generative AI space. In terms of revenues, the data infrastructure market segment is leading. The market showed extraordinary growth, reaching $49 billion in 2023, a substantial 182% increase from the previous year. NVIDIA emerged as the dominant player, commanding an impressive 92% market share. Despite challenges such as price hikes and supply constraints, the demand for data centre GPUs remains robust. NVIDIA's quarterly revenue in 2023 saw a staggering 272% increase, surging from $4.3 billion in Q1 to a projected $16 billion in Q4. The second market segment in terms of revenues is composed of foundational model companies. This market underwent a transformative shift with the public release of ChatGPT in late 2022, reaching revenues of $3.0 billion in 2023. Projections indicate a promising future, with enterprises investing heavily in Generative AI implementations and applications. OpenAI

dominates this market with a 39% share, closely followed by Microsoft at 30%. Microsoft's Azure AI, leveraging OpenAI's LLMs, distinguishes itself by emphasizing enhanced data security and customizable AI applications. AWS, with an 8% market share, introduced the Bedrock service in September 2023, providing access to models from various AI companies, including Anthropic, AI21 labs, and Cohere, each holding a 2% market share. Bedrock combines these models with developer toolsets, facilitating the construction and scalability of Generative AI applications. The foundation models and platforms market continue to evolve, with significant growth expected in the coming years.

COMPETITION ISSUES

Although most of the efforts by governments are on general regulation of AI with some specific reference to GAI such as in the EU AI Act, initiative, the AI and GAI governance framework includes also the leg of competition policy and anti-trust. While GAI has just emerged as a new market segment in the last two years, the fast pace of development has already raised concerns about ensuring competition and a level playing field. Big Tech, in fact, possess some advantages that may lead them to dominate the new GAI market. As recently as the end of February 2024, for instance, Microsoft deal with French tech start-up Mistral AI has provoked outcry in the European Union, with lawmakers demanding an investigation into what they see as a concentration of power by the tech giant[33]. Antitrust authorities are already looking at Microsoft's partnership with ChatGPT-maker OpenAI, with the European Commission earlier warning the companies' relationship could be in breach of EU competition rules. When facing regulatory pressure over its multi-billion-dollar investment in OpenAI late last year, Microsoft sought to clarify that it did not in fact own a stake in the company, and therefore could not control it. Yet, Microsoft has gained access to OpenAI's cutting edge models and rights to share OpenAI's profits in return for the investment, which is under antitrust scrutiny in both EU and the U.S. So, the issues of competition and preserving a level playing field are high on the political agenda.

In the United Kingdom, the Competition and Markets Authority (CMA) has launched a review of AI Foundation Models (FMs) to understand how the market works[34]. In particular, the CMA examines FMs entry barriers and their impact on competition in other markets. But it is especially in the US where consumer protection agencies and the Federal Trade Commission (FTC), in particular, have expressed serious concerns about the risk of competition[35]. FTC Chair Lina Khan recently warned on national news that "AI could be used to turbocharge fraud and scams" and the FTC is watching to ensure large companies do not use AI to "squash competition"[36]. The FTC has produced several analyses, and below we report the main contents of a blog titled Generative AI Raises Competition Concerns[37].

According to the FTC's blog there are three main issues that may reduce competition, matched by the possibility of players adopting unfair competition methods. First, control over data may create entry barriers to the advantage of big players who also own digital platforms and have already cumulated large amounts of data. So, according to the FTC, the volume and quality of data required to pre-train a GAI model from scratch may impact the ability of new players to enter the market. Second, a fundamental input to build GAI model is talent. Since such kind of talent is scarce, large and rich companies have an advantage and may end up excluding new entrants from accessing the needed talent. Third, access to the required computational power might also become an entry barrier in favour of large companies rich in financial resources and/or owning their own computational power and cloud service offering. Whereas the FTC recognises that small new entrants can simply rent computational power from cloud services, it also stresses that such services are expensive and may become more expensive in the future and also create lock-ins, which brings us to what we consider the most relevant element in the FTC's analysis, namely the risk of possible unfair methods of competition. Big players may engage in bundling and/or tying. Bundling occurs when businesses offer multiple products in one single package. So, the FTC envisages a number of risks that are worth reporting in full below:

> Incumbents that control key inputs or adjacent markets, including the cloud computing market, may be able to use unfair methods of competition to entrench their current power or use that power to gain control over a new generative AI market.
> […]
> Incumbents may be able to link together new generative AI applications with existing core products to reduce the value of competitors' standalone generative AI offerings, potentially distorting competition. Incumbents that offer a range of products and services as part of an ecosystem may also engage in exclusive dealing or discriminatory behaviour, funnelling users toward their own generative AI products instead of their competitors' products. Further, incumbents that offer both compute services and generative AI products—through exclusive cloud partnerships, for instance—might use their power in the computing services sector to stifle competition in generative AI by giving discriminatory treatment to themselves and their partners over new entrants. A related scenario exists where an incumbent offers both their own products leveraging generative AI as well as offering APIs allowing other companies to leverage their generative AI capabilities. In such circumstances, there is a risk that incumbent firms will offer their APIs on terms which exist to protect their incumbent position. Incumbent firms could also use M&A activity in the generative AI space to consolidate market power in the hands of a few players. Large firms already active in generative AI—or that already control a critical input—may try to buy up critical applications and cut off rival access to core products. Market leaders may also try to buy up complementary applications and bundle them together. Additionally, incumbents may be tempted to simply buy up nascent rivals instead of trying to out-compete them by offering better products or services.[38]

A somehow less pessimistic picture can be found in a recent Bruegel report[39], according to which the markets for GAI models are currently still competitive with many providers and a fair degree of openness. The entry barriers are not yet insurmountable, although also this report at the end stresses the same kind of risk of unfair methods of competition illustrated by the FTC. According to this report, the markets for LLMs are competitive and dynamic with steady releases of new models of different dimensions and with various degrees of openness from a large variety of developers. Two factors may reduce the need of financial resources to train and run the models, namely first the development of SLMs (Small Language Models) and of the fine-tuning techniques, and second the increased availability of a growing community of developers and open-access research that may reduce the human resources required. At the level of access to data some players (i.e., Google) do benefit from having proprietary data to train and run the models. Yet, developers can use the increasing number of open-source datasets available on the web[40]. Furthermore, in Europe developers seeking access to data can invoke the Digital Market Act [DMA, Regulation (EU) 2022/1925, Article 6(11)], which imposes on players identified as 'gatekeepers' the obligation to share search data (ranking, query, click and view data). So, according to the Bruegel's report, the entry barriers at the level of access to data are not currently prohibitive. At computing power level, it is recognized that players who both offer computing power/cloud services and develop their own model have an advantage. Large cloud providers might enjoy a competitive advantage due to economies of scale, access to large financial resources, and the provision of their own models exclusively on their cloud services. The cloud service market, however, is still competitive and offers small developers the possibility to rent the needed computing power in a scalable way for finetuning and working with smaller models, without the need of large upfront financial investments. Also in this area, the development of SLMs and fine-tuning techniques will lower the need for computing power. Hence, the Bruegel report concludes that even in this case entry barriers are low.

Finally, there is the issue of open-source that may have ambivalent and uncertain consequences. Very rapidly the capability of open-source image generation emerged, eclipsing those of the

proprietary base models that inspired them. In principle, open-source could open the playing field, once base models become public. There is, however, also the risk that open-source is misused. Imagine an open-source AI image generation tool released with built-in restrictions on the types of images that could be generated. It is possible that malicious users remove these protections and use the models to create non-consensual intimate images. Furthermore, we have seen in the past the practice of 'open first, closed later'. Anticipating here the discussion on the EU AI Act, this ambivalence of open source explains why the European Parliament was not fully satisfied with the exemption granted to open-source models that is contained in the latest compromise on the AI Act agreed in December 2023 (see infra, Section 4.3).

ECONOMIC IMPACTS

Earlier generations of automation technology excelled at automating data management tasks linked to the collection and processing of information. GAI, with its natural-language capabilities, does enhance the automation potential for these cognitive activities. The most significant influence of GAI is anticipated in knowledge work, especially in activities requiring decision-making and collaboration, which historically had lower automation potential. The steep rise in automation potential is largely attributed to GAI's proficiency in understanding and utilising natural language across various activities and tasks. This surge in automation potential implies that numerous work activities involving communication, supervision, documentation, and general interaction with people have the potential to be automated by GAI.

McKinsey conducted an in-depth analysis of GAI's potential impact on the global economy by scrutinizing the work activities essential in approximately 850 occupations[41]. The analysis involved modelling scenarios to estimate when GAI could proficiently execute over 2,100 specific work activities, ranging from 'communicating with others about operational plans or activities' to other nuanced tasks within these occupations across the global

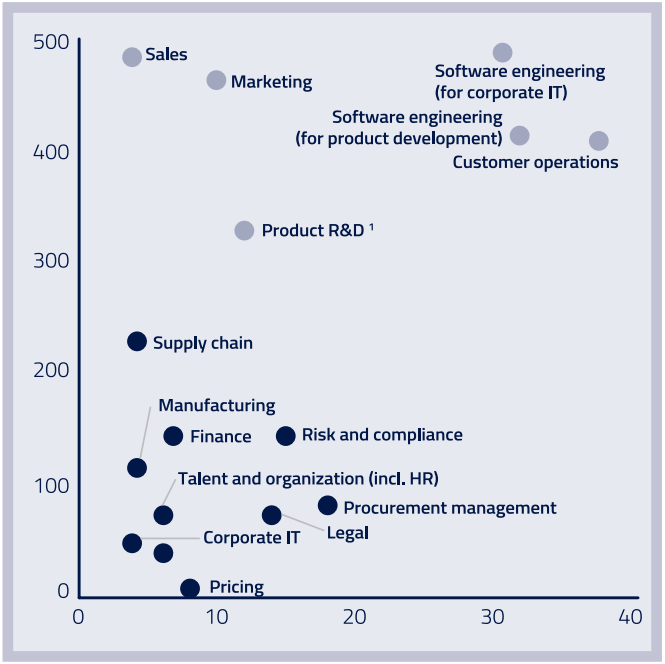

Figure 4 Impact of Generative AI by business function, Source: Chui et al (2023) see footnote 39.

economy. Through this approach, McKinsey gauged the potential influence of current GAI capabilities on labour productivity across the entire spectrum of work undertaken by the global workforce. The outcome reveals substantial economic benefits, estimating an annual increase of productivity valued in the range of $6.1 trillion to $7.9 trillion. This represents a notable 35% to 70% increase in productivity, underscoring the transformative potential of GAI on a global scale. According to McKinsey's economic analysis, of 16 business functions that will be impacted by GAI, four (customer operations, marketing and sales, software engineering, and research and development) could account for approximately 75% of the total annual value from Generative AI use cases. Therefore,

using GAI in these functions could drive most of the economic impact of this technology.

A recent publication looking at the labour market impact potential of LLMs models[42] reveals that around 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of LLMs, while approximately 19% of workers may see at least 50% of their tasks impacted. This study analysed more than 1,000 occupations and 19,000 tasks, and the authors observe that most occupations exhibit some degree of exposure to LLMs, with higher-wage occupations generally presenting more tasks with high exposure. In terms of skills, in line with the Figure above from the cited McKinsey report, the study findings indicate that occupations requiring critical thinking skills and science-related occupations are less likely to be impacted by current LLMs. Conversely, occupations involving programming and writing skills are more susceptible to being influenced by LLMs. Recent experimental evidence on the impact of GAI on productivity is also revealing about the future impact of this technology. As already mentioned, software engineers seem to be the workers who may be mostly affected by the deployment of Generative AI. Software engineers can use Generative AI in pair programming and to do augmented coding and train LLMs to develop applications that generate code when given a natural-language prompt describing what that code should do. An experimental study used a controlled trial of GitHub Copilot, an AI pair programmer that suggests code and entire functions in real-time based on context. The study results show that the group using Generative AI completed the task 55.8% faster compared to the group without access to it[43]. Another compelling experimental study investigates the impact of GAI on management consulting, a traditional knowledge-intensive sector.

Using data on 758 consultants from Boston Consulting Group, a global management consulting firm, the experiment delves into the performance implications of AI on tasks that are realistic, intricate,

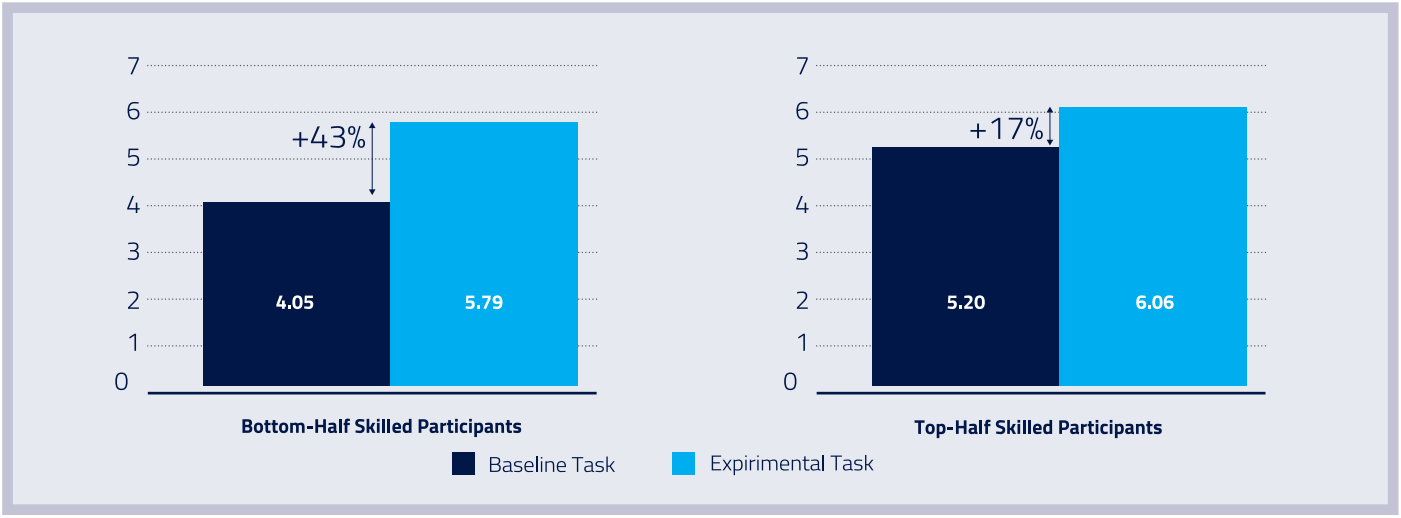

Figure 5 Impact of Generative AI on low-skilled and high-skilled knowledge workers, Source: Dell'Acqua et al., (2023) see footnote 45.

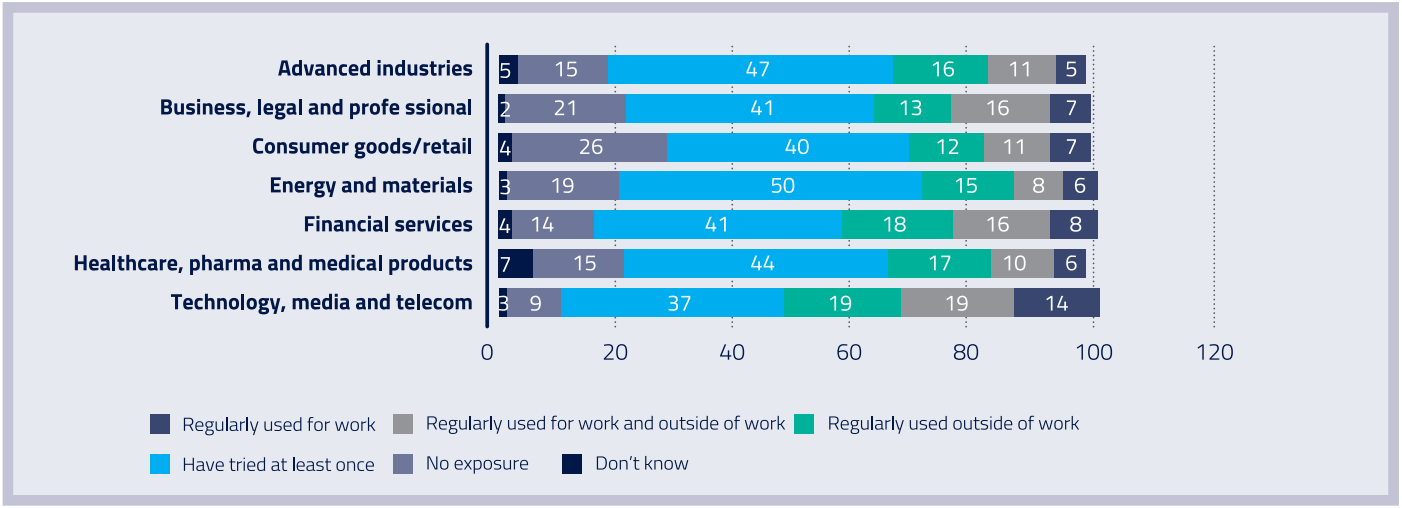

Figure 6 Share of respondents using Generative AI in 2023, by industry, Source: McKinsey & Company, see footnote 45

and knowledge-intensive[44]. The study shows that for each realistic consulting task within the frontier of AI capabilities, consultants using AI were significantly more productive (they completed 12.2% more tasks on average, and completed tasks 25.1% more quickly), and produced significantly higher quality results (more than 40% higher quality compared to a control group). However, for a task selected to be outside the frontier of AI capabilities consultants using AI were 19% less likely to produce correct solutions compared to those without AI. Moreover, as shown in the Figure below, the study reveals that low-productive and novel workers tend to benefit more from Generative AI compared to experienced and highly skilled workers. This finding is confirmed by another experimental study, which focuses on another sector, and analysed the impact of a Generative AI-based conversational assistant using data from 5,179 customer support agents[45]. While the study shows that AI assistance increases worker productivity overall, resulting in a 14% increase in the number of chats that an agent successfully resolves per hour, the productivity impacts of AI assistance are highly uneven. Less-skilled and less-experienced workers improve significantly across all productivity measures. In contrast, the study finds minimal impacts on the productivity of more experienced or more skilled workers.

