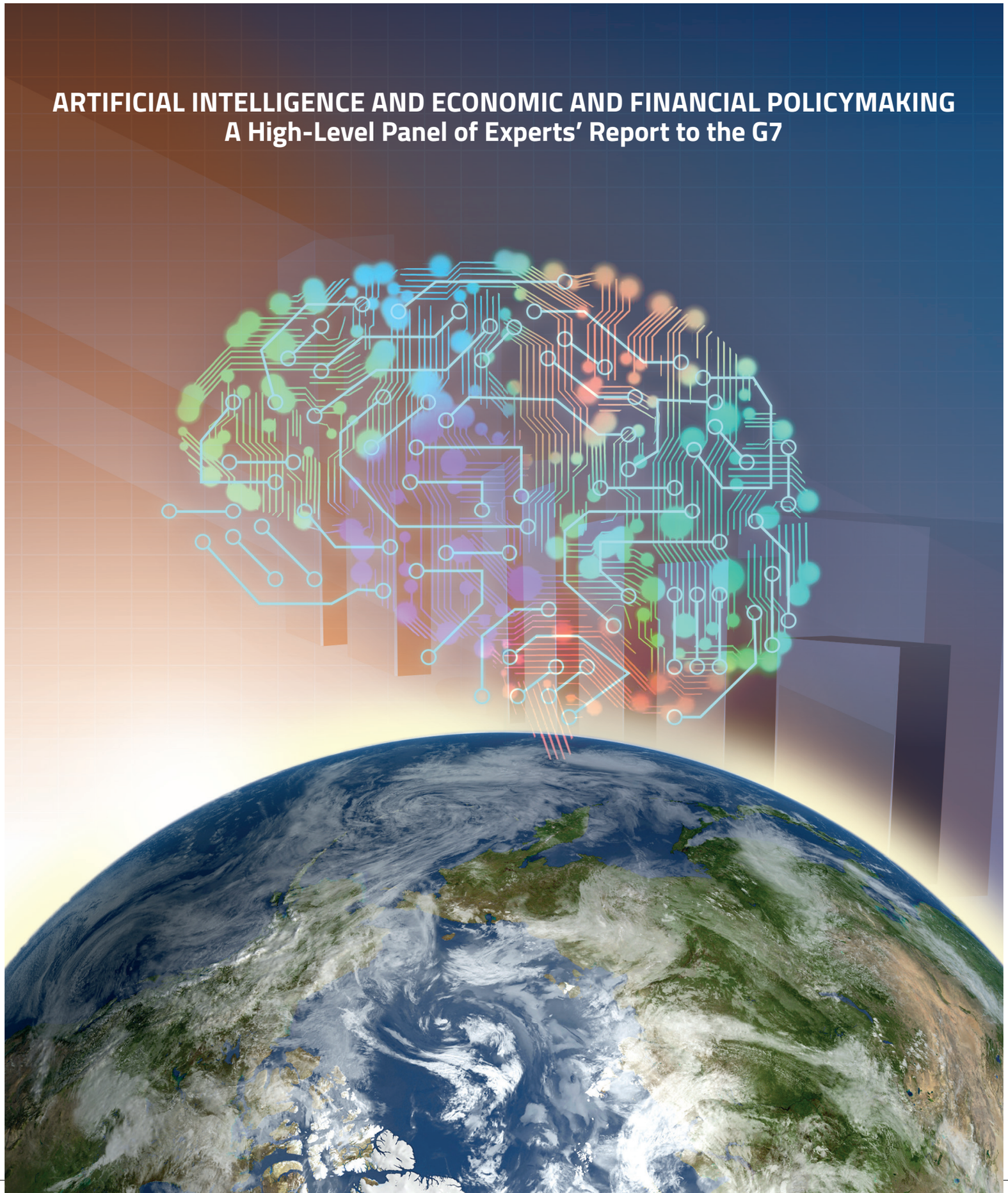
G7 ITALIA 2024

ARTIFICIAL INTELLIGENCE AND ECONOMIC AND FINANCIAL POLICYMAKING
A High-Level Panel of Experts' Report to the G7



ARTIFICIAL INTELLIGENCE AND ECONOMIC AND FINANCIAL POLICY MAKING
A High-Level Panel of Experts' Report to the G7

Report prepared by:

Luis Videgaray (Chair), Philip Aghion, Barbara Caputo, Tracey Forrest, Anton Korinek, Katja Langenbucher, Hiroaki Miyamoto and Michael Wooldridge

DECEMBER 2024

Panel Composition

High-Level Panel of Experts:

Dr. Luis Videgaray. Massachusetts Institute of Technology (Chair)
Prof. Philip Aghion. INSEAD, Special Committee on AI
Prof. Barbara Caputo. DAUIN, Politecnico di Torino
Prof. Tracey Forrest. Centre for International Governance Innovation CIGI
Prof. Anton Korinek. University of Virginia, Brookings Institution
Prof. Dr. Katja Langenbucher. Goethe University Frankfurt
Prof. Hiroaki Miyamoto. Policy Research Institute, Hitotsubashi University
Prof. Michael Wooldridge. Alan Turing Institute.

This report represents a synthesis of the panel members' expertise. However, individual members may not endorse every point or recommendation contained herein.

Acknowledgements

Special thanks are extended to WTO Deputy Director-General Johanna Hill, Tomasz Kozieł, Emmanuelle Ganne, and Eddy Bekkers at the WTO for their willingness to share views and ideas during an interview with the High-Level Panel of Experts, and to the members of the G7 Drafting Team: Claudio Battiati, Oscar Borgogno, Ines Buono, Eva Vittoria Cammerino, Fabio Di Vittorio, and Gabriele Palomba. We wish to express our sincere gratitude to Mr. Phil Suttle for his invaluable contribution as reviewer of the final version of the report.

Outline

Executive Summary

This Report is the **outcome of the work of the High-Level Panel of Experts on Artificial Intelligence (AI)**. The report is **intended to serve as a reference for policymakers**. The Panel was established by the mandate of the Finance Ministers and Central Bank Governors of the Group of 7 (G7), with the objective of exploring the implications of AI for the economy and the financial sector. The full implications of AI are unclear, and we know that there will be winners and losers. Opinions on the best approach to the technology will thus inevitably diverge. There were indeed some differences in views and emphasis among the members of the Panel in the preparation of this Report. Similarly, G7 governments and central banks will likely prefer different approaches to AI, depending on culture, legal systems, national priorities, perception of risk and specific needs. There is no correct one-size-fits-all approach.

AI has emerged as a transformative technology across a spectrum of activities with potential to lead to significant changes in economic structure and financial systems. Being a general-purpose technology with a potential for rapid innovation across a broad array of economic processes, **AI calls for careful management and proactive policies**. Governments have three roles to play in AI development: AI enablement (R&D, education, infrastructure, and financing); the use of AI in government itself; and the enactment of laws and regulations for the private sector, ensuring that the use of AI technologies facilitates governments' objectives of economic growth, stability, equity and well-being.

The speed of innovation and adoption of AI, its subsequent impact on the economy, labor markets, finance, and sustainability is unknown, but could

be rapid. As a result, **policymakers should remain vigilant to AI developments and be prepared to adapt policy flexibly**. AI already poses several urgent questions for policymakers. These include whether scaling laws will continue to hold and how that might impact concentration of market power; the trade-offs between proprietary and open-source systems; the availability of data for AI training; and the sustainability of AI business models due to the high associated capital and energy costs. Regulatory authorities should remain vigilant about market developments, as excessive concentration could reduce innovation and slow the spread of productivity gains.

AI's **macroeconomic impact** will show up in productivity and employment. Estimates of the productivity impact of AI vary widely between 0.1 and 1.2 percentage points per year, depending on AI's influence on task automation and innovation. AI could affect employment through three main potentially interacting channels: a displacement channel, where AI-driven automation replaces jobs; a productivity channel, where AI increases overall efficiency; and a reinstatement channel, where AI creates new job opportunities, particularly in industries where human labor maintains a comparative advantage. Whether AI will complement or replace workers is an open issue. As AI capabilities advance, ever more sophisticated tasks could be automated, potentially increasing job losses in both cognitive and manual occupations and further exacerbating income and wealth inequalities.

Financial agencies and authorities could ultimately use AI in many ways, enhancing tax systems and policy design. Government budgeting

and expenditure could also benefit from AI-driven enhanced data-driven budgeting, fraud detection and program evaluations.

Rapid advances in AI also accelerate known challenges to **financial stability**.[1] Widespread reliance on similar AI systems in finance could lead to correlated failures and amplified shocks. Automated herding behavior could also emerge, leading to increased market volatility and cyclicality. In addition, AI's ability to process large amounts of data could enable sophisticated forms of manipulation and algorithmic collusion, or even cyber-attacks. Finally, a divide between AI-leading and lagging countries may result in financial imbalances, capital inflows, and exchange rate volatility. However, AI in finance could also bring opportunities, such as increased efficiency in financial markets, risk management, fraud identification, and an enhanced ability to identify cybersecurity threats and respond more quickly to such threats.

Against this background, **the key guiding principle informing policymakers should be policy preparedness**. Governments and central banks must be ready to act promptly, based on alternative AI development scenarios. In the concluding Chapter of this Report, we outline three possible scenarios: a “Conventional Wisdom”, in which gradual AI progress generates limited economic disruption; an “Intermediate Artificial General Intelligence (AGI)” scenario, in which AGI will near human-level performance in 20 years; an “Accelerated AGI” scenario, with rapid AGI development (3-5 years) leading to dramatic economic changes. Based on these scenarios, and their very different implications, we make the following advisory/non-binding

[1] Gensler and Bailey (2020).

recommendations: i) adopt a proactive and flexible policy stance; ii) understand policy trade-offs (both technical and incentive); iii) rapidly augment AI expertise; iv) continuously explore new AI capabilities; v) develop a coherent procurement strategy for AI tools; vi) adopt modern government data-management tools; vii) act on cybersecurity enhancement immediately; viii) create guidelines for AI tools; ix) monitor labor market impacts; x) enhance cross-border cooperation. Finally, we propose a policy preparedness matrix, as an aid to think through AI policy options.

Introduction

1. Why should economic and financial policymakers think about AI policy?

AI has emerged as a transformative technology across a wide spectrum of activities, from routine and repetitive tasks to knowledge-based and creative, and sectors, including finance and economics. As development and dissemination of this technology continue to expand at an unprecedented rate, it will lead to significant changes in economic structure and financial systems. AI is a general-purpose technology that will likely influence—and perhaps transform—most economic processes and activities, introducing new policy challenges. AI will also provide new tools that will enhance the capacity of institutions to design and implement economic policies. Many policy issues related to AI are not new, but it is now necessary to provide new answers to old questions since AI is such a transformative technology. Moreover, policymakers need to consider AI policy to ensure that technological advancements contribute positively to economic growth, stability, equity and well-being in general. By proactively addressing AI-related challenges and opportunities, policymakers will be in a better position to guide the integration of AI in a manner that maximizes benefits while mitigating risks.

In the current fast-changing environment, a high priority should be given to ensuring a positive impact on growth, employment and the labor market, maintaining the stability of the financial system, fostering inclusive AI development and deployment, and facilitating the technological transition (accompanied by workers' support). That said, AI's evolution and the future uses are surrounded by great uncertainty, both from the perspective of technology and of market structure. A central point to this Report is that AI policy in the economic and financial realm should be designed by fully acknowledging what we don't know yet about AI itself and the industrial organization of AI deployment. At a time when even experts disagree on fundamental questions about the immediate and long-term future of AI, policymakers do not have the luxury of waiting to learn what the future will be. Policymakers should prepare for different scenarios and preserve policy optionality.

2. Objectives of AI policy in finance and economics

Promoting economic growth, fostering human well-being, enhancing financial market efficiency, improving risk management mechanisms to prevent economic crises, protecting financial consumers and investors, enhancing data privacy, and mitigating biases, and upholding regulatory compliance are just as important in the age of AI as before. However, AI is poised to transform productivity, consumption, investment, income and wealth distribution as well as labor markets, all of which could have direct and indirect effects on financial stability. AI's widespread adoption could also enhance firms' ability to quickly adjust prices in response to macroeconomic changes, with repercussions for inflation dynamics. Therefore, while AI does not fundamentally change policy objectives, it creates new challenges and amplifies existing ones toward achieving those objectives. However, AI also introduces new tools for policymaking to achieve its objectives, with the potential to significantly enhance its effectiveness.

3. Who should read this?

This Report is written for a non-technical audience. It is an interdisciplinary Report written for decision makers by computer scientists, economists, and policy experts. The goal is for it to be a useful resource for policymakers to think through the policy challenges and options available. While the focus is primarily on the policy challenges faced by G7 finance ministries and central bankers, the Report is intended to be useful to anybody interested in issues related to the impact of AI on economics and finance.[2]

4. Foundation framework

As a G7 document, this Report builds upon the **G7 Hiroshima AI Process Comprehensive Policy Framework,** launched in May 2023, in the aftermath of the 2023 G7 Leaders' Summit under the Japanese Presidency, with the aim of promoting safe, secure and trustworthy AI. The Hiroshima Framework started a process to establish a set of international rules that act as the basis for an inclusive global governance of AI, in order to maximize its innovative opportunities, while mitigating the risks and challenges arising from advanced AI systems. It is currently composed by

[2] This report is also not intended to be an exhaustive primer on AI concepts and definitions. We focus on the most relevant concepts and direct readers looking for additional context to the references section.

the “Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI system”, addressed to all relevant actors in the AI life cycle, and the “Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems”, more specifically addressed to AI developers (see Annex A).

The Italian 2024 G7 Presidency also choose AI as one of its key priorities. The **Apulia G7 Leaders’ Communique** (see Annex B) emphasizes the importance of promoting safe, secure, and trustworthy AI and of pursuing an inclusive, human-centered, digital transformation that underpins economic growth and sustainable development. Moreover, G7 Leaders pledged to ensure that AI enables increased productivity, quality jobs, and decent work and, acknowledging the key role that emerging technologies can play in economic growth, committed to enhancing cooperation to bolster the adoption and development of new technologies, including AI, among micro, small, and medium enterprises, thereby fostering inclusive economic growth.

This Report from the G7 High-Level Panel of Experts on Artificial Intelligence draws its mandate from the discussion on the implications of AI for the economy and the financial sector that the G7 Finance Ministers and Central Bank Governors (FMCBGs) had during the **Stresa and Washington Meetings** (see Annex C). They agreed to analyze the issue of AI to ensure that economic and financial institutions are well equipped to deal with AI, with a focus on several key topics: macroeconomic impact of AI, competition in the AI sector, measurement challenges, financial stability, cybersecurity in the financial sector, fiscal policy, environmental sustainability, and a series of general, government-wide factors. These issues are the focus of this Report.

More broadly, this Report is also inspired by existing documents on AI policy at the international level. For example, the **OECD AI Principles**[3], first adopted in 2019 and updated in 2024, are one of the most relevant among these early documents on AI. This OECD Council recommendation includes five value-based principles (Inclusive growth, sustainable development, and well-being; Human rights and democratic values, including fairness and privacy; Transparency and explainability; Robustness,

[3] OECD (2019), Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

security and safety; Accountability) and five related recommendations for policymakers. The **UNESCO Recommendation on the Ethics of Artificial Intelligence**[4] was adopted by all 193 Member States in November 2021. It includes extensive Policy Action Areas, to help translate the values and principles it promotes into action with respect to data governance, environment and ecosystems, gender, education and research, and health and social wellbeing. The **Bletchley Declaration**[5] on AI safety was signed by 28 countries in 2023 as an outcome of the first global meeting on AI safety. This Declaration acknowledges the collective necessity to understand and manage the potential risks of AI and to ensure that it is developed and used safely, responsibly and to the benefit of the global community.

5. Report focus on implications for the economy and the financial sector

The mandate from the G7 FMCBGs is strictly focused on issues involving the economic system and the financial sector. Consequently, this Report is not intended to assess general or purely technical aspects of the development and implementation of AI systems. It is concerned with the key policy challenges that emerge from AI for economy and finance, aiming to give an overview of the current debates in this area and offer a policy preparedness framework for economic and financial policy decision makers.

6. Report Chapters

The remainder of this Report is structured as follows: Chapter 1 will address the issue of the technical and market structure uncertainties that policymaking has to face; Chapter 2 will assess the macroeconomic impact of AI, including on innovation and growth, jobs and labor productivity, and the distributional effects, while considering measurement challenges; Chapter 3 will discuss the use of AI by financial agencies and authorities; Chapter 4 is about AI and financial stability; and Chapter 5 presents the policy preparedness framework based on scenario analysis.

[4] UNESCO, Recommendation on the Ethics of Artificial Intelligence, Recommendation on the Ethics of Artificial Intelligence - UNESCO Digital Library.

[5] The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023 - GOV.UK.

1. Policymaking under uncertainty: unanswered questions about AI

AI is now top of mind everywhere. While AI has been attracting major academic and industry attention since the emergence of deep learning as the dominant paradigm in the early 2010s, the release of OpenAI's ChatGPT in 2022 was a tipping point that captured the world's imagination. Today, it is hard to find an academic, industry or government gathering in which AI is not part of the conversation. Almost every organization is expected to have a plan to seize the opportunities and address the challenges brought by AI. Public and private financial markets allocate an unprecedented amount of capital towards AI innovation and deployment. It is therefore remarkable that, while we think and discuss AI almost everywhere, there are fundamental things that we don't fully understand about AI—at least not yet. The height of interest in AI is also a high point in experts' disagreements about where AI is going—and how. In this Chapter we summarize some of these critical unanswered questions, with a focus on those that introduce relevant uncertainties for economic and financial policymakers.

1.1. The non-linear emergence of AI

Artificial Intelligence is far from a new field. It originated just after the end of the Second World War, concurrently with the advent of the first digital computers. By the mid-1950s, it had become an established scientific discipline, and it has been studied continually since then. But progress in AI has been far from steady or predictable. Asked to evaluate the rate of progress at any point between 1970 and 2005, most honest evaluators would have assessed progress as "incremental" at best. But since around 2005, progress in AI appears to have accelerated. Since 2020, it seems that the rate of acceleration has accelerated. What does this mean and where is it going?

When we talk about AI in 2024, the main category of systems that come to mind are based on neural networks. Neural networks attempt to replicate in computer software some of the nerve cell structures that we see in human and animal brains. It was obvious by the 1960s that this was possible, but progress was held by partly for scientific reasons (we didn't know how to "train" these artificial neural networks), but more mundanely, because to do anything useful, neural networks need to be *big*—and

computers in the 1960s simply weren't powerful enough. By the 1980s, scientific advances had shown us how to train neural networks, and as a consequence there was a spurt of activity around the technology. Progress stalled again, however, because computers still were not powerful enough.

That started to change this century, as computer power became sufficiently cheap that it became feasible to build neural networks big enough to solve problems of practical importance (e.g., recognising faces in photographs, or identifying tumours on X-ray scans). In 2012, progress intensified when it was realised that Graphics Processing Units (GPUs), a technology originally developed to enable high-quality graphics in computer games, were ideally suited to doing the computations required to configure neural networks (Krizhevsky et al, 2017). At this point, massive investment started to flow from the world's leading technology companies, who could see the potential of the new technology and were eager to secure a competitive advantage. A flood of applications followed. An example, was the AlphaFold system, which predicts the shape of proteins (its creators received the Nobel Prize this year). This accelerated the rate of progress in other areas of science[6]. Current activity in AI has been made possible by the combination of very cheap computer power and data drawn from social media and the World-Wide Web.

By 2020, it became clear that one particular AI technology—Large Language Models—responded unexpectedly well to this race to scale. In June 2020, OpenAI released GPT-3, the third in their series of GPT-class AI systems[7]. While its predecessor, GPT-2, had been released 18 months previously, it was clear that GPT-3 represented a significant *and unexpected* step change in ability. Put simply, it was much more capable than its predecessor. And its ability caught the AI community—and the world's richest companies—by surprise. GPT-3 was the technology underneath ChatGPT, the first general purpose AI tool to reach a global audience, and the most rapidly adopted online tool in history[8].

---

[6] AlphaFold - Google DeepMind

[7] GPT-3 powers the next generation of apps | OpenAI

[8] ChatGPT

The history of AI was thus marked by a long period of very little progress; about 15 years of increased progress; and the past five years of unexpectedly rapid progress.

It is important to understand the enablers of the current progress. The investment required to build a cutting-edge foundation model is extraordinary[9]. According to the Stanford AI Index 2024, the training costs of state-of-the-art AI foundation models have reached unprecedented levels. For example, OpenAI's GPT-4 used an estimated $78 million worth of compute to train, while Google's Gemini Ultra cost $191 million for compute.[10] This cost derives from the scale of the computer resources required to train the neural networks (tens of thousands of AI GPUs, each costing tens of thousands of dollars, running for months); the scale of data required (the training data for GPT-3 included a large fraction of the data available on the World-Wide Web); and of course, the (highly paid) staff required to code the model, build it, and test it[11].

As AI is being rapidly developed and deployed, it is key for policymakers to acknowledge and understand that there are many unanswered questions about it. Some of these are fundamental uncertainties about the future of the technology itself, while others are about the impact of AI in the economy. The following sections summarize some of the most relevant AI uncertainties.

1.2. Technology questions

i. The race to scale: how far will AI improve by increasing the size of AI systems?

In September 2024, there were reports that OpenAI, one of the world's leading AI developers, intends to build data centers each with power requirements of five gigawatts[12]. For some perspective, this is approximately the amount of power capacity

[9] An AI foundation model is a large, versatile AI system trained on vast amounts of data, designed to perform a wide range of tasks and be adapted for specific applications. The term was introduced by Bommasani et. al. (2021).

[10] Stanford University (2024), AI Index Report 2024 – Artificial Intelligence Index.

