INTERNATIONAL MONETARY FUND

The Economic Impacts and the Regulation of AI: A Review of the Academic Literature and Policy Actions

Mariarosaria Comunale, Andrea Manera

WORKING PAPER

2024
MAR

IMF Working Paper

[Research Department]

**The Economic Impacts and the Regulation of AI:
A Review of the Academic Literature and Policy Actions.**

Prepared by Mariarosaria Comunale and Andrea Manera

Authorized for distribution by Chris Papageorgiou
March 2024

ABSTRACT: We review the literature on the effects of Artificial Intelligence (AI) adoption and the ongoing regulatory efforts concerning this technology. Economic research encompasses growth, employment, productivity, and income inequality effects, while regulation covers market competition, data privacy, copyright, national security, ethics concerns, and financial stability. We find that: (i) theoretical research agrees that AI will affect most occupations and transform growth, but empirical findings are inconclusive on employment and productivity effects; (ii) regulation has focused primarily on topics not explored by the academic literature; (iii) across countries, regulations differ widely in scope and approaches and face difficult trade-offs.

The Economic Impacts and the Regulation of AI: A Review of the Academic Literature and Policy Actions

Prepared by Mariarosaria Comunale and Andrea Manera[1]

1 Introduction

This review paper investigates how Artificial Intelligence (AI) affects the economy and how the technology has been regulated, relying on academic and policy sources through early 2024. We cover insights on employment and wage effects, productivity, and economic growth from the economic literature. In the policy realm, we summarize the regulatory actions undertaken in different regions, detailing their rationales, approaches and areas of coverage. Given the rapid evolution of AI technologies and the related literature, the paper aims to provide a structure to organize the latest contributions for the use of policymakers, economists, researchers, and industry stakeholders.

Before delving deeper into the content of our paper, we shall clarify its scope with some key definitions. Professor John McCarthy, one of the organizers of the 1956 Dartmouth research project that started AI as a field, defined "AI" as "the science and engineering of making intelligent machines" (McCarthy 2007).[1] In this review, we concern ourselves with the economic impact and regulation of recent advances in AI, such as machine learning (ML), and its sub-fields of deep learning, generative-AI (gen-AI) and Large Language Models (LLMs).[2,3] Broadly, we cover empirical studies focusing on either ML applications excluding gen-AI ("pre-gen-AI", until late 2022) or on the latest gen-AI and LLMs ("post-gen-AI," mostly since 2023). This distinction is motivated by the fact that LLMs and gen-AI have come to the fore of the economic and policy debate after the release of Dall-E 2 and ChatGPT by Open AI in late 2022, and related data, applications and research remain relatively scarce. Accordingly, we refer to "ML" as machine learning or deep learning for prediction, image and pattern recognition, text analysis and data analysis. When using the term "ML," we exclude gen-AI, which is treated separately due to the wide availability of text- and image-generating LLM tools. We will clarify the type of AI covered by each paper when the context does not make it clear. Consistent with this discussion, we will use "ML" to denote research that applies to machine learning and excludes gen-AI, usually for lack of available date. We instead denote studies as "gen-AI" if they focus exclusively on gen-AI. When we use the term "AI," we refer to contexts encompassing both ML and gen-AI technologies. This distinction is mostly relevant to empirical or experimental papers, since theoretical studies often do not distinguish gen-AI from ML. In the case of regulation, there is no global consensus on a definition of "AI" or the technologies covered by "AI" regulation.[4]

We start by describing the impact of AI on labor markets, where the literature has focused on employment and wages of various occupational groups. Most of the studies we review adopt—implicitly or explicitly—the task frameworks of Zeira (1998), Acemoglu and Autor (2011), and Acemoglu and

---

[1] The original document coining the term AI was the 1955 proposal for the Darthmouth Research Project coauthored by McCarthy, Minsky, Rochester and Shannon, see http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

[2] ML studies how computer agents can improve their perception, knowledge, thinking, or actions based on experience or data.

[3] LLMs are broadly defined as neural networks that learns context and meaning by tracking relationships in sequential data. See https://www.nvidia.com/en-us/glossary/data-science/large-language-models/.