GAI could impact a broad number of functions across different industries. Its capacity to autonomously generate content, solve complex problems, and mimic human creativity has the potential to transform any sector. However, certain industries can experience more profound shifts, as the technology's versatility promises to unlock new dimensions of efficiency, creativity, and problem-solving. As shown in a recent study, the overall impact of LLMs on firms' values is higher in those sectors with occupations that are mostly exposed to GAI[46]. It seems, however, that the impact is widespread and still difficult to predict. For instance, the latest annual McKinsey Global Survey on the state of AI shows that most people in nearly all industries around the world have tried a GAI tool at least once in 2023[47]. In these early days, expectations for Generative AI's impact are high: three-quarters of all respondents expect Generative AI to cause significant or disruptive change in

their industry's competition in the next three years. The Figure above shows the share of respondents using GAI in 2023 by industry. In general, those in technology, media, and telecom were far likelier to have used GAI tools regularly. This is perhaps unsurprising as these fields are the most likely to involve the use of GAI in business practices.

As previously highlighted, the disruption brought about by GAI is not uniform across all industries. Industries with knowledge-intensive activities are likely to witness more transformations, potentially creating substantial benefits. In contrast, industries rooted in manufacturing, such as aerospace, automotives, and advanced electronics, might experience comparatively less disruption. This departure from the historical pattern of technology waves predominantly impacting manufacturing arises from Generative AI's proficiency in language-based activities rather than those reliant on physical labour. Below we will focus, for exemplificative purpose only, on two specific sectors, for which we discuss the benefits and the challenges brought by GAI.

In the remaining part of this section we focus on two specific sectors, delving into the nuanced ways in which this technology can impact sectors with varying degrees of intensity. We start first looking at the healthcare sector, and then we move to the energy sector.

Healthcare

The advent of GAI signals a potential breakthrough for healthcare operations, particularly in dealing with the abundance of unstructured data present in clinical notes, diagnostic images, medical charts, and recordings. These unstructured datasets, whether utilised independently or in conjunction with extensive structured ones like insurance claims, present a promising frontier for transformative applications in healthcare. Like its role in clinician documentation, various use cases for GAI in healthcare are emerging, generating excitement but also concerns among

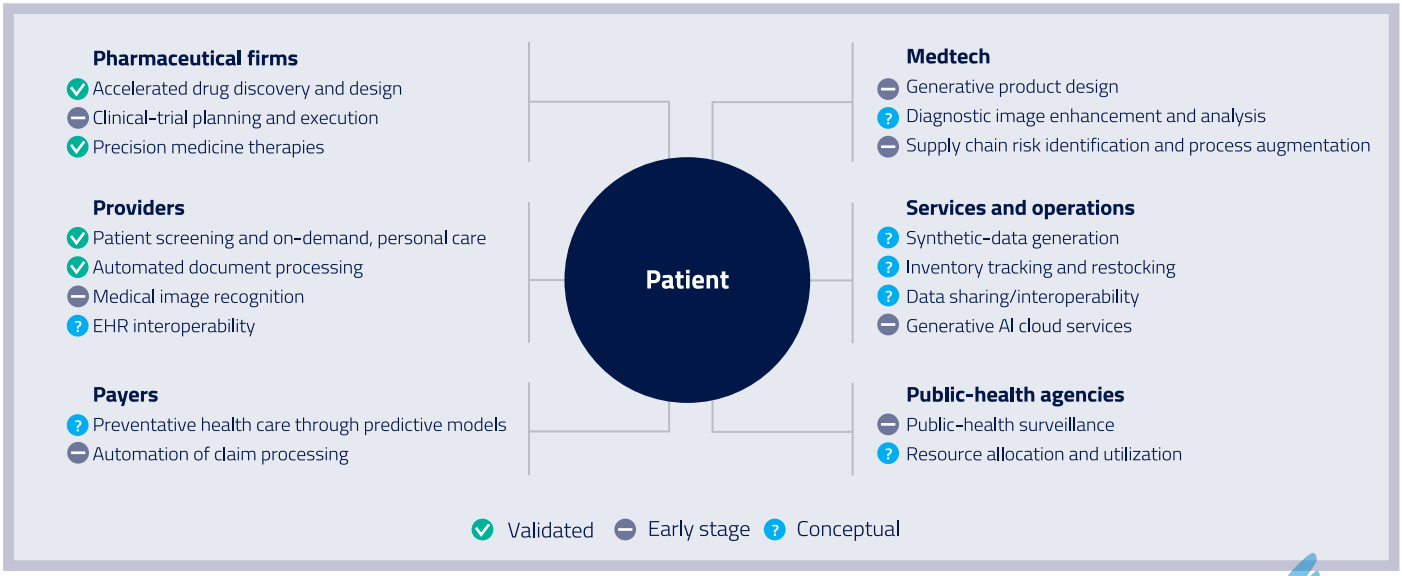

Figure 7 Potential use case across all healthcare segments, Source: Boston Consulting Group

healthcare professionals. In healthcare, safeguarding patient information is paramount due to its sensitivity. Given the occasional inaccuracies in GAI responses, close facilitation and monitoring by healthcare practitioners are essential to mitigate risks and ensure reliable outcomes. While AI technology has been employed in healthcare for years, with applications such as adverse-event prediction and operating-room scheduling optimisation, GAI can transform new operations unlocking an important industry potential value[48]. As shown in the Figure below extracted from an analysis carried out by Boston Consulting Group, GAI generates numerous use cases, and new potential ones will emerge over the next few years.

In this section, we focus on the potential applications in two specific areas: clinical operations and clinical decision-making. Later we also discuss some of the main risks associated with the introduction of this technology in the healthcare sector.

1. Clinical operations. Clinical operations present a promising arena for the potential efficiencies that GAI could introduce. Currently, hospital providers and administrative staff grapple with the extensive paperwork associated with patient care, including numerous forms, post-visit notes, and administrative tasks, contributing to significant time-consuming tasks and potential burnout. This administrative burden is equally felt by physician groups. GAI, under clinician oversight, holds the promise of generating discharge summaries or instructions in a patient's native language for better comprehension. Additionally, it could synthesise care coordination notes, shift-hand-off notes, and real-time checklists, as well as lab summaries from physician rounds and clinical orders. The technology's capacity to generate and synthesise language also presents an opportunity to enhance Electronic Health Records (EHRs), substituting manual input requirements and minimising the risk of human error. Hospitals and physician groups are actively exploring Generative AI applications, ranging from prepopulating visit summaries in the EHR to refining documentation and providing research for decision support, with some health systems already incorporating this technology into pilot programs. The box below shows an example of the application of Generative AI in clinical operations.

GAI was one of the key topics discussed at the Healthcare Information and Management Systems Society (HIMSS) Annual Conference and Exhibition[49], which is one of the largest and most influential healthcare IT events in the world. One breakout session focused on how GAI has the potential to change the way medicine is practiced. With use cases from data entry and processing to clinical decision support and patient engagement, GAI offers many opportunities to create a more efficient, effective, and sustainable healthcare system, especially when paired with automation. Demonstrating its capabilities, the technology showcased how a healthcare clinician could seamlessly leverage new platforms to transform patient interaction into structured clinician notes within seconds. Using the AI platform's mobile app, a clinician records a patient visit. In real-time, the platform incorporates the patient's information, intelligently identifying any missing details and prompting the clinician to fill them in. This transforms the initial dictation into a meticulously structured note with a natural conversational flow. Post-visit, the clinician accesses the AI-generated notes on a computer, where they can effortlessly edit the content through voice commands or typing. The finalized notes are then swiftly submitted to the patient's electronic health record (EHR). This near-instantaneous process renders the manual and time-consuming notetaking, a once obligatory task for every patient interaction, seemingly archaic in comparison.

Box 1 Example of practical application of Generative AI in healthcare

2. Clinical decision-making. The impact of LLMs on clinical decision-making is also expected to be significant. The technology can provide a sophisticated interface to organise, retrieve, and synthesize intricate medical facts, notes, and research presenting a paradigm shift in the way physicians navigate and interpret complex healthcare data. Leveraging the capabilities of Generative

AI, physicians can potentially streamline their interactions with vast amounts of drug information. Through simple queries, physicians can leverage the power of LLMs to access comprehensive and up-to-date drug-related data, enhancing their ability to make informed decisions about patient care. Furthermore, Generative AI opens up possibilities for more personalised patient care. By empowering physicians with LLMs, tailored insights can be generated, enabling a deeper understanding of individual patient needs and treatment options. While the integration of Gen-AI in clinical decision-making represents a departure from traditional workflows, the potential benefits in terms of improved information access, enhanced decision support, and personalized patient care underscore the transformative impact of this technology in shaping the future of healthcare.

3. Risk. While the adoption of GAI in healthcare brings several opportunities, there are potential risks to mitigate, especially as the technology continues to evolve. Some crucial considerations should be discussed. First, there are data security concerns, as the level of security provided by open-source Gen-AI tools may fall short of the stringent requirements demanded by the healthcare industry. The risk of data breaches poses a significant threat, potentially compromising sensitive patient information. Second, the reliance on data sets that over-index certain patient populations pose a substantial risk in the generation of patient care plans. If the models are trained on biased data, the generated care plans may also be wrong, providing inaccurate, unhelpful, or even harmful information to patients. Ensuring diverse and representative training data becomes imperative to mitigate this risk. Third, given the potential for GAI to produce incorrect but plausible answers, also referred to as "hallucinations" or "confabulations" in the context of Generative AI, maintaining a human in the loop remains critical. The technology, while powerful, is not infallible, and human oversight is essential to validate and contextualize AI-generated insights, especially in critical healthcare decision-making scenarios. GAI has promising applications in health care, but potential ethical issues need to be addressed proactively to prevent harm[50]. Clear rules and legal boundaries are needed to properly allocate liability and protect users. Humanistic ethics concerns arise from the potential disruption of the physician-patient relationship, humanistic care, and issues of integrity. Overreliance on artificial intelligence (AI) can undermine compassion and erode trust.

In conclusion, the introduction of GAI in healthcare signals a transformative breakthrough, particularly in managing unstructured data. GAI's applications in clinical operations and decision-making show promise for improving efficiency and personalized patient care. However, concerns about data security, biases in training data, and ethical considerations underscore the need for careful implementation. Striking a balance between innovation and ethical standards is crucial for realizing the full potential of GAI in shaping the future of healthcare.

Energy

The energy industry is confronted with a trilemma of challenges, demanding solutions that ensure energy security, affordability, and a transition to cleaner energy sources. In response, the International Energy Agency (IEA) highlights the transformative potential of digitalization, particularly through GAI, as a catalyst for cost reduction, efficiency improvement, and emissions reduction across the energy sector. As the industry grapples with this digital transformation, the complexities of power systems are increasing exponentially. Power systems are undergoing significant transformations, driven by the growing demand for electricity and intensified efforts towards decarbonization. Unlike traditional grid structures that directed energy from centralized power stations, modern power systems must now accommodate multi-directional flows of electricity between distributed generators, the grid, and end-users. The proliferation of grid-connected devices, ranging from electric vehicle charging stations to residential solar installations, adds a layer of unpredictability to energy flows. Additionally, the deepening integration between the power system and various sectors, including transportation,

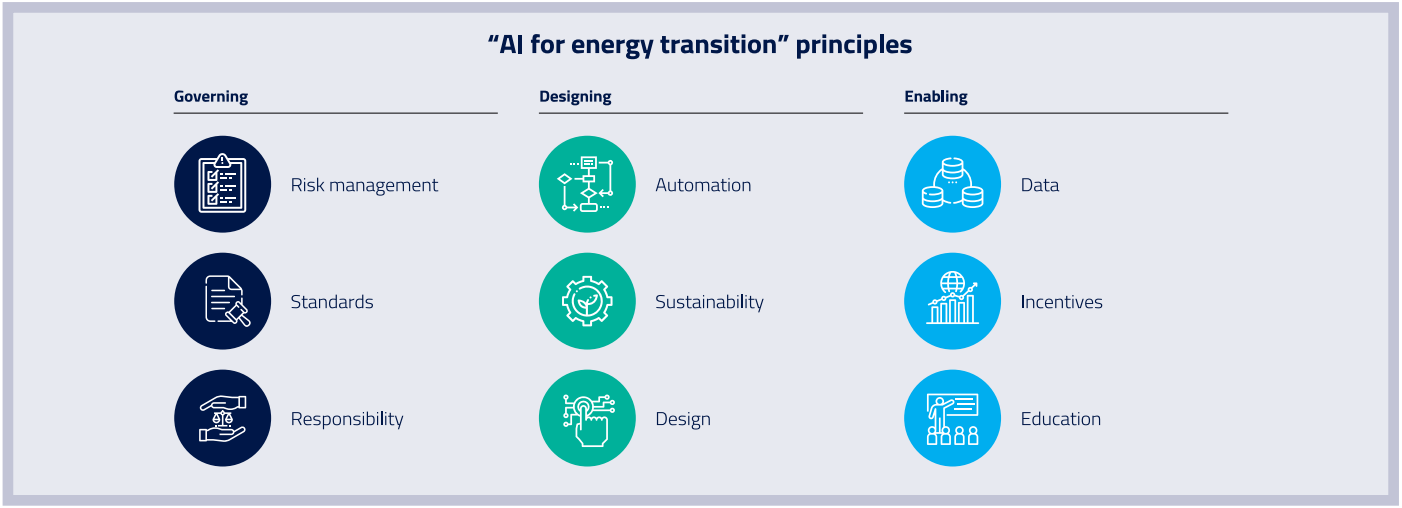

Figure 8 Generative AI for the energy transition, Source: World Economic Forum

industry, and buildings, necessitates a substantial increase in information exchange and more sophisticated tools to plan and operate evolving power systems.

This complexity underscores the vital role of GAI in boosting efficiency and innovation within the energy sector. Power systems are becoming more data-intensive, with smart grids producing vast amounts of data. Smart meters, for example, generate several thousand times more data points than their analogue counterparts, while new devices for monitoring grid power flows significantly surpass the data output of their predecessors. The global fleet of wind turbines alone is estimated to produce over 400 billion data points annually. Energy firms recognize GAI as a critical resource, already serving over 50 different purposes within the energy system. The market for GAI technology in the sector is anticipated to reach up to USD 13 billion, reflecting its increasing importance in addressing the challenges posed by the evolving energy landscape[51].

AI holds far greater potential to accelerate the global energy transition. The World Economic Forum has identified nine 'AI for the energy transition principles' aimed at the energy industry, technology developers and policymakers. If adopted, these principles would accelerate the uptake of AI solutions that serve the energy transition by creating a common understanding of what is needed to unlock AI's potential and how to adopt AI safely and responsibly in the energy sector. Among the several applications that GAI can have in the energy transition process, below we provide two concrete examples to understand how GAI can impact the energy sector.

1. Energy optimisation. The first relevant application for GAI is energy optimization, which is a crucial process aimed at enhancing efficiency and curbing energy wastage across diverse sectors. Traditionally, technical solutions like equipment upgrades, sensor installations, and the integration of renewable sources have been employed for energy optimization. However, these approaches face challenges such as high operational costs, limited scalability,

compatibility issues, and regulatory hindrances in participating in energy markets. Gen AI emerges as a transformative force in addressing these challenges through its advanced capabilities. Through comprehensive data analysis, Gen AI combines information from various sources, including smart meters, weather forecasts, and user preferences. Leveraging its analytical insights, Gen AI offers actionable recommendations to reduce energy costs, enhance overall energy utilization efficiency, and implement strategies for a more sustainable energy footprint by lowering carbon emissions. Furthermore, Gen AI empowers a dynamic energy ecosystem by facilitating peer-to-peer energy trading and active participation in demand response programs, contributing to a more resilient and responsive energy landscape.

2. Smart Grids and Demand Response. The second application for GAI is the transformation and enhancement of smart grids. Led by pioneers in the field, this innovative approach focuses on creating AI-driven generative models for customer load data. The application of these generative models extends to grid modelling and training algorithms for energy tech startups. By training on existing data, these models generate realistic data, enabling stakeholders to explore "what-if" scenarios beyond the limitations of current datasets. For instance, the generated data can predict the potential load on the grid with the adoption of solar technologies by a specific number of households and how this load might fluctuate throughout the day. The integration of Generative AI models with smart grid technologies offers a virtual testing ground, allowing for the development and testing of hardware and software products for scalability and interoperability. This transformative use of Generative AI holds the promise of creating a more sustainable and resilient future for the energy sector.

We have only reported two examples, but the future of Generative AI in the energy industry looks promising. AI has the potential to transform the energy sector, making it more efficient, reliable, and sustainable. As the technology continues to evolve, we can expect to see more advanced AI algorithms, improved data analytics, and greater integration with other promising digital technologies such as neuromorphic hardware and quantum computing. AI has the potential to accelerate the transition to a low-carbon economy by enabling the development of new renewable energy technologies and improving the efficiency of existing ones.