[11] We note that robotic AI is nowhere near the level of systems like ChatGPT. General purpose household robots, for example, are not imminent, and multiple major scientific and technological challenges would need to be solved to make them a reality.

[12] OpenAI reportedly wants to build 5-gigawatt data centers, and nobody knows who could supply that much power | Fortune.

needed for a major city the size of Miami. Data center complexes of such size would require an investment amount in the neighbourhood of USD 100 Billion. Why is an AI company pursuing such an unprecedented venture? The answer lies in the "scaling laws" of AI.

Scaling laws in AI suggest that increasing model size, dataset size, and computing resources leads to improved performance, often following a power law relationship. This principle has driven much of the progress in large language models (LLMs) over the past few years. As noted above, early AI systems struggled to generalize beyond specific tasks due to limited model complexity and constrained datasets. In contrast, the adoption of the scaling paradigm allowed researchers to build increasingly sophisticated neural networks capable of learning from vast amounts of data.

If scaling laws continue to hold, AI systems will continue to show accelerated progress and develop capabilities that will be increasingly transformative of financial markets and the economy. However, scaling laws are primarily empirical regularities rather than a fully developed theoretical framework. The current understanding of scaling laws is a mix of empirical regularities and partial theoretical explanations, with a comprehensive theory still being an active area of research. Not surprisingly, the benefit of future scaling is highly controversial among AI experts, which introduces a significant source of uncertainty for AI policy.

If those who are pursuing massive scaling of AI systems turn out to be right, near-term advances could lead to highly capable and reliable AI agents, which could plan and reason—and eventually to Artificial General Intelligence (AGI)[13]. This is a scenario that would require policy and regulatory frameworks to evolve rapidly. In contrast, if diminishing returns to scale kick-in (as some expect, e.g., Kaplan et al., 2020), the economic disruptions introduced by AI would be slower and less profound. The prevalence of scaling laws would also imply rapidly increasing capital expenditures needed to enter the market, potentially leading to concentration and market power in the global industrial organization of AI, whereas the eventual emergence of decreasing

[13] Bowen (2024).

returns to scale would tamper down the need for capital expenditures and favour a more competitive AI industry.

ii. Proprietary vs open-source AI systems

The question of whether AI systems—particularly frontier foundation models—should be open-source or proprietary remains one of the most contested issues in the field. The term open-source AI systems typically refers to those whose code and models are freely available for use, modification, and distribution[14]. Proprietary systems, by contrast, are owned and controlled by private entities, often with restricted access and usage rights. Both approaches have advantages and drawbacks, and their implications extend beyond technical considerations to touch on ethics, innovation, security, and economic policy. The unsettled nature of this issue reflects the rapid evolution of AI technology and its far-reaching implications.

Proponents of open-source AI argue that transparency and accessibility drive innovation, democratize technology, and enhance trust. Open-source projects foster collaboration by allowing researchers, developers, and institutions to build on each other's work. This approach can accelerate progress, reduce duplication of effort, and enable smaller organizations, startups, and academics to compete in a field increasingly dominated by well-funded corporations. For example, open-source foundation models like Meta's LLaMA series have enabled widespread experimentation and contributed to a vibrant ecosystem of applications. Open-source advocates also highlight ethical considerations. Transparent AI systems—some argue—allow for public scrutiny, making it easier to identify and mitigate biases, errors, and harmful behaviors.[15] Open-source proponents claim that open access to algorithms and datasets helps ensure that societal values—such as fairness and accountability—are embedded into AI systems, although this remains a debated topic (Luna, 2024).

---

[14] Open-source is a spectrum rather than a binary concept, reflecting varying degrees of accessibility and transparency across different components of AI systems. It is worth noting that the term 'open-source AI' is a defined legal term in some jurisdictions, including in the European AI Act.

[15] Open Source Initiative – The steward of the Open Source Definition, setting the foundation for the Open Source Software ecosystem.

Proprietary systems are often defended on the grounds of security, control, and financial sustainability. Organizations like OpenAI and Google argue that proprietary frameworks are necessary to protect against misuse of powerful foundation models. For example, unrestricted access could enable malicious actors to create misinformation, automate cyberattacks, or develop dangerous technologies. By keeping foundation AI systems proprietary, developers retain the ability to monitor, control, and safeguard their use. Another consideration is the rapidly increasing cost of developing advanced AI systems, as discussed above. Proprietary settings help organizations recoup these costs and incentivize further innovation. Critics of open-source AI argue that the "free-for-all" nature of open systems may undermine intellectual property rights and disincentivize investment in cutting-edge research.

The open-source vs. proprietary debate reflects a broader tension between competing priorities: innovation vs. control, democratization vs. safety, and collaboration vs. competition. While open-source AI has proven transformative in areas such as education and research, the risks associated with its misuse are becoming more apparent as models grow increasingly powerful. At the same time, the dominance of proprietary AI systems raises concerns about monopolization, reduced competition, and the concentration of power in the hands of a few corporations.

iii. Data uncertainty: are we running out of training data?

Over the last decade, the progress of notable and foundation AI models has been fuelled by the increasing availability of massive datasets. For example, models such as GPT-3 were trained on hundreds of billions of words scraped from the internet, requiring extensive volumes of text data to achieve state-of-the-art results (Brown et al., 2020). However, concerns are growing that we may be approaching the limits of freely available training data. Recent research has suggested that we are reaching a saturation point for high-quality, publicly accessible text data on the internet. The pool of diverse, high-quality information that can be legally scraped is becoming depleted, making further scaling of AI models through existing data sources increasingly challenging (Xu, 2022).

Additionally, the practice of web scraping is facing growing resistance, both from legal challenges and heightened enforcement of data privacy regulations. Platforms are tightening access to their content in response to data scraping practices, and some jurisdictions are introducing stricter data protection measures, making it legally and logistically more difficult to gather data in large quantities (Tiederich, 2024). This dwindling availability of web data raises questions about how future models will be trained.

While synthetic data has been proposed as a potential alternative, it has significant limitations. Synthetic data often lacks the nuanced, context-rich characteristics of real-world information, leading to concerns about its usefulness in training deep models (Shumailov et al, 2024). These challenges suggest that data generation and monetization could become lucrative opportunities in the future. Companies might invest heavily in creating proprietary datasets, and the emergence of a marketplace for high-quality data could significantly influence the direction of AI research and development, where the scarcity of data could make it a key commodity.

iv. Is the promise of AI on the edge real?

AI on the edge is the deployment of artificial intelligence algorithms on devices that are close to the source of data generation. This represents a significant leap forward in the deployment of artificial intelligence by pushing computation away from centralized cloud servers and directly to edge devices, such as sensors, smartphones, and embedded systems. This shift has the potential to reshape how data is processed, making it faster, more efficient, and more secure (by minimizing the need to send sensitive information over networks, which decreases the risk of data interception or breaches during transmission). Edge AI is positioned to become an important technology for industries requiring ultra-low latency and localized decision-making, such as healthcare, autonomous driving, industrial automation, and smart cities. For instance, in healthcare, real-time patient monitoring can be critical, and edge AI facilitates on-the-spot analysis without depending on cloud connections, which may experience delays.

Among the notable trends currently shaping the landscape of edge AI, energy-efficient AI chips are most important for the future development of this field. Advances in hardware, particularly AI accelerators, are playing a critical role in enhancing edge AI's capabilities. These energy-efficient chips, such as those based on neuromorphic computing architectures, are designed to perform complex AI computations without demanding high power consumption. They are crucial for battery-powered edge devices like wearables or IoT sensors. The use of AI to design such chips might open the way to a new generation of edge computing devices highly optimized for AI algorithms and extremely efficient in their energy consumption, ultimately facilitating faster deployment across various industries. Moreover, the rise of 5G technology is set to boost the capabilities of edge AI by providing faster and more reliable connectivity, enabling seamless communication between edge devices and other network components.

1.3. Market structure uncertainties: competition or market concentration?

The uncertainties about AI are not only about the technology itself, but about the industrial organization of the AI stack.

i. Concentration in foundation model providers

The market for foundation models is currently characterized by intense competition and rapid technological shifts, exemplified by OpenAI's recent leap forward with its O1 Pro model. While currently dynamic, the market's economic characteristics—in particular the prevalence of scaling laws and the concurrent acceleration of capital expenditure needs—suggest a potential for future concentration, with leading players employing competing strategies in a landscape shaped by economies of scale.

The current market for foundation models is characterized by intense competition and rapid technological advancements.[16] Throughout most of 2024, the capabilities of leading foundation model providers were closely clustered, with minimal differentiation among top performers (see, e.g., Korinek and Vipra, 2024). However, the leading

[16] The intensity of competition among foundation AI models is evident in rapidly falling prices for users. The cost of AI intelligence has decreased dramatically, with GPT-4 equivalent intelligence from OpenAI dropping 240x in 18 months, from $180 per million tokens to less than $1 (Jaipuria, 2024).

model has regularly changed in recent months. This illustrates the volatility of the market and the potential for rapid changes in competitive dynamics.

As of November 25, 2024, the LMSYS leaderboard, a widely used benchmark for model performance that uses an ELO-style mechanism, reflects this new competitive landscape:

Lab	Country	Top Model	Released	LMSYS
Google DeepMind	USA/UK	Gemini-1.5-Pro-Exp	11/25/2024	1365
OpenAI	USA	ChatGPT-4o-latest	11/20/2024	1361
xAI	USA	Grok-2	8/13/2024	1289
01 AI	China	Yi-Lightning	10/16/2024	1287
Anthropic	USA	Claude 3.5 Sonnet	10/22/2024	1282

OpenAI and Google DeepMind are neck and neck – the difference in LMSYS scores reported in the last column of the table was not statistically significant. This underscores the uncertainty about the evolution of relative technological leads of the companies involved and future market shares.

Major players in the foundation model space employ competing strategies to gain or maintain competitive advantage. OpenAI, Google DeepMind, xAI, 01 AI, and Anthropic each leverage their unique strengths, whether it's first-mover advantage, vast computational resources, or extensive datasets. The market currently exhibits Bertrand-like competition, with prices barely covering variable costs as providers vie for market share and technological supremacy. Moreover, while the motives of Meta to open source its highly capable LLaMA series are debated, it has effectively pushed the price for most model capabilities to zero. Simultaneously, the world’s largest technology companies are making significant investments in leading AI labs, exacerbating the intensity of competition. These investments may provide crucial resources for AI development but also raise questions about market power and vertical integration.

Despite the intense level of competition that currently prevails, the potential prevalence of scaling laws may generate a force towards future concentration. So far, the foundation model market exhibits significant economies of scale and scope, which may drive future concentration (RAND Corporation, 2024). The fixed costs for pre-training

models are growing rapidly, with estimates suggesting they could exceed one billion dollars per frontier model by 2027. This escalation in costs may lead to a shrinking number of players that the market can sustainably support. In other words, the continuity of scaling laws going forward would translate into strong returns to scale and, potentially, natural monopoly characteristics in the production function of AI (Hopkins et al., 2023).

Furthermore, there is potential for an "intelligence feedback loop," where leading labs could progressively separate themselves from competitors (Korinek and Vipra, 2024). This dynamic bears similarities to the early days of digital platform markets, raising concerns about the risk of market tipping. Just as with early digital platforms, the foundation model market might experience a period of intense competition followed by a shake-out, potentially resulting in a highly concentrated market structure.

ii. Concentration in microprocessor providers

The microprocessor market encompasses two main product categories, Central Processing Units (CPUs) and Graphics Processing Units (GPUs). CPUs, the traditional "brains" of computing systems, are used in personal computers, servers, and numerous other devices. GPUs, originally designed for rendering graphics, have become instrumental in AI and machine learning applications due to their parallel processing capabilities.

In the CPU market, competition has intensified in recent years. Intel, historically dominant, has faced challenges as competitors such as AMD have gained market share, particularly in data centers and PCs. The mobile processor segment, led by companies such as Qualcomm with its ARM-based designs, has also become increasingly significant, accounting for over 45% of the microprocessor market revenue in 2023.

GPUs are of paramount importance for AI applications. This segment exhibits remarkably high concentration, with a single company, Nvidia, commanding over 90% market share in frontier GPUs (Fernandez et al., 2023). Nvidia's dominance is further reflected in its extraordinary gross profit margin of approximately 75% in mid-2024. While some competitors, such as Google with its Tensor Processing Units (TPUs),

have developed alternative AI-specific chips, their market presence remains limited compared to Nvidia's.

The creation of cutting-edge chips involves a sophisticated process with substantial R&D costs, resulting in a complex supply chain that is also highly concentrated. ASML (Netherlands), is a key player in this chain as it is the sole supplier of Extreme Ultraviolet (EUV) lithography machines. These are crucial for fabricating advanced semiconductor nodes used in modern GPUs. TSMC (Taiwan Semiconductor Manufacturing Company) is the world's only advanced semiconductor foundry capable of manufacturing frontier GPUs for Nvidia, AMD, and many other companies.

This high level of concentration in the GPU market and its supply chain presents potential risks for the AI industry, including supply bottlenecks and geopolitical vulnerabilities. It also raises important questions about market power and its implications for innovation and pricing in the rapidly evolving field of AI as a growing fraction of the funding raised by foundation model providers goes to pay for GPUs.

iii. Concentration in data providers

The landscape of data provision for artificial intelligence, particularly for foundation models, is undergoing significant shifts that create significant uncertainties surrounding the future value and scarcity of data that may have important implications for market dynamics and economic policy.

The current generation of foundation models was made possible by the vast amounts of freely available data on the internet. This democratization of information allowed both established tech giants and new entrants to develop sophisticated AI models. However, as these models become more advanced, the nature of data requirements is evolving, potentially altering the competitive landscape. In particular, as data is acknowledged valuable and free data is harder to procure, the cost structure of AI may be altered in a relevant way.

Large technology companies with diverse ecosystems currently hold a natural advantage in accessing varied and extensive datasets. For instance, Google's ecosystem provides access to data from search indexing, YouTube, and even Google Docs, while Elon Musk's xAI benefits from data generated on X (formerly Twitter) and

Tesla. This data advantage could potentially lead to market concentration if left unchecked. New players are emerging in the data provision landscape, exemplified by companies such as Scale AI, which employs a large global workforce to provide data services to AI companies.

However, the trajectory of AI development is shifting, calling into question whether access to vast amounts of real-world data will remain a critical competitive edge. There is a growing trend towards using synthetic data for training models. Thompson (2024) estimates that future models may rely on synthetic data for up to 70% of their training needs. This shift may eventually diminish the reliance on vast real-world datasets for training foundation models.

Simultaneously, as foundation models are increasingly fine-tuned for specific industry applications, the importance of sector-specific training data may be growing. This trend has significant implications for traditional companies across sectors, who find themselves in possession of valuable proprietary data, and could create the potential for highly specialized AI models using this data and new opportunities across sectors. Balancing innovation with concerns about data privacy and security may become increasingly complex, particularly in sensitive sectors. As data becomes increasingly crucial for AI development, policies governing international data flows may also have significant implications for global competitiveness in AI.

1.4. The AI talent gap

Competition for AI talent is intensifying globally, driven by the rapid advancement of artificial intelligence technologies and the growing demand for skilled professionals. Companies are not only vying for top-tier data scientists and machine learning engineers but also for a more diverse talent pool that includes professionals from different educational backgrounds. The landscape of AI talent is becoming increasingly competitive, particularly as companies recognize the critical role that AI will play in their future operations. This competition is characterized by soaring salaries and aggressive recruitment strategies, as businesses seek to secure the best minds in the field. For instance, top tech firms are reportedly offering salaries that can reach into the hundreds of thousands, alongside lucrative stock allocations and benefits. In Europe, the

struggle for AI talent is notably heating up. European firms are facing challenges in attracting skilled workers, partly due to competition from tech hubs in the U.S. and Asia. Additionally, the shortage of AI talent is exacerbated by the limited number of educational programs specifically focused on artificial intelligence, which means that many companies are compelled to look abroad to fill their talent needs. Startups are playing a pivotal role in this talent battle. They are often more agile and willing to experiment with innovative recruitment methods, making them attractive to younger professionals who seek dynamic work environments. Many talented individuals are gravitating towards startups that promise not only competitive compensation but also opportunities to work on cutting-edge technologies.

The current shortage of AI talent is at least partly driven by the speed at which AI recently developed. While market signals are already at work and AI related fields are attracting a large share of students electing majors and graduate degrees, this process will inevitably take time (Coffey, 2024). Within this context, actively cultivating a future AI workforce is becoming crucial. Educational initiatives must be significantly increased to prepare more graduates for careers in AI, mainly with significant investment in STEM education but also with a significant push for diversity, as companies are increasingly recruiting talents from various backgrounds, thus fostering innovation and creativity within AI development.

1.5. AI for whom? The challenge of diffusion

The widespread and responsible adoption and diffusion of AI is crucial for realizing the potential economic benefits and welfare gains promised by this technology – a technology can only benefit us if we use it safely. Historical experience with previous general-purpose technologies underscores the slow and uneven nature of this process. From the steam engine to electricity to the internet, the full transformative impact of these technologies often took decades to materialize. Robert Solow's famous productivity paradox observed that the impact of computers took a decade to be reflected in productivity statistics. This historical context suggests that the diffusion of AI may also face challenges, yet the degree of these remains uncertain.

Within G7 nations and other advanced economies, shortages in digital infrastructure, human capital, and organizational readiness can lead to uneven adoption rates across regions and industries. Small and medium enterprises often lag behind larger corporations in AI implementation due to resource constraints and limited access to expertise. Moreover, regulatory uncertainties and concerns about data privacy and security can impede adoption in some sectors, holding back potential economic benefits.

On a global scale, the challenges are even more pronounced. Developing countries often lack the necessary digital infrastructure, making it difficult to deploy and utilize AI technologies effectively. The scarcity of AI talent and limited access to high-quality data further exacerbate this divide. Additionally, many developing nations face resource constraints that make it challenging to invest in AI research, development, and implementation. These factors may contribute to the risk of an emerging "artificial intelligence divide," where countries that lag in AI adoption may see declining competitiveness and unfulfilled economic potential.

The barriers to AI adoption extend beyond infrastructure and resources. Cultural and organizational resistance to change can significantly slow the diffusion process. Many organizations struggle to integrate AI into existing workflows and business models, requiring substantial changes in organizational structure and decision-making processes. Furthermore, concerns about job displacement and ethical implications of AI may lead to societal resistance, potentially slowing adoption rates.

Despite these challenges, there are factors that may accelerate the diffusion of AI compared to previous general-purpose technologies (Bailey et al., 2023). The digital nature of AI technologies allows for faster and more flexible deployment, often without the need for extensive physical infrastructure changes[17]. The increasing integration of AI into existing software platforms lowers adoption barriers, making it easier for

[17] According to a survey by McKinsey, AI adoption worldwide has increased dramatically over the past year. 72% of companies surveyed in March 2024 said they had adopted AI in at least one business function (up from 55% in 2023). The proportion of companies that say they are using generative AI has almost doubled, from 33% in 2023 to 65% in 2024. See The state of AI in early 2024 | McKinsey. According to the AI Index Report 2024, by Stanford University, the number of newly funded AI companies in the world jumped from 1289 in 2022 to 1812 in 2024 (AI Index Report 2024 – Artificial Intelligence Index).

organizations to incorporate AI capabilities into their operations. Additionally, the ability to interact with many AI systems using natural language reduces the learning curve for users, potentially speeding up adoption across various sectors and demographics.

There is significant scope for policy to influence the pace and inclusivity of AI diffusion through targeted investments in digital infrastructure, education, and workforce development initiatives. Supporting research and development, particularly in areas that address specific local needs and challenges, can help ensure that AI solutions are relevant and accessible to a wider range of users. Furthermore, international initiatives for knowledge sharing and technology transfer can help bridge the global AI divide, fostering more inclusive development and deployment of AI technologies. The localization and adaptation of AI solutions to different cultural contexts and languages is another important consideration. Ensuring that AI technologies are truly adapted to local needs and values can significantly enhance their relevance and acceptance across diverse nations. Furthermore, data governance policies can be implemented to enable safe data access while ensuring compliance with personal data protection laws, thereby fostering competition and innovation and mitigating the risk of market concentration in the data market for training AI models.

1.6. A double-edged sword? AI and environmental sustainability

Global data centre electricity consumption of 460 terawatt-hours (TWh) in 2022[18] (almost 2% of global electricity demand) is set to rise with rapid demand growth for AI, particularly energy-intensive generative AI. While AI currently consumes less than 10%[19] of overall data centre energy consumption, it's expected to drive future computing-related energy growth.

Each stage of the AI lifecycle (from equipment manufacturing, through model training and use, to e-waste disposal) consumes energy, water, and minerals, resulting in the emission of greenhouse gases. Estimates have shown that emissions derived from

[18] International Energy Agency (IEA), Electricity 2024 – Analysis.

[19] Ibid.

training, fine-tuning, using models form the primary contribution to AI's carbon footprint (Wu et al, 2022; Patterson et al., 2022; Kirkpatrick, 2023).

The use of the AI model (inference) is estimated to outweigh the carbon impact of training and fine-tuning; however, the relative proportion varies depending on the size of the model and the level of usage, amongst other factors (Luccioni and Garcia, 2023). The embodied emissions that arise from the manufacturing of hardware used for the models comprise roughly one-quarter to half of the model's overall carbon footprint (Faiz et al., 2024). Furthermore, AI models have a sizeable water footprint of many millions of Liters of freshwater withdrawn or consumed for electricity generation and data centre cooling[20].