[4] See O'Shaughnessy (2022) and also Appendix A.1

Restrepo (2018) that have been applied to study human-robot substitution in the automation literature. As we briefly discuss, these frameworks model output as a bundle of tasks carried out by either workers or capital, and obtain that employment and wages of different occupational or demographic groups are directly related to the quantity of tasks assigned to each group. Accordingly, the earlier pre-gen-AI literature and much of the post-gen-AI literature focused on computing "task exposures"–the share of tasks that can potentially be replaced by AI–to give estimates of the potential impact of AI on groups of workers. Following the earlier automation literature, these works generally associate a higher task exposure with larger potential displacement for affected groups of workers. Most studies agree that white-collar, higher-skilled occupations have higher task exposures and therefore face stronger employment risk from AI adoption (e.g., on gen-AI, Eloundou et al. (2023)). Other researchers instead separated task exposure and employment risk, highlighting the augmentation potential of AI technologies or other "shielding factors" (e.g., Cazzaniga et al. (2024)). Empirical studies on ML show that employment effects might overall be nil or positive (Acemoglu, Autor, Hazell, et al. 2022; Albanesi et al. 2023), while experimental papers on gen-AI highlight productivity gains for low-skilled workers that could harbor in wage compression (Noy and W. Zhang 2023; Brynjolfsson, Li, and Raymond 2023). Ultimately, theory suggests that the final verdict on labor market impacts of AI will depend on the race between job displacement and productivity increases, resulting from direct worker complementation or economy-wide gains from AI. The findings that we present provide some elements to evaluate this trade-off, but leave us far from a definitive understanding. In particular, all studies seem to agree that exposure is pervasive, but remain inconclusive on how such exposure may translate in substitution of complementation of workers. We close our discussion of labor market impacts of AI with a discussion of potential policy to tackle the potential labor displacement brought about by this technology.

Next, we proceed to survey the more limited studies concerning productivity and growth effects of AI, which encompass theory, firm-level studies on ML, and the experimental evidence on genAI cited above. The work in this area agrees on the theoretical possibility of large gains from AI adoption, but their actual magnitude remains uncertain. Firm-level studies on pre-gen-AI suggest that adopting firms may see sales per worker increase between 0 and 6.8%. In this discussion, we also include estimates of AI adoption—a crucial conduit to realize productivity gains. We close by assessing the potential impact of AI on emerging markets and developing economies (EMDEs), where some studies suggest that growth and productivity spillovers may be more relevant than dis-employment effects.

The rest of the review covers the regulatory challenges and responses emergent in the wake of AI's speedy deployment. We start by enumerating the rationales for regulation that have attracted most attention in the literature and regulatory action to date: market competition; privacy concern; intellectual property protection; military uses and national security; ethical issues and algorithmic bias; and financial stability risks. We also briefly discuss the interplay of AI and political systems. Insights on optimal AI

regulation from the literature are collected in a dedicated section. We then detail how different regions such as the European Union (the EU), the United States (the US), and China have initiated steps to regulate AI and incorporated the various rationales, notwithstanding substantial ambiguity in the definition of AI for regulatory purposes. These definitions are the object of Appendix A.1, which also contains the links to each regulatory document. We then briefly describe other cases in Advanced Economies (AEs) and Emerging Markets (EMs). We close this section with a look at ongoing multilateral actions and fora. In summary, only a few countries considered in this review have covered financial stability aspects in their AI regulation, while the majority explicitly include obligations on monopoly, ethical, and privacy risks. Current regulations are not clear yet on copyright of AI-generated (or co-generated) material, and in some cases (US, China) they defer to lower courts the decision to apply existing copyright laws. The US stand out covering national security explicitly, while other entities like the EU have been more ambiguous. Overall, countries have taken very different approaches to AI regulation: an ex-ante risk-based one prevails in the EU and Brazil; a decentralized approach based on guidelines is taking shape in the US; China has focused on algorithm recommendation and ethical reviews; the UK espoused a context-based view in recent a white paper; finally, Japan and India maintain a largely deregulated and flexible view on AI. Ultimately, all regulators face important trade-offs, which require them to balance first-mover advantages from AI innovation and development with potential risks.