In conclusion, the energy industry is facing challenges demanding solutions for security, affordability, and a shift to cleaner sources. The International Energy Agency recognizes GAI as a transformative catalyst for cost reduction, efficiency improvement, and emissions reduction. As power systems evolve in complexity, GAI becomes crucial for efficiency and innovation. The data-intensive nature of smart grids and the growing market for GAI in the energy sector highlight its significance. Two key applications underscore GAI's potential to optimize processes. Looking forward, GAI holds promise in enhancing efficiency, reliability, and sustainability, facilitating the transition to a low-carbon economy through advanced algorithms and integration with emerging technologies.

FROM **RISKS TO REGULATION** AND EU PERSPECTIVES

In this chapter we start discussing some of the risks of GAI that to some extent provide the rational for ongoing regulatory initiatives surveyed at Global level (Section 4.2) and discussed in more detail at EU level (Section 4.3).

RISKS OF GENERATIVE AI

GAI promises to produce transformations across various domains, fundamentally reshaping many aspects of our daily lives. However, the advent of this new technology also brings some potential risks. In the literature one can find discussions of a very wide range of risks going from Intellectual Property Infringements to Cybersecurity, and many more. Below we selectively focus only on a few risks, and particularly on those that have a clearer and potentially more sizeable socioeconomic and political effects.

While discussions surrounding AI risks are not novel, the rapid advancements in GAI require a fresh perspective from researchers and policymakers. The discussion around the risks associated to GAI became increasingly popular after the publication of a ground-breaking paper by AI researchers in March 2021 titled "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?"[52]. The paper resulted in two of the co-authors losing their jobs at Google, and a subsequent protest by Google employees[53]. According to the authors who coined the expression "Stochastic Parrots", LLMs are just repeating or assembling phrases based on probabilities and statistical patterns learned from vast datasets of text, without real understanding or awareness. The paper is the first attempt to consider the risks of very large language models, regarding their environmental and financial costs, their inscrutability leading to unknown dangerous biases, the inability of the models to understand the concepts underlying what they learn, and the potential for using them to deceive people. In the last two years, there have been very comprehensive analyses of the risks emerging from GAI[54]. Drawing on multidisciplinary literature from computer science, linguistics, and social sciences, we can identify and structure the risk landscape associated with this technology[55].

The first risk area pertains to potential societal discrimination. GAI models may introduce algorithmic bias due to imperfect training data or decisions made during model development. Specific hazards encompass this risk[56]. LLMs have the potential to perpetuate and propagate unfair discrimination, as biases present in the training data can be learned and replicated by the model. Often, language resources on which AI language models are trained reflect data about dominant social groups, which can raise diversity and inclusion concerns. These risks largely emanate from the selection of training datasets that incorporate biased language and disproportionately represent specific social identities. In some cases, training data reflect historical patterns of systemic injustice when they are gathered from contexts in which inequality is the status quo. In other cases, training data can be biased because some communities are better represented in the training data than others, generating content that fails to represent the language of those who are marginalised, excluded, or less often recorded. There are several concrete examples from the literature. For instance, gender and representation biases were found in fictional stories generated by GPT-3, where female-sounding names were more often associated with stories about family and appearance and described as less powerful than masculine characters[57]. Therefore, using balanced training datasets that

represent different groups more equally is essential to minimise biases of AI language models.

The second risk area centres on the dissemination of false or misleading information by LLMs. Language models and other GAI tools can enable the production of fake news, deepfakes, and other forms of manipulated content that may be impossible to distinguish from real ones. Even advanced large-scale LMs do not reliably predict true information, as these models emit detailed and correct information in some circumstances but then provide incorrect information in others. This can happen because LLMs predict the likelihood of different next words based on prior words. Yet, whether a sentence is likely does not reliably indicate whether the sentence is also factually correct. As a result, it is not surprising that LLMs frequently assign high likelihoods to false or nonsensical predictions. LLMs that often provide correct information may lead users to overly trust the predictions of the model, thus exacerbating risks from users relying on these models where they are unreliable or unsafe. Arguably, an LLM that gives factually correct predictions 99% of the time, may pose a greater hazard than one that gives correct predictions 50% of the time, as it is more likely that people would develop heavy reliance on the former. Therefore, the combination of AI and disinformation can lead to deception at a wide scale that traditional approaches like fact-checking, user education and media literacy, or detection tools would be challenged to address. This poses a significant threat to the fairness of democratic processes, as it can lead to the proliferation of disinformation, potentially influencing public opinion and decision-making.

As discussed in Section 3.1, the economic and organisational impact is a multifaceted risk, with Generative AI potentially affecting the workforce and disproportionately impacting specific groups and local communities negatively. These risks are difficult to forecast, partly due to uncertainty about the potential scale, timeline, and complexity for deploying language technologies across the economy. Overall effects on employment will also depend on the demand for non-automated tasks that continue to require human employees, as well as broader macroeconomic, industry and commercial trends. However, as already mentioned, LLM-based automation may impact job quality and undermine aspects of the creative economy.

Privacy is another important consideration for GAI. Privacy-related issues may arise if user input information becomes identifiable in model outputs, raising concerns about the inadvertent disclosure of sensitive data. Privacy violations may occur at the time of inference even without the individual's private data being present in the training dataset. Similar to other statistical models, GAI may make correct inferences about a person purely based on correlational data about other people, and without access to information that may be private about the particular individual. Predictions of sensitive data may require only minimal personal information, such as who a user "follows" on Twitter. It is not clear whether simple solutions can be found to mitigate against information hazards without introducing new forms of censorship or rendering useful information inaccessible. Yet, the potential threat to privacy deriving from GAI models can erode individual rights, especially if sensitive data about individuals affect the public domain and democratic processes, as for instance in the scenario where sensitive personal information are disseminated during electoral campaigns.

REGULATION: A SELECTIVE GLOBAL REVIEW

Over the past few years, several AI governance frameworks have been published around the world aimed at providing high-level guidance for safe and trustworthy AI development. A variety of multilateral organizations have published their principles, such as the OECD's "Principles on Artificial Intelligence" (2019), the EU's "Ethics Guidelines for Trustworthy AI" (2019), and UNESCO's "Recommendations on the Ethics of Artificial Intelligence" (2021). The development of Generative AI, however, has led to new guidance, including the OECD's recently published "G7 Hiroshima Process on Generative Artificial Intelligence" (2023). These

voluntary principles and frameworks often serve as guidance for regulators and policymakers around the world. As of 2023, more than 60 countries in the Americas, Africa, Asia, and Europe published national AI strategies[58].

While no country has passed comprehensive general AI or GAI specific regulation to date, leading legislative efforts include those in Brazil, China, the European Union, India, Singapore, South Korea, and the United States. The approaches taken by the different countries vary from broad AI regulation supported by existing data protection and cybersecurity regulations (the European Union and South Korea) to sector- specific laws (the United States) and more general principles or guidelines-based approaches (Brazil, Singapore, and the United States). Each approach has its own benefits and drawbacks, and some markets will move from principles-based guidelines to strict legislation over time[59]. The exemplificative figure below provides a glance.

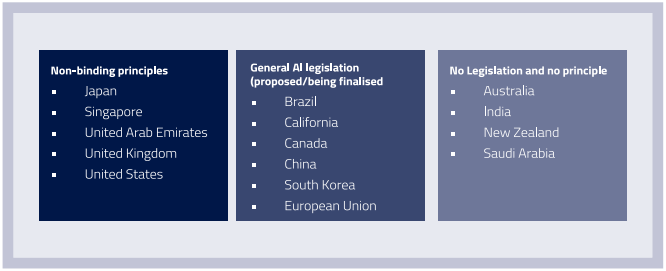

Figure 9 AI related regulation in the world, December 2023, selective, Source: our elaboration from Kremer et al (2023, see footnote XX); OECD.AI Policy observatory https://oecd.ai/en/ ; https://www.information-age.com/how-generative-ai-regulation-shaping-up-around-world-123503911/ ; https://www.insideprivacy.com/artificial-intelligence/from-washington-to-brussels-a-comparative-look-at-the-biden-administrations-executive-order-and-the-eus-ai-act/

While the approaches vary, common themes in the regulatory landscape have emerged globally:

- **Transparency.** The goal is the traceability and clarity of AI output and that users are informed when they engage with any AI system.
- **Human agency and oversight.** AI systems should be developed and used as tools that serve people, uphold human dignity and personal autonomy, and function in a way that can be appropriately controlled and overseen by humans.
- **Accountability.** Ensure awareness of responsibilities, accountability, and potential redress regarding AI systems.
- **Technical robustness and safety.** Minimize unintended and unexpected harm by ensuring that AI systems are robust, meaning they operate as expected, remain stable, and can rectify user errors.
- **Diversity, non-discrimination, and fairness.** Another goal is to ensure that AI systems are free of bias and that the output does not result in discrimination or unfair treatment of people.
- **Privacy and data governance.** Development and usage of AI systems should follow existing privacy and data protection rules while processing data that meet high standards in quality and integrity.
- **Social and environmental well-being.** AI should be sustainable, for people and the environment.

We review below, with different degree of detail, the initiative in some countries (or administrative entity, as in the case of California).

We start from the US and look at it in comparison with Europe[60]. The main difference between the EU and US approach is in the fact that, while the AI Act aims to introduce a new and systematic regulation of AI, the Executive Order signed by President Biden on 30 October 2023 (see more below) does not create new legislative obligations, and rather introduces several directions for government agencies. In the US, several guidance documents and voluntary frameworks have emerged in the past few years, such as the "AI Risk Management Framework" from the US National Institute of Standards and Technology (NIST), a voluntary guidance published in January 2023, and the White House's "Blueprint for an AI Bill of

Rights," a set of high-level principles published in October 2022. However, the two main proposed legislation pieces at the federal level are the "Algorithmic Accountability Act" and the "AI Disclosure Act," both of which are under discussion. The first one aims to create protections for individuals who are subject to algorithmic decision-making in areas like housing, credit, education and more. The legislation tasks the Federal Trade Commission to create regulations that provide companies operating these AI systems with concrete assessment instructions and procedures for ongoing evaluation. Moreover, the AI Disclosure Act would require including on any output generated by AI the following notice: "Disclaimer: this output has been generated by artificial intelligence." More recently, on 30 October 2023, President Biden issued the "Executive Order Safe, Secure, and Trustworthy Artificial Intelligence" (Biden EO)[61]. While differing from the EU AI Act approach, the Biden EO shares a focus on "dual-use foundation models," particularly those with high-performance capabilities in tasks posing security, economic, public health, or safety risks. These risks encompass areas such as cyber threats, chemical/biological/radiological/nuclear weapons, and deception/manipulation. The Biden EO introduces "red-teaming" measures for AI foundation model developers to assess potential risks and vulnerabilities, especially crucial for dual-use applications. Developers are required to report red-teaming results to the government, detailing how they address the most significant risks and vulnerabilities in their systems. The US position confirms that foundation models require a distinct approach from the other AI applications.

China's strategy towards AI, instead, is marked by a dual focus on fostering AI innovation and maintaining state oversight of the technology. This commitment is evident in key policy documents, such as the New Generation Artificial Intelligence Development Plan, outlining China's ambition to lead the global AI landscape by 2030. Contrary to the EU AI Act, which adopts a comprehensive, horizontal approach with adaptable standards across diverse AI applications, China employs a more targeted, vertical method using distinct laws to address specific AI challenges. Currently, China's AI regulations have concentrated on two distinct issues: AI-driven recommendation algorithms and deep synthesis tools, commonly employed in the creation of deepfakes. Regarding AI-driven recommendation algorithms, regulations mandate service providers in this sector to curb discrimination, mitigate the dissemination of negative information, and address exploitative work conditions for delivery workers. This legislation also empowers Chinese consumers with the right to disable algorithmic recommendations and receive explanations when algorithms significantly affect their interests. In the realm of deep synthesis, China's regulations encompass the use of algorithms to artificially generate or modify online content. This legislation necessitates that deep synthesis content adheres to information controls, is clearly labelled as synthetically generated, and mandates providers to implement measures preventing misuse. Additionally, it includes vaguely defined censorship requirements and mandates that deep synthesis providers register their algorithms. While China extensively employs AI in its law enforcement and surveillance apparatus, regulations have been introduced to govern the technology's usage by non-governmental entities. According to these regulations, facial recognition tools are permitted only for specific purposes when alternative tools are inadequate, and their deployment in public spaces must prioritize public safety.

In Australia the 2023 Federal Budget saw the governmental announcement of a Responsible AI Network, as well as funding of AU$41.2m ($26.9m) towards responsible roll-out of multiple AI technologies across the country. In addition, talks are ongoing among regulators regarding possible changes to the more general Privacy Act, in order to cover a lack of transparency that can occur with AI model training via feedback loops without the presence of human supervision. There are also discussions around the use of analytics and biometric data to train models, which could call for additional privacy rules to be put in place.

In Brazil the legislation process regarding AI was advanced by a legal framework that was approved by government in September

2022, but widely criticised for being too vague. The legislation process regarding AI in Brazil was advanced by a legal framework that was approved by government in September 2022, but widely criticised for being too vague. Following the release of ChatGPT in November 2022, discussions among lobbyists led to a report being sent to the Brazilian government, detailing recommendations around how to regulate artificial intelligence. This document is currently being debated across the Brazilian government, with dates for further drafting yet to be announced.

Regulating AI innovation in California with the hotbed that is Silicon Valley is set to be an ever-present challenge for regulators, with the likes of OpenAI and major investors Microsoft, along with Google being headquartered in the state and overseeing heavy involvement. To overcome this pressing challenge, regulators are planning a sweeping AI proposal. Drawing from the national AI Bill of Rights framework, the pending legislation looks to prevent discrimination and harms across private sectors including education, utilities, healthcare, and financial services. Annual impact assessments submitted to the California Civil Rights Department by developers and users – which would detail types of automated tools involved – and being publicly accessible, have both been suggested as safeguards. Alongside this, there are plans to ask developers to implement a governance framework detailing how the tech is being used and possible impacts.

An Artificial Intelligence and Data Act (AIDA) is planned in Canada for 2025 at the earliest, with drafting having begun under the Bill C-27, the Digital Charter Implementation Act, 2022. government, aligns itself with similar regulations in the US and the EU, with plans to build on existing Canadian consumer protection and human rights law to recognise the need for "high-impact" AI systems to meet human rights and safety legislation. Additionally, it's said that the Minister of Innovation, Science, and Industry would be responsible for ensuring that regulation keeps up with tech evolution, and that new law provisions could be created for malicious use. Six main obligation areas have been identified, for high-impact systems to adhere to: accountability, fairness and equity, human insight and monitoring, safety, transparency, validity and robustness.

In India the government announced in March 2021 that it would apply a "light touch" to AI regulation in the aim of maintaining innovation across the country, with no immediate plans for specific regulation currently. Opting against regulation of AI growth, this area of technology was identified by the Ministry of Electronics and IT as "significant and strategic", but the agency stated that it would put in place policies and infrastructure measures to help combat bias, discrimination, and ethical concerns. Voluntary frameworks have been proposed by the Indian government for the management of AI. Its 2018 National Strategy for Artificial Intelligence considered five key areas of AI development: agriculture, education, healthcare, smart cities, and smart mobility. Then in 2020, ethical uses of AI were detailed in a draft of the National Artificial Intelligence Strategy, calling for all systems to be transparent, accountable, and unbiased. In as swift change compared to previous development, on March 1, 2024, the Indian government issued an advisory asking platforms to seek the explicit permission of the Ministry of Electronics and Information Technology before deploying any "unreliable Artificial Intelligence model(s) /LLM/Generative AI, software(s) or algorithm(s)" for "users on the Indian Internet."[62]. Additionally, it asks intermediaries or platforms to ensure that their systems do not permit any bias or discrimination or threaten the integrity of the electoral process and label all synthetically created media and text with unique identifiers or metadata so that it is easily identifiable.

In South Korea the AI Act is currently in its final phases of drafting, with votes to be made within the National Assembly. The law, as it stands, looks to clarify that regulations must allow any user to create new models without needing to obtain any government pre-approval, with systems considered "high-risk" regarding the lives of citizens required to gain long-term trust. The pending bill holds a prominent focus on national innovation with ethics in mind,

with businesses using Generative AI set to receive governmental support on how to responsibly develop systems. Additionally, the country's Personal Information Protection Commission has announced plans to create a taskforce dedicated to rethinking biometric data protection, in light of Generative AI developments.
As it stands, regulation of Generative AI in the United Kingdom is set to be kept in the hands of sector regulators where AI is being used, with no general law planned beyond the UK GDPR. The Government has opted for a "pro-innovation approach" in official announcements around this topic, with the country looking to take the lead in the global AI race. However, questions around how Generative AI risks such as system breaches, misinformation and bias remain. To help mitigate this, an Impact Assessment has been published by the UK government, which aims to determine suitable and fair regulation of AI developers. This measure comes as part of the wider National AI Strategy, with its summary stating: "A number of market failures exist (information asymmetry, misaligned incentives, negative externalities, regulatory failure), meaning AI risks are not being adequately addressed." "The UK government is best placed to put forward a suitable cross-sectoral regulatory regime to achieve these goals." Objectives laid out include driving AI SME growth, increasing public trust, and maintaining or improving the UK's position in Stanford Global AI Index. The Competition and Markets Authority (CMA), meanwhile, has launched a review into AI foundational models, examining development of tools including ChatGPT for competition and consumer protection considerations. AI developers are being called to demonstrate alignment with five overarching principles: safety, security, and robustness; appropriate transparency and explainability; fairness; accountability and governance; contestability and redress.