Current trends show the demand for AI compute exceeding improvements in hardware and algorithmic efficiencies. This differential is significant: data centre energy use has been growing by 20-40% annually over the last several years[21]. While estimates vary, AI, along with cryptocurrency, are expected to almost double the global electricity consumption of data centres from 460 TWh in 2022 to more than 1 000 TWh by 2026[22]. Historically, the increased demand for compute was largely offset through efficiency measures that followed Moore's Law and the 'hyperscale shift' which gave rise to super-efficient data centres. However, further efficiency increases through hardware approaches will be increasingly difficult to attain. In recent years, AI has been driven by a race towards larger models and larger datasets, with commensurate demand for more compute. Despite the dramatic improvements in AI algorithmic efficiency over the past decade (Pilz et al., 2024), the current "more compute" approach is challenged by the factors mentioned above, along with limited data availability and performance considerations associated with scaling transformer-based LLMs[23].

AI's escalating power demand is concerning, since it comes at a time when the electricity grid is challenged by broader electrification trends and mandates for

[20] Data Center Water Usage: A Comprehensive Guide - Dgtl Infra.

[21] Data centres & networks - IEA.

[22] International Energy Agency (IEA), Electricity 2024 – Analysis.

[23] See Jones (2024) or Lohn and Musser (2022).

increased renewable energy supply. The push to accelerate AI deployment has threatened climate change targets set by government and technology firms. In the US (which houses the largest global share of data centres), technology firms are shoring up long-term electricity contracts, driving the re-opening of formerly closed nuclear and fossil fuel power sites while also increasing investment in emerging energy technologies. The rapid acceleration to dominate AI compute is already having environmental consequences. Google, Microsoft and Meta have all reported significant increases in emissions due to AI (e.g., for Google, that included an almost 50% increase from 2019 to 2023[24]). Given that the method by which they use to self-report has been called into question, the emissions may in fact be much higher[25].

Apart from its environmental cost, AI also has the potential to drive efficiency improvements. For example, AI has been used to increase cooling performance in data centres and optimize renewable energy supply. Beneficial applications of AI for the environment and financial authorities abound, including satellite monitoring to advance the monitoring of global climate impacts and progress on targets, and productivity and risk measures relevant to financial authorities (e.g., pollution and land use changes as proxies for development and climate risk).

As AI capabilities grow, so too does the likelihood of breakthrough advances in materials science and energy that would transform the AI sector and beyond. However, such breakthrough advances will take time to diffuse, and their implementation time horizon may not align with the net-zero imperative. Furthermore, increased efficiency through the adoption of AI can also lead to increased demand for AI computing, further exacerbating AI's carbon footprint (Varoquaux et al., 2024), while at the same time incentivizing innovation across the AI supply chain.

[24] Google (2024), 2024 Environmental Report - Google Sustainability.

[25] The Guardian (2024), Data center emissions probably 662% higher than big tech claims. Can it keep up the ruse? | Technology | The Guardian.

2. The macroeconomic impact of AI

2.1. Measurement challenges in the age of AI

Before we address the macroeconomic impact of AI, there is a cautionary note on our ability to measure it properly. Conventional macroeconomic indicators and economic statistics may not adequately capture the full extent of AI's influence on the economy, presenting a complex challenge for policymakers. Much like the "productivity paradox" seen during the early days of computing, AI's contributions to productivity and economic growth might not immediately appear in traditional metrics like GDP. AI frequently creates value in nontraditional ways, such as quality improvements and efficiency gains, that conventional indicators fail to recognize (or recognize with long lags). Many services in the digital economy have a price of zero, so they are not picked up by GDP calculations. For instance, innovations such as Google's enhanced translation services significantly boost utility without directly impacting GDP calculations because the translation service is offered for free and therefore produces no monetary transactions subject to record in national accounts. This invisible value creation highlights the limitations of traditional measures in reflecting AI's true economic contributions.

Moreover, AI is driving the emergence of entirely new economic activities and business models that do not fit neatly into existing statistical frameworks, further complicating efforts to measure its impact. Metrics such as GDP and traditional productivity indicators struggle to capture advancements in service industries, knowledge work, and quality-of-life improvements driven by AI. As a result, these measures risk underestimating AI’s effects. The unique characteristics of AI, including its capacity to enhance efficiency and create novel outputs, demand a rethinking of economic measurement tools to better align with the realities of a technology-driven economy.

In sum, policymakers should be mindful that our ability to observe and measure the effects of AI on the economy is not perfect. This should prompt caution when making

strong claims and should motivate the development and use of alternative measurement approaches to complement longstanding techniques and metrics[26].

2.2. Innovation and growth

The integration of AI into our economies has the potential to sharply boost productivity and drive innovation, leading to significant prosperity. However, the scale of these gains and how they will be distributed across sectors remains uncertain, depending largely on the speed of AI adoption and the necessary investments to support it. In the next decade, AI could generate value comparable to entire industries, leading to productivity gains similar to those from the adoption of electricity and digital technologies.

The Productivity Challenge: Historical Context

Since the post-World War II economic boom, known as the "Thirty Glorious Years," most developed economies have experienced a gradual decline in growth rates. Some economists argue that this slowdown is inevitable, referring to it as "secular stagnation,"[27] where advanced economies reach a plateau in productivity gains due to several factors, such as diminishing returns on innovation, demographic changes, institutional and regulatory frameworks, and labor reallocation to relatively less productive sectors. AI, however, represents a potential breakthrough that could reverse this trend, at least partially, driving a new wave of growth. This resurgence could occur through two primary mechanisms:

1. Increased productivity: AI's ability to automate and enhance the production of goods and services.
2. Accelerated innovation: AI’s potential to generate new ideas, products, and forms of organization, thereby fostering long-term economic growth.

[26] Several studies, including Brynjolfsson and Collis (2019), have proposed alternative measurement techniques, but those innovative approaches are still to be adopted in national accounts methodologies.

[27] See for example, Gordon (2012, 2015), Cowen (2011), and Jones (2009).

AI's Role in Enhancing Productivity

AI's most immediate and tangible impact will be its ability to boost, to varied degrees, productivity across a wide array of sectors. Similar to how mechanization transformed agriculture, the assembly line revolutionized manufacturing, and digitization modernized the economy, AI stands to dramatically increase the efficiency with which we produce goods and, especially, services. The automation of simpler tasks frees up human labor for more complex and creative work, thereby raising overall productivity.

One of the first pieces of evidence supporting this channel comes from a microeconomic study by Brynjolfsson et al. (2023). The authors analyzed the impact of generative AI on worker productivity at a U.S. customer service firm. The firm gradually introduced an AI tool that helped employees respond to customer queries through automated suggestions. The results showed a significant productivity boost: employee output increased by 14% in the first month of using the AI assistant and stabilized at around 25% after three months. Similar results have been observed in more specialized, high-skill professions. In the U.S., for instance, studies have shown that AI tools like ChatGPT can increase the productivity of consultants and managers by between 25% and 35% (Noy and Zhang, 2023; Dell'Aqua et al., 2023). This demonstrates that AI has the potential to deliver productivity gains across a wide range of occupations, both low- and high-skill.

Moreover, the findings are echoed in surveys conducted in France. According to a 2023 survey by Pôle Emploi titled "Les employeurs face à l'intelligence artificielle," 72% of employers who have integrated AI into their operations reported a positive impact on employee performance. Key benefits included the reduction of tedious, repetitive tasks (cited by 63% of employers) and a decrease in error rates (51%). This wide-ranging impact on productivity suggests that AI will play a central role in driving future economic growth[28].

[28] The uncertainty around the impact of AI in productivity is also highlighted in a recent study by the OECD, see Filippucci et al. (2024).

Accordingly, the WTO's *Report on AI and Trade* highlights that the AI's most significant impact will be on services trade[29], driven by four factors. First, barriers to intermediate services trade are primarily technological, with minimal regulation of back-office services. Second, digital technology is rapidly lowering these barriers. Third, AI technologies, such as machine translation and upcoming speech translation, are making domestic and foreign workers more interchangeable. Finally, generative AI will accelerate this by transferring the skills of high-skilled workers from developed economies to emerging economy workers, making their output more similar.

As argued in Baldwin et al. (2023), the enhanced productivity and innovation driven by AI can lead to increased trade in specific services, which in turn could result in higher trade volumes and greater economic interconnectedness on a global scale. For example, AI has been shown to increase the number of foreign users of AI-driven mobile applications by an average of tenfold (Sun and Trefler, 2023). Additionally, projections using the WTO Global Trade Model suggest that services in sectors such as education, human health, recreation, and finance could experience substantial growth in trade[30]. This means that AI is set to boost trade in services previously thought of as "non-tradeable" by enabling global collaboration through tools such as video conferencing, transcription, translation, and virtual reality. These advancements enhance ICT (Information and Communication Technology) services, overcoming geographical barriers and enabling real-time interactions, negotiations, and decision-making, thus facilitating trade and reducing the need for physical travel. These findings underline the potential for AI to significantly boost trade in services and foster a more interconnected global economy.

At the same time, it is becoming clear that global efforts to mitigate regulatory fragmentation related to the use of AI—ranging from sector-specific regulations to

[29] World Trade Organization, 21 November 2024 (Report "Trading with intelligence - How AI shapes and is shaped by international trade"), p 24-28.

[30] The WTO Global Trade Model was developed to project the potential impact of AI on global trade patterns. This model is a recursive recursive dynamic computable general equilibrium (CGE) model, allowing for long-term projections extending through 2040. World Trade Organization Report "Trading with intelligence - How AI shapes and is shaped by international trade", pp 29-34.

intellectual property and data governance—could reduce economic costs, particularly for small businesses (OECD and WTO, 2024).

How large will the economic gains from AI be?

One of the key questions for policymakers is the scale of economic gains AI could bring. To address this, recent academic research has moved beyond individual firm case studies to assess AI's impact on broader economic growth.

Using a task-based framework that accounts for automation and task complementarities, Acemoglu (2024) argues that AI's economic effects will primarily come from cost savings and productivity improvements at the task level. His analysis suggests that GDP and aggregate productivity gains will depend on how many tasks are affected and the average cost savings achieved. According to current estimates of AI exposure and task-level improvements, Acemoglu projects a relatively modest increase in total factor productivity—no more than 0.07% per year over the next decade.

In contrast, Aghion and Bunel (2024) offer two alternative approaches for estimating AI's potential growth impact. The first draws a parallel between the AI revolution and past technological revolutions, such as the advent of electricity or digitalization. Based on this historical analogy, they estimate that AI could boost aggregate productivity growth by 0.8 to 1.3 percentage points annually over the next decade[31], potentially adding between 250 and 400 billion euros to GDP by 2034. Their second approach closely follows Acemoglu's task-based framework but incorporates more recent empirical data. They find that, when considering new empirical evidence, AI could increase aggregate productivity growth by between 0.07 and 1.24 percentage points per year, with a median estimate of 0.68 percentage points over 10 years[32]. This result

[31] Based on historical analogies, there are two potential scenarios for AI-driven productivity growth. If it is drawn a parallel with the electricity revolution of the 1920s, AI could increase productivity growth by 1.3 percentage points per year from 2024 onward. Alternatively, if it is compared to the digital revolution of the late 1990s and early 2000s in the United States, productivity growth could rise by around 0.8 percentage points annually.

[32] The baseline scenario taking into account (i) the share of tasks exposed to AI in developed countries estimated at 60% (Pizzinelli et al., 2023), (ii) the share of exposed tasks for which it will be profitable to

illustrates, first and foremost, the uncertainty surrounding the quantification of the impact of AI on aggregate productivity growth, depending on the studies from which the estimates are drawn. However, it also aligns with the productivity effects seen in other general purpose technologies, providing a more optimistic outlook compared to Acemoglu's conclusions[33].

This estimate in turn may be seen as a lower bound to the extent that it does not account for the fact that AI could also automate the production of ideas, leading to an ongoing cycle of innovation and sustained economic growth. On the other hand, it does not take into account potential barriers to growth, in particular those associated with the lack of competition in the upstream segments of the AI value chain.

AI and the Generation of New Ideas

In addition to increasing productivity, AI has the potential to transform the way we generate new ideas and innovations. Historically, key scientific and technological breakthroughs have had a profound and lasting impact on growth. For example, the invention of calculus in the 17th century revolutionized physics, while advancements in glass polishing techniques enabled the development of the microscope, leading to critical discoveries in medicine.

AI could serve a similar role by automating the process of idea generation, making it easier to innovate and solve complex problems. For example, AlphaFold, an AI model, has already revolutionized the field of protein folding, while GNoME suggests new materials for industrial and everyday use. Although the full impact of AI on science and innovation is difficult to quantify, the potential is vast. If AI continues to aid researchers in generating new hypotheses, designing experiments, and conducting research, the

use AI estimated at 50% due to a 22% annual decline in computing costs (Besiroglu and Hobbhahn, 2022) and (iii) productivity gains enabled by AI estimated at 40% based on three benchmark studies (Peng et al., 2023; Noy and Zhang, 2023; Brynjolfsson et al., 2023).

[33] Following Bresnahan and Trajtenberg (1995), general-purpose technologies are defined by their "potential for pervasive use in a wide range of sectors and by their technological dynamism". Agrawal et al. (2023) offer a comparison between the key features of AI and previous general-purpose technologies.

pace of innovation could accelerate, leading to a permanent increase in productivity growth.

The Historical Parallel: Time Lags in Technological Adoption

The economic gains from AI, however, may take time to emerge. A historical parallel can be drawn with the adoption of electricity, where productivity gains did not materialize until decades after the technology was first introduced. In the early 20th century, despite the availability of electric power, factories continued to organize themselves around central drive shafts, a legacy of the water mill and steam-powered era. It wasn't until innovations such as the electric motor and assembly line were fully implemented that productivity surged.

Similarly, AI's widespread adoption will likely require fundamental changes in organizational structures, business processes, and additional investments in infrastructure. However, the adoption curve for AI could be shorter, especially with the rapid spread of generative AI tools that are relatively easy to integrate into various sectors. For example, in the video game industry, AI can now generate game concepts in two months rather than six, illustrating the speed at which it can enhance productivity. As the technological foundation for AI has already been in place for a decade, we may see the productivity gains from AI materialize much sooner than those from past technological revolutions.

The Role of Institutions and Competition in AI-Driven Growth

While AI holds enormous potential for growth, realizing these gains depends on the right institutional and competitive framework. A key lesson from the digital revolution is that without strong competition policies, a small number of companies can dominate markets, stifling innovation and curbing long-term growth. The rise of companies like Google (now Alphabet), Amazon, Facebook (now Meta), Apple, and Microsoft (GAFAM) initially contributed to productivity growth in the late 1990s and early 2000s.

However, their arguably dominant positions eventually hindered the entry of new competitors.

The AI revolution presents a similar risk. As discussed in Chapter 1, the GAFAMs and other large firms already control critical segments of the AI value chain, including access to data and computing power. Ensuring that the AI revolution leads to widespread growth, rather than reinforcing the dominance of a few companies, will require adapting competition policies and fostering an environment where new, innovative companies can thrive.

2.3. Jobs and labor productivity

Recent advancements in robotics and AI technologies present opportunities and challenges for the global labor market. While these innovations are expected to enhance productivity and foster new job creation, they also raise concerns about potential job displacement and growing inequality (Brynjolfsson and McAfee, 2014; Ford, 2015; Harari, 2016)[34]. For example, a Goldman Sachs report highlights that generative AI could boost global productivity but may expose 300 million jobs to automation, prompting significant labor market disruption[35]. Policymakers face the challenge of promoting AI-driven progress while implementing protective measures for workers and consumers against the associated risks.

The impact of technological change on employment is a longstanding issue, with a history that can be traced back to the Industrial Revolution. The fear of technology destroying jobs and displacing workers is not a new one, as evidenced by the protests of the Luddites in England. An extensive literature on creative destruction (Aghion and Howitt, 1994, 1998; Mortensen and Pissarides, 1998; Postel-Vinay, 2002) explores how innovation can lead to job displacement. Simultaneously, evidence suggests that technological progress reduces unemployment (Blanchard and Wolfers, 2000; Pissarides and Vallanti, 2007; Miyamoto and Takahashi, 2011). Over the past few

[34] Frey and Osborne (2013) estimate that 47% of total US employment is at risk of losing jobs to automation over the next decade. A widespread concern is that AI-assisted machines can be used to automate more and more jobs (Acemoglu, 2022, Acemoglu and Johnson, 2023).

[35] Generative AI could raise global GDP by 7% | Goldman Sachs.

decades, the labor market has polarized between high-skilled and low-skilled jobs, leaving medium-skilled positions most vulnerable to technological advances like computers (Autor, 2015).

The effects of AI and automation on employment can be categorized into three main channels: displacement, productivity, and reinstatement effects. AI may displace workers in specific sectors (displacement) while increasing productivity, potentially creating new job opportunities, particularly in industries where human labor maintains a comparative advantage (reinstatement, Acemoglu and Restrepo, 2018).

If policymakers wish to support workers in benefiting from technological advancements, they can encourage conditions under which the productivity and reinstatement effects can flourish[36].

Experts remain divided over the potential impacts of AI on employment, with both optimistic and pessimistic perspectives. The pessimistic view focuses on the displacement effect, with estimates suggesting that nearly 40% of global jobs are at risk due to AI-driven automation (Cazzaniga et al., 2024). In contrast, optimists argue that the risk that AI will result in a substantial loss of jobs is limited, as AI's productivity and reinstatement effects will compensate for any job losses, ultimately driving overall job growth.

The effect of new technologies, such as automation and robotics, on employment largely depends on whether these technologies substitute for or complement human labor (e.g., Autor 2015; Acemoglu and Restrepo, 2018; Kudoh and Miyamoto, 2024). Additionally, even if AI complements labor, workers may still face significant changes in skill and task requirements[37]. This shift could result in the emergence of new specializations and professions, while others may disappear altogether.

Whether AI complements or substitutes workers is an empirical question. So far, the evidence is mixed. Historically, automation primarily affected routine task-intensive jobs, but AI has the potential to disrupt abstract task-intensive jobs and, thus, high-

[36] While achieving this balance is not guaranteed, historical evidence suggests that labor demand has shown resilience in the long run, and demographic changes may also reduce labor supply pressures.
[37] It is important to note that this is not the sole determinant. The net effect of AI adoption also hinges on the elasticity of demand for the goods and services produced more efficiently with AI.

skilled white-collar professions. As AI capabilities advance, even more sophisticated tasks may become automated, potentially increasing job losses across both cognitive and manual occupations[38]. This could further suppress wages and labor's share of income while exacerbating income and wealth inequality.

It is still early to understand empirically the economic effects of AI. So far, the findings are mixed (Acemoglu et al., 2022; Copestake et al. 2023). This variation in findings partly stems from the fact that different types of AI systems can have distinct labor market effects. Moreover, each study uses different approaches, which makes extracting common lessons hard. Most empirical research on AI's employment impact has been conducted at the firm or establishment level. Some studies find positive effects of the adoption of AI on employment (Brynjolfsson et al., 2023; Babina et al., 2024). Others have adopted a macro perspective and identified negative effects on overall employment (Bonfiglioni et al., 2024; Huang, 2024). There is also evidence that generative AI can affect the employment and remuneration of freelance workers, displacing them in performing small- and medium-sized tasks (Hui et al., 2023). A study by the ILO (Gmyrek et al., 2023) focuses on the possible effects of the adoption of AI on tasks, arguing that there are more jobs that can be improved by AI (13.4%) than jobs that can be replaced by AI (5.1%). The IMF published a similar study, with a higher estimated percentage of jobs exposed to AI (60% in total), one half of which is at risk of displacement (Pizzinelli et al., 2023). The variation in results highlights the complexity of measuring AI's labor market effects and suggests that the impacts may depend on factors such as jobs type, industry, firm size, and geographic region.

The impact of AI on wage structures also remains equally uncertain. On one hand, AI could widen wage differentials, as companies increasingly seek highly educated, technical professionals whose skills align with AI adoption, and who generally earn above-average salaries. On the other hand, studies by Brynjolfsson et al. (2023) and Noy and Zhang (2023) indicate that AI can boost productivity for less qualified workers,

[38] As skilled employment is prevalent in high-income nations, the impact of AI could be significant in these countries, particularly for women, who hold a large proportion of clerical roles (Cazzaniga et al., 2024).

potentially empowering them to negotiate higher wages, which could help reduce inequality.

The potential for job displacement due to automation to be unequally distributed between men and women is a critical area that requires further research and understanding. Historically, industries such as manufacturing, where men are more heavily concentrated, have been most vulnerable to automation, initially placing men at a higher risk of job loss (UNESCO, OECD, ID, 2022). However, women, who tend to perform more codifiable and routine tasks across sectors, may face a slightly higher risk of displacement in fields such as clerical support and retail, where automation technologies are making significant inroads (e.g., Brussevich et al. 2019). A study by the ILO (Gmyrek et al., 2023) signals that women hold more jobs at risk of replacement by AI (3.5%) than men (1.6%). This potential disparity underscores the need for more in-depth research and analysis.

The disproportionate negative effect on disadvantaged groups and communities is a major concern. In the case of women, underrepresentation in AI-related roles is a growing concern. With AI technologies driving demand for new skills, men currently dominate in AI professions, holding the majority of positions in coding, engineering, and programming. Women make up only 22% of AI professionals globally, and this disparity threatens to perpetuate gender inequalities in the workforce[39]. This is compounded by a problem of underrepresentation and discrimination in the data sets and in the developer teams, affecting not only women but also marginalized communities, and leading to skewed outcomes and biased systems. It is crucial that policymakers take immediate action to remove barriers to education and skills development for women. Failure to do so could reinforce existing gender stereotypes and further widen the digital gender divide.