In 2023, international efforts to govern and safeguard AI have gained momentum through a series of impactful multistakeholder initiatives, shaping a unified global approach to address pressing challenges. The UN Secretary General's AI Advisory Body, co-chaired by Carme Artigas and James Manyika, and convened by Secretary-General António Guterres, is poised to publish its final report ahead of the anticipated Summit of the Future in the summer of 2024, providing invaluable insights into responsible artificial intelligence (AI) development[63]. Simultaneously, the Global Partnership on Artificial Intelligence (GPAI) held its 2023 Summit in New Delhi, fostering collaboration on AI-related priorities across diverse sectors[64]. The G7 leaders' agreement on International Guiding Principles on Artificial Intelligence, along with a voluntary Code of Conduct for AI developers under the Hiroshima AI process, marked a significant stride in global AI governance. The eleven Guiding Principles provide comprehensive guidance for organizations developing and deploying advanced AI systems, aiming to promote safety, trustworthiness, and responsible use. The G7 Hiroshima Artificial Intelligence Process, established in May 2023, contributes to a broader international discourse on guardrails for AI, aligning with discussions at the OECD, GPAI, and within the context of the EU-U.S. Trade and Technology Council and the EU's Digital Partnerships[65]. Moreover, the AI Safety Summit 2023 at Bletchley Park and the REAIM 2023 in The Hague, Netherlands, served as major international platforms for collaboration, addressing risks and opportunities associated with AI. These collective initiatives underscore the world's commitment to navigating the responsible development and deployment of AI technologies on a global scale.

REGULATION: THE EU AI ACT

The EU AI Act, proposed by the European Commission in April 2021, did not explicitly cover GAI. After the phenomenal take off ChatGPT in November 2022, the legislative process and negotiations on the AI Act basically had to restart and GAI and foundation models have become the key point of controversy among the European Institutions (European Commission, European Parliament, and the Council, the so called Trilogue). On 9 December 2023, EU policymakers reached a political agreement on the AI Act, although as we shall see lots still need to be done and the AI Act will not become fully implemented until 2026.

When ChatGPT surged in 2022, policy circles in Brussels started

to be concerned with the fact that the structure of the AI Act largely reflected traditional EU product liability legislation, which is inadequate for GAI. Unlike stable products with specific intended purposes, GAI can be deployed for several purposes unforeseen during development; this placed GAI outside the scope of the AI Act as it was drafted until the end of 2022. A related issue lies in the European Commission's risk-based approach, assuming that AI systems can be confined to certain risk classes. The central part of the AI Act in the initial Commission proposal[66] is the regulation of high-risk AI systems. High-risk AI systems are either part of a safety component of a product, or is a product itself, or specific applications in a sector designated as high-risk. However, GAI possesses the capacity to perform diverse tasks and learn new ones without predetermined use. Therefore, it cannot qualify as high-risk, implying that developers of GAI would only become responsible for compliance if they significantly fine-tuned or adapted the AI system for high-risk use. Another problem is that actors developing the foundation models behind GAI applications would profit from a distant downstream application while avoiding any corresponding responsibility, even though there are several examples showing that models often carry inherent risks embedded in the underlying data. Conscious of these problems, in April 2023, the research organization AI Now Institute published a report signed by more than 50 experts and institutions arguing that general-purpose AI systems "carry serious risks and must not be exempt under the forthcoming EU AI Act." The report argued that there were risks inherent in the development of foundation models, such as potential privacy violations committed in order to collect the data required to train a model, which can only be addressed by regulating the models themselves rather than their application.[67]

Given these challenges, both the European Council and the European Parliament advanced specific proposals between April and July 2023 to adapt the AI Act to cover Generative AI and its underlying foundation models. In its position[68] the Council introduced the new category of general purpose AI systems defined as: "general purpose AI system means an AI system that - irrespective of how it is placed on the market or put into service, including as open-source software - is intended by the provider to perform generally applicable functions [...] a general purpose AI system may be used in a plurality of contexts and be integrated in a plurality of other AI systems;"[69]. The Council proposed that the rules on high-risk systems apply to general purpose AI. The Parliament's[70] position, instead, focused on foundation models and introduced a separate risk category. It defined foundation models as "an AI model trained on a wide range of data at scale, [which] is designed for the generality of output and can be adapted to a wide range of specific tasks"[71]. The amendments proposed by the Parliament required that all providers of foundation models perform basic due diligence on their offerings[72] and met three obligations: a) Risk identification. Even though it is not possible to identify in advance all potential use cases of a foundation model, providers are typically aware of certain vectors of risk. OpenAI knew, for instance, that the training dataset for GPT-4 featured certain language biases because over 60% of all websites are in English. The European Parliament would make it mandatory to identify and mitigate reasonably foreseeable risks, in this case, inaccuracy and discrimination, with the support of independent experts; b) Testing. Providers are obliged to make adequate design choices to guarantee that the foundation model achieves appropriate levels of performance, predictability, interpretability, corrigibility, safety, and cybersecurity. Since the foundation model functions as a building block for many downstream AI systems, it should meet certain minimum standards. For instance, companies that draw upon PaLM 2 for their AI product should be certain that the building block fulfils basic cybersecurity requirements; c) Documentation. Providers of foundation models would be required to provide substantial documentation in the form of data sheets, model cards and intelligible instructions for use. This is essential not only to help downstream AI system providers better understand what exactly they are refining or fine-tuning but also to enable them to comply with any regulatory requirements.

The provisional agreement on the AI Act reached by the EU negotiators 9 December 2023 and approved by the European

Parliament on 13 March 2024, took up some of the proposals made by the Parliament[73]. Such agreement was earlier opposed by France, Germany, and Italy that pushed for a lighter regulatory regime for models such that support General Purpose AI systems like ChatGPT and Bard. These three countries asked for limiting the rules in this area to codes of conduct, as not to hamper European start-ups like Mistral AI and Aleph Alpha that might challenge American companies. However, the European Parliament was united in asking for hard rules for these models. The compromise was based on a tiered approach, with horizontal transparency rules for all models and additional obligations for compelling models deemed to entail a systemic risk. The compromise refers to the term of general-purpose AI ("GPAI") systems/models and distinguishes between obligations on two tiers: (1) a number of horizontal obligations that apply to all GPAI models, and (2) a set of additional obligations for GPAI models with systemic risk. With respect to the first tier, all GPAI model providers will have to adhere to transparency requirements by, inter alia, drawing up technical documentation. They will also need to comply with EU copyright law and provide detailed summaries about the content used for training. GPAI models of the lower tier will be exempt from the transparency requirements while they are in the R&D phase or if they are open source. With respect to the second tier, GPAI models are set to be designated as encompassing systemic risk when meeting certain criteria. GPAI models that have been classified as systemic risk will be subject to more stringent obligations, which include "conducting model evaluations, assessing and mitigating systemic risks, conducting adversarial testing, reporting to the Commission on serious incidents, ensuring cybersecurity and reporting on their energy efficiency. GPAI models with systemic risk may comply with the EU AI Act by adhering to codes of practice, at least until harmonized EU standards are published. The codes of practices will be developed by industry, the scientific community, civil society, and other stakeholders together with the Commission. Furthermore, a scientific panel of independent experts will issue alerts on systemic risks and support the classification and testing of the models. This compromise is clearly less stringent on GPAI than in the original proposal of the EP.

The agreement reached on December 9, however, was not yet 100% finalised and required some additional technical work. The complexity of the law meant its technical refinement took more than one month. This technical work entails the finalisation of 95 recitals and especially of art. 5 that concerns those systems that should be in principle prohibited but will be exempted by the prohibition on the grounds that they are needed for some special public interests. On 24 January, the Belgian presidency of the Council of EU Ministers presented the final version of the text. Most Member States maintained reservations at the time as they did not have enough time to analyse the text comprehensively. These reservations were finally lifted with the adoption of the AI Act by the Committee of Permanent Representatives on February 2, 2024, when EU Ambassadors unanimously voted on the adoption of the EU AI Act confirming the political agreement reached in December 2023[74]. Subsequently, on February 13, 2024, the new version of the AI Act was approved by the European Parliament's Internal Market and Civil Liberties Committees[75]. This will be followed by a plenary vote by the European Parliament expected for 10-11 April 2024. The formal adoption will then be complete with endorsement at the ministerial level. Hence, the AI Act is expected to be published in the Official Journal of the European Union (OJEU) between May and June 2024, and will enter into force 20 days after publication in the official journal. The prohibitions will become applicable six months after the Act entered into force, and the provisions for 'Foundation Models' (GAI) after 12 months. The transition period for the systems included in Annex 2 will last until 2026. While some provisions of the AI Act will already apply shortly after the adoption of the Regulation, others (for example, some requirements on the high-risk AI systems) will only be applicable at the end of a transitional period. In this context, the Commission is initiating the AI Pact[76], seeking the voluntary commitment of industry to anticipate the AI Act and to start implementing its requirements ahead of the legal deadline. After the AI Act is enforced, there will still be a lot to do, including setting up the AI Office of the Commission, hiring experts, drafting guidelines, and especially formulating horizontal and vertical standards by

2025. Such standards are of crucial importance since they must turn general statements into technical statements that can be measured and checked. For instance sentences such as "suitable risk management measures" will have to be operationalised and made concrete. And the Act is full of such general and vague statements that need to be backed by clear technical standards. While they are deemed 'technical', once formulated they may heavily also impact the legislative implications of the Act. Lots of things can still happen during the definition of such standards and the organisation of regulatory oversight, and a lot of water will have passed under the bridge by 2026, when the Act will be fully implemented. The technological landscape between now and 2026 may be radically changed. So, from now until 2026 the AI Act remains a moving target still surrounded by uncertainties about how it will be operationally implemented and what its effect and applicability will be given the pace of technological change.

The main sources of dissatisfaction of the European Parliament (EP) with the last version of the agreement reached are mainly three[77]. First, from the very beginning the EP was in favour of a general bill of rights and ethical guidelines applicable to all AI systems (without the risk stratification approach), to be then operationalised at vertical level for healthcare, media, creative industries, etc and complemented by a new liability law. The EP was and is not satisfied with the horizontal approach chosen by the Commission as it allegedly raises several ambiguities and problems due to the lack of vertical specific context. This could be solved only by introducing very robust and clear horizontal and vertical technical standards. Second, the EP would have preferred a much more centralised approach to enforcement. In the current solution issues concerning AI in relation to the Act will have to be brought first to the National Competent Authorities, which in turn will simply coordinate with the Commission's AI Office. The AI Office will basically work as a coordinating body and a platform to share best practices. This solution, according to the EP, could lead to the same lack of unified implementation across the EU that characterises the GDPR, and would leave autonomy to those Member States where Big Tech giants are headquartered to be lenient with them. Third, the EP consider the statements and articles on 'Foundation Models' (GAI) too vague and weak, as a result of last-minute lobbying by companies backed by France and Germany. In particular, according to the EP, the procedure about designation (as GPAI with systemic risks) described in article 8 leaves loopholes for big tech companies to challenge decisions on designation and bring them to court. Finally, the EP disapproves the exemption provided for open-source system with respect to transparency and documentation, as it is counter-intuitive and against the spirit of open-source. It is worth noting that European companies operating in the open-source space include France's Mistral and Germany's Aleph Alpha, both of which had criticised the Act as we report below. It is also important to recall that later on Mistral entered into a partnership with Microsoft. So, Mistral had presented itself as the European champion and lobbied to minimize regulation of GAI in the AI ACT as to enable European start-ups to compete with US tech giants to later partner with one of them . Many EU policymakers now, with hindsight, think they have been played and that Mistral played as the front for Microsoft.

According to Metakides[78], the work on the AI Act has been completed despite fierce lobbying, that has been strong and came on the radar at the very last minute (30 October 2023) especially against regulating 'Foundation Models' (GAI). Big tech used not only the self-declared lobbying budget but enrolled SMEs, consumers associations, think tanks, NGOs, consulting companies to lobby against strict rules in general and especially against the regulation of GAI. Still according to Metakides, the Council remains the black box of EU Policy making. The least transparent of EU institutions as far as lobbying is concerned. Difficult for citizens to know who lobbies the representatives of their government in the Council. This time European companies lobbied together with US Big Tech against the regulation of GAI. The earlier cited French company Mistral, where in charge of EU relations is Cedric O (former state secretary for digital) launched an open letter campaign signed by more than 150 SMEs stating that: "The AI Act would jeopardise

EU's competitiveness and technological sovereignty" that was echoed by the French Ministry. A similar position was taken by German Aleph Alpha which was also echoed by the German Ministry.

Having provided an account of the past process, of the current situation and of possible future development of the AI Act, below we very selectively report a few opinions and positions among the many that have been expressed in the past twelve months.
The first mainstream opinion is that proposed by the likes of Anu Bradford. According to Bradford, the EU defied the often-voiced argument that AI cannot be regulated and that it is too early to intervene given the fast-evolving nature of this technology. On the contrary, argues Bradford, the AI Act sets a powerful example, and many governments are already looking at it. We may observe a Brussels effect as several AI developers are likely to conform to the AI ACT also outside of Europe. After all, they may not want to train different models for each individual market in which they operate[79]. So, this is the 'Brussels Effect' narrative version 3.0 (after GDPR and the DMA).

On the other hand, a second position is that of authors arguing that enforcing all obligations uniformly on every foundation model provider, regardless of size, could stifle innovation and reinforce market dominance for leading firms like OpenAI, Anthropic, and Google Deepmind[80]. It may be challenging for smaller companies to compete and catch up with these market leaders. To address this, several stakeholders proposed a distinction between systemically important and non-systemically important foundation models, with considerably lighter burdens for the latter. This proposal would align with the EU Digital Services Act's approach, recognizing the importance of adapting due diligence obligations based on the type, size, and nature of the service. Differentiating between systemic and non-systemic foundation models, and imposing full obligations only on the former, would be justified by the larger resources of firms developing systemic models, making them better equipped for regulatory compliance. Additionally, the potential harm caused by a deviation from compliance by a small firm with few customers is likely to be significantly less than that posed by a systemic foundation model. Such a tiered approach would require some specific criteria to separate 'systemic' or 'high-impact' foundation models from the rest. There are hints from scholars as to criteria that might be used to identify the different types of model, such as the data sources used, or the computing resources required to initially train the model[81].

A third position expresses outright scepticism on the need to horizontally regulate especially GAI. According to Soete[82], for instance, regulation of AI is made difficult by the speed of development of commercial AI platforms and tools, and regulatory sandboxes resemble 'quicksand boxes'. Soete argues that regulators should instead focus on opening 'the black box of data used to train AI and support algorithms', making sure that the data are made public and transparent, 'so that the potential for its much broader application with further improvements can be unleashed, based on trial and error'[83]. He cites the approach proposed by Mons[84] of making data inputs "Fully AI Ready" in any further AI research (an alternative wordplay on the original FAIR acronym of Findable, Accessible, Interoperable and Reusable, for machines as well as people[85]). Mons's position is that 'Attempts to regulate these tools, and the concomitant hype, may only play into the commercial interests of their creators'[86]. He suggests that the hype on this model may implode on itself, and that rather than top-down regulation, the focus should be on good data stewardship. He concludes that 'seeking a top-down approach to curbing such outcomes would drive machine learning into areas where regulations cannot reach. Trustworthy input and the consistent exposure and critique of hype are vital – as are the principles of leading by example and feeding models with proper substrates and conceptual constraints'[87].

EU AND DUTCH PERSPECTIVES ON GENERATIVE AI

EU LEVEL PERSPECTIVES

Regardless of the regulatory debate and of how the AI Act will take shape once technical standards are defined, the hard facts are that Europe currently lags behind the global competition and international leadership on GAI. In this monumental shift, Europe is lacking a cohesive, ambitious EU-wide AI initiative and is struggling to assert itself with 'Made in Europe' AI products or infrastructure. The current global AI landscape accentuates Europe's vulnerability, with only 6% of AI funding going to EU27 start-ups, contrasting with the 61% allocated to US companies and 17% to Chinese ones. Notably, Generative AI-related companies based in the United States secured a staggering $23.8 billion from 2019 to 2023 (with $12.3 billion only for OpenAI), constituting 75% of total investments in such companies during that period. Patent applications have a similar trend, with 34% from US-based inventors, 22% from Chinese inventors, and only 11% from EU27 inventors. Despite this, Europe possesses significant AI talent, scientific excellence, and a commitment to responsible AI development, offering the potential to carve a path toward trustworthy AI.

In line with the Figure above, a recent estimate shows that 73% of AI foundation models have come from the U.S., where development is mainly driven by large technology companies, and 15% from China[88]. European start-ups face challenges in fundraising, accessing necessary skills, utilizing data in less commonly spoken languages, and obtaining sufficient computing infrastructure. There are, however, some noticeable European start-ups and initiatives Start-ups such as France's LightOn and Mistral, Germany's Aleph Alpha that are developing alternative LLMs that safeguard data and incorporate European values. As anticipated at the start of section 3.2, however, Mistral, despite being earlier presented as a European Champion, has recently entered in partnership with Microsoft that invested 16 million Euro in the French company and would soon make the company's AI models available via its Azure cloud computing platform[89]. Similarly, Open GPT-X, an initiative by ten German organizations from business, science, and media, is developing the European answer to GPT-3. The German Federal Ministry for Economic Affairs and Climate Action is funding the Open GPT-X project within the Gaia-X funding initiative with around €15 million. Under the leadership of the Fraunhofer Institutes for Intelligent Analysis and Information Systems (IAIS) and for Integrated Circuits (IIS), the OpenGPT-X project goal aims to develop a large AI language model for Europe that offers data protection as well as European language diversity. However, if compared to the $11 billion raised by OpenAI, these emerging LLM models in Europe have a limited amount of funding. Another noteworthy European initiative is Poro, an Open Language Model developed in Finland by Silo AI, the largest private AI lab in Europe, Together with the University of Turku and HPLT. This marks an important step for the efforts to strengthen European digital sovereignty and democratize access to large language models (LLMs) for all European languages. The model is evidence of the successful application of a novel method to train LLMs for low-resource languages.