Older workers are particularly vulnerable to job displacement as AI and automation become more integrated into the workforce. Research shows that older employees face greater challenges in adapting to new technologies, leading to higher risks of unemployment. For instance, Bessen et al. (2023) found that the adverse effects of

[39] The World Economic Forum's 2018 Global Gender Gap Report (World Economic Forum).

automation tend to be more pronounced for older and middle-educated workers by using Dutch micro-data. Older workers could experience disproportionately negative impacts from AI-driven automation without timely reskilling efforts. As G7 nations grapple with aging populations, AI presents risks and opportunities for managing labor market challenges. Countries such as Japan and Italy, where more than 37% of the population will be 65 or older by 2050, must examine how AI can support, rather than displace, older workers.

2.4. Distributional effects

AI has the potential to significantly impact income and wealth inequality, particularly by widening the gap between workers who can effectively harness AI and those who cannot. Workers, who are more likely to complement AI, may experience increased productivity and wages, while those more likely to be substituted by AI may struggle to keep pace (e.g., Eloundou et al., 2024)[40], although Generative AI, in contrast to earlier automation technologies that predominantly impacted blue-collar jobs, is expected to disrupt a distinct set of 'cognitive' and 'nonroutine' tasks, particularly in middle- to higher-income professions (e.g. Kinder at al., 2024). Moreover, Machine Learning pattern recognition has an inbuilt distributional effect in that it can create groups or achieve statistical discrimination. This polarization could lead to a disproportionate increase in labor income for those already earning more, exacerbating inequality. Additionally, the productivity gains AI brings to firms are likely to boost returns to capital, further benefiting high earners and contributing to a shift in income away from labor.

In the long run, the increasing automation of tasks may reduce the overall labor share of income, with capital owners reaping most of the rewards. To counteract these trends, policymakers should invest in AI literacy and access for a broader population. Comprehensive social safety nets and retraining programs will also be important. Strengthening social safety nets and investing in education and training programs will

[40] There is also potential for AI to benefit less experienced or lower-skilled workers, helping them enhance productivity. For example, Autor (2024) argues that AI, if used well, can assist with restoring the middle-skill.

enable workers to acquire skills that complement AI, ensuring that the economic rewards of technological progress are shared more broadly across society.

Policymakers must strike a delicate balance between fostering AI innovation and ensuring equitable distribution of its benefits. While avoiding excessive taxes or regulations that could hinder AI's development, it is essential to implement policies that support a fair income distribution and labor market adaptability.

In more transformative scenarios where AI approaches or achieves AGI, the implications for labor markets and inequality could be particularly stark. Under such scenarios, AI systems that match or exceed human capabilities across virtually all cognitive and physical tasks could fundamentally alter the role of human labor in the economy. This could lead to widespread displacement of workers across skill levels, as machines become perfect substitutes for human labor. Although output and productivity would likely grow rapidly in such scenarios, the benefits may accrue primarily to owners of capital and AI technologies, potentially creating unprecedented levels of income concentration. The traditional relationship between education, skills, and wages could break down as AI systems make many high-skilled jobs obsolete. This suggests that current social insurance and income distribution mechanisms, which are largely tied to employment, might need to be fundamentally reimagined.

3. Use of AI by financial agencies and authorities

In this Chapter, we turn our attention to the role of governments as *users* of AI technology. We start by assessing the improvements that AI can provide to two key governmental functions—tax collection and expenditure control. We then move on to discuss some of the challenges that governments will face in AI implementation. Specifically, these are: how to manage the future of work in government agencies; how to handle the environmental impact of the use of AI by financial and economic agencies; how to overcome the barriers to AI use in government; and how best to address cybersecurity risks in AI-driven government functions[41].

Tax design and collection

The adoption of AI in government systems has introduced the possibility for an era of innovation, including in tax policy and administration. AI's ability to process vast amounts of data, recognize patterns, and automate complex tasks positions it as a powerful tool for addressing the challenges faced by tax authorities. AI offers a range of opportunities to modernize tax systems from improving efficiency in tax collection to combating fraud and supporting data-driven policy decisions. However, these advancements must be balanced against concerns about data privacy, implementation costs, and the need for skilled oversight. This section delves into the potential of AI to transform taxation by focusing on applications in collection, compliance, policy design, international cooperation, and fraud prevention, alongside its broader implications for governance.

Integrating AI into tax systems potentially can enhance tax collection and compliance. Traditional methods of detecting tax evasion often rely on manual audits and random sampling, which are time-consuming, resource-intensive, and prone to human error. AI can analyze large datasets to identify patterns indicative of fraud or non-compliance. It can also cross-reference taxpayer data with economic indicators, credit histories, and spending behaviours to flag potential discrepancies. As agencies explore the deployment use of such capabilities, decision makers should be mindful that current

[41] As for central banks, discussions at the G7 Central Banks' Digitalization Working Group and in other fora such as the BIS are showing that G7 central banks are becoming more sophisticated adopters of AI. Since the learning curve remains steep, there is a need to keep sharing concrete experiences and practices.

AI systems are probabilistic constructs, prone to errors and biases. The deployment of current AI capabilities into tax collection and compliance should therefore complement human judgment and should be guided by the general principles contained in the G7 Hiroshima framework and preceding documents.

AI can also facilitate the tax filing process, making it more accessible and accurate for both individuals and businesses. Filing taxes can be a complex and intimidating task, especially for those unfamiliar with the nuances of tax law. AI-powered systems can simplify this process by automating tax calculations and submissions. These systems analyze financial data, calculate liabilities, and identify eligible deductions, minimizing errors and ensuring compliance. In addition, AI-driven virtual assistants may provide real-time support to taxpayers by answering questions about filing requirements, deductions, and regulations. This not only enhances the taxpayer experience but may also reduce the likelihood of errors that could trigger audits or penalties. By automating routine tasks and providing personalized assistance, AI could reduce the administrative burden on both taxpayers and tax authorities.

Beyond administration, AI has the potential to play an increasingly important role in shaping tax policy. Policymakers must consider a wide range of factors when designing tax systems, including revenue needs, economic growth, and fairness. AI tools provide valuable insights into these considerations by enabling more accurate revenue forecasting and behavioral modeling. Predictive analytics, for example, can analyze historical tax data alongside real-time economic indicators to project future revenues under different policy scenarios. This allows governments to craft policies that align with fiscal goals while minimizing economic disruption. By leveraging these insights, governments can design tax systems that are both equitable and efficient.

Looking further into the future, a promising application of AI in taxation may be real-time reporting, which would reduce compliance costs and enhance transparency. Traditional tax reporting relies on periodic filings, creating opportunities for errors and omissions. AI, particularly when integrated with blockchain technology, could create

tamper-proof, real-time records of financial transactions[42]. These records could provide a more accurate basis for calculating tax liabilities, reducing disputes and improving trust in the system. Continuous auditing, powered by AI, would further ensure that discrepancies are identified and addressed promptly. For businesses, this would mean lower compliance costs and fewer disruptions, while governments would benefit from more reliable revenue streams.

Despite its potential, the implementation of AI in taxation is not without challenges. Data privacy is a critical concern, as tax authorities must ensure that the vast amounts of personal and financial information processed by AI systems are protected against misuse. Moreover, the integration of AI into existing tax systems requires substantial investments in infrastructure, training, and governance, while facing cultural and organizational opposition from within the tax collection organizations. Policymakers must also consider the ethical implications of using AI, particularly in cases where decisions about audits or penalties are based on algorithmic assessments. Transparent and accountable frameworks will be essential to address these concerns and build public trust in AI-augmented tax systems.

Expenditure effectiveness and efficiency

Traditional budgeting methods, which often rely on historical data and manual oversight, can fall short in addressing the complexity of modern public finance. AI may offer a solution by enabling data-driven decision-making, real-time monitoring, and optimized resource allocation. As governments face increasing demands for efficiency and accountability, AI can be a tool to improve how public funds are managed and spent.

One of the most promising applications of AI in public finance is its ability to enhance budgeting and forecasting through the analysis of vast datasets. Traditional budget forecasts often rely on static models that struggle to incorporate real-time data or adapt to rapidly changing economic conditions. AI, by contrast, has the potential to excel at

[42] Given the private and confidential nature of tax data, public blockchains would not be suitable in this context. Permissioned blockchains with robust data protection systems, and most likely under the governance of the tax collection agency, would be better suited.

predictive analytics, which could enable governments to analyze historical data alongside current indicators. This could allow for more accurate predictions of economic trends and expenditures. For instance, AI can help predict infrastructure costs and social service demands with a higher level of precision. AI-driven scenario analysis would further augment these capabilities by simulating the impacts of changes in public spending priorities. These simulations would help policymakers understand the implications of their decisions, ensuring that budget allocations align with strategic goals. Additionally, AI tools could optimize resource allocation by enhancing the identification of areas where funds will have the greatest impact.

AI also has the potential to play a critical role in improving the efficiency of public spending. Waste and inefficiencies in government expenditure are common challenges, often stemming from outdated processes and lack of oversight. AI tools could improve these issues by monitoring expenditure patterns in real-time and detecting irregularities that could indicate fraud, waste, or abuse. For example, anomaly detection algorithms could flag unusual transactions or deviations from established spending norms, which could enable authorities to take corrective action promptly. AI-driven systems could also streamline procurement processes by automating elements of tasks such as vendor selection, contract management, and pricing negotiations. Real-time monitoring capabilities further enhance spending efficiency by tracking the progress of government projects and programs. These systems would alert decision-makers when projects exceed their budgets or fall behind schedule, allowing for more timely interventions to bring them back on track.

Another potential benefit of AI in public finance could be its ability to improve program evaluation and performance monitoring. Governments often struggle to assess whether their initiatives are achieving desired outcomes, leading to misallocated resources and ineffective programs. AI helps addressing this challenge by enhancing data collection capabilities and through continuous performance tracking through the analysis of key performance indicators. For instance, AI tools could monitor the effectiveness of social welfare programs, ensuring that aid reaches the intended beneficiaries and identifying areas for improvement. This data-driven approach could support a shift toward outcome-based budgeting.

In addition to improving the management of existing programs, AI can reduce operational costs across government agencies. Routine administrative tasks, such as financial reporting, data entry, and document processing, are often time-consuming and resource intensive. AI can help automate these tasks, freeing up staff to focus on more strategic activities. For example, governments could adopt AI-powered systems to automate financial reporting, reducing errors and speeding up the budgeting process. Similarly, AI-powered chatbots and virtual assistants could handle citizen inquiries, reducing the need for human operators and lowering the costs of government-run helplines. These cost-saving measures could enable governments to stretch their budgets further while maintaining or even improving the quality of public services.

Transparency and accountability are fundamental to good governance, and AI could enhance these principles by making government spending more accessible to citizens and oversight bodies. Open data platforms powered by AI can provide real-time updates on government expenditures, allowing stakeholders to monitor how public funds are being used. These platforms could help promote public trust by enabling citizens, journalists, and watchdog organizations to scrutinize spending patterns and hold governments accountable. AI can also aid in detecting corruption by analyzing financial data for irregularities that may indicate misconduct.

While many benefits may result from the use of AI in budgeting, there are potential drawbacks. The successful implementation of AI would require careful planning, robust governance frameworks, and a commitment to ethical principles. Moreover, some of the technical characteristics of AI, such as the limited ability to explain the outcomes of highly non-linear and high-dimensional AI systems may conflict with fundamental principles governing the public sector, such as the right to explanation[43].

[43] Consumer Financial Protection Circular 2022-03 Consumer Financial Protection Circular 2022-03: Adverse action notification requirements in connection with credit decisions based on complex algorithms | Consumer Financial Protection Bureau.

The future of work in financial agencies and authorities

One of the most pressing challenges for developed countries in the coming decades is the aging of their populations and the consequent decline in the supply of skilled labor, particularly in government service. Aging populations mean that a larger proportion of citizens will require services such as healthcare, pensions, and social support, increasing the demand for efficient and effective government operations. Simultaneously, the supply of experienced professionals to deliver these services is dwindling. Retirement rates among government employees are rising, and younger generations often prefer careers in the private sector, attracted by higher salaries, more flexible work environments, and opportunities for innovation[44]. This dual pressure creates a critical gap in the ability of governments to meet public needs while maintaining high service standards.

AI offers a promising solution to mitigate these challenges by automating processes traditionally handled by skilled, experienced government employees. By leveraging AI, governments can enhance efficiency, reduce reliance on a shrinking human workforce, and redirect remaining personnel toward tasks requiring human judgment, empathy, or political discretion. Automation can also help preserve institutional knowledge as experienced workers retire, embedding their expertise into AI systems that can carry out routine but essential functions.

For example, AI can automate how governments identify and address cases of tax fraud or evasion. Traditionally, this task requires experienced auditors who are adept at analyzing financial records, identifying irregularities, and interpreting complex tax codes. AI systems capable of processing vast amounts of financial data to detect patterns indicative of noncompliance could allow governments to maintain robust tax enforcement even as the number of skilled auditors declines. In the management of government benefits, AI could increase the productivity of government employees in processing applications, verifying eligibility, and ensuring compliance with program rules, as AI systems equipped with natural language processing and advanced data

[44] Rendon (2024) finds empirical evidence that Generation Z adults are less likely to aspire to work in the public sector than non-Generation Z (https://scholarsjunction.msstate.edu/cgi/viewcontent.cgi?article=7175&context=td).

analytics can automate significant portions of these workflows. AI can review applications for completeness, cross-reference data with other government records to verify eligibility, and even answer routine inquiries from applicants through AI-powered chatbots. This would allow governments to handle growing caseloads with fewer staff while improving service speed and accuracy.

One significant concern resulting from the automation of government processes is the potential displacement of government employees, particularly those in the later stages of their careers. Many of these workers have spent decades developing specialized expertise and may find it difficult to transition to new roles or industries if their positions are automated. This could lead to increased job insecurity and dissatisfaction among public sector workers, further complicating governments' ability to attract and retain talent.

Policymakers can address this risk by creating pathways for displaced workers to transition into new roles. This could include reskilling programs tailored to the needs of the public sector, such as training employees in managing and interpreting AI systems or shifting them into roles requiring human-centric skills, such as policy advising or community engagement. Governments should also consider phased automation approaches, allowing employees nearing retirement to complete their careers with dignity while enabling younger staff to adapt to new, AI-enhanced roles.

Preparing for and accelerating the adoption of AI in government requires strategic investment and institutional readiness. Governments must prioritize developing internal AI expertise by establishing dedicated teams to oversee AI projects and integrating AI training into professional development programs for existing staff. Collaborating with academic institutions and private-sector partners can help governments access cutting-edge AI technologies and insights while mitigating risks through shared learning.

Additionally, governments should adopt flexible regulatory frameworks to facilitate experimentation with AI in a controlled and ethical manner. Pilot programs can be used to test AI applications in specific areas, such as tax audits or benefits management, providing valuable insights before scaling these technologies nationwide.

Transparency and public trust are critical; governments must communicate the benefits of AI adoption while ensuring that data privacy and ethical standards are rigorously upheld.

Minimizing the carbon footprint of using AI in financial governmental functions

The energy-intensity of AI necessitates a careful approach by governments to ensure alignment of AI adoption with their environmental objectives. Measures to lower the carbon footprint of the use of AI by the public sector come with cost savings.

Agencies yield can lower their environmental footprint related to AI in several ways. These include: i) selective adoption of generative AI for use-cases where there is a demonstrable improvement in performance over standard methods (to avoid the unnecessary environmental burden of generative AI); ii) using task-specific rather than multi-task models, where possible (using fine-tuned models that operate within a specific knowledge domain that maintain high performance but low energy consumption; see Luccioni et al., 2024); iii) powering AI computing with “green” data centres (or cloud service providers) that provide specialized low-energy hardware, high-performance cooling and heat re-use, and loads matched to renewable energy supply; iv) careful measurement and optimization of AI model energy consumption through tools such as Code Carbon, MLCO2, LLMCarbon and CentML; and v) experimentation with innovative approaches related to the aforementioned levers, including for example, co-development of specialized hardware and software with industry for high performance and resource efficiency (Wu et al., 2022), engagement in green software trials with other public authorities, and investment in more efficient data science techniques that reduce the size of training data sets[45].

[45] Tony Blair Institute for Global Change (2024), Greening AI: A Policy Agenda for the Artificial Intelligence and Energy Revolutions.

Barriers to AI adoption in Government

Early evidence suggests that the adoption of AI in government is lagging behind its implementation in the private sector, largely due to a range of structural, legal, financial, and organizational challenges[46].

A major barrier to AI adoption in the public sector is a financial constraint. A recent study, conducted by SAS in partnership with Civil Service World found that two-thirds of surveyed officials cited financial constraints as the primary roadblock to invest in the infrastructure, talent, and maintenance required for effective AI deployment[47]. AI systems often require significant upfront costs, including the acquisition of advanced computing hardware, the development of custom software solutions, and ongoing operational expenses. These challenges are particularly acute in smaller or resource-constrained jurisdictions where funding for technological innovation competes with other pressing priorities such as healthcare, education, and infrastructure.

In addition to financial challenges, governments face a widespread skills deficit that impedes AI adoption. AI development and implementation demand a high level of technical expertise, including knowledge of data science, machine learning, and software engineering. However, many government agencies lack in-house technical teams capable of handling these complex tasks. Such a gap in skills becomes binding when adoption requires customization of general-purpose systems (for example, via retrieval augmented generation or through fine tuning), or even in procurement decisions. Compounding this issue is the difficulty of attracting and retaining top AI talent, as government salaries and work environments often cannot compete with the private sector. This shortage of expertise not only slows the pace of AI adoption but also increases reliance on external contractors, which can raise costs and complicate project management.

Another critical obstacle is the prevalence of outdated infrastructure in many government agencies. Legacy systems, often designed decades ago, are ill-equipped to support modern AI technologies. These systems frequently lack the processing

[46] Government Technology (2024), What Government Can Learn from the Private Sector About AI.

[47] SAS (2022), Data & AI in the UK government: Overcoming barriers and realising potential | SAS UK.

power, scalability, and flexibility required for AI integration, making upgrades a necessary but expensive undertaking. Moreover, the incompatibility of legacy systems with new technologies can create inefficiencies and bottlenecks, further delaying AI deployment. Governments must balance the need to modernize their IT infrastructure with the fiscal and logistical challenges of doing so, a task that requires careful planning and prioritization.

Regulatory complexity adds another layer of difficulty to AI adoption in government. Ensuring the ethical, secure, and transparent use of AI is a significant challenge, particularly given the sensitive nature of many government functions. Citizens and stakeholders expect AI systems to be fair, unbiased, and accountable, but achieving these goals requires robust regulatory frameworks and enforcement mechanisms. Governments must navigate a delicate balance between fostering innovation and safeguarding public trust, a process that demands proactive policymaking and continuous oversight.

Compounding these challenges are issues related to data quality and accessibility. AI systems rely on large volumes of high-quality data to function effectively, but many government agencies struggle with fragmented, outdated, or incomplete datasets. Poor data quality not only reduces the accuracy and reliability of AI models but also limits their applicability to real-world problems. Additionally, the lack of standardized data-sharing protocols between agencies creates silos that hinder collaboration and innovation. Improving data infrastructure is therefore a prerequisite for unlocking the full potential of AI in government.

To overcome these barriers, governments must adopt a multifaceted approach that addresses financial, technical, and organizational challenges. One critical strategy is prioritizing investment in AI infrastructure and talent. Governments must recognize that upfront expenditures on modernizing IT systems, training staff, and acquiring AI tools are essential for long-term gains in efficiency and service delivery. By allocating dedicated funding for AI initiatives, governments can create the foundation needed for successful implementation.

Cybersecurity challenges in AI-Driven Government Functions

As governments adopt AI for critical roles such as tax collection and public expenditure management, ensuring robust cybersecurity is essential. AI systems, which process vast amounts of sensitive data and influence significant decisions, are particularly vulnerable to a range of evolving threats. Addressing these risks proactively is crucial to safeguarding the integrity and reliability of AI in public governance.

A significant risk is data breaches, as AI systems handle extensive personal and financial information, making them attractive targets for cybercriminals. Weak encryption, inadequate access controls, and outdated security protocols can lead to unauthorized access and theft. Additionally, adversarial attacks pose a unique threat to AI systems, where malicious actors manipulate models to produce incorrect results or extract proprietary algorithms, compromising their functionality and accuracy.

Other threats include malware and ransomware, which can disrupt AI-driven services, hijack resources, or gain unauthorized access to government networks. Model poisoning, where attackers introduce malicious data into AI training sets, further exacerbates risks by skewing predictions and undermining decision-making processes. Collaborations with external vendors and the use of third-party tools also introduce vulnerabilities, particularly when security standards are inconsistent, or compliance varies across jurisdictions.

To mitigate these risks, governments must implement a combination of technical, procedural, and organizational measures. Strong encryption, strict access controls, and advanced monitoring systems are essential for protecting sensitive data. AI-specific security protocols should address adversarial manipulation and model poisoning, while continuous monitoring and regular vulnerability assessments can help detect and respond to emerging threats.

Secure development practices, including rigorous testing and adherence to governance frameworks, are critical for ensuring the integrity of AI systems. Data minimization and retention policies can reduce exposure by limiting the amount of sensitive information collected and stored. For third-party risks, governments should vet external tools thoroughly and enforce compliance with security standards.

Training staff on AI-related cybersecurity risks and best practices can reduce human errors, while comprehensive incident response plans ensure readiness for potential breaches.

4. AI and financial stability

The financial sector has historically been at the forefront of adopting new technologies– from the abacus to modern AI. This pioneering role provides early insights into the impacts of automation and AI on the broader economy as well as on financial stability. As such, the financial sector can enjoy significant potential to improve efficiency and productivity (Leitner et al., 2024). AI can enhance information processing, increase the accuracy of quantitative predictions, streamline operational processes, strengthen risk management functions, reduce cyber vulnerabilities and better match products with customers. Financial institutions and their service providers are actively exploring a range of AI applications, including customer support, fraud detection, market analysis, document processing, information retrieval, and software development (FSB, 2024).

However, this phenomenon also implies that the financial sector has traditionally been among the first to experience the challenges posed by new technologies (Aldasaro et al, 2024).

The ongoing rapid advances in AI, including the current rise of AI agents, mark a new phase in the automation of the financial sector, introducing systems capable of executing more complex tasks without human involvement. As AI systems become faster and more powerful, there is a growing difficulty in maintaining effective human oversight, presenting new challenges for financial stability.

It is important to note that many of the risks discussed depend on a significant share of companies adopting AI. The lack of comprehensive data on AI adoption by financial services firms complicates assessing use cases (FSB, 2024). While some surveys suggest firms are quick to move from experimentation to deployment (Bank of England and FCA, 2024) the data is mixed. High adoption rates often come from surveys of larger firms, while broader economic indicators show low adoption, especially for customer-facing activities. This gap may be due to competitive pressures, with firms eager to showcase AI adoption through marketing. As to banks, evidence from European Central Bank (2023) shows that 60% of players in the Euro area are already using AI, with more use cases in development.

4.1. Systemic Risks

The speed and scale of AI-driven decisions in financial markets introduce new systemic risks that range from the homogenization of decision-making processes and hidden points of failure to greater herding and the risk of market manipulation.

Also, the concerns regarding financial stability addressed in the following sub-sections largely stem from risks associated with market concentration. Specifically, these risks arise from monocultures and from the fact that many financial actors base their AI systems on the same foundation models (e.g. for trading and investing decision), chip producers, and cloud computing providers, as well as the dependence on a small number of providers for financial data. Each of these factors has the potential to create systemic risks.

i. Homogeneity, Cybersecurity, and Hidden Points of Failure

The increasing adoption of AI in finance raises concerns about the homogenization of decision-making processes (Gensler and Bailey, 2020; Bommasani et al., 2022). If multiple financial institutions rely on similar systems, it could lead to a monoculture in the financial system. This uniformity increases the risk of correlated failures or simultaneous adverse reactions to market events (Carlson, 2007).

The main danger of this homogeneity lies in its potential to amplify shocks to the financial system. If many institutions use a single AI system or similar AI systems with similar blind spots, then they may all fail to anticipate or respond appropriately to certain types of market events, especially out-of-distribution events. This could lead to synchronized behavior that exacerbates market movements, turning small fluctuations into major disruptions.

Moreover, the complexity of advanced AI systems introduces the risk of hidden points of failure, including broader operational and cybersecurity risks, that may not be apparent until a crisis occurs. For example, AI-based liquidity management systems may outperform traditional methods during normal times but exacerbate liquidity crises. Unlike simpler rule-based systems, deep learning models can develop intricate internal representations that are difficult for humans to interpret or audit. This opacity makes it

challenging to identify potential failure modes or biases in advance or during crises events.

The opacity of AI systems also raises concerns about accountability. In the event of AI-driven market disruptions, it becomes challenging to attribute responsibility or understand the root causes of the problem. This complicates the task of developing effective preventive measures and crisis response strategies.

ii. Automated Herding, Pro-cyclicality, and Market Manipulation

AI agents, operating at speeds far beyond human capability, can make rapid, large-scale decisions that may lead to unintended market movements. This speed, combined with the potential for multiple AI systems to react similarly to market signals, creates a risk of automated herding behavior and greater cyclicality (OECD, 2021; Shabsigh and Boukherouaa, 2023).

Historical precedents, such as the 1987 stock market crash, demonstrate how automated systems can exacerbate market volatility. In that instance, computerized trading programs led to a cascade of sell orders, triggering the largest single-day global market decline in history. The 2010 flash crash in US equity markets represents another example of how algorithms—then already significantly faster—can create havoc through herding behavior, with major stock indices dropping by 7% within 15 minutes. What is instructive about the 2010 flash crash is that it took regulators five months to identify the causes of the market drop and issue a report, which is still contested.

Moreover, AI systems could enable more sophisticated forms of market manipulation. Their ability to process vast amounts of data and identify subtle patterns could be exploited to create or exploit market inefficiencies at scales and speeds that are difficult for human regulators to detect or counter. Furthermore, these expanded algorithmic capabilities also create the risk of undesired algorithmic collusion, which competition law is currently poorly equipped to address (Calvano et al., 2020).

iii. Impact of Competition on AI Risks

The competitive landscape in the financial sector has complex implications for financial stability risks, with both positive and negative potential outcomes.

On the one hand, competition can drive innovation and greater efficiency, potentially leading to a more robust and effective resource allocation. Competition in AI deployment can reduce economic rents and promote innovation. As financial institutions vie for market share, they are incentivized to develop more sophisticated, accurate, and efficient AI systems. This competition can speed up advancements in AI capabilities, potentially improving risk management and overall market efficiency. Additionally, competition can drive down costs, leading to a democratization of access to financial services and, in the best case, to a more diverse and resilient financial system.

On the other hand, institutions acting in a highly competitive environment may prioritize the speed of deployment and short-term performance gains over thorough testing and risk assessment of their AI systems. This could lead to the proliferation of insufficiently tested or poorly understood AI models, and creates the risk of institutions spending large amounts of capital without obtaining returns. When there are externalities, competition provides strong incentives to race to the bottom and forces companies to ignore their external effects on safety and system stability to stay ahead of their competitors, unless they are forced to do so by their regulators. By its very nature, systemic risk involves externalities. This implies that competition may exacerbate the systemic risks described above.

The development of advanced AI capabilities requires substantial resources, potentially leading to a concentration of AI capabilities in a few large financial institutions. This concentration could create new forms of systemic risk, where the failure or compromise of a single institution's AI system could have outsized effects on the entire financial system. Smaller institutions, unable to compete in AI development, might either become overly reliant on third-party AI providers, or being crowded out by larger players, further concentrating risk. Moreover, as AI technology becomes more central to financial activities, large technology companies may increase their

participation in financial markets, raising concerns about market power and concentration (e.g. Grout, 2021).

Opacity and financial supervision

The increasing complexity and opacity of AI systems pose significant challenges for financial supervision and stability. Many advanced AI models, particularly those based on foundation models and other deep learning systems, operate as black boxes, making it difficult to explain their decision-making processes. This lack of explainability complicates risk assessment and regulatory oversight.

The opacity of modern AI systems manifests itself in several critical areas within finance. Traditional risk models are typically based on well-understood statistical methods with clear assumptions and limitations. In contrast, newer AI models may identify complex, non-linear relationships in data that are not easily interpretable. This makes it challenging for risk managers to assess the robustness and reliability of these models, especially under stressed or unprecedented market conditions. The opacity also complicates regulatory compliance, as many financial regulations require institutions to explain and justify their decision-making processes, which is difficult to do in a meaningful manner as AI systems increasingly drive decision-making.

Furthermore, the opacity of AI systems could create challenges for systemic risk assessment and audit processes. Regulators tasked with maintaining financial stability may struggle to understand how different parts of the financial system might interact, especially during times of stress, when these interactions are driven by opaque AI systems. Traditional auditing methods become more difficult with complex AI systems, challenging the ability of both internal and external auditors to verify the integrity and compliance of these systems.

The opacity problem is further exacerbated by the rapid pace of AI development. As new AI techniques and architectures emerge, the complexity and inscrutability of these systems often increase. This could create a moving target for regulators and risk managers, who must continually update their understanding and assessment methods. Moreover, the proprietary nature of many AI systems in finance adds another layer of opacity. Financial institutions may be reluctant to fully disclose the details of their AI

models to protect their competitive advantage, further complicating efforts to understand and regulate these systems at a systemic level.

Cybersecurity in the financial sector

AI technologies present both opportunities and challenges for cybersecurity in the financial sector. On one hand, AI can enhance defensive capabilities, improving the detection and prevention of cyber threats. AI-powered systems can analyze vast amounts of data to identify anomalies and potential security breaches more quickly and accurately than traditional methods.

On the other hand, AI also enables more sophisticated cyberattacks. Adversarial AI can be used to create more convincing phishing attempts, automate the discovery of software vulnerabilities, or launch more effective distributed denial-of-service (DDoS) attacks. The financial sector, with its high-value targets and critical role in the economy, is likely to be a primary focus for such advanced cyber threats.

The interconnectedness of the financial system has the potential to amplify cybersecurity risks. The financial impacts of a successful attack on one institution could potentially spread rapidly through the system, causing widespread disruption. As AI systems become more integrated into core financial operations, ensuring their security becomes increasingly critical for maintaining overall financial stability.

Real Disruptions and Financial Implications

The impact of AI extends beyond the financial sector itself, with potential disruptions in the real economy having significant implications for financial stability. Like all technological changes, advances in AI may create winners and losers (Korinek and Stiglitz, 2019). If AI capabilities advance rapidly, as in some of the more disruptive scenarios that we described, then there is a risk that the sources of income of certain categories of high-income households and large corporations may be undermined.

In the context of households, as discussed in Chapter 2, there are scenarios in which AI may negatively impact labor markets. If AI capabilities advance rapidly, there is potential for significant job displacement across sectors. From a financial stability perspective, loan defaults would increase if a meaningful fraction of the workforce sees

their skills devalued and experiences prolonged unemployment or underemployment. For example, whereas the credit risk of lending to software engineers seemed low just a few years ago, recent advances in automated coding have made the possibility that demand for software engineers may soon decline more plausible.

In the corporate sector, rapid AI advancement scenarios could lead to significant shifts in revenue streams and market dominance, on a previously unseen scale. Companies that successfully leverage AI technologies may experience rapid growth, while those that lag behind could face equally rapid obsolescence. This dynamic could generate increased credit risk for traditional businesses as they face declining revenues and profitability, potentially leading to defaults on corporate loans and bonds. Entire industries could be disrupted in relatively short time frames, making traditional credit risk models and sector-based diversification strategies less effective as AI-native companies might quickly dominate sectors.

The resulting shifts in corporate fortunes could have cascading effects on financial markets, potentially leading to stark changes in asset valuations. Such rapid and significant changes in valuations could create financial stability risks through several channels. We may see increased volatility in equity and corporate bond markets as investors try to price in the uncertain future impacts of AI. Rapid shifts in valuations could lead to liquidity crunches, especially if there's a flight to quality during periods of AI-driven market uncertainty. Banks and other financial institutions may face growing difficulties in accurately assessing credit risks and setting appropriate loan loss reserves in a rapidly changing, AI-driven economy.

The international dimension adds another layer of complexity to these financial stability risks. An "artificial intelligence divide" could emerge, where countries or regions with advanced AI capabilities might see rapid economic growth and appreciation of their financial assets, while others may lag behind. This divide could also lead to exchange rate volatility. Countries perceived as AI leaders might experience large capital inflows, with the associated financial stability risks, while countries falling behind might face capital flight and increased borrowing costs.

5. Policy Recommendations

In the previous Chapters, we discussed how the rapid evolution of AI presents a rather complex landscape of opportunities and risks, particularly within the realms of finance and economic policy. As AI technology becomes increasingly integrated into global financial systems and economic decision-making, policymakers face the monumental task of crafting strategies that harness AI's benefits while safeguarding against potential disruptions. As emphasized throughout this Report, this challenge is compounded by the diverse and sometimes unpredictable ways AI may evolve, making it imperative to develop policy frameworks that are both robust and adaptive.

At the heart of effective AI policy lies the necessity for guidance informed by rigorous scientific understanding. As outlined by Bommasani et al. (2024)[48], the science of AI risk assessment is in its nascent stages, but it must play a central role in shaping regulatory measures. Policymakers must ground their decisions in data-driven analysis, yet they face a dilemma: AI science, despite its strides, remains far from reaching a definitive consensus. Fundamental questions about the long-term impact of AI on economies and societies remain subjects of active debate among experts.

These uncertainties pose a formidable challenge. The diverse projections and conflicting expert opinions make it difficult to craft one-size-fits-all policy prescriptions. In the absence of an established consensus, decision-makers must recognize that policy cannot be indefinitely postponed or paralyzed by debates over AI's eventual trajectory. Indeed, waiting passively to see how AI developments "play out" is a luxury that governments cannot afford, particularly when global financial stability and economic prosperity are at stake.

Given these complexities, the guiding concept for AI governance should be one of policy preparedness. This approach emphasizes readiness and flexibility, ensuring that policies can be adjusted and refined as new evidence emerges and technology matures. Preparedness entails anticipating various scenarios, from AI-driven financial innovation that could revolutionize economic growth to potential crises that may require swift and effective mitigation measures. It is about building institutional and regulatory

[48] A Path for Science- and Evidence-based AI Policy.

frameworks that are resilient to economic and societal changes brought by the deployment of AI and capable of rapid response.

This Report proposes a policy preparedness approach based on two pillars. First, we propose a set of ten immediate policy recommendations that would advance the preparedness of economic and financial authorities facing uncertainty about the future of AI technology. Second, we recommend that each government develops its own policy preparedness matrix, as the basis for monitoring AI policy preparedness. This Report recognizes that each country and each jurisdiction may prefer different policy mixes, depending on culture, legal systems, national priorities, perception of risk and specific needs. We are not prescribing a specific policy mix for all G7 countries to follow, but instead a framework to help each country develop its own policy path.

AI Scenarios

An AI policy framework based on preparedness starts by identifying the key uncertainties and organizing them into scenarios. While an infinite number of scenarios is always possible, the goal is to identify a limited number of scenarios that make the framework tractable, while representing the variability of outcomes. This Report proposes three distinct scenarios that allow policymakers to develop their own "policy preparedness matrix," in which each policy objective is matched by viable policy options for each scenario.

As explained in Chapter 1, there is a striking divergence in experts' predictions about the future of AI. On one side, skeptics argue that the impact of AI will be more gradual and far less transformative, following at best patterns similar to previous general-purpose technologies such as the internet. On the other, AI experts and industry leaders forecast radical, transformative change that could fundamentally reshape our economy and society. Geoffrey Hinton, one of the pioneers of deep learning, predicted in 2023 that human-level AI, also termed AGI, will be reached in "5 to 20 years but without much confidence. We live in very uncertain times." As discussed throughout this Report, this view is far from consensus, and the wide divide in expert opinions presents a significant challenge for policymakers attempting to prepare proactive policies for the future.

In light of this uncertainty, scenario planning is one of the most valuable tools at our disposal (Korinek, 2023). By considering a range of possible futures, we can develop more robust strategies and policies adaptable to various outcomes. This approach allows us to hedge against the risks associated with committing too heavily to any single prediction while ensuring we are prepared for a spectrum of possibilities.

Our strategy for setting scenarios aims to span the full range of uncertainty, from the most conservative estimates to the most techno-optimistic predictions. This approach ensures that we consider both incremental change and potential radical disruption. By doing so, we can prospectively stress-test our economic models and policy frameworks against a variety of futures, identifying areas where they may need to be reinforced or redesigned.

Based on this strategy, we propose three distinct scenarios for consideration.

1. Conventional Wisdom Scenario: This scenario aligns with the view that technological change is always gradual, and AI will follow a similar trajectory to earlier general purpose technologies. Under this scenario, AI may lead to a modest increase in productivity growth, perhaps in a similar range to the computer and internet boom of 1995-2024 but would not cause radical disruption to existing economic structures. The development and deployment of AI capabilities continue at a steady but manageable pace, allowing for gradual adaptation of workforce skills and business models.

2. Intermediate AGI Scenario: This scenario incorporates predictions of AGI and transformative change, but on a more extended timeline of approximately 20 years, which was the upper bound of Hinton's prediction. In this future, we see a progressive advancement of AI capabilities across domains, gradually approaching human-level performance in increasing tasks. This scenario allows for more time to adapt economic policies and social structures but may still generate significant challenges by the 2040s.

3. Accelerated AGI Scenario: This scenario envisions the development of AGI within a 3 to 5-year timeframe (2027 to 2029), reflecting the most optimistic predictions from the Silicon Valley community (e.g., Aschenbrenner, 2024) as well as the lower bound of Hinton's time frame. This rapid progress could be driven by breakthrough algorithms,

exponential increases in computing power, or unexpected synergies between existing AI technologies. Under this scenario, we would see a swift and dramatic transformation of the economy, with AI systems quickly matching or surpassing human capabilities across a wide range of cognitive tasks. This could lead to unprecedented productivity gains but also pose significant challenges for labor markets and economic and social stability.

While the "Accelerated AGI" scenario may seem unlikely to some, the potential magnitude of its impact makes it crucial to consider it in our planning. Optimal preparedness requires us to take seriously even scenarios that some consider low probability but that would have extreme consequences.

By examining this range of scenarios, we can better understand the possible economic impacts and policy challenges that may arise from advances in AI (Korinek and Suh, 2024). This approach allows us to develop more flexible and robust strategies for managing the economic transition, regardless of which path AI development ultimately takes.[49] Moreover, even if AGI is reached technically, there is significant uncertainty about its adoption and rollout. This uncertainty will have first-order implications for the economic and social impact of AGI.

AI Scenarios: impact on the economy and financial systems

The three scenarios described above have distinct consequences for the economy and the financial system. In the "Conventional Wisdom" Scenario, gradual AI progress implies that the financial stability challenges of AI are largely limited to the technical challenges from integrating AI into the financial system. Under this scenario, financial institutions adopt AI technologies incrementally, continuing the experience of prior decades, enhancing efficiency and decision-making processes without radical disruption. Regulatory frameworks must adapt progressively to manage risks

[49] In addition to the uncertainty about when AGI will be reached that could automate human cognitive tasks, there is also uncertainty about complementary advances in robotics that could automate physical tasks that used to be performed by humans. At present, many experts see advances in robotics lagging recent advances in AI, but perhaps not by much. The development of AGI would likely also lead to intelligent machines designing far more capable robots.

effectively. If managed well, this scenario could see improvements in risk management and market efficiency, with limited disruption to existing financial structures.

In the “Intermediate AGI Scenario”, there is a gradual development of AGI over the coming two decades. Here, AI capabilities progressively approach human-level performance in all areas, including finance. In this scenario, we might see the emergence of a growing number of AI agents that automate financial functions that have hitherto been reserved for humans, including in financial intermediation, insurance, asset management, and liquidity management (Aldasaro et al., 2024). These AI agents may autonomously design, market, and sell complex financial products and services, challenging regulators to keep pace with these advancements. The potential for increased market volatility would grow as AI systems become more sophisticated. As AGI approaches, this scenario could also be associated with significant shifts in demand for workers, potentially causing some of the disruptions discussed in the preceding sections.

The “Accelerated AGI Scenario” posits that AGI will be achieved within 3-5 years, leading to rapid and transformative changes in both the financial sector and the broader economy. This scenario presents the most acute challenges for financial stability. AI systems would quickly surpass human comprehension in financial decision-making, exacerbating the opacity issues detailed earlier and presenting extreme challenges for regulatory oversight. The scenario could potentially lead to a loss of human control over the financial system, with heightened risks of severe market instability due to the speed and scale of AI-driven decision-making. In the real economy, we could see a dramatic disruption of labor markets, potentially causing widespread loan defaults as outlined earlier. Rapid and significant shifts in asset valuations could occur, amplifying financial stability risks.

Immediate recommendations for AI Policy Preparedness in Economic and Financial Policy

The first element of this Report’s policy preparedness framework consists of ten immediate policy recommendations to guide G7 nations in navigating the challenge of

uncertainty around the future of AI while fostering a stable, inclusive, and prosperous economic landscape.

i. Adopt a proactive and flexible policy stance

Policymakers face the dual challenge of addressing immediate AI-related issues while remaining adaptable to new developments and unforeseen risks. AI technologies evolve rapidly, and rigid regulatory frameworks can quickly become obsolete. A proactive stance requires policymakers to anticipate emerging challenges and seize opportunities while simultaneously embedding flexibility into their strategies.

For instance, finance ministries and central banks could institute annual reviews to evaluate the adequacy of AI-related policies in the face of technological advancements. These reviews should include stakeholder consultations to capture diverse perspectives, from financial institutions to technology developers. Flexibility also means adopting experimental approaches, such as pilot programs for AI applications in finance, to test solutions in controlled environments before scaling them up. Such a strategy ensures that regulatory frameworks remain relevant and effective in a dynamic landscape.

ii. Navigate policy trade-offs with transparency

As with most policies, AI policies inevitably involve trade-offs, particularly when addressing conflicting objectives such as innovation versus consumer protection or fairness versus privacy (Yang et al., 2024). Policymakers must go beyond simply stating AI principles and confront the tensions that inevitably arise from trying to achieve multiple policy goals simultaneously. Moreover, policymakers should be prepared to communicate the rationale behind their decisions clearly to build public trust.

Balancing explainability and accuracy in AI systems in the context of lending or insurance underwriting is one concrete example of such a trade-off. While explainable AI enhances accountability, highly accurate but opaque models may deliver better outcomes and foster financial inclusion. Policymakers need to define thresholds for acceptable trade-offs in different contexts, informed by empirical research and stakeholder engagement. In parallel, governments should invest in public outreach

initiatives to educate citizens about these trade-offs, fostering greater acceptance and understanding of AI policies.

iii. Rapidly augment AI expertise within institutions

The effective governance of AI requires a deep understanding of its capabilities, limitations, and risks. Governments should continue building internal expertise by promoting AI literacy across all levels of policymaking and investing in advanced skill development for specialized teams. This could include allocating dedicated budgets for training programs, hiring technical experts, and fostering partnerships with academic institutions and industry leaders.

For example, finance ministries could establish AI research units tasked with analyzing emerging trends and advising on policy implications. External advisory boards comprising AI researchers and technologists could further supplement internal expertise. Moreover, governments could institutionalize mechanisms for continuous learning, such as AI training for senior policymakers, to stay abreast of rapid advancements in the field.

iv. Explore AI capabilities to enhance public institutions

As discussed in Chapter 3, AI offers transformative potential to improve the efficiency and effectiveness of public sector functions, including within central banks, finance ministries, and regulatory agencies. Policymakers should actively explore innovative AI applications, leveraging experimental approaches such as regulatory sandboxes to test new tools in real-world settings.