Examining patents from the past eight years reveals a concentrated landscape in AI innovation, with over 25% of all European AI patent applications originating from four specific regions: inner London, Île-de-France (Paris), Noord-Brabant (Eindhoven), and Ober Bayern (Munich). This concentration highlights significant AI talent

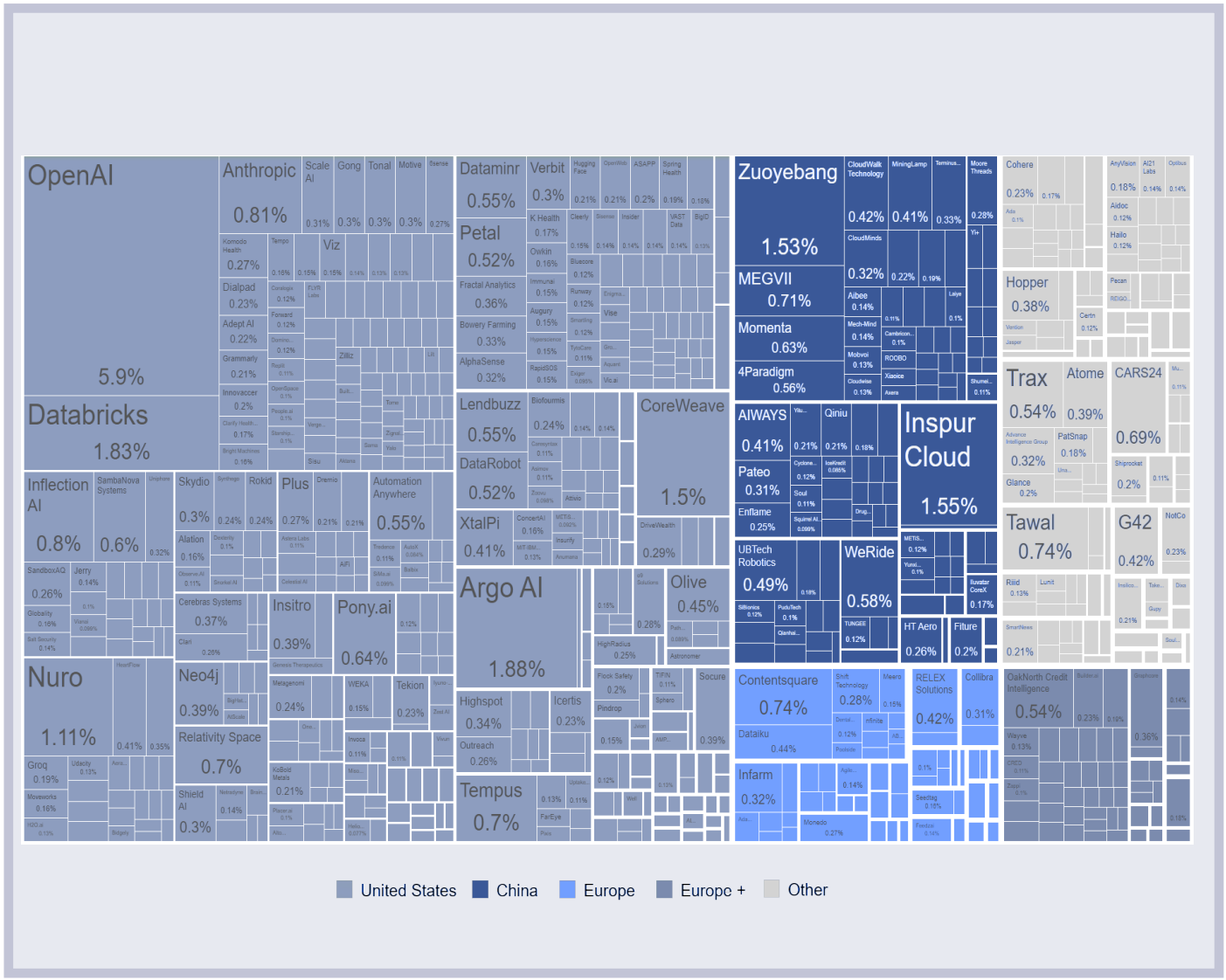

Figure 10 The share of venture capital (VC) funding received by AI start-ups, Source: Crunchbase

and innovation potential in selected European cities and regions. Such concentration would suggest substantial investments in AI research and development with a targeted approach to support and foster these innovation hubs, so that they can become global magnets for funding and talent. Moreover, Europe is currently lacking inter-regional cohesion, as most AI patent connections within Europe remain confined within national borders. To meet global demands for sophisticated AI technologies, Europe should prioritize and enhance inter-hub connectivity for a collaborative and interconnected AI ecosystem[90].

The development of the European Generative AI ecosystem is interconnected with the discussion around the AI Act and how the new regulation would impact emerging European AI companies. Research conducted by a group of European AI associations at the end of 2022 found that 50% of start-ups surveyed thought the AI Act would slow down AI innovation in Europe, while 16% said they were considering stopping developing AI or relocating outside the EU[91]. More recently, the French, German, and Italian governments have been advocating for a "balanced and innovation-friendly" approach to regulating AI that is risk-based but also reduces "unnecessary administrative burdens on companies that would hinder Europe´s ability to innovate." The French and German governments have actively shown their commitment to fostering innovation within their respective AI industries. In June, President Macron pledged €500 million to support AI "champions," while the German government, in August, announced plans to nearly double public funding for AI research to approximately €1 billion in the next two years. Both governments share concerns that excessive regulation may hinder the growth of their domestic AI sectors and suggest that the EU AI Act should focus on regulating AI uses rather than the underlying foundation models.

As this report was being finalised, on January 24, 2024, the European Commission released the Communication on 'on boosting start-ups and innovation in trustworthy artificial intelligence'[92], launching an AI innovation package to support Artificial Intelligence start-ups and SMEs. Among the most noteworthy aspects of this package we outline the following:

- An amendment of the EuroHPC Regulation to set up AI Factories, a new pillar for the EU's supercomputers Joint Undertaking activities. This includes:
  - Acquiring, upgrading and operating AI-dedicated supercomputers to enable fast machine learning and training of large General Purpose AI (GPAI) models.
  - Facilitating access to the AI dedicated supercomputers, contributing to the widening of the use of AI to a large number of public and private users, including start-ups and SMEs.
  - Offering a one-stop shop for start-ups and innovators, supporting the AI start-ups and research ecosystem in algorithmic development, testing evaluation and validation of large-scale AI models, providing supercomputer-friendly programming facilities and other AI enabling services.
  - Enabling the development of a variety of emerging AI applications based on General Purpose AI models.
- Financial support from the Commission through Horizon Europe and the Digital Europe programme dedicated to Generative AI. This package will generate an additional overall public and private investment of around €4 billion until 2027.
- Accompanying initiatives to strengthen EU's Generative AI talent pool through education, training, skilling and reskilling activities.
- Further encourage public and private investments in AI start-ups and scale-ups, including through venture capital or equity support (including via new initiatives of the EIC accelerator Programme and InvestEU).
- The acceleration of the development and deployment of Common European Data Spaces made available to the AI community, for whom data is a key resource to train and improve their models. A new Staff Working Document on common European data spaces has also been published today, providing the latest state of play.

- Several Member States will join forces through the Alliance for Language Technologies European Digital Infrastructure Consortium (ALT-EDIC) initiative, which will provide centralised access to language resources for the development of European LLM's. As advanced models can effectively handle multiple types of data simultaneously (e.g. text, audio, video, images, code), ALT-EDIC will also open possibilities for more holistic and comprehensive AI applications across various domains.
- The 'GenAI4EU' initiative, which aims to support the development of novel use cases and emerging applications in Europe's 14 industrial ecosystems, as well as the public sector. Application areas include robotics, health, biotech, manufacturing, mobility, climate, and virtual worlds.

This innovation package is to be considered as the third leg of an AI governance framework at EU Level, the first one being the AI Act and the second competition policy and anti-trust. The Communication calls for public investments and is part of the European Strategy of making available to everyone the three basic ingredients of GAI: computer power (network of AI factories), data (single market), and talent (training). So, seemingly public investments will go hand in hand with regulation.

THE CASE OF THE NETHERLANDS

The Netherlands is known for its strong performance in digital infrastructure. The country has an impressive coverage of high-capacity network connectivity, and the semiconductor and quantum computing sectors in the Netherlands are seeing significant growth in research and development[93]. The Netherlands has a digitally skilled population, with 79% of adults possessing at least basic digital skills, significantly higher than the EU average of 54%. In the workforce, 7.2% are ICT specialists, surpassing the EU wide average of 4.6%. Despite this, the Netherlands falls behind the EU average in terms of ICT graduates, with 3.7% compared to the EU average of 4.2%. Efforts are also underway to enhance digital skills in foundational and higher education. Dutch SMEs tend to have high digital engagement, surpassing the EU average with an 80% score in digital intensity. The country also leads in adopting advanced technologies like big data, cloud services, and AI, with respective usage rates of 27%, 60%, and 13%, all of which are higher than the EU average. To further boost digital innovation, six European Digital Innovation Hubs (EDIH) started in 2023. The Netherlands' robust start-ups ecosystem is evident from its 24 unicorns and 39 potential future unicorns. TechLeap.nl and InvestNL help technology start-ups and scale-ups expand and become successful on a global scale. The Amsterdam-Delta region and the Brainport region (Noord-Brabant) are recognized as world-leading regional start-ups ecosystems[94 95].

Artificial Intelligence in the Netherlands

The AI-ecosystem in the Netherlands has seen significant developments in recent years, driven by strategic collaborations and investments. A report[96] recently published by Google states that Generative AI alone could boost the Netherlands' GDP by €80-85 billion. Around 600 thousand Dutch jobs are likely to be fully or partially displaced, and another 6.5 million Dutch jobs are likely to be augmented by GAI, and the AI-powered economy will also help create new jobs. In 2019, the Dutch government unveiled its Strategic Action Plan for AI (SAPAI)[97], aiming to capitalize on the societal and economic opportunities presented by AI while safeguarding public interests. This plan recognized the transformative potential of AI across various sectors and sought to establish the Netherlands as a leader in AI knowledge and application. Part of this strategy has been the establishment of the Netherlands AI Coalition (NL AIC), a public-private partnership launched in October 2019. The NL AIC initiated a network of 7 AI Hubs to facilitate the regional connectivity and the deployment of national programs. The Dutch ambitions on AI are underpinned by a €204.5 million budget from the National Growth Fund for the AiNed program. AiNed aims at strengthening the competitive position of the Netherlands in the field of AI, enhancing, and promoting the use of AI across various sectors, including industry, healthcare, energy, and mobility. It supports excellent research

in AI to bring the Netherlands at the forefront of AI technology, encouraging partnerships between academia and industry to drive innovation and practical applications of AI, and boosting economic growth, and support international collaboration with European partners. To support AI research and talent development, a national collaborative initiative for ICAI (Innovation Centre for Artificial Intelligence) labs was established by the University of Amsterdam in 2018. The ICAI Lab network, which currently exists of over 50 labs at 13 university campuses, encourages engagement of companies with the academic AI research community, thereby maintaining a cutting edge in AI research and applications. ROBUST, an ICAI initiative to further strengthen AI research in the Netherlands, has received an investment of €25 million. The Hybrid Intelligence Zwaartekracht programme has received €20 million funding from NWO to develop theories and methods for intelligent systems to collaborate with humans, adapt to dynamic conditions and explain their actions. GPT-NL is a publicly-funded (€13 million) project by SURF, TNO and NFI to a language model tailored specifically for understanding and generating natural language text in Dutch. It is trained on a diverse range of text corpora in Dutch to grasp the nuances of the language, cultural context, and the subtleties that come with it. The purpose of GPT-NL is to provide a tool that can assist in various tasks such as translation, content creation, and language learning, while also aiding businesses and developers to create applications that can communicate effectively with Dutch-speaking users.

EU research and innovation programs play a significant role in strengthening the AI position of the EU and its Member states. International AI research networks CLAIRE and ELLIS have established bases in the Netherlands, recognizing recognising the Netherlands' excellence and leading role in these networks. Some developments, such as playing an impactful role in Generative AI, are too big to do from the Netherlands alone. In this global playing field, which is dominated by America and China, a European response is needed. For example, the Netherlands is collaborating with other Member States to establish a common infrastructure to

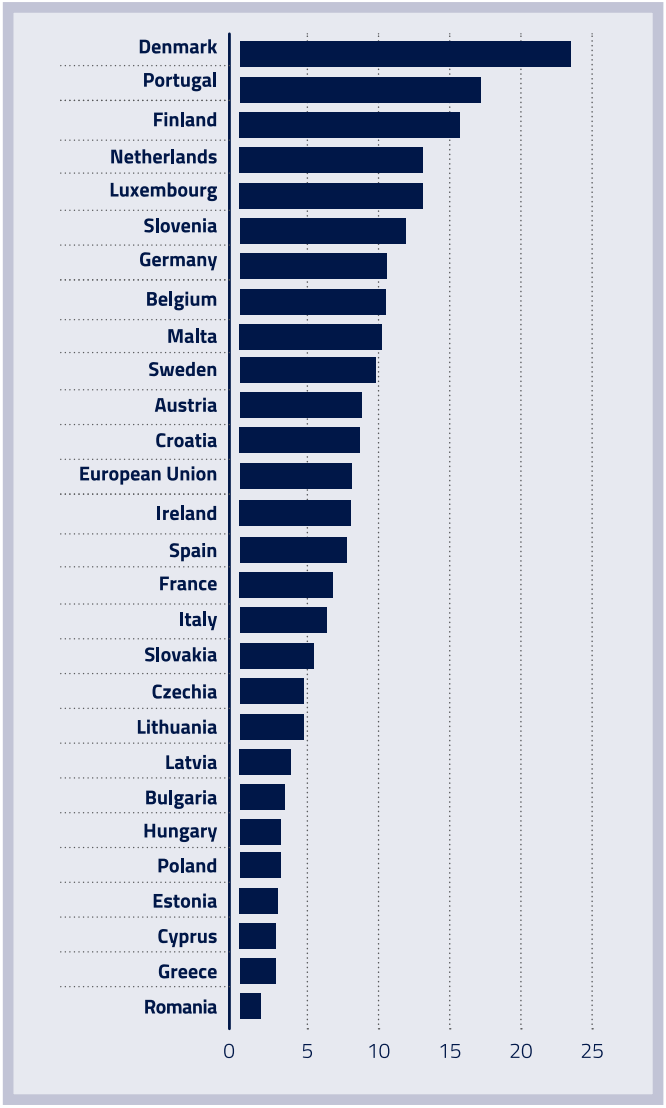

Figure 11 Artificial Intelligence adoption by enterprises , Source: https://digital-decade-desi.digital-strategy.ec.europa.eu/datasets/desi/charts

develop large frontier models for natural language processing and multi-language models, in the Alliance for Language Technologies EDIC. While the start-ups ecosystem is robust, scaling these start-ups to compete on an international level remains a challenge. In the Netherlands, businesses across various sectors are increasingly integrating AI technologies to enhance operational efficiency, customer experience, and competitive edge and the country score relatively high in the DESI index (see below).

Legislation and regulatory oversight of AI in the Netherlands cannot be seen in isolation from European efforts to steer AI in a human centred direction. Digital legislation such as the AI Act, GDPR, DMA and DSA are essentially determined in Brussels. The Netherlands has a strong emphasis on developing and implementing AI responsibly and ethically. Dutch institutions actively engage in discussions and research on the societal impacts of AI, data privacy, and AI governance, for example in AiNed ELSA Labs. The Netherlands have been a strong proponent of regulatory sandboxes, which are currently included as an instrument for regulatory innovation as part of the AI Act. Whereas regulatory oversight for the GDPR is mandated to the Dutch Data Protection Authority, The Dutch Authority for Digital Infrastructure (RDI) is responsible for coordinating regulatory oversight on the upcoming AI Act. Together with the European Commission and UNESCO, RDI initiated a collaboration to design ethical AI governance frameworks. This project aims to develop effective supervision of AI in Europe, aligning with the EU AI Act and UNESCO's Ethics of AI recommendations, includes producing tools, case studies, and training for better AI supervision and ethical development.

CONCLUSIVE COMPARATIVE CONSIDERATIONS

While Europe is home to some of the world's best universities and talents, progress is being hampered by emphasis on separate national AI strategies, slow decision-making, too little investment in AI by Europe's private sector and bureaucratic procedures in the public sector. As the adoption of these Large Language Models (LLMs) gains momentum, Europe faces the prospect of growing reliance on foreign AI models, posing a potential threat to the overall competitiveness of the European economy. But it is not just economics at stake. As we have discussed in the previous sections, Generative AI is expected to have an important impact on society, and one area at risk is culture. The sheer dominance of the big US and Chinese AI companies is staggering. By not having viable European AI platforms, European users will have to accept

	Overall	Talent	Infrastructure	Operating	Research	Development	Government Strategy	Commercial	Scale	Intensity
1	United States	1	1	28	1	1	8	1	1	5
2	China	20	2	3	2	2	3	2	2	21
3	Singapore	4	3	22	3	5	16	4	10	1
4	United Kingdom	5	24	40	5	8	10	5	4	10
5	Canada	6	23	8	7	11	5	7	7	7
6	South Korea	12	7	11	12	3	6	18	8	6
7	Israel	7	28	23	11	7	47	3	17	2
8	Germany	3	12	13	8	9	2	11	3	15
9	Switzerland	9	13	30	4	4	56	9	16	3
10	Finland	13	8	4	9	14	15	12	13	4
11	Netherlands	8	16	15	10	13	28	20	11	8
12	Japan	11	5	10	20	6	18	23	6	25
13	France	10	11	25	15	18	13	10	9	20

Figure 12 Global Artificial Intelligence ranking: top 10 performers, Source: https://www.tortoisemedia.com/intelligence/global-ai/#rankings

cultural and ethical embedding from AI developed in other parts of the world and European diversity of culture and languages may suffer from it. As the ranking below shows, only four EU countries (Germany, Finland, Netherlands, and France) make it to the top 10 AI ranking (see below).