For instance, AI-driven workflows could revolutionize tax collection by automating complex data analysis and fraud detection processes. Chain-of-thought reasoning capabilities in advanced AI systems could streamline end-to-end workflows in areas like anti-money laundering (AML) compliance or financial forecasting. By embracing experimentation, governments can identify scalable solutions that enhance public sector performance without compromising stability or accountability.

v. Establish AI procurement strategies

The acquisition and deployment of AI tools in public institutions should prioritize ethical standards, cost-effectiveness, and sustainability. The approach must ensure that AI technologies align with principles like fairness, transparency, and inclusivity, as articulated in the Hiroshima Framework and OECD guidelines.

Policymakers should develop an approach for evaluating the risks and benefits of AI tools before implementation. These evaluations should also account for the environmental impact of AI systems, particularly their energy and water usage. By adopting responsible practices, governments can demonstrate best practices for the private sector to follow regarding ethical AI deployment and demonstrate their commitment to sustainability.

vi. Modernize government data management practices

Effective data management is critical for realizing AI's potential while safeguarding privacy and security. Governments should adopt modern data management tools and practices, including sovereign cloud integrations, where appropriate, to enhance the efficiency and reliability of their data infrastructure.

Sovereign cloud solutions can ensure that sensitive government data and/or data important for national security purposes remains secure and compliant with privacy regulations while enabling seamless collaboration across agencies[50]. Advanced data management systems going forward could also incorporate features like real-time analytics and interoperability to support complex decision-making processes. By modernizing their data ecosystems, governments can create a solid foundation for AI innovation.

vii. AI cybersecurity:

Given the rise of AI-enhanced cybersecurity threats governments should be prepared to safeguard critical infrastructure and financial systems. As malicious actors

[50] A sovereign cloud ensures data remains within a specific jurisdiction, complying with that region's data residency and privacy laws. Key aspects of sovereign AI clouds include control over data residency, infrastructure and access management (Zeichick, 2023).

increasingly deploy AI to create sophisticated attacks, from automated phishing schemes to advanced network infiltration techniques, governments must ensure they are prepared to counter these threats.

Immediate actions may include establishing dedicated AI-focused cybersecurity task forces to develop and implement countermeasures. Governments could also invest in AI-driven threat detection systems capable of analyzing patterns in real time to identify and neutralize emerging risks. Cross-border cooperation is critical in this domain, as cybersecurity threats often transcend national boundaries. Collaborative platforms for intelligence sharing and joint response initiatives must be prioritized to create a unified defense against AI-driven cyberattacks. Policymakers could also mandate regular cybersecurity audits and penetration testing for public institutions to identify vulnerabilities and strengthen resilience against evolving threats.

viii. Establish guidelines for the reliance on general AI tools in financial activities

As financial institutions increasingly rely on general AI tools, such as pre-trained models, regulators should provide clear guidelines to ensure their responsible use. These guidelines should address issues such as transparency, technical assurances, and adherence to terms and conditions that protect consumers and the financial system.

For instance, generalist pre-trained AI models used as foundations of AI-driven tools in credit scoring or investment decision-making should be subject to rigorous validation processes to ensure their accuracy and fairness. Regulators should also require providers to disclose the limitations of these models to foster greater accountability and consumer trust.

ix. Monitor labor market impacts and prepare for disruptions

The scale and direction of labor market impacts are very uncertain, and ongoing monitoring will be important. AI's potentially transformative impact on labor markets necessitates measures to mitigate workforce disruptions. Policies that support reskilling and adaptability can help ensure that workers can transition into new roles as the demand for some traditional skills declines. Governments should also

incentivize industries to adopt AI responsibly, balancing efficiency gains with the need to preserve employment opportunities. These efforts will help build a more resilient workforce capable of thriving in an AI-driven economy.

However, policymakers should also prepare for more transformative scenarios in which AI capabilities approach or exceed human-level performance across most tasks, potentially leading to widespread disruption of labor markets. In such scenarios, traditional workforce development approaches and social safety net may need to be fundamentally reimagined.

x. Enhance cross-border cooperation on AI governance

The global nature of AI technologies calls for international engagement to address shared benefits, risks and challenges. While each country or region will set its own policies, policymakers can benefit from cross-border collaboration in areas such as financial market regulation, tax collection, trade, and cybersecurity to ensure consistent standards and practices.

Establishing frameworks for knowledge sharing and joint monitoring of AI systems deployed and used in financial systems across countries and regions is particularly critical. Initiatives like the BIS' proposal for a "community of practice" can facilitate knowledge exchange and foster stronger international partnerships[51]. Similarly, existing international fora, such as the OECD, can play a key role in promoting convergence and alignment through transparency, dialogue, and the exchange of best practices. By working together, nations can harness and share AI benefits, manage AI's global risks more effectively and promote its responsible use across borders.

The AI Policy Preparedness Matrix

Beyond the above immediate recommendations, we suggest the adoption of a policy preparedness matrix. The policy matrix would list the appropriate types of AI policy to be pursued under the alternative AI scenarios, and the status of readiness in each. Developing a policy preparedness matrix allows for identification of policies that work

[51] BIS Annual Economic Report 2024, III. Artificial intelligence and the economy: implications for central banks.

in all scenarios as well as policy approaches that work in one scenario but not in others. Therefore, the policy matrix is a tool to create or preserve policy optionality.

Each country should define the content of its own matrix. However, given the magnitude of the challenges arising in the most accelerated scenarios, the policy mix should not be only incremental. We encourage governments to consider all policy options, including those that may require increased budget allocations or changes in regulations and laws.

As mentioned in the Introduction, the tools of AI policy can be categorized into three main areas: government enablement, the use of AI in government, and laws and regulations. Each of these tools represents a strategic domain through which governments can shape the development, deployment, and impact of AI within their economies and societies.

First, policy plays a key role in enabling AI innovation and adoption. This encompasses investment in research and development (R&D), which is critical for fostering foundational discoveries and creating new technologies. Governments can fund academic institutions, private enterprises, and research centers to drive AI advancements. Education is another vital aspect of public investment, as it contributes to the workforce being equipped with the skills needed to thrive in an AI-driven economy. Investments in STEM education, reskilling programs, and AI-focused curricula are essential for building human capital. Infrastructure spending, such as the development of data centers, high-speed connectivity, and AI-specific testing facilities, provides the backbone for AI implementation. Additionally, governments can use financial incentives, including subsidies and tax breaks, to encourage private sector investment in AI. Governmental venture capital initiatives further catalyze innovation by funding startups and small enterprises focused on AI technologies, enabling risk-taking and entrepreneurial growth.

Second, as discussed in Chapter 3 of this Report, the use of AI in government itself is a tool for policy. AI can enhance the efficiency and effectiveness of public service delivery by automating processes, improving resource allocation, and providing data-

driven insights. In this Report we have focused on the use of AI to enhance tax design and collection, government expenditures and regulatory oversight and enforcement.

Third, there is a set of economic policies aimed at modulating the impact of AI on the economy and the financial markets, such as incentives, social transfers and other redistribution programs, and the expansion of the social safety net.

Lastly, laws and regulations are fundamental to the governance of AI, ensuring that its development and deployment align with societal values, ethical principles, and economic goals. Regulatory frameworks can establish guidelines for data protection, algorithmic transparency, and accountability in AI systems. These regulations are intended to mitigate risks such as bias, misuse, and unintended consequences. AI laws and regulations in finance encompass a wide array of areas, including:

- Data protection and privacy, including in the context cross-border data flows.
- Bias, fairness, and non-discrimination in areas such as consumer protection in credit and insurance markets credit or investment recommendations.
- Transparency of AI systems and explainability of AI-driven decisions, including those related to lending and insurance underwriting.
- Deception, abusive persuasion and manipulation of financial consumers and investors.
- As discussed in Chapter 4, a fundamental regulatory concern is the impact of advanced AI systems in financial markets stability, fragility and systemic risk.
- Regulations may establish standards for AI's use in anti-money laundering (AML) and combatting the financing of terrorism (CFT) frameworks.
- Cybersecurity.

The AI policy preparedness matrix is built by defining a desired use of the AI policy tool kit for each of the AI scenarios described in this Report (or other scenarios considered relevant). The table below illustrates the structure of the matrix. As already emphasized, each country could create—and regularly update—its own policy

preparedness matrix based on its culture, legal systems, national priorities, perception of risk and specific needs[52].

The AI Policy preparedness Matrix Structure			
	Scenarios		
	Conventional Wisdom Scenario	Intermediate AGI Scenario	Accelerated AGI Scenario
Policy tool			
Investment in AI enablement			
R&D investment			
Education			
Infrastructure			
Incentives (subsidies, tax breaks)			
Governmental venture capital			
Use of AI in government			
Tax design and collection			
Decision making in government			
Government workforce enablement			
Other economic policy tools			
Competition policies			
Employment incentives			
Redistribution/social safety redesign			
Laws, regulations and enforcement			
Privacy and data protection			
Bais, fairness and non-discrimination			
Explainability and transparency			
Accountability and liability allocation			
Persuasion and manipulation			
Financial stability and systemic risks			
AML, CTF			
Cybersecurity			

[52] For example, European Union member countries need to comply with the recently enacted European AI Act.

List of Abbreviations

5G	5th Generation
AGI	Artificial General Intelligence
AI	Artificial Intelligence
AML	Anti-Money Laundering
ASI	Artificial Superintelligence
AWS	Amazon Web Services
CPUs	Central Processing Units
DDoS	Distributed Denial-of-Service
EUV	Extreme Ultraviolet
FMCBG	Finance Ministers and Central Bank Governors
G7	Group of Seven
GAFAM	Google, Amazon, Facebook, Apple and Microsoft
GDP	Gross Domestic Product
GNoME	Graph Networks for Materials Exploration
GPT	Generative Pre-trained Transformer
GPUs	Graphics Processing Units
ILO	International Labour Organization
IMF	International Monetary Fund
IoT	Internet of Things
IT	Information Technology
KPIs	Key Performance Indicators
LLaMA/Llama	Large Language Model Meta AI
LMSYS	Large Model Systems Organization
OECD	Organization for Economic Co-operation and Development
PCs	Personal Computers
R&D	Research and Development
STEM	Science, Technology, Engineering and Mathematics
TWh	Terawatt-hour
TPUs	Tensor Processing Units
UNESCO	United Nations Educational, Scientific and Cultural Organization
US	United States
WTO	World Trade Organization

References

1. Acemoglu, Daron (2024), “The Simple Macroeconomics of Artificial Intelligence”, Economic Policy (2024), eiae042, https://doi.org/10.1093/epolic/eiae042;
2. Acemoglu, Daron (2022), "Harms of AI" In Justin B. Bullock and others (eds.), The Oxford Handbook of AI Governance, Oxford Handbooks (online edition, Oxford Academic), https://doi.org/10.1093/oxfordhb/9780197579329.013.65;
3. Acemoglu, Daron, and Simon Johnson (2023), “Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity”, John Murray Press;
4. Acemoglu, Daron, David Autor, Jonathan Hazell, and Pascual Restrepo (2022), "Artificial Intelligence and Jobs: Evidence from Online Vacancies", Journal of Labor Economics 40:S1, S293-S340, https://doi.org/10.1086/718327;
5. Acemoglu, Daron, and Pascual Restrepo (2018), "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment", American Economic Review 108 (6): 1488–1542, https://doi.org/10.1257/aer.20160696;
6. Aghion, Philippe and Simon Bunel (2024) “AI and Growth: where do we stand?”, Mimeo, https://www.frbsf.org/wp-content/uploads/AI-and-Growth-Aghion-Bunel.pdf;
7. Aghion, Philippe and Peter Howitt (1998), “Capital Accumulation and Innovation as Complementary Factors in Long-Run Growth”, Journal of Economic Growth 3, 111–130, https://doi.org/10.1023/A:1009769717601;
8. Aghion, Philippe and Peter Howitt (1994), "Growth and Unemployment", The Review of Economic Studies, 61 (3), 477–494, https://doi.org/10.2307/2297900;
9. Agrawal, Ajay, Joshua S. Gans and Avi Goldfarb (2023) “Similarities and Differences in the Adoption of General Purpose Technologies”, NBER Working Paper No. w30976, https://ssrn.com/abstract=4364724;

10. Aldasoro, Iñaki, Leonardo Gambacorta, Anton Korinek, Vatsala Shreeti and Merlin Stein (2024), "Intelligent financial system: how AI is transforming finance", BIS Working Papers 1194, Bank for International Settlements, https://www.bis.org/publ/work1194.htm;
11. Ananthaswamy, Anil (2024), “How Close is AI to human-level intelligence?”, Nature 636, 22-25, https://doi.org/10.1038/d41586-024-03905-1;
12. Aschenbrenner, Leopold (2024),“Situational Awareness”, Available at situational-awareness.ai.pdf;
13. Autor, David (2024) "Applying AI to Rebuild Middle Class Jobs", NBER Working Paper 32140, https://doi.org/10.3386/w32140;
14. Autor, David H. (2015), "Why Are There Still So Many Jobs? The History and Future of Workplace Automation", Journal of Economic Perspectives 29 (3): 3–30, https://doi.org/10.1257/jep.29.3.3;
15. Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson (2024), "Artificial Intelligence, Firm Growth, and Product Innovation", The Journal of Financial Economics 151: 1–26, https://doi.org/10.1016/j.jfineco.2023.103745;
16. Bank of England and FCA (2024), “Third survey of artificial intelligence and machine learning in UK financial services”, https://www.bankofengland.co.uk/report/2024/artificial-intelligence-in-uk-financial-services-2024;
17. Bailey, Martin, Erik Brynjolfsson and Anton Korinek (2023), “Machines of mind: The case for an AI-powered productivity boom”, Brookings, https://www.brookings.edu/articles/machines-of-mind-the-case-for-an-ai-powered-productivity-boom/;
18. Berg, Andrew, Edward F. Buffie, and Luis-Felipe Zanna (2018), "Should We Fear the Robot Revolution? (The Correct Answer Is Yes)", Journal of Monetary Economics 97: 117–148, https://doi.org/10.1016/j.jmoneco.2018.05.014;
19. Besiroglu, Tamay and Marius Hobbhahn (2022), “Trends in GPU Price-Performance,” https://epoch.ai/blog/trends-in-gpu-price-performance;

20. Bessen, James, Maarten Goos, Anna Salomons, and Wiljan van den Berge (2023), “What Happens to Workers at Firms that Automate?", The Review of Economics and Statistics. https://doi.org/10.1162/rest_a_01284;

21. BIS (2024), “Annual Economic Report”, III. Artificial intelligence and the economy: implications for central banks, https://www.bis.org/publ/arpdf/ar2024e3.htm;

22. Blanchard, Olivier, and Justin Wolfers (2000), "The Role of Shocks and Institutions in the Rise of European Unemployment: The Aggregate Evidence", Economic Journal 110: C1–C33, https://www.jstor.org/stable/2565720;

23. Bommasani, Rishi, et al. (2021), “On the opportunities and risks of foundation models”, https://doi.org/10.48550/arXiv.2108.07258;

24. Bommasani, Rishi, Kathleen Creel, Ananya Kumar, Dan Jurafsky, and Percy Liang (2022); “Picking on the Same Person: Does Algorithmic Monoculture lead to Outcome Homogenization?”, https://doi.org/10.48550/arXiv.2211.13972;

25. Bommasani, Rishi, Sanjeev Arora, Yejin Choi, Li Fei-Fei, Daniel E. Ho, Dan Jurafsky, Sanmi Koyejo, Hima Lakkaraju, Arvind Narayanan, Alondra Nelson, Emma Pierson, Joelle Pineau, Gaël Varoquaux, Suresh Venkatasubramanian, Ion Stoica, Percy Liang, Dawn Song (2024), “A Path for Science- and Evidence-based AI Policy”, https://understanding-ai-safety.org/;

26. Bonfiglioni, Alessandra, Rosario Crinò, Gino Gancia, and Ioannis Papadakis (2024), "Artificial Intelligence and Jobs: Evidence from US Commuting Zones" *Economic Policy*, eiae059, https://doi.org/10.1093/epolic/eiae059;

27. Bowen, Xu (2024) “What is Meant by AGI? On the Definition of Artificial General Intelligence”, Department of Computer and Information Sciences, Temple University, Philadelphia, USA, https://arxiv.org/html/2404.10731v1

28. Bresnahan, Timothy, Trajtenberg Manuel (1995). “General purpose technologies ‘Engines of growth’?”; Journal of Econometrics; Volume 65, Issue 1, January 1995, Pages 83-108, Elsevier, https://www.sciencedirect.com/science/article/pii/030440769401598T.

29. Brown, Tom, Mann Benjamin, Ryder Nick et al. (2020), Language Models are Few-Shot Learners. https://arxiv.org/abs/2005.14165;

30. Brussevich, Mariya, Era Dabla-Norris, and Salma Khalid (2019), "Is Technology Widening the Gender Gap? Automation and the Future of Female Employment.", IMF Working Paper, https://www.imf.org/en/Publications/WP/Issues/2019/05/06/Is-Technology-Widening-the-Gender-Gap-Automation-and-the-Future-of-Female-Employment-46684;

31. Brynjolfsson, Erik, and Andrew McAfee (2014), "The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies", WW Norton & Company;

32. Brynjolfsson, Erik, Avinash Collis at al. (2019). "GDP-B: Accounting For The Value Of New And Free Goods In The Digital Economy", National Bureau of Economic Research Working Papers No. 25695, w25695.pdf:

33. Brynjolfsson Erik, Li Danielle and Raymond Lindsey (2023), "Generative AI at Work", National Bureau of Economic Research Working Papers No. 31161, https://doi.org/10.3386/w31161;

34. Business Standard (2024) "OpenAI pitched White House on need of massive 5GW data center buildout", https://www.business-standard.com/world-news/openai-pitched-white-house-on-need-of-massive-5gw-data-center-buildout-124092500102_1.html;

35. Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello (2020) "Artificial Intelligence, Algorithmic Pricing, and Collusion", American Economic Review, 110 (10): 3267–97, https://doi.org/10.1257/aer.20190623;

36. Carlson, Mark A. (2007) "A Brief History of the 1987 Stock Market Crash with a Discussion of the Federal Reserve Response", FEDS Working Paper No. 2007-13, https://www.federalreserve.gov/pubs/feds/2007/200713/200713pap.pdf;

37. Cazzaniga, Mauro, Florence Jaumotte, Longji Li, Giovanni Melina, Augustus J Panton, Carlo Pizzinelli, Emma J Rockall, and Marina Mendes Tavares (2024),

"Gen-AI: Artificial Intelligence and the Future of Work", IMF Staff Discussion Notes, https://doi.org/10.5089/9798400262548.006;

38. Coffey, Lauren, (2024) "Data Science Major Takes Off", Inside Higher Ed, https://www.insidehighered.com/news/tech-innovation/teaching-learning/2024/01/25/data-science-major-takes-across-college-campuses;
39. Copestake, Alex, Ashley Pople, Katherine Stapleton, and Max Marczinek (2023), "AI and Services-Led Growth: Evidence from Indian Job Adverts", Working Paper, https://copestake.info/workingpaper/akai/AKAI.pdf;
40. Cowen, Tyler (2011), "The Great Stagnation: How America Ate All the Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better", Dutton Adult;
41. Dell'Acqua, Fabrizio, Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani (2023), "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality", Working Paper, https://www.hbs.edu/ris/Publication%20Files/24-013_d9b45b68-9e74-42d6-a1c6-c72fb70c7282.pdf;
42. Dgtl Infra (2024), Data Center Water Usage: A Comprehensive Guide, https://dgtlinfra.com/data-center-water-usage/;
43. European Central Bank (2023), "Take-aways from the horizontal assessment of the survey on digital transformation and the use of fintech". https://www.bankingsupervision.europa.eu/ecb/pub/pdf/Takeaways_horizontal_assessment~de65261ad0.en.pdf;
44. Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Roc. (2024) "GPTs Are GPTs: Labor Market Impact Potential of LLMs", Science 384: 1306–1308, https://doi.org/10.1126/science.adj0998;
45. Faiz, Ahmad, et al. (2023), "Llmcarbon: Modeling the end-to-end carbon footprint of large language models." https://doi.org/10.48550/arXiv.2309.14393;

46. Fernandez, Joaquin, Knud Lasse Lueth, and Philipp Wegner (2023) “Generative AI Market Report 2023–2030”, IoT Analytics, 14 December 2023, https://iot-analytics.com/product/generative-ai-market-report-2023-2030/;

47. Filippucci, Francesco et al. (2024), “The impact of Artificial Intelligence on productivity, distribution and growth”, OECD, https://www.oecd.org/en/publications/the-impact-of-artificial-intelligence-on-productivity-distribution-and-growth_8d900037-en.html

48. Ford, Martin (2015), “The Rise of the Robots”, Basic Books, New York;

49. Financial Stability Board (2024). “The Financial Stability Implications of Artificial Intelligence”, https://www.fsb.org/uploads/P14112024.pdf

50. Frey, Carl Benedikt, and Michael A. Osborne (2017), "The Future of Employment: How Susceptible Are Jobs to Computerisation?", Technological Forecasting and Social Change 114: 254–280, https://doi.org/10.1016/j.techfore.2016.08.019;

51. Gensler, Gary and Lily Bailey (2020), “Deep Learning and Financial Stability”, SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3723132.

52. Gmyrek Pawel, Janine Berg and David Bescond (2023), “Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality”, (August 21, 2023), ILO Working Paper 96 (Geneva, ILO), https://doi.org/10.54394/FHEM8239;

53. Google (2024), “Environmental Report”, https://sustainability.google/reports/google-2024-environmental-report/;

54. Gordon, Robert J. (2015), “Secular stagnation: A supply-side view”, American Economic Review, 105(5), 54-59, https://doi.org/10.1257/aer.p20151102;

55. Gordon, Robert J. (2012), “Is U.S. Economic Growth Over? Faltering Innovation Confronts the Six Headwinds”, NBER Working Papers, No. 18315, https://doi.org/10.3386/w18315;

56. Government Technology (2024), “What Government Can Learn from the Private Sector About AI”, https://www.govtech.com/biz/data/what-government-can-learn-from-the-private-sector-about-ai;

57. Grout, Paul A. (2021), “AI, ML, and competition dynamics in financial markets”, Oxford Review of Economic Policy, https://academic.oup.com/oxrep/article/37/3/618/6374677?login=false

58. The Guardian (2024), “Data center emissions probably 662% higher than big tech claims. Can it keep up the ruse?”, https://www.theguardian.com/technology/2024/sep/15/data-center-gas-emissions-tech

59. Harari, Yuval Noah (2016), “Homo Deus: A Brief History of Tomorrow”. Random House;

60. Hopkins, Aspen, Andrew Ilyas, Aleksander Madry, Isabella Struckman, and Luis Videgaray (2023), “Who will provide AI to the world?”, Thoughts on AI Policy; A newsletter from MIT & Harvard researchers on regulating AI technologies, https://aipolicy.substack.com/p/supply-chains-3;

61. Hui, Xiang, Oren Reshef, and Luofeng Zhou (2023), "The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market"; CESifo Working Paper Series 10601, CESifo, https://www.cesifo.org/DocDL/cesifo1_wp10601.pdf;

62. Huang, Yueling (2024), "The Labor Market Impact of Artificial Intelligence: Evidence from US Regions", IMF Working Papers, 199 (2024), https://doi.org/10.5089/9798400288555.001;

63. International Energy Agency (2024), Report “Electricity 2024”: https://www.iea.org/reports/electricity-2024.

64. Jaipuria, Tanay, 2024, “The plummeting cost of intelligence”, Wing, https://www.wing.vc/content/plummeting-cost-ai-intelligence

65. Jones, Benjamin F. (2009), "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?", The Review of Economic Studies, Vol. 76, pp. 283-317, https://doi.org/10.1111/j.1467-937X.2008.00531.x;

66. Kaplan, Jared, Sam McCandlish, Tom Henighan, et al. (2020); “Scaling Laws for Neural Language Models”, https://arxiv.org/abs/2001.08361;

67. Kinder, Molly, Xavier de Souza Briggs, Mark Muro, and Sifan Liu (2024); "Generative AI, Generative AI, the American worker, and the future of work", Brookings, https://www.brookings.edu/articles/generative-ai-the-american-worker-and-the-future-of-work/.
68. Kirkpatrick, Keith, (2023), "The Carbon Footprint of Artificial Intelligence", Commun. ACM 66, 8 (August 2023), 17–19. https://doi.org/10.1145/3603746;
69. Korinek, Anton (2023), "Scenario Planning for an A(G)I Future", IMF Finance & Development Magazine 60(4), pp. 30-33, https://www.imf.org/en/Publications/fandd/issues/2023/12/Scenario-Planning-for-an-AGI-future-Anton-korinek;
70. Korinek, Anton, and Donghyun Suh (2024), "Scenarios for the Transition to AGI". SSRN Electronic Journal, https://doi.org/10.48550/arXiv.2403.12107;
71. Korinek, Anton and Jai Vipra (2024), "Concentrating Intelligence: Scaling Laws and Market Structure in Generative AI', National Bureau of Economic Research Working Papers No. http://www.nber.org/papers/w33139.
72. Korinek, Anton, and Joseph E. Stiglitz (2019), "Artificial Intelligence and Its Implications for Income Distribution and Unemployment", The Economics of Artificial Intelligence: An Agenda, Ajay Agrawal, Joshua Gans, and Avi Goldfarb (eds.), University of Chicago Press, pp. 349–90.
73. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton (2017), "ImageNet classification with deep convolutional neural networks". Commun. ACM 60, 6 (June 2017), 84–90. https://doi.org/10.1145/3065386
74. Kudoh, Noritaka, and Hiroaki Miyamoto (2024), "Robots and Unemployment." Mimeo, https://dx.doi.org/10.2139/ssrn.4854753;
75. Leitner, Georg, Jaspal Singh, Anton van der Kraaij, Balázs Zsámboki, (2024) "The rise of artificial intelligence: benefits and risks for financial stability," Financial Stability Review, European Central Bank, vol. 1. https://www.ecb.europa.eu/press/financial-stability-publications/fsr/special/html/ecb.fsrart202405_02~58c3ce5246.en.html

76. Lohn, Andrew and Micah Musser (2022), “AI and Compute: w Much Longer Can Computing Power Drive Artificial Intelligence Progress?”, CSET Issue Brief, Center for Security and Emerging Technology, https://doi.org/10.51593/2021CA009;

77. Luccioni, Alexandra Sasha and Alex Hernandez-Garcia (2023). "Counting carbon: A survey of factors influencing the emissions of machine learning." https://doi.org/10.48550/arXiv.2302.08476.

78. Luccioni Alexandra Sasha, Yacine Jernite and Emma Strubell (2024), "Power hungry processing: Watts driving the cost of AI deployment?", The 2024 ACM Conference on Fairness, Accountability, and Transparency, https://doi.org/10.1145/3630106.3658542,

79. Luna, Angela (2024) “Open-Source AI: The debate that could redefine AI innovation”, American Action Forum, https://www.americanactionforum.org/insight/open-source-ai-the-debate-that-could-redefine-ai-innovation/.

80. Jones, Nicola (2024), “The AI revolution is running out of data. What can researchers do?”, Nature, 636, 290-292, https://doi.org/10.1038/d41586-024-03990-2.

81. Miyamoto, Hiroaki, and Yuya Takahashi (2011), "Productivity Growth, On-the-job Search, and Unemployment", Journal of Monetary Economics 58: 666–680, https://doi.org/10.1016/j.jmoneco.2011.11.007;

82. Mortensen, Dale T., and Christopher A. Pissarides (1998), "Technological Progress, Job Creation, and Job Destruction." Review of Economic Dynamics 1: 733–753, https://doi.org/10.1006/redy.1998.0030;

83. Noy, Shakked and Whitney Zhang (2023), “Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence”, Science, 381,187-192(2023), https://doi.org/10.1126/science.adh2586;

84. Organization for Economic Cooperation and Development (2019), Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

85. Organization for Economic Cooperation and Development (2021), “OECD Business and Finance Outlook 2021: AI in Business and Finance”, OECD Publishing, Paris, https://doi.org/10.1787/ba682899-en;

86. Organization for Economic Cooperation and Development and World Trade Organization (2024), “A Song of Data Flows and Trust – A Quantitative Analysis of Cross-border Data Flow Regulation”. Paris: OECD.

87. Patterson, David, et al. (2022), "The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink," Computer, 55 (7), 18-28 https://doi.org/10.1109/MC.2022.3148714;

88. Patterson, David, Joseph Gonzalez, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier and Jeff Dean (2021), “Carbon Emissions and Large Neural Network Training”, https://arxiv.org/abs/2104.10350;

89. Peng Sida, Kalliamvakou Eirini, Cihon Peter and Demirer Mert (2023), “The impact of ai on developer productivity: Evidence from github copilot”, arXiv preprint arXiv:2302.06590, https://arxiv.org/abs/2302.06590;

90. Pilz Kostantin, Heim Lennart, Brown Nicholas (2024), “Increased Efficiency and the Diffusion of AI Capabilities”, https://arxiv.org/pdf/2311.15377;

91. Pissarides, Christopher A., and Giovanna Vallanti (2007), “The Impact of TFP Growth on Steady-State Unemployment.” International Economic Review 48, no. 2: 607–40. http://www.jstor.org/stable/4541982.;

92. Pizzinelli, Carlo, Augustus Panton, Marina Mendes Tavares, Mauro Cazzaniga and Longii Li (2023), “Labor Market Exposure to AI: Cross-country Differences and Distributional Implication”, IMF Working Paper 23/216, https://www.imf.org/en/Publications/WP/Issues/2023/10/04/Labor-Market-Exposure-to-AI-Cross-country-Differences-and-Distributional-Implications-539656;

93. Postel-Vinay, Fabien (2002), “The Dynamics of Technological Unemployment”, International Economic Review, vol. 43, no. 3, 2002, pp. 737–60. JSTOR, http://www.jstor.org/stable/827040;

94. RAND Corporation (2024), Strategic competition in the age of AI, https://www.rand.org/pubs/research_reports/RRA3295-1.html

95. Rendon, Andrew S. (2024), “Generation Z’s perceptions of public sector culture moderate their aspirations towards public service”, https://scholarsjunction.msstate.edu/cgi/viewcontent.cgi?article=7175&context=td

96. SAS (2022), “Data & AI in the UK government: Overcoming barriers and realising potential”, https://www.sas.com/en_gb/whitepapers/data-and-ai-in-the-uk-government-overcoming-barriers-and-realising-potential-114054.html;

97. Shabsigh, Ghiath and El Bachir Boukherouaa (2023), “Generative Artificial Intelligence in Finance: Risk Considerations”, Fintech Notes, 2023(006), https://doi.org/10.5089/9798400251092.063;

98. Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson (2024); “The curse of recursion: training on generated data makes models forget”, https://arxiv.org/pdf/2305.17493;

99. Stanford University (2024), AI Index Report 2024, https://aiindex.stanford.edu/report/;

100. Strubell, Emma, Ananya Ganesh, and Andrew McCallum (2019), “Energy and Policy Considerations for Deep Learning in NLP”, https://arxiv.org/abs/1906.02243;

101. Sun, R. and Trefler, D. (2023) "The Impact of AI and Cross-Border Data Regulation on International Trade in Digital Services: A Large Language Model". NBER Working Paper No. 31925, Cambridge (MA): National Bureau of Economic Research (NBER), https://www.nber.org/papers/w31925;

102. Tiederich Lee (2024), “The AI data scraping challenge: how can we proceed responsibly?”, https://oecd.ai/en/wonk/data-scraping-responsibly;

103. Tony Blair Institute for Climate Change (2024), “Greening AI: A Policy Agenda for the Artificial Intelligence and Energy Revolutions”, https://institute.global/insights/climate-and-energy/greening-ai-a-policy-agenda-for-the-artificial-intelligence-and-energy-revolutions.

104. Thompson, Alan (2024), “What’s in GPT-5? A Comprehensive Analysis of Datasets Likely Used to Train GPT-5.”, LifeArchitect.AI, https://lifearchitect.ai/whats-in-gpt-5/;

105. UNESCO (2021), Recommendation on the Ethics of Artificial Intelligence, https://unesdoc.unesco.org/ark:/48223/pf0000381137;

106. UNESCO, OECD, IDB (2022). The Effects of AI on the Working Lives of Women, https://unesdoc.unesco.org/ark:/48223/pf0000380861;

107. Varoquaux, Gaël, Alexandra Sasha Luccioni, and Meredith Whittaker (2024) "Hype, Sustainability, and the Price of the Bigger-is-Better Paradigm in AI". https://doi.org/10.48550/arXiv.2409.14160.

108. World Economic Forum (2028), “2018 Global Gender Gap Report”. World Economic Forum

109. World Trade Organization (2024), Report “Trading with intelligence - How AI shapes and is shaped by international trade”. WTO | Trading with intelligence: How AI shapes and is shaped by international trade

110. Wu, Carole-Jean, et al. (2022), "Sustainable AI: Environmental implications, challenges and opportunities", Proceedings of Machine Learning and Systems 4, 795-813, https://proceedings.mlsys.org/paper_files/paper/2022/file/462211f67c7d858f663355eff93b745e-Paper.pdf;

111. Xu Tammy (2022), “We could run out of data to train AI language programs”, https://www.technologyreview.com/2022/11/24/1063684/we-could-run-out-of-data-to-train-ai-language-programs/

112. Yang, Yuzhe, Haoran Zhang, Judy W. Gichoya et al. (2024), “The limits of fair medical imaging AI in real-world generalization”, Nat Med 30, 2838–2848, https://doi.org/10.1038/s41591-024-03113-4;

113. Zeichick, Alan, (2023) “What Is a Sovereign Cloud? Why Is It Important?”, Oracle Cloud Infrastructure, Oracle, https://www.oracle.com/cloud/sovereign-cloud/what-is-sovereign-cloud/;

Annexes

A. Hiroshima Framework

G7 Leaders' Statement on the Hiroshima AI Process
October 30, 2023

We, the Leaders of the Group of Seven (G7), stress the innovative opportunities and transformative potential of advanced Artificial Intelligence (AI) systems, in particular, foundation models and generative AI. We also recognize the need to manage risks and to protect individuals, society, and our shared principles including the rule of law and democratic values, keeping humankind at the center. We affirm that meeting those challenges requires shaping an inclusive governance for artificial intelligence. Building on the progress made by relevant ministers on the Hiroshima AI Process, including the G7 Digital & Tech Ministers' Statement issued on September 7, 2023, we welcome the Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems and the Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems (Attached). In order to ensure both documents remain fit for purpose and responsive to this rapidly evolving technology, they will be reviewed and updated as necessary, including through ongoing inclusive multistakeholder consultations. We call on organizations developing advanced AI systems to commit to the application of the International Code of Conduct.

We instruct relevant ministers to accelerate the process toward developing the Hiroshima AI Process Comprehensive Policy Framework, which includes project based cooperation, by the end of this year, in cooperation with the Global Partnership for Artificial Intelligence (GPAI) and the Organisation for Economic Co-operation and Development (OECD), and to conduct multistakeholder outreach and consultation, including with governments, academia, civil society, and the private sector, not only those in the G7 but also in the economies beyond, including developing and emerging economies. We also ask relevant ministers to develop a work plan by the end of the year for further advancing the Hiroshima AI Process.

We believe that our joint efforts through the Hiroshima AI Process will foster an open and enabling environment where safe, secure, and trustworthy AI systems are designed, developed, deployed, and used to maximize the benefits of the technology while mitigating its risks, for the common good worldwide, including in developing and emerging economies with a view to closing digital divides and achieving digital inclusion. We also look forward to the UK's AI Safety Summit on November 1 and 2.

Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI system

The International Guiding Principles for Organizations Developing Advanced AI Systems aims to promote safe, secure, and trustworthy AI worldwide and will provide guidance for organizations developing and using the most advanced AI systems, including the most advanced foundation models and generative AI systems (henceforth "advanced AI systems"). Organizations may include, among others, entities from academia, civil society, the private sector, and the public sector.

This non-exhaustive list of guiding principles is discussed and elaborated as a living document to build on the existing OECD AI Principles in response to recent developments in advanced AI systems and are meant to help seize the benefits and address the risks and challenges brought by these technologies. These principles should apply to all AI actors, when and as applicable to cover the design, development, deployment and use of advanced AI systems.

We look forward to developing these principles further as part of the comprehensive policy framework, with input from other nations and wider stakeholders in academia, business and civil society.

We also reiterate our commitment to elaborate an international code of conduct for organizations developing advanced AI systems based on the guiding principles below.

Different jurisdictions may take their own unique approaches to implementing these guiding principles in different ways.

We call on organizations in consultation with other relevant stakeholders to follow these actions, in line with a risk-based approach, while governments develop more enduring and/or detailed governance and regulatory approaches. We also commit to develop proposals, in consultation with the OECD, GPAI and other stakeholders, to introduce monitoring tools and mechanisms to help organizations stay accountable for the implementation of these actions. We encourage organizations to support the development of effective monitoring mechanisms, which we may explore to develop, by contributing best practices.

While harnessing the opportunities of innovation, organizations should respect the rule of law, human rights, due process, diversity, fairness and non-discrimination, democracy, and human-centricity, in the design, development and deployment of advanced AI systems.

Organizations should not develop or deploy advanced AI systems in a way that undermine democratic values, are particularly harmful to individuals or communities, facilitate terrorism, enable criminal misuse, or pose substantial risks to safety, security, and human rights, and are thus not acceptable.

States must abide by their obligations under international human rights law to promote that human rights are fully respected and protected, while private sector activities should be in line with international frameworks such as the United Nations Guiding Principles on Business and Human Rights and the OECD Guidelines for Multinational Enterprises.

Specifically, we call on organizations to abide by the following principles, commensurate to the risks:

1. Take appropriate measures throughout the development of advanced AI systems, including prior to and throughout their deployment and placement on the market, to identify, evaluate, and mitigate risks across the AI lifecycle.

This includes employing diverse internal and independent external testing measures, through a combination of methods such as red-teaming, and implementing appropriate mitigation to address identified risks and vulnerabilities. Testing and mitigation measures should for example, seek to ensure the trustworthiness, safety and security of systems throughout their entire lifecycle so that they do not pose unreasonable risks. In support of such testing, developers should seek to enable traceability, in relation to datasets, processes, and decisions made during system development.

2. Patterns of misuse, after deployment including placement on the market.

Organizations should use, as and when appropriate commensurate to the level of risk, AI systems as intended and monitor for vulnerabilities, incidents, emerging risks and misuse after deployment, and take appropriate action to address these. Organizations are encouraged to consider, for example, facilitating third-party and user discovery and reporting of issues and vulnerabilities after deployment. Organizations are further encouraged to maintain appropriate documentation of reported incidents and to mitigate the identified risks and vulnerabilities, in collaboration with other stakeholders. Mechanisms to report vulnerabilities, where appropriate, should be accessible to a diverse set of stakeholders.

3. Publicly report advanced AI systems' capabilities, limitations and domains of appropriate and inappropriate use, to support ensuring sufficient transparency, thereby contributing to increase accountability.

This should include publishing transparency reports containing meaningful information for all new significant releases of advanced AI systems.

Organizations should make the information in the transparency reports sufficiently clear and understandable to enable deployers and users as appropriate and relevant to interpret the model/system's output and to enable users to use it appropriately, and that transparency reporting should be supported and informed by robust documentation processes.

4. Work towards responsible information sharing and reporting of incidents among organizations developing advanced AI systems including with industry, governments, civil society, and academia.

This includes responsibly sharing information, as appropriate, including, but not limited to evaluation reports, information on security and safety risks, dangerous, intended or unintended capabilities, and attempts AI actors to circumvent safeguards across the AI lifecycle.

5. Develop, implement and disclose AI governance and risk management policies, grounded in a risk-based approach – including privacy policies, and mitigation measures, in particular for organizations developing advanced AI systems.

This includes disclosing where appropriate privacy policies, including for personal data, user prompts and advanced AI system outputs. Organizations are expected to establish and disclose their AI governance policies and organizational mechanisms to implement these policies in accordance with a risk based approach. This should include accountability and governance processes to evaluate and mitigate risks, where feasible throughout the AI lifecycle.

6. Invest in and implement robust security controls, including physical security, cybersecurity and insider threat safeguards across the AI lifecycle.

These may include securing model weights and algorithms, servers, and datasets, such as through operational security measures for information security and appropriate cyber/physical access controls.

7. Develop and deploy reliable content authentication and provenance mechanisms, where technically feasible, such as watermarking or other techniques to enable users to identify AI-generated content.

This includes, where appropriate and technically feasible, content authentication such provenance mechanisms for content created with an organization's advanced AI system. The

provenance data should include an identifier of the service or model that created the content, but need not include user information. Organizations should also endeavor to develop tools or APIs to allow users to determine if particular content was created with their advanced AI system such as via watermarks.

Organizations are further encouraged to implement other mechanisms such as labeling or disclaimers to enable users, where possible and appropriate, to know when they are interacting with an AI system.

8. Prioritize research to mitigate societal, safety and security risks and prioritize investment in effective mitigation measures.

This includes conducting, collaborating on and investing in research that supports the advancement of AI safety, security and trust, and addressing key risks, as well as investing in developing appropriate mitigation tools.

9. Prioritize the development of advanced AI systems to address the world's greatest challenges, notably but not limited to the climate crisis, global health and education.

These efforts are undertaken in support of progress on the United Nations Sustainable Development Goals, and to encourage AI development for global benefit.

Organizations should prioritize responsible stewardship of trustworthy and human-centric AI and also support digital literacy initiatives.

10. Advance the development of and, where appropriate, adoption of international technical standards

This includes contributing to the development and, where appropriate, use of international technical standards and best practices, including for watermarking, and working with Standards Development Organizations (SDOs).

11. Implement appropriate data input measures and protections for personal data and intellectual property

Organizations are encouraged to take appropriate measures to manage data quality, including training data and data collection, to mitigate against harmful biases.

Appropriate transparency of training datasets should also be supported and organizations should comply with applicable legal frameworks.

Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems

On the basis of the International Guiding Principles for Organizations Developing Advanced AI systems, the International Code of Conduct for Organizations Developing Advanced AI Systems aims to promote safe, secure, and trustworthy AI worldwide and will provide voluntary guidance for actions by organizations developing the most advanced AI systems, including the most advanced foundation models and generative AI systems (henceforth "advanced AI systems").

Organizations should follow these actions in line with a risk-based approach.

Organizations that may endorse this Code of Conduct may include, among others, entities from academia, civil society, the private sector, and/or the public sector.

This non-exhaustive list of actions is discussed and elaborated as a living document to build on the existing OECD AI Principles in response to the recent developments in advanced AI systems and is meant to help seize the benefits and address the risks and challenges brought by these technologies. Organizations should apply these actions to all stages of the lifecycle to cover, when and as applicable, the design, development, deployment and use of advanced AI systems.