Europe's current position should not be mistaken for a lack of talent or a lack of technological potential. Several European hubs do indeed harbour a rich pool of AI talent, scientific excellence, and commitment to responsible AI development, which can forge a path towards trustworthy AI, rooted in humanitarian and democratic values. Looking at cities and regions, the picture is much less alarming than at the country level. Top AI start-ups in Paris, for instance, are receiving similar funding to those in Boston, and more than those based in Pittsburgh, Seattle, Chicago, Shenzhen, Guangzhou, or Hangzhou. What is missing in Europe is a 'mission oriented industrial policy', something similar to the Airbus mission to boost European excellence in AI[98].

Moving to the Netherlands, the AI-ecosystem is marked by a strong academic foundation and excellent educational basis. The Netherland has adopted a collaborative approach to innovation, a commitment to ethical AI development, and is characterised by a relatively high technology adoption rate for citizens and companies. This makes the country a recognised significant player in the European AI landscape, and in AI rankings it scores reasonably high in comparison to similar European economies; it is outperformed by the US, China, and UK. Compared to these globally competitive countries, the Netherlands falls short in translating AI research results into technological AI innovations and applications that contribute to economic earning capacity and the transitions needed to solve major societal challenges. Few Dutch AI start-ups grow into global players, and opportunities for AI in transitions are not sufficiently seized. Despite the availability of excellent education and research institutions, the Netherlands is unable to produce, retain and attract sufficient AI talents and create AI technological innovations that are embraced by the market. The Netherlands has difficulty valorising research results and sees too few start-ups maturing into significant market players. It has a strategy for research and development, and for emerging innovations, but the absence of an overarching strategy for managing the transition to an AI-integrated economy is a clear deficiency. The Netherlands requires breakthrough projects where relevant stakeholders collaborate to create AI system solutions to support specific transitions, in healthcare, energy and other sectors. Furthermore, it needs an increase in the number of start-ups and scale-ups that are further professionalized and are able to scale up to become international providers in Europe and globally. Focus is required on emerging AI fields such as Generative AI. It is crucial to retain and attract top AI researchers in Generative AI, while ensuring they have access to state-of-the-art facilities. Trying to find the right balance between opportunities and risks - technology and industrial development on the one hand, and legislation, regulatory oversight, and ethical discourse on the other - is a risk in itself, as it can lead to a lack of focus and paralyze the required efforts to bring innovation and development of AI to global competitiveness. The Netherlands has the potential to be among the frontrunners in AI and keep pace with key players such as Germany, France, and the UK, but this requires a strategic approach. For example, with the arrival of frontier AI models such as ChatGPT late 2022, a European lobby emerged to protect citizens from the impact of frontier models by expanding the scope of the AI Act. Germany and France subsequently initiated a counter-lobby to bolster the innovative strength of their country's start-ups and corporations that are potential globally competitive players in the field of frontier models. Conversely, the Netherlands is not involved as it lacks a significant position in the development of frontier models.

TECHNOLOGY, SOCIOECONOMIC DYNAMICS, AND REGULATION: **THE SCENARIOS**

MAIN AREAS OF UNCERTAINTIES

In the previous chapters we have presented Generative AI as a technology (Chapter 2), its socioeconomics dynamics (Chapter 3) , the main risk of society and democracy and the regulatory process (Chapter 4), followed by an assessment of the state-of-play in both the EU and The Netherlands (Chapter 5). The technology development, the socioeconomic dynamics, the regulatory process and the corresponding level of public investments are all a source of uncertainty.

The technology behind Generative AI can be seen as the main driver of change that impact both the socioeconomic dynamics and regulation. In the current technology landscape there is a focus on generalist, monolithic and closed source models, which require large-scale and concentrated resources in terms of investment, data, and computing power. However uncertainties exist regarding the future development of this technology landscape since there exist alternatives in these three dimensions: generalist vs specialist models, monolithic vs federative models and open-source vs closed-source approaches to model development and training. These technological developments could either drive more or less competition, impacting business models and revenue streams. The technology could also either make regulation redundant/outdated or generate new opportunities associated to new risks that would require new regulation. If technological development will reduce the need for large amount of data to train models and, thus, also reduce the computing power requirements, this will ensure low entry barriers, vibrant competition, and there will be less opportunities for unfair competition from incumbent big players. Conversely, if the technology will develop in such a way as to require ever larger amounts of data and computing power, this will hamper competition and incentivise bundling unfair competitive practices that will further increase market concentration. Another uncertainty partly related to technology and partly to the implementation of existing regulation, is the extent to which cumulated proprietary data by incumbents running a platform will be an insurmountable advantage, or whether smaller players will be able to effectively and efficiently pre-train and fine tune their model using the increasing number of publicly accessible databases; and whether, in Europe, according to the provision of the DMA, smaller players will be granted access to the data possessed by big incumbents. Technological developments matched by self-regulation might also make regulation redundant if future models reduce the occurrence of the societal risks described in section 4.1.

The socioeconomic dynamics depends on market trends, on investment decisions, on companies competitive strategies, on how anti-trust authorities will keep large incumbents in check, on labour market developments, as well as on the actions of consumers and citizens, both in terms of adoption and of complaints and request of redressing. In this respect, it is worth noting that there is a difference in the attitudes and behaviours in our different role as 'citizens' and as 'consumers/users'. Let us look, for instance, at the case of large online platforms and digital services. In Europe they have been regulated (by the DMA and also by the Digital Service Act, DSA) to protect our rights as citizens, yet many of us behaving as consumers/users continue to use them with little attention for the new regulatory provisions, out of habits (status quo bias), lock-in (network of contacts and reputational capital), and convenience. The same might occur with Generative AI applications and solutions,

if technological development is such that produces convenience and great user experience for users/consumers, while those voicing concerns and critique as citizens would remain an active but increasingly small minority, thus, reducing support for regulatory intervention. The socioeconomic dynamic also depends on success or failure in strategic domain of applications such as for instance healthcare and energy. It is uncertain whether investments will continue to focus mostly on model makers or spread funding also to other layers of the GAI ecosystem, and also importantly if such investments will continue to be skewed toward the US and China or will also reach increasingly the EU. Also, it is still uncertain if the revenues will concentrate for the most successful early entrants and for other incumbents or will spread more evenly. The future competitive landscape is also very uncertain, with factors that may favour oligopolistic concentration and countervailing trends that may keep a level playing field. It is uncertain to what extent big incumbent players will adopt unfair competition methods (bundling and tying) and anti-trust authorities will be able to stop them. Adoption by firms and consumers will greatly shape the further development of the technology and its social and economic impact, above all on productivity and growth, but also on occupational level. Finally, as the development and adoption of the technology progresses inevitably, until the technology improve itself and or is regulated, some of the risk described earlier will materialise and have consequences, and it remains to be seen to what extent consumers and citizens will mobilise to defend their rights and obtain redress. One of such risk is very imminent and concerns the next European Election of 2024, when the AI Act will not be fully enforceable yet, and GAI systems may be maliciously used to orchestrate fake news and disinformation campaign to affect the results of the elections.

A lot of uncertainty still surrounds AI governance, defined as comprising regulation (AI Act), public investment (Innovation Package), as well as competition policy and anti-trust. Here we refer to the AI Act both in general and with specific regard to GAI. The work is not finished yet and the Act will be fully implemented in 2026. In 2025 horizontal and vertical technical standards and guidelines must be produced. The provision for GAI currently in the Act remains vague. For instance the procedure for the designation of GPAI model as tier 1 (with less obligations) or tier 2 (GPAI models with systemic risks subjects to more obligation) is vague and leaves loophole for company to challenge it in court. Furthermore, GPAI models with systemic risk may comply with the EU AI Act by adhering to codes of practice, at least until harmonized EU standards are published probably but not certainly by 2025. Finally, GPAI models of the lower tier will be exempt from the transparency requirements while they are in the R&D phase or if they are open source. It is safe to conclude that the regulation of GAI is far from certain and clear. Also worth noting that, as anticipated, some analysts have argued that enforcing all obligations uniformly on every foundation model provider, regardless of size, could stifle innovation and reinforce market dominance for leading firms like OpenAI, Anthropic, and Google Deepmind[99]. So, at this stage it is highly uncertain the extent to which the Act will be rigidly and strictly enforced in close cooperation between national authorities and the Commission with clear technical standards, or if it will remain to some extent loose, dealt with mostly at national level, and will require a lot of additional interpretation and possibly clarification in courts. It is also worth noting that the AI Act does not have any provision on competition issues, and so the regulation dimension does not incorporate competition policy and anti-trust initiatives. In this context it is important to stress the crucial role that standardisation will have in making many of the currently general and generic provisions more concrete and operational. The standards will play a strategic role in the AI ACT and much will depend on which players will participate and the extent to which European SMEs will manage to influence them. In addition, the size and focus of future public investments remains uncertain, as the Innovation package presented by the Commission in January 2024 for now remain a plan yet to be implemented and deployed. Finally, beside regulatory initiatives such as the AI Act, there are uncertainties regarding competition policy and anti-trust initiatives within the governance framework of GAI. Risks of unfair competition practices by incumbents such as bundling products or engaging in exclusive dealing exist and it is still uncertain to what extent such practices will be implemented by companies

and countered/mitigated by anti-trust authorities. Depending on how competition policy and anti-trust will deal with such issues, the potential entry barriers for new entrants in terms of access data, talent, and computational power may either diminish or increase.

Evidently, these three dimensions interact with each other, and to render such complex reciprocal interactions a sophisticated system dynamic modelling exercise would be required. With some conceptual simplification we select socioeconomic dynamics and regulation as the two dimensions of uncertainty shaping the scenarios and look at them from a high-level perspective. Technology will be used to characterise the identified scenario. The technological uncertainties about the future need of proprietary computation infrastructure and data, and supremacy of closed, monolithic model architectures verses more open, federated models, require continuous monitoring to anticipate future needs and opportunities. On the other hand, whether or not public investments will support AI innovators and start-up to offset the dominance of the incumbents remains uncertain and will affect the competitive structure of the market.

THE PROPOSED SCENARIOS

In view of the discussion above we selected as the two uncertainties representing the axes of the scenarios depicted below: 'the socioeconomic dynamics' and 'regulation'. The technology itself, as explained, is the main driver that will impact both the regulatory process and the socioeconomic dynamics and, thus, will be included inside the scenarios storylines.

The socioeconomic dynamics vary between two extremes deemed inclusive on the right-hand side and exclusive on the left-hand side. This characterisation of the socioeconomic dynamics is related both to the level of competition in general, but also to the kind of players that will be able to stay in the market, as well as to what extent there will be closure or openness for European market players. So, for instance, a more inclusive dynamics would mean that European SMEs are able to compete and European start-ups manage to scale-up, whereas an exclusive dynamics would entail

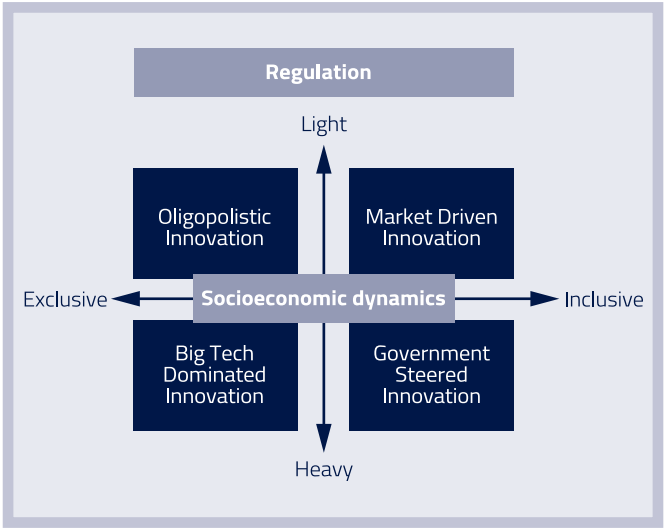

Figure 13 The scenarios, Source: Authors' elaboration

dominance by incumbents, either in the form of big techs or big tech plus a limited number of new companies that manage to consolidate, with less space for EU SMEs and start-ups.

The regulation varies between the two extremes of heavy and light, to reflect the uncertainties still present in the AI Act, especially for GAI, and the unclear relation with competition policy. A light regulatory regime will also be loose and flexible and will leave space for market players to innovate; but it also brings perduring uncertainties as to implementation and enforcement that may require interpretation and possibly court ruling. A heavy regulatory regime entails, not only a stronger and steady scrutiny of suppliers with various requirements and sanctions (provide transparent documentation, audits, prohibitions, and fines), but also stricter and clear-cut definition of tiers of risk, standards, guidelines, and enforceable procedure. This regime would certainly bring certainty of law but may also stifle innovation by start-ups, SMEs, and the academic community, or may be made obsolete by new technological

developments. It is also assumed that a stricter regulatory regime will be matched by public investments exactly to counter the stifling effect on innovation.

In the tradition of foresight, it is assumed that the future will be different from any of the four scenarios but may contain elements present in each of them. So, taken collectively with their storylines and their assessment in terms of their potential impacts on key dimensions, the scenarios may bring out relevant implications and recommendations for policy makers.

SCENARIOS' STORYLINES

Market Driven Innovation (MDI).
In this scenario the socioeconomic dynamics is inclusive and the regulatory regime light. The lack of heavy regulation does not cause dominance by incumbents because the plurality of players that are able to stay in the market boost technological developments that lower barriers in terms of data and computing power requirements.

In absence of the administrative burden that goes together with heavy regulation, start-ups, scale-ups, SMEs, and the research community can stay in the market alongside Big Tech and early entrants already consolidated. Venture capital funding spread more widely across different kind of players and across countries, further boosting an open innovation ecosystem driven by the logic of a competitive market. The plurality of players also entails increase and diversification of R&D with steady development of new models without dominance of, and lock-in with, the models brought to the market by the early entrants.

In such context the number of players and talents at work will greatly contribute to improving the technology itself, and this trend increasingly involve European entities, from large companies, to SMEs, start-ups, and the academic community, which is positive for strategic autonomy. The success of open-source approaches to model development and innovation could also be a contributing factor for European players to become more relevant. So, the technology itself will improve, greatly reducing the instances of 'hallucinations' but also the actual occurrence of the kind of societal risks reviewed in Section 4.1, which contribute to social cohesion and fairness. This contribution will be reinforced by the fact that voluntarily as a result of self-regulation efforts the GAI solutions will be transparent and will accommodate for redress of consumers' complaints, if any.

Moreover, the potential success in the further development of federated approaches or Small Language Models (SLMs) and of fine-tuning techniques of existing model such as Low-Rank Adaptation (LoRA) technique could lower the data and computing power requirements compared to monolithic approaches and contribute to keep the new GAI market a level playing field. Technological improvements may also make more efficient model pre-training and fine-tuning using publicly available data and from those obtained from incumbent platforms according to the provision of the DMA. In this scenario the cumulated proprietary data owned by incumbents running a platform are no longer an insurmountable advantage.

Less data and computing power requirements reduce the opportunities and incentives for unfair competition practices such as bundling (especially with clouds services) by Big Tech. Moreover, while regulation is light, anti-trust authorities will monitor the competitive behaviour of big tech incumbents and prevent the adoption of such practices.

As a result of the growing number of solutions brought to the market of the increasing presence of many market players, not only efficiency and productivity increase contributing to overall growth, but potentially the impact on social cohesion and democratic fairness will be positive. As the potential applications of GAI increase and many players are active in the market, employment levels may remain stable or even increase, both within GAI producing and GAI using firms. Second, the improvement in the models

and in the capacity to use less but more quality data can reduce both 'hallucinations' and societal discrimination produced by the algorithm. New technological developments in GAI itself and in the differential privacy can reduce risks for privacy and for data breaches. Third, self-regulation integrated with better technological solutions that block malicious reuse of GAI generated contents, images, and videos would also reduce the risk of deep fake and misinformation distorting democracy and the public sphere.

Government Steered Innovation (GSI).
This scenario there is a presence of strict regulation combined with inclusive socio-economic dynamics that is made possible by sizeable public investments in support of AI innovators and start-ups. Such a scenario allows the government to take an active role in shaping the market, by imposing regulation and orchestrating concerted investments e.g. for stimulating European scale-ups and providing public computing infrastructure which may yield positive spinoff effects in terms jobs and revenues. Strong regulation can be instrumental to creating a more inclusive market when it reduces the barrier of entry of new market players, promoting experimentation and exploration, while also allowing players to grow through innovation policies. This balancing act requires vision and strategy from a steering government to develop a balanced technology and industry position, and a targeted investment policy that complements regulatory actions, in order to create an orchestrated market. Innovation takes a form of integration between large incumbents that incorporate in various form smaller innovative players from SMEs, start-ups and from the academic community. If in such a scenario for example the European government successfully drives innovation through effective public-private partnerships, this scenario allows for achieving a good balance between regulation and industry leadership.