This document will be reviewed and updated as necessary, including through ongoing inclusive multistakeholder consultations, in order to ensure it remains fit for purpose and responsive to this rapidly evolving technology.

Different jurisdictions may take their own unique approaches to implementing these actions in different ways.

We call on organizations in consultation with other relevant stakeholders to follow these actions, in line with a risk-based approach, while governments develop more enduring and/or detailed governance and regulatory approaches. We also commit to develop proposals, in consultation with the OECD, GPAI and other stakeholders, to introduce monitoring tools and mechanisms to help organizations stay accountable for the implementation of these actions. We encourage organizations to support the development of effective monitoring mechanisms, which we may explore to develop, by contributing best practices.

In addition, we encourage organizations to set up internal AI governance structures and policies, including self-assessment mechanisms, to facilitate a responsible and accountable approach to implementation of these actions and in AI development.

While harnessing the opportunities of innovation, organizations should respect the rule of law, human rights, due process, diversity, fairness and non-discrimination, democracy, and human-centricity, in the design, development and deployment of advanced AI systems.

Organizations should not develop or deploy advanced AI systems in ways that undermine democratic values, are particularly harmful to individuals or communities, facilitate terrorism, promote criminal misuse, or pose substantial risks to safety, security and human rights, and are thus not acceptable.

States must abide by their obligations under international human rights law to ensure that human rights are fully respected and protected, while private sector activities should be in line with international frameworks such as the United Nations Guiding Principles on Business and Human Rights and the OECD Guidelines for Multinational Enterprises.

Specifically, we call on organizations to abide by the following actions, in a manner that is commensurate to the risks:

1 Take appropriate measures throughout the development of advanced AI systems, including prior to and throughout their deployment and placement on the market, to identify, evaluate, and mitigate risks across the AI lifecycle.

This includes employing diverse internal and independent external testing measures, through a combination of methods for evaluations, such as red-teaming, and implementing appropriate mitigation to address identified risks and vulnerabilities. Testing and mitigation measures, should, for example, seek to ensure the trustworthiness, safety and security of systems throughout their entire lifecycle so that they do not pose unreasonable risks. In support of such testing, developers should seek to enable traceability, in relation to datasets, processes, and decisions made during system development. These measures should be documented and supported by regularly updated technical documentation.

This testing should take place in secure environments and be performed at several checkpoints throughout the AI lifecycle in particular before deployment and placement on the market to identify risks and vulnerabilities, and to inform action to address the identified AI risks to security, safety and societal and other risks, whether accidental or intentional. In designing and

implementing testing measures, organizations commit to devote attention to the following risks as appropriate:

> Chemical, biological, radiological, and nuclear risks, such as the ways in which advanced AI systems can lower barriers to entry, including for non-state actors, for weapons development, design acquisition, or use.

> Offensive cyber capabilities, such as the ways in which systems can enable vulnerability discovery, exploitation, or operational use, bearing in mind that such capabilities could also have useful defensive applications and might be appropriate to include in a system.

> Risks to health and/or Safety, including the effects of system interaction and tool use, including for example the capacity to control physical systems and interfere with critical infrastructure.

> Risks from models of making copies of themselves or “self-replicating” or training other models.

> Societal risks, as well as risks to individuals and communities such as the ways in which advanced AI systems or models can give rise to harmful bias and discrimination or lead to violation of applicable legal frameworks, including on privacy and data protection.

> Threats to democratic values and human rights, including the facilitation of disinformation or harming privacy.

> Risk that a particular event could lead to a chain reaction with considerable negative effects that could affect up to an entire city, an entire domain activity or an entire community.

Organizations commit to work in collaboration with relevant actors across sectors, to assess and adopt mitigation measures to address these risks, in particular systemic risks.

Organizations making these commitments should also endeavor to advance research and investment on the security, safety, bias and disinformation, fairness, explainability and interpretability, and transparency of advanced AI systems and on increasing robustness and trustworthiness of advanced AI systems against misuse.

2 Identify and mitigate vulnerabilities, and, where appropriate, incidents and patterns of misuse, after deployment including placement on the market.

Organizations should use, as and when appropriate commensurate to the level of risk, AI systems as intended and monitor for vulnerabilities, incidents, emerging risks and misuse after deployment, and take appropriate action to address these. Organizations are encouraged to

consider, for example, facilitating third-party and user discovery and reporting of issues and vulnerabilities after deployment such as through bounty systems, contests, or prizes to incentivize the responsible disclosure of weaknesses. Organizations are further encouraged to maintain appropriate documentation of reported incidents and to mitigate the identified risks and vulnerabilities, in collaboration with other stakeholders. Mechanisms to report vulnerabilities, where appropriate, should be accessible to a diverse set of stakeholders.

3 Publicly report advanced AI systems' capabilities, limitations and domains of appropriate and inappropriate use, to support ensuring sufficient transparency, thereby contributing to increase accountability.

This should include publishing transparency reports containing meaningful information for all new significant releases of advanced AI systems.

These reports, instruction for use and relevant technical documentation, as appropriate as, should be kept up-to-date and should include, for example;

> Details of the evaluations conducted for potential safety, security, and societal risks, as well as risks to human rights,

> Capacities of a model/system and significant limitations in performance that have implications for the domains of appropriate use,

> Discussion and assessment of the model's or system's effects and risks to safety and society such as harmful bias, discrimination, threats to protection of privacy or personal data, and effects on fairness, and

> The results of red-teaming conducted to evaluate the model's/system's fitness for moving beyond the development stage.

Organizations should make the information in the transparency reports sufficiently clear and understandable to enable deployers and users as appropriate and relevant to interpret the model/system's output and to enable users to use it appropriately; and that transparency reporting should be supported and informed by robust documentation processes such as technical documentation and instructions for use.

4 Work towards responsible information sharing and reporting of incidents among organizations developing advanced AI systems including with industry, governments, civil society, and academia

This includes responsibly sharing information, as appropriate, including, but not limited to evaluation reports, information on security and safety risks, dangerous intended or unintended capabilities, and attempts by AI actors to circumvent safeguards across the AI lifecycle.

Organizations should establish or join mechanisms to develop, advance, and adopt, where appropriate, shared standards, tools, mechanisms, and best practices for ensuring the safety, security, and trustworthiness of advanced AI systems.

This should also include ensuring appropriate and relevant documentation and transparency across the AI lifecycle in particular for advanced AI systems that cause significant risks to safety and society.

Organizations should collaborate with other organizations across the AI lifecycle to share and report relevant information to the public with a view to advancing safety, security and trustworthiness of advanced AI systems. Organizations should also collaborate and share the aforementioned information with relevant public authorities, as appropriate.

Such reporting should safeguard intellectual property rights.

5 Develop, implement and disclose AI governance and risk management policies, grounded in a risk-based approach – including privacy policies, and mitigation measures.

Organizations should put in place appropriate organizational mechanisms to develop, disclose and implement risk management and governance policies, including for example accountability and governance processes to identify, assess, prevent, and address risks, where feasible throughout the AI lifecycle.

This includes disclosing where appropriate privacy policies, including for personal data, user prompts and advanced AI system outputs. Organizations are expected to establish and disclose their AI governance policies and organizational mechanisms to implement these policies in accordance with a risk based approach. This should include accountability and governance processes to evaluate and mitigate risks, where feasible throughout the AI lifecycle.

The risk management policies should be developed in accordance with a risk based approach and apply a risk management framework across the AI lifecycle as appropriate and relevant, to address the range of risks associated with AI systems, and policies should also be regularly updated.

Organizations should establish policies, procedures, and training to ensure that staff are familiar with their duties and the organization's risk management practices.

6 Invest in and implement robust security controls, including physical security, cybersecurity and insider threat safeguards across the AI lifecycle.

These may include securing model weights and, algorithms, servers, and datasets, such as through operational security measures for information security and appropriate cyber/physical access controls.

This also includes performing an assessment of cybersecurity risks and implementing cybersecurity policies and adequate technical and institutional solutions to ensure that the cybersecurity of advanced AI systems is appropriate to the relevant circumstances and the risks involved. Organizations should also have in place measures to require storing and working with the model weights of advanced AI systems in an appropriately secure environment with limited access to reduce both the risk of unsanctioned release and the risk of unauthorized access. This includes a commitment to have in place a vulnerability management process and to regularly review security measures to ensure they are maintained to a high standard and remain suitable to address risks.

This further includes establishing a robust insider threat detection program consistent with protections provided for their most valuable intellectual property and trade secrets, for example, by limiting access to proprietary and unreleased model weights.

7 Develop and deploy reliable content authentication and provenance mechanisms, where technically feasible, such as watermarking or other techniques to enable users to identify AI-generated content

This includes, where appropriate and technically feasible, content authentication and provenance mechanisms for content created with an organization's advanced AI system. The provenance data should include an identifier of the service or model that created the content, but need not include user information. Organizations should also endeavor to develop tools or APIs to allow users to determine if particular content was created with their advanced AI system, such as via watermarks. Organizations should collaborate and invest in research, as appropriate, to advance the state of the field.

Organizations are further encouraged to implement other mechanisms such as labeling or disclaimers to enable users, where possible and appropriate, to know when they are interacting with an AI system.

8 Prioritize research to mitigate societal, safety and security risks and prioritize investment in effective mitigation measures.

This includes conducting, collaborating on and investing in research that supports the advancement of AI safety, security, and trust, and addressing key risks, as well as investing in developing appropriate mitigation tools.

Organizations commit to conducting, collaborating on and investing in research that supports the advancement of AI safety, security, trustworthiness and addressing key risks, such as prioritizing research on upholding democratic values, respecting human rights, protecting children and vulnerable groups, safeguarding intellectual property rights and privacy, and avoiding harmful bias, mis- and disinformation, and information manipulation. Organizations also commit to invest in developing appropriate mitigation tools, and work to proactively manage the risks of advanced AI systems, including environmental and climate impacts, so that their benefits can be realized.

Organizations are encouraged to share research and best practices on risk mitigation.

9 Prioritize the development of advanced AI systems to address the world's greatest challenges, notably but not limited to the climate crisis, global health and education

These efforts are undertaken in support of progress on the United Nations Sustainable Development Goals, and to encourage AI development for global benefit.

Organizations should prioritize responsible stewardship of trustworthy and human-centric AI and also support digital literacy initiatives that promote the education and training of the public, including students and workers, to enable them to benefit from the use of advanced AI systems, and to help individuals and communities better understand the nature, capabilities, limitations, and impact of these technologies. Organizations should work with civil society and community groups to identify priority challenges and develop innovative solutions to address the world's greatest challenges.

10 Advance the development of and, where appropriate, adoption of international technical standards

Organizations are encouraged to contribute to the development and, where appropriate, use of international technical standards and best practices, including for watermarking, and working with Standards Development Organizations (SDOs), also when developing organizations' testing methodologies, content authentication and provenance mechanisms,

cybersecurity policies, public reporting, and other measures. In particular, organizations also are encouraged to work to develop interoperable international technical standards and frameworks to help users distinguish content generated by AI from non-AI generated content.

11 Implement appropriate data input measures and protections for personal data and intellectual property

Organizations are encouraged to take appropriate measures to manage data quality, including training data and data collection, to mitigate against harmful biases.

Appropriate measures could include transparency, privacy-preserving training techniques, and/or testing and fine-tuning to ensure that systems do not divulge confidential or sensitive data.

Organizations are encouraged to implement appropriate safeguards, to respect rights related to privacy and intellectual property, including copyright-protected content.

Organizations should also comply with applicable legal frameworks.

B. Apulia Agenda

Excerpt from the Apulia G7 Leaders' Communiqué

Artificial Intelligence, Science, Technology, and Innovation

Artificial Intelligence can play a crucial role in promoting progress and development in our societies. We will promote safe, secure, and trustworthy AI. We will pursue an inclusive, human-centered, digital transformation that underpins economic growth and sustainable development, maximizes benefits, and manages risks, in line with our shared democratic values and respect for human rights.

In this regard, we recognize the need for approaches to AI governance that foster inclusion, to help us harness the potential of AI in a way that reflects these values and promotes its development while mitigating risks, including with respect to human rights and avoiding governance fragmentation. We will work toward these objectives by actively cooperating with other stakeholders, organizations and initiatives as relevant, such as the Global Partnership on AI (GPAI) and the OECD. We will build on the outcomes of the AI Seoul Summit and upcoming milestones, including this year's UN Summit of the Future and the AI Action Summit in 2025.

Recognizing the importance of advancing the Hiroshima AI Process outcomes, we welcome support from the countries and organizations beyond the G7, as demonstrated by its Friends Group.

We will step up our efforts to enhance interoperability amongst our AI governance approaches to promote greater certainty, transparency and accountability while recognizing that approaches and policy instruments may vary across G7 members. We will take a risk-based approach in these efforts as we seek to foster innovation and strong, inclusive, and sustainable growth. To achieve this goal, we will step up our coordination around the evolution of our governance and regulatory frameworks, including by sharing best practices. We will enhance our regular consultations. We are also committed to deepening coordination between our respective institutes and offices focused on AI, to work towards shared understanding of risk management and advance international standards for AI development and deployment.

We welcome our Industry, Tech, and Digital Ministers' efforts to advance the Hiroshima AI Process outcomes released last year, including the development of a reporting framework for monitoring the International Code of Conduct for Organizations Developing Advanced AI Systems. We look forward to the pilot of the reporting framework, developed in cooperation with the OECD, in view of the Industry, Tech, and Digital Ministers' Meeting in October. We

will work towards developing a brand that can be used to identify organizations that are voluntarily participating in and implementing the Code's forthcoming reporting framework.

We welcome the G7 Toolkit for Artificial Intelligence in the Public Sector, that can help governments to deliver better services to our economies and societies, while protecting human rights and fundamental freedoms.

We will work to ensure that AI enables increased productivity, quality jobs, and decent work; empowers workers; fosters inclusiveness and equal opportunities in the world of work; and enhances active labor market policies, including by fostering dialogue and transparency with workers organizations. To achieve these goals, we will launch an action plan on the use of AI in the world of work. We ask our Labor Ministers to develop the action plan, envisaging concrete actions to fully leverage the potential of AI to enable decent work and workers' rights and full access to adequate reskilling and upskilling, while addressing potential challenges and risks to our labor markets. We emphasize the need to anticipate future skills needs, provide higher education opportunities and equip workers and employers with the skills and competencies needed to design, adopt, and work with a human-centric, safe, secure and trustworthy AI. Against this background, we also emphasize the importance of innovative education, international talent mobility, digital competencies, and personalized lifelong learning to meet the demand for a qualified workforce. We recognize and encourage the work of the private sector to address skills gaps, including through the Global Partnership on Artificial Intelligence's Working Group on the Future of Work.

Acknowledging the key role that emerging technologies can play in economic growth, we commit to enhancing cooperation to bolster the adoption and development of new technologies, including AI, among micro, small, and medium enterprises, thereby fostering inclusive economic growth.

As we look to the future of emerging technologies, we encourage transparency and adherence to international workers' rights and labor standards at each stage of the AI supply chain. We also encourage our competition authorities to monitor the development of the AI industry, with a view to addressing potential competition issues, and prevent adverse effects at an early stage.

We will also work, including with developing countries and emerging economies, towards closing digital divides, including the gender digital divide, and achieving digital inclusion. We welcome the UN General Assembly Resolution on Seizing the Opportunities of Safe, Secure and Trustworthy AI Systems for Sustainable Development, which advances international

conversations on AI, including to promote equitable access to the benefits of AI for all. We will leverage the benefits of AI for SDGs by closing gaps in technologies for development, and by strengthening research and development ecosystems. To this end, G7 countries seek to promote safe, secure and inclusive practices, tools and solutions to make the benefits of AI and advanced computing available to partners to advance their development. In this regard, we welcome the Italian Presidency's decision to establish the AI Hub for Sustainable Development, in collaboration with UNDP. The Hub aims to enable multistakeholder partnerships to support local AI digital ecosystems, strengthen capacities to advance AI for sustainable development, and complement existing initiatives including the AI for Development Donors Partnership.

We reiterate the importance of operationalizing Data Free Flow with Trust (DFFT) to enable trustworthy cross-border data flows, and invigorate the digital economy as a whole, while preserving governments' ability to address legitimate public interest. We value OECD's leading role in advancing DFFT and we welcome its expert community.

Acknowledging the crucial role of resilient and reliable global semiconductor supply chains, we welcome the establishment of a Semiconductors G7 Point of Contact Group. This Group will bolster our coordination in addressing issues impacting the semiconductor industry. We will also advance our cooperation on secure and resilient undersea cable connectivity, in particular for strategic routes such as the Arctic and the Pacific. Our efforts will include better coordination on technical security requirements and advancing research on the economic and environmental sustainability of cable connectivity.

Given the potential of quantum technology developments, we will adopt or implement our respective quantum strategies. We also affirm our commitment to promoting responsible innovation of biotechnology, including its convergence with AI.

We support the development of frontier science, emerging technologies and research infrastructures to solve global challenges, including a better understanding of the ocean-climate-biodiversity nexus. We also emphasize the importance of promoting international talent mobility and circulation in emerging technologies among the G7 and partners, as well as

cooperation with low- and middle-income countries. In this regard, we reiterate our commitment to open science and research security and integrity.

We support further efforts to promote and reinforce research security and integrity, together with like-minded partners. We welcome the Extension of the G7 Virtual Academy to non-G7

like-minded partners to share best practices and policies on research security and integrity, and we welcome that Italy will host a G7 conference on these topics this year.

We recognize the impact of AI on the military domain and the need for a framework for responsible development and use. We welcome those who have endorsed the Political Declaration on Responsible Military Use of AI and Autonomy (REAIM) and the REAIM Call to Action, and we encourage more States to do so to ensure that military use of AI is responsible, complies with applicable international law, particularly international humanitarian law, and enhances international security.

We strongly support the implementation of the International Guidelines adopted at the UN Committee on the Peaceful Use of Outer Space as urgent and necessary. We welcome national efforts to develop further solutions for space debris mitigation and remediation, including further research and development of orbital debris mitigation and remediation technologies, and the development of space sustainability standards and regulations.

We welcome the establishment of the G7 Venice Justice Group, which will serve as a coordinative function to address global challenges, including AI, using our judicial and enforcement expertise. Against this background, we recognize the impacts of deploying AI within the justice sector and that the use of AI systems must not interfere with the decision-making power of judges nor judicial independence.

C. Stresa and Washington Finance Ministers and Central Bank Governors' discussions

Excerpt from the G7 Finance Ministers and Central Bank Governors' Communiqué

Stresa, 23-25 May 2024

Financial Sector Issues

11. We remain committed to strengthening cyber resilience in the financial sector, also against the background of heightening geopolitical tensions and in the context of hybrid threats. Cyber threats are evolving rapidly and becoming increasingly sophisticated. Emerging technologies, like Artificial Intelligence (AI) and quantum computing, provide new opportunities but also pose new challenges that are not yet fully understood. Besides regulation and supervision, it is crucial to promote sound cooperation and information sharing among relevant public and private stakeholders. Adoption of guidelines, deeper analysis of emerging risks, and cyber exercises are important building blocks of an effective strategy and should also inform the international coordination of responses and sharing of key information. In this respect, we welcome the successful completion of the cross-border coordination exercise that the G7 Cyber Expert Group (G7 CEG) conducted on 16 and 17 April 2024, and we ask the G7 CEG to further advance its work to improve the financial sector's preparedness and response capacity to cyber threats. (…)

Artificial Intelligence

15. We recognize the potentially transformative role of AI for our economies and societies. AI offers novel opportunities for productivity growth, but it also brings new risks and policy challenges, notably for the labour market and financial stability for example the potential for herd behaviour and an increase in the frequency of exogenous financial shocks.

We will advance our discussion on AI's economic potential, and on how to leverage AI to increase productivity and growth while supporting a human-centred approach and enhancing well-being. AI also offers new opportunities to improve the quality of public services, the efficiency of the public administration, and the effectiveness and fairness of tax collection. We will share experiences among finance ministries and central banks to forge a common G7 view on how to leverage the potential of AI while closely monitoring and mitigating adverse effects and risks. We will continue work to deepen our understanding of how AI affects the economy and the financial sector and how to ensure that our institutions are equipped to deal with AI, with a focus on the following shared policy agenda: macroeconomic impact and scenarios,

measurement challenges, impact on fiscal policy and on financial stability, implications on the required skills of the labour force, and environmental sustainability.

16. Furthermore, we recognize that AI offers significant opportunities for social and economic progress, including in developing countries, in crucial sectors like health, education, and agriculture. However, we also recognize that there is a risk of further divergence among countries if competition, distributional and displacement effects are not properly addressed and, in particular, if enabling conditions are not in place. We are committed to continue these discussions and we invite International Financial Institutions to keep working on these issues, within their mandates, to assess the macroeconomic implications of AI adoption and to help developing countries take advantage of the opportunities of AI while mitigating risks.

Excerpt from the G7 Finance Ministers and Central Bank Governors' Statement

Washington, DC, 25 October 2024

Artificial Intelligence

We remain committed to advancing our discussion on how to leverage AI in a safe, secure, and trustworthy way to increase productivity and growth while minimising the risks to the financial system and the wider economy. Following up on our Stresa shared policy agenda, we set up a High-Level Panel of Experts to identify the opportunities and challenges for economic and financial policymaking arising from the development and use of AI and to prepare a Report for the G7. The Panel is focusing on the implications of AI for policymakers on areas deemed at the core of the G7 Finance Track, including macroeconomic impact, the potential use of AI by governments and financial agencies, financial stability considerations, implications for skills of the labour force, and environmental sustainability. We look forward to the Panel's assessment of how to harness the benefits of AI while mitigating the associated risks. We welcome the Panel Chair's update on the ongoing work and look forward to the Report on AI and Economic and Financial Policymaking.