In this scenario, the market dynamics remains inclusive without monopolistic/oligopolistic closure for two reasons. First, public investments manage to support the strongest AI innovators and SMEs helping them compete and stay in the market despite the presence of strong incumbents. Second, the market dynamics invites incumbents to help smaller players deal with the more rigid regulation, and in exchange benefit from technological improvements and access to talent. This is a situation where, as the case of Mistral partnership with Microsoft shows, for European innovators and SMEs partnering up with US Big Techs is the only way for to scale up fast . So, it remains to some degree an inclusive scenario because Big Techs do not entirely close the market and allow many initiatives linked to them to develop and create additional jobs, and public investment also help small innovative players stay in the market. So, increases in efficiency and productivity do not lead to loss of employment. From smaller payers, incumbent also benefit in terms of technological improvement for reducing 'hallucinations' and the actual occurrence of societal risk. Governments can also actively orchestrate the market by breaking up Big Tech monopolies into smaller entities, forcing fair market competition. In this scenario European entities will retain a space and given opportunities to grow but in a complementary way compared to the open cohesive innovation scenario.

Big Tech Dominated Innovation (BDI).
This scenario portrays a socioeconomic dynamics with a heavier and stricter regulatory approach that enforce all obligations uniformly on every foundation model provider, regardless of size. Big Tech with their legal powerhouses and organisational capacity will cope successfully with such a regime both in terms of the associated administrative burden and by finding loopholes and, when needed, challenging regulatory decision in court. Conversely, heavy regulation will seriously reduce the capacity to stay on the market of start-ups, scale-ups, SMEs. This in turn will reduce the diversity of players and, as a consequence, stifle technological innovation and development.

The reduced set of players involved in bringing new and efficient solutions and applications to the market means that the technology will develop along the lines that are more advantageous for early entrants and Big Tech, which means a focus on foundation models

requiring ever increasing amount of data and computing power. As a result, the market will be dominate by a few Big Tech players that will acquire the relatively large GAI early entrant and will become the main funders/users of innovation produced by the research community. This scenario can lead to directly opposing positions between Big Tech players and governments, where governments shape increasingly stricter regulations to constrain Big Tech. Paradoxically this may lead to regulatory capture working in favour for Big Tech and governments are left with few to none instruments to orchestrate the market.

In this scenario with strict regulation and a market dominated by a few Big Tech players operating in generic domains with monolithic foundation models, there are few incentives for those incumbents to expand further into niches or push for federated development of models, or to keep investing in open-source development, if this does not bring them a direct competitive advantage relative to the other big players, as bottom-up competition from challengers will be mostly absent. This may lead to a continuation of the current monolithic, generic, and closed-source approach to building foundation models despite the fact that exploration of alternative approaches might lead to better models overall.

Because of the lack of the plurality of players and talents and because of the more restrictive interests of a few big players, the technology will improve but not as much as in the Market Driven Innovation scenario, especially in reducing 'hallucinations' and the occurrence of the societal risk reviewed in Section 4.1. Given their lobbying and legal power, big tech incumbents may take an adversarial role and go to court when issue of societal discrimination, misinformation, and privacy are brought up by consumers and/or citizens. This will not positively contribute to social cohesion and fairness. As a result of concentration and the smaller number of solutions brought to the market, while productivity grows, there is the risk of a drop in employment because many more human related task are created by this wave of innovation, thus weakening social cohesion. This scenario will be dominated by US tech giants and Chinese companies, and there will be little space left for European entities.

Oligopolistic Innovation (OI).
This scenario portrays a socio-economic dynamics with a regulatory approach that will enable start-ups, SMEs, and academic scholars to innovate GAI without having to comply with strict obligations and requirements. In this inclusive context the number of players and talents at work contribute to improving the technology itself, particularly in a technological context where federated and open-source innovation outpaces monolithic closed source efforts. This trend may increasingly involve European entities, from large companies, to SMEs, start-ups, and the academic community as new players in the market. However, as a result of early growth and market dominance, and of competition methods not adequately checked by anti-trust, big incumbents are likely to continue dominating the market as scale in compute, data, and risk-prone equity remains advantageous. This can be disadvantageous from a European perspective in terms of economic benefits, as well as strategic autonomy, possibilities for EU data and model regulation and oversight, as renting infrastructure creates dependencies.

In a context where big incumbents are able to stay competitive and attract resources without an open data and open science approach, all relevant innovation may happen behind closed moats and the incumbent advantages will extend even further. Smaller players including many EU entities will then be forced to collaborate with these large incumbents. As a result of concentration and the smaller number of solutions brought to the market, while productivity grows, there is the risk of a drop in employment as many more human related tasks are automated by this wave of innovation.

This scenario is likely to further strengthen the market domination of US tech giant and Chinese companies, with limited competitive opportunities for European companies and initiatives.

SCENARIOS ASSESSMENT AND CONCLUSIONS

The figure below presents the assessment of the four scenarios with respect to five dimensions of potential impacts that we explain below.

First, we use growth and assume that it will result from the productivity and efficiency gains that GAI innovation will produce. Second, social cohesion refers to possible effects on employment levels and on societal discrimination. Third, under Democratic fairness we include the possible effects of misinformation (with impact on the democratic process) and that of privacy mishandling. Strategic Autonomy measures the extent to which in each scenario European entities can play a leading role and preserve the continent's autonomy vis-à-vis US tech giants and large Chinese corporations. Finally, the dimension risk containment assesses the degree to which each scenario includes regulatory or self-regulatory measures to reduce/minimise the risks described in Section 4.1.

We consider growth to be slightly higher under Market Driven Innovation (henceforth MDI) because of the increased level of competition and of offerings. On the other hand, we assume that contribution of growth will be high also for the Big Tech Dominated Innovation (henceforth BDI) and the other two scenarios Government Steered Innovation (henceforth GSI) and Oligopolistic Innovation (henceforth OI). Our assumption is that this technology in one form or the other will greatly impact growth positively across the scenarios, but that the scenarios differ with respect to the other dimensions. MDI is clearly superior to BDI in terms of social cohesion, fairness, and strategic autonomy, whereas it performs worse in terms of legally compliant management of risk because of the uncertainty that a more flexible approach to regulation entails. The Government Steered Innovation, henceforth GSI scores a bit better in terms of social cohesion, fairness, and strategic autonomy compared to BDI, whereas the latter and the OI score very similarly.

From the discussion presented throughout the report we draw the following policy relevant conclusions both at the level of the EU and at that of the Netherlands.

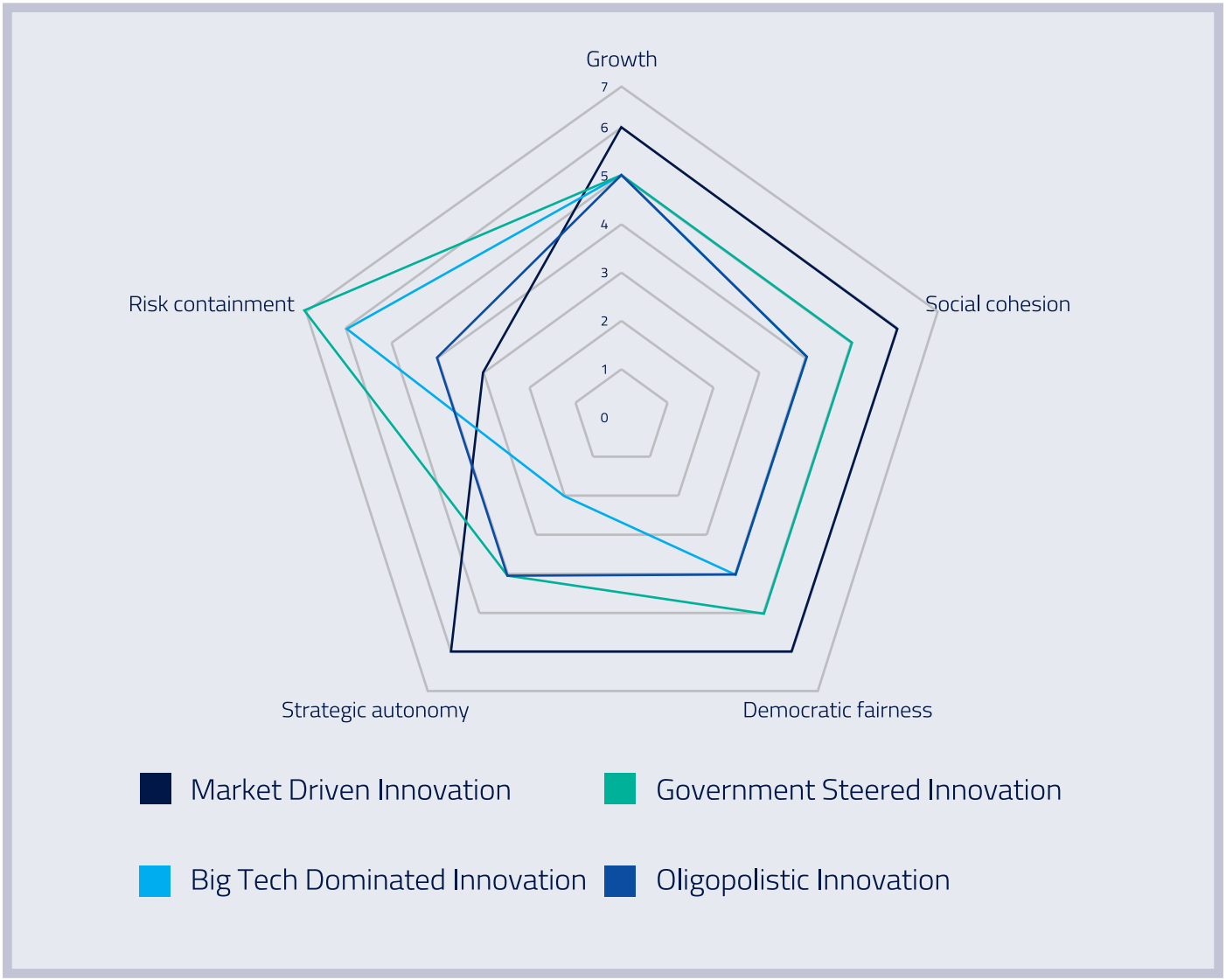

Figure 14 Radar diagram assessment: objective dimensions , Source: Own elaboration

CONCLUSIONS AND RECOMMENDATION

EU CONCLUSIONS

The outlook of Generative AI in Europe presents a landscape marked by transformative potential, regulatory challenges, and the quest for industry leadership. As Europe navigates the evolving dynamics of Generative AI, several key factors can be derived from the analysis and the scenarios presented in this report:

- Generative AI holds promises of enhancing productivity, quality of output, and potentially improving work quality across various job tasks. While uncertainties persist regarding job displacement and societal implications, early indications suggest positive impacts on economic growth and job functions
- Europe currently lags behind global competition in Generative AI, with a significant disparity in funding allocation to European start-ups compared to US and Chinese counterparts. The absence of a cohesive EU-wide AI initiative poses challenges for Europe to assert itself with 'Made in Europe' AI products and infrastructure.
- Besides the more tangible economic implications, without a strong position in GAI Europe's strategic autonomy would be weakened and, at the same time, European countries would have to accept, and adapt to, GAI solutions culturally and ethically embedded in other parts of the world. This, among other things, could impact the language richness of Europe for even when adapted GAI solutions will be English-driven and would lose the linguistic nuances of the many other European languages.
- Boosting funding is essential, but it should target open ecosystems and collaboration, not just individual big tech players. Initiatives like the proposed GenAI4EU, which creates hubs across sectors, can help democratize access to resources. The packages unveiled by the European Commission on January 24, 2024, to support Artificial Intelligence start-ups and SMEs is a first step but more needs to be done and the funds envisaged are not sufficient.
- Leading only on regulation and on the expected Brussels effect that this produce is clearly not sufficient to fill the current gaps. Regulation is important to create certainty and ensure reduction of risks to society, but without a supporting strategy of investments might potentially further increase the existing gaps. Policymakers are urged to adopt proactive measures that balance innovation with ethical considerations. Collaborative platforms can facilitate dialogue on mitigating risks and maximizing opportunities associated with AI technologies. While regulation is crucial to mitigate risks, over-regulation may stifle innovation, especially for smaller startups. It would be useful to create regulatory sandboxes and test beds for European Generative AI researchers and companies to experiment responsibly, alongside the EU AI Act.
- Beside regulatory initiatives such as the AI Act, there is also a need to address uncertainties regarding competition policy and anti-trust initiatives within the regulatory framework of Generative AI. Risks of unfair competition practices by incumbents in the Generative AI market, such as bundling products or engaging in exclusive dealing must be mitigated. Potential entry barriers for new entrants must be lowered by providing access to talent and computational power.

Going into more specific aspects, we present the following five conclusions:

1. **Regulation as co-creation.** Work on regulating GAI, including operationalizing legislation through standards, norms and regulatory oversight, should continue but should better involve the emerging industries in a co-creation process based on learning-by-doing and steady fine-tuning through sandboxes and other forms of experimentation.

2. **Mission-oriented approach at EU level.** The future of GAI impacts European sovereignty and therefore must be integral to an overall EU strategy rather than relying on disconnected initiatives by Member States. A coordinated GAI industrial policy with ad hoc funding mechanisms should be put in place to support and connect the already existing poles of excellence in Europe. The focus should be holistic, driven by a long-term vision on AI, lead to globally competitive excellence and include model development, computing infrastructure, data, skills, and research and innovation.
3. **New R&D instruments.** Traditional R&D instruments do not work in the fast-moving world of AI; instruments should be modernised and more results driven. For example a true large scale Generative AI Challenge should be set up with a budget of at least 0.5 Billion Euro.
4. **High Performance Computing.** The European High Performance Computing Joint Undertaking (EuroHPC JU) has had so far very positive impacts in Europe. The focus, however, has been mostly on HPC for the so called 'hard sciences' (physics, chemistry, biology, medicine, etc.), whereas there is the need of a HPC initiative specifically targeting computer scientists, GAI model developers, and industry.
5. **Beyond LLMs.** GAI is more than just scaling LLMs. More efforts should go into combining foundation models with semantic modelling techniques to support more explicit modelling of the reasoning state, for example to enable reasoning explanations, as well as effort and research into alternative foundation models architectures such as federated specialised smaller scale models.

NETHERLANDS CONCLUSIONS/ RECOMMENDATIONS

As this report is a co-production between EIT Digital and AiNed, the public-private programme to boost AI in the Netherlands, we have a specific perspective on GAI in the Netherlands. Many of the above conclusion and recommendations considered from a European perspective are also applicable for the Dutch situation. Nevertheless, we also add some recommendations specific for the Netherlands. The regulatory framework for AI, adopted on March 13, 2024 by the European Parlement provides a context to GAI development that is not specific for the Netherlands, and neither are the accompanying standards, methods for regulatory oversight and regulatory sandboxes which are all more or less determined at European or global level. However, where rules are not sufficient or unspecifiable, a national approach to developing and validating tools and methods for trustworthy GAI is feasible, for example by the Dutch ELSA labs network. Similarly, the benefits of GAI for improving productivity and quality accross all sectors holds as much for the Netherlands as it does for Europe, but the Netherlands may benefit particularly when applied to Dutch key sectors such as health and care, mobility, energy and sustainability, and technical industry.

For a relatively small European country, with a limited ecosystem and industry position, it is important for the Netherlands to build out strengths together with European partners and connect to EU initiatives, particularly in fields where global competition plays a significant role. For example, by creating synergies between the Dutch open language model GPT-NL and the European Alliance for Language Technologies EDIC. To benefit from this collaboration, it is vital for the Netherlands to involve and support Dutch GAI industry and create new private or public-private initiatives where the Netherlands needs a stronger strategic position, for example initiatives for frontier model developers similar to the French Mistral and the German Aleph Alpha. Folllowing the scenario analysis and policy conclusions the recommendations below can be drawn for the Netherlands:

1. Although there are a couple of general AI initiatives like the NL AIC and the National Growth Fund AiNed program, there is little dedicated effort on Generative AI other than a relatively small investment in GTP-NL and the Dutch participation in the ALT EDIC. Also, the investments from the Dutch industry in collaborative initiatives are rather limited. Therefore, to strengthen the Dutch technology and industry position with

respect to Generative AI, the Dutch government together with the Dutch industry should significantly step up the Generative AI investments in a public-private innovation partnership.

2. The Netherlands has a strong start-up ecosystem, also when it comes to AI start-ups and start-ups focussing on (the application of) Generative AI, Large Language Models cq. Foundation Models. These start-ups however suffer from lack of resources (talent, capital, data, computing power) to grow to internationally operating companies, or are incorporated or bought by non-Dutch companies and investors before or during scaling internationally. To support these start-ups and thus contribute to more technological sovereignty, a national task force should be tasked with the coordination of special public and private investments in high potential start-ups in the domain of Generative AI should be mobilised by combining regional, national, and European investors and investment instruments.
3. Access to High Performance Computing Infrastructure is a bottleneck for Dutch SMEs and start-ups and to some degree even to Dutch multinationals. The Dutch government should therefore in collaboration with Dutch industry establish a national Generative AI test facility to equally support the Dutch industry (notably the SME and start-ups community) as well as the Dutch research community to develop and experiment with Generative AI technology. Such a facility should be embedded in the EU EuroHPC Joint Undertaking.
4. Where the Dutch R&D instruments are well developed, the innovation instruments and the transition instruments are less developed. This results in a strong knowledge infrastructure, but in a relatively weak innovation infrastructure, especially in the digital domain. The more traditional R&D instruments work well to establish and maintain long term knowledge development but are less suited to serve the needs of innovators in fast-moving domains like digital and especially (generative) AI. The Dutch government should therefor revise its innovation approach, policy, and instruments in order to create strong technology and industry positions in emerging and fast-moving digital domains, such as Generative AI. Key ingredients of such an approach should be: tighter integration of the knowledge institutes and innovation communities, e.g. through orchestration and direct funding of academic spin-offs, attracting and retaining excellent talent, better alignment of investment instruments, result-based value-driven and risk-prone financing, stronger focus on economic impact and growth.

ACKNOWLEDGEMENTS

This report on Generative AI: Europe's Quest for Regulation and Industry Leadership is part of the EIT Digital Makers & Shapers report series. These reports address specific aspects of digital technologies and developments. They follow a scenario-based approach, grounded on the developments and state of the art in the specific field, and analyse the impact of the different scenarios on specific predefined indicators.

This report that has been developed together with AiNed and with the support of the EIT Digital-led EIT AI Community. The report dives into Generative AI (GAI), exploring its rapid rise and influence on the wider realm of AI. Through scenario-based analysis, the report assesses potential outcomes across various GAI trajectories. In the midst of evolving regulatory landscapes, such as the EU's efforts to integrate GAI into frameworks like the Artificial Intelligence Act, the report offers actionable insights for decision-makers, helping them navigate the potential impacts of policy decisions on the trajectory of GAI and the broader AI landscape.

Cristiano Codagnone was contracted to support the study and write this report under the guidance and supervision of EIT Digital and AiNed. We acknowledge the contribution of Cristiano Codagnone, assisted by Giovanni Liva, by providing breadth and depth to the study via interdisciplinary stakeholder discussions, as well as extensive literature analysis and survey. Special thanks go to Stefan Leijnen from AiNed. Not at last, we would like to thank the members of our experts' rountable that took place on March 6 and 7 2024 for reviewing the report and providing invaluable input to this study: Arash Hajikhani, Research Team Leader - Quantitative science and technology studies, VTT - Technical Research Centre of Finland; Bálint Árpád Pataki, Accredited Parliamentary Assistant to the Member of the European Parliament Eva Maydell, European Parliament; Benjamin Larsen, Lead - Artificial Intelligence and Machine Learning, World Economic Forum; Brando Benifei, Member of the European Parliament; Daniel Abbou, Managing Director, German AI Association; Professor Dominic Chalmers, University of Glasgow; Edwin M. Colella, Chief Sales and Marketing Officer, Wallife; Emanuela Girardi, President, Adra - AI, Data and Robotics association; Frauke Goll, Managing Director, Applied AI; Professor Fredrik Heintz, Linköping University; George Metakides, President, Digital Enlightenment Forum; Professor of Artificial Intelligence and Computer Science Holger Hoos, RWTH Aachen University; Igor Carron, CEO and Co-founder, Lighton; Jeannette Gorzala, Vice President, European AI Forum; Jim Spohrer, Member of the Board of Directors of the non-profit ISSIP and ServCollab; Kai Zenner, Head of Office and Digital Policy Adviser to Member of the European Parliament Axel Voss, European Parliament; Niels Taatgen, Research Director, University of Groningen; Onno Zoeter, Principal Data Scientist, Booking.com; Orestis Trasanidis, Coordinator, EIT AI Community; Philipp Hutchinson, Senior AI Strategist, Applied AI; Sabine Demey, Program Director, Flanders Artificial Intelligence Research Program; Slavina Ancheva, Team Leader and Policy Adviser to Member of the European Parliament Eva Maydell, European Parliament; Tero Ojanperä, Co-Founder and Executive Chair, Silo AI; Professor of Artificial Intelligence and Computer Science Tom Heskes, Radboud University; Vangelis Karkaletsis, Director of the Institute of Informatics & Telecommunications, NCSR Demokritos.

Brussels, March 2024

REFERENCES

1. Knott, A., Pedreschi, D., Chatila, R. et al. (2023). Generative AI models should include detection mechanisms as a condition for public release. Ethics Inf Technol 25, 55 (https://doi.org/10.1007/s10676-023-09728-4).
2. Eloundou, T., S. Manning, P. Mishkin and D. Rock (2023) 'GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models', mimeo, available at http://arxiv.org/abs/2303.10130
3. Hatzius, J., J. Briggs, D. Kodnani and G. Pierdomenico (2023) 'Global Economics Analyst The Potentially Large Effects of Artificial Intelligence on Economic Growth', Global Economics Analyst, 26 March, Goldman Sachs Economics Research
4. Ambrosetti & Microsoft. (2023). AI 4 Italy: impatti e prospettive dell'intelligenza artificiale generativa per l'Italia e il Made in Italy (https://www.ambrosetti.eu/news/ai-4-italy-impatti-e-prospettive-dellintelligenza-artificiale-generativa-per-litalia-e-il-made-in-italy/).
5. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38.
6. OECD (2023). AI Language Models: Technological, Socio-Economic and Policy Consideration', OECD Digital Economy Papers 352 (https://doi.org/10.1787/13d38f92-en).
7. In six experiments, participants (N = 4,600) were unable to detect self-presentations generated by state-of-the-art AI language models in professional, hospitality, and dating contexts, see Jakesch, M., Hancock, J. T., & Naaman, M. (2023). Human heuristics for AI-generated language are flawed. Proceedings of the National Academy of Sciences, 120(11), e2208839120. Similar results are reported also in Waltzer, T., Cox, R. L., & Heyman, G. D. (2023). Testing the Ability of Teachers and Students to Differentiate between Essays Generated by ChatGPT and High School Students. Human Behavior and Emerging Technologies, 1923981, doi:10.1155/2023/1923981
8. Knott, A., Pedreschi, D., Chatila, R. et al. (2023). Generative AI models should include detection mechanisms as a condition for public release, op. cit., p. 2.
9. Wang, C., Liu, S., Yang, H., Guo, J., Wu, Y., & Liu, J. (2023). Ethical Considerations of Using ChatGPT in Health Care. J Med Internet Res, 25, e48009. doi:10.2196/48009.
10. In March 2023 the first brief from the Parliament anticipated what was to come later: see Madiega, T. General Purpose Artificial Intelligence, EPRS | European Parliamentary Research Service, March 2023 (https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/745708/EPRS_ATA(2023)745708_EN.pdf). The brief stressed the risks of Generative AI and the need to regulate them. The whole legislative process of the AI Act and the new perspective were then presented in another brief released in June 2023: see Madiega, T., EU Legislation in Progress: Artificial Intelligent Act, EPRS | European Parliamentary Research Service, June 2023 (https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf).
11. See: Madiega, T., Parliament's negotiating position on the artificial intelligence act, EPRS | European Parliamentary Research
12. https://arxiv.org/abs/1706.03762
13. https://www.asimovinstitute.org/overview-neural-network-zoo/
14. Rae, J.W., S. Borgeaud, T. Cai, K. Millican, J. Hoffmann, F. Song ... G. Irving (2021). Scaling Language Models: Methods, Analysis & Insights from Training Gopher (http://arxiv.org/abs/2112.11446).
15. Schick, T. and H. Schütze,(2021). 'It's Not Just Size That Matters: Small Language Models Are Also Few- Shot Learners', Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (https://doi.org/10.18653/v1/2021.naacl-main.185).
16. Hu, E.J., Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang and W. Chen (2021). LoRA: Low-Rank Adaptation of Large Language Models (http://arxiv.org/abs/2106.09685). It is worth mentioning also QLoRA which is a more recent technique for efficient fine-tuning. It significantly reduces memory requirements for fine-tuning by quantizing the model's weights into lower width bits.
17. The developers used used the open-source model LLaMA and fine-tuned it on high-quality open-source datasets containing dialogue data from the web, including dialogue with LLMs, such as ChatGPT. See: Geng, X., A. Gudiband, H. Liu, E. Wallace, P. Abbeel, S. Levine and D. Song (2023). Koala: A Dialogue Model for Academic Research', Berkeley Artificial Intelligence Research, 3 April (https://bair.berkeley.edu/blog/2023/04/03/koala/).
18. Data reported in Carugati, C. (2023). Competition in generative artificial intelligence foundation models, Working Paper 14/2023, Bruegel (https://www.bruegel.org/

sites/default/files/2023-09/WP%2014.pdf), p. 7.
19. Carugati, C. (2023). Competition in generative artificial intelligence foundation models, op. cit., p. 7.
20. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., ... & Natarajan, V. (2022). Large language models encode clinical knowledge. arXiv preprint arXiv:2212.13138.
21. Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Harvard Business School Technology & Operations Mgt. Unit Working Paper, (24-013).
22. Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Available at SSRN 4375283.
23. Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of ai on developer productivity: Evidence from github copilot. arXiv preprint arXiv:2302.06590.
24. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation, op.cit.
25. Mons, B. (2023). Does the hype of Generative AI need top-down regulation, or will it implode? Frontiers | Policy Labs (https://policylabs.frontiersin.org/content/commentary-does-the-hype-of-generative-ai-need-top-down-regulation-or-will-it-implode).
26. Ibid.
27. Ibid.
28. https://a16z.com/who-owns-the-generative-ai-platform/
29. Source: Dealroom.co.
30. https://www.economist.com/the-world-ahead/2023/11/13/generative-ai-will-go-mainstream-in-2024
31. https://info.kpmg.us/news-perspectives/technology-innovation/kpmg-usexecutives-genai-2023.html
32. https://iot-analytics.com/leading-generative-ai-companies/
33. https://www.reuters.com/technology/microsofts-deal-with-mistral-ai-faces-eu-scrutiny-2024-02-27/.
34. See CMA press release of 4 May 2023 (https://www.gov.uk/government/news/cma-launches-initial-review-of-artificial-intelligence-models).
35. See Sheppard Mullin Richter & Hampton LLP, AI Under the Antitrust Microscope: Competition Enforcers Focusing on Generative AI from All Angles, August 9 2023, Lexology (https://www.lexology.com/library/detail.aspx?g=5604adda-62c0-4ffe-8007-94b802d4b81c).
36. Dickerson, J. and Novak, A. FTC Chair Lina Khan says AI could "turbocharge" fraud, be used to "squash competition", CBS News July 27, 2023 (https://www.cbsnews.com/news/lina-khan-ftc-chair-federal-trade-commission-ai-turbocharge-fraud-squash-competition/).
37. Federal Trade Commission (FTC), Generative AI Raises Competition Concerns, June 29, 2023 (https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns).
38. Federal Trade Commission (FTC), Generative AI Raises Competition Concerns, op. cit.
39. Carugati, C. (2023). Competition in generative artificial intelligence foundation models, op. cit.
40. As reported in Carugati (2023), around 37,000 open-source datasets are available on Hugging Face (https://huggingface.co/datasets?sort=downloads).
41. Chui, M. et al (2023). The economic potential of Generative AI. McKinsey & Company (https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/the%20economic%20potential%20of%20generative%20ai%20the%20next%20productivity%20frontier/the-economic-potential-of-generative-ai-the-next-productivity-frontier-vf.pdf?shouldIndex=false).
42. Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. arXiv preprint arXiv:2303.10130.
43. Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of ai on developer productivity: Evidence from github copilot. arXiv preprint arXiv:2302.06590.
44. Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Harvard Business School Technology & Operations Mgt. Unit Working Paper, (24-013).
45. Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work (No. w31161). National Bureau of Economic Research.
46. Eisfeldt, A. L., Schubert, G., & Zhang, M. B. (2023). Generative ai and firm values (No. w31222). National Bureau of Economic Research.
47. McKinsey & Company. (2023). The state of AI in 2023: Generative AI's breakout year (https://www.mckinsey.com/~/media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai%20in%202023%20generative%20ais%20breakout%20year/the-state-of-ai-in-2023-generative-ais-break-

out-year-v3.pdf?shouldIndex=false).
48. https://www.mckinsey.com/industries/healthcare/our-insights/tackling-health-cares-biggest-burdens-with-generative-ai
49. https://www.himss.org/global-conference
50. Wang, C., Liu, S., Yang, H., Guo, J., Wu, Y., & Liu, J. (2023). Ethical Considerations of Using ChatGPT in Health Care, op. cit.
51. https://www.latitudemedia.com/news/seven-ways-utilities-are-exploring-ai-for-the-grid
52. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? . In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623).
53. https://www.theverge.com/2020/12/5/22155985/paper-timnit-gebru-fired-google-large-language-models-search-ai
54. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.
55. Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., ... & Gabriel, I. (2021). Ethical and social risks of harm from language models. arXiv preprint arXiv:2112.04359.
56. OECD (2023). AI language models. Technological, socio-economic and policy considerations. OECD Digital Economy Papers.
57. Lucy, L., & Bamman, D. (2021). Gender and representation bias in GPT-3 generated stories. In Proceedings of the Third Workshop on Narrative Understanding.
58. Stanford University (2023): https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf
59. Kremer, A. et al (2023). As gen AI advances, regulators—and risk functions—rush to keep pace. McKinsey & Company (https://www.mckinsey.com/~/media/mckinsey/business%20functions/risk/our%20insights/as%20gen%20ai%20advances%20regulators%20and%20risk%20functions%20rush%20to%20keep%20pace/as-gen-ai-advances-regulators-and-risk-functions-rush-to-keep-pace_final.pdf?shouldIndex=false).
60. Drake, M. et al. From Washington to Brussels: A Comparative Look at the Biden Administration's Executive Order and the EU's AI Act, October 30 2023, Inside Privacy (https://www.insideprivacy.com/artificial-intelligence/from-washington-to-brussels-a-comparative-look-at-the-biden-administrations-executive-order-and-the-eus-ai-act/).
61. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/
62. See https://www.techpolicy.press/the-many-questions-about-indias-new-ai-advisory/ and https://dig.watch/updates/indias-it-ministry-issues-advisory-on-approval-and-labelling-of-ai-tools
63. https://www.un.org/en/ai-advisory-body
64. https://gpai.ai/
65. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_5379
66. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX-:52021PC0206.
67. https://ainowinstitute.org/publication/gpai-is-high-risk-should-not-be-excluded-from-eu-ai-act
68. https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf
69. Ibid., Art 3(1b).
70. https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf.
71. Ibid., Article 3(1c).
72. https://oecd.ai/en/wonk/foundation-models-eu-ai-act-fairer-competition
73. The following is based on the contents of the European Parliament press release (https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai) as the text of the agreement has not been published yet.
74. See: https://artificialintelligenceact.com/updates/#:~:text=Representatives%20from%20member%20states%20unanimously,agreement%20reached%20in%20December%202023.
75. Ibid.
76. https://digital-strategy.ec.europa.eu/en/policies/ai-pact
77. According Kai Zenner in his lecture at the cited Digital Humanism Lecture Series of 19 December 2023. (https://www.youtube.com/digitalhumanism).
78. George Metakides, in his lecture at the cited Digital Humanism Lecture Series of 19 December 2023. (https://www.youtube.com/digitalhumanism).
79. This is part of a statement solicited by George Metakides and read in his lecture at the cited Digital Humanism Lecture Series (https://www.youtube.com/digitalhumanism). Anu Bradford expressed similar views also in various other public statements, see for instance "EU Establishes World-Leading AI Rules, Could That Affect Everyone?", VOA, 11 December 2023 (https://www.voanews.com/a/eu-establishes-world-leading-ai-rules-could-that-affect-everyone-/7393780.html).

80. https://www.bruegel.org/analysis/adapting-european-union-ai-act-deal-generative-artificial-intelligence#footnote4_30gc309
81. See for instance the examples from Stanford and the OECD available here: https://crfm.stanford.edu/2023/06/15/eu-ai-act.html#fn:1 and here https://oecd.ai/en/wonk/foundation-models-eu-ai-act-fairer-competition.
82. Soete, L. (2023). FAIR use in Artificial Intelligence? Access to data for the benefit of all, Frontiers | Policy Labs (https://policylabs.frontiersin.org/content/commentary-fair-use-in-artificial-intelligence-access-to-data-for-the-benefit-of-all).
83. Ibid.,
84. Mons, B. (2023). Does the hype of Generative AI need top-down regulation, or will it implode? Op. cit.
85. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 160018. doi:10.1038/sdata.2016.18.
86. Mons, B. (2023). Does the hype of Generative AI need top-down regulation, or will it implode? Op. cit.
87. Ibid.
88. https://leam.ai/feasibility-study-leam-2023/.
89. See https://www.reuters.com/technology/microsofts-deal-with-mistral-ai-faces-eu-scrutiny-2024-02-27/ .
90. Balland, P. & Renda, A. (2023). Forge ahead or fall behind. Why we need a United Europe of Artificial Intelligence, CEPS Explainer (https://cdn.ceps.eu/wp-content/uploads/2023/11/CEPS-Explainer-2023-13_United-Europe-of-Artificial-Intelligence.pdf).
91. https://www.appliedai.de/en/hub-en/ai-act-impact-survey.
92. COM(2024) 28 final, Brussels, 24.1.2024: https://ec.europa.eu/newsroom/dae/redirection/document/101621. See also: https://ec.europa.eu/commission/press-corner/detail/en/ip_24_383.
93. Digital Decade Country Report 2023 https://digital-strategy.ec.europa.eu/en/library/2023-report-state-digital-decade.
94. European Deep Tech Report 2023 https://dealroom.co/reports/the-european-deep-tech-report-2023.
95. ESIR 2023.
96. https://implementconsultinggroup.com/article/the-economic-opportunity-of-ai-in-the-netherlands
97. https://www.digitaleoverheid.nl/wp-content/uploads/sites/8/2019/11/RapportSAPAI.pdf.
98. Balland, P. & Renda, A. (2023). Forge ahead or fall behind, op. cit., p. 5.
99. https://www.bruegel.org/analysis/adapting-european-union-ai-act-deal-generative-artificial-intelligence#footnote4_30gc309.

Publisher
EIT Digital
Rue Guimard 7
1040 Brussels
Belgium
www.eitdigital.eu

Contact
info@eitdigital.eu

ISBN 978-91-87253-71-3

Co-funded by the
European Union