We provide a comprehensive discussion of the limitations and future avenues for research and policy in a dedicated section, which also delineates our three main takeaways from the review. First, there is a lack of consensus in the academic literature on the effects of AI, which we attribute primarily to a lack of adequate and timely data. Second, we detect a disconnect between policy and research and a need for more research to inform areas and actions of interest to the regulators. Third, we note that regulations differ widely in their approach and scope, and face difficult trade-offs that may be addressed by increased multilateral cooperation.

The structure of the paper is as follows. Section 2 focuses on the potential effects of AI on the labor market, including adaptation of the workforce and public policies. Section 3 looks at the impact that AI might have on productivity and growth, based on the most recent estimates. Section 4 gets covers the regulatory challenges related to AI and related actions. In Section 5, we highlight open issues and gaps in the economic literature and regulation of AI and provide our key takeaways. Section 6 concludes.

2   AI and the Labor Market

AI technologies may induce dramatic change and profound dislocations in the labor market. While official statistics denote AI adoption as still in its infancy, both national strategies and projections indicate that the landscape will change rapidly (see Section 3). A comparison with previous technological revolutions

and with automation in manufacturing suggests that AI deployment may benefit certain individuals or groups at the detriment of others. With an eye to identifying these groups, the economic literature has concentrated on the computation of "task exposures," the share of tasks that AI might carry out autonomously in each job, finding that white-collar high-skilled workers are most exposed. Part of the literature adopts a view of AI as an automation technology—a technology that replaces workers, like industrial robots. In this view, occupations with high task exposure will face higher displacement, and exposed workers will see reduced employment opportunities and wages. Another strand of the literature instead tries to assess the actual consequences of such exposure, finding that AI may augment workers in exposed occupations. We close this section with a description of policy options to cope with potential worker displacement that research has proposed.

2.1 The Task Framework of Automation and Empirical Studies on Robots

In its capacity of "intelligent machine," AI can carry out typically human tasks, like pattern recognition, text analysis, and prediction without supervisions. As such, it is akin to other forms of labor-substituting automation technologies studied by the vast literature recently surveyed by Restrepo (2023). As discussed in detail there, the "task model of automation" provides a simple foundation to assess the effect of labor-displacing technological change. The task model postulates that producing goods and services requires the completion of tasks that can be assigned to groups of workers, defined by their skills and abilities, or robots and machines. The crucial takeaway of this framework is that the *relative* labor demand for groups of workers whose tasks are taken over by robots falls with automation, and the *level* of labor demand for these groups might even fall depending on the productivity gains that come with robot adoption, and workers' ability to transition to occupations unaffected by task displacement. Acemoglu and Restrepo (2022) derive explicit closed-form log-linear formulas that summarize the effect of automation on wages of affected worker groups. Aggregate productivity and the demand for labor in non-automated tasks that workers can easily reallocate positively impact wages, while the share of tasks displaced by automation directly lower wages. The overall effect of automation on labor market outcomes thus hinges on whether productivity improvements and their demand spillovers to other sectors outweigh direct substitution effects. When applied to the data, this framework explains 50% of the labor share declines across US industries over the period 1987-2016.

We refer the interested reader to Restrepo (2023) for a full discussion of the automation literature concerning industrial robots. For the purposes of our review, we can summarize the empirical literature in this strand as showing that groups experiencing large task displacement from robot adoption saw decreasing employment and wages, which is consistent with US evidence that firms deployed these technologies primarily with the aim of substituting human labor (Acemoglu, Anderson, et al. 2022). Similarly, firm-level studies in various countries generally found that automation led to a reduction in

the labor share and an increase in sales per worker, although sometimes accompanied by an employment expansion. However, as highlighted by Restrepo (2023), these effects are hard to interpret and translate to the aggregate level as adopting firms might just be expanding at the expense of competitors. Accordingly, the impact of this first wave of automation on economy-wide productivity remain uncertain.

The above discussion clarifies why the emerging literature on AI has focused on finding a measure of AI task exposure for different occupations, defined as the share of tasks that could be carried out by AI in the relevant occupation or job. Under the lenses of the automation literature, this measure should give a sufficient statistic to predict relative wage changes, as more exposed workers would see relative wage losses relative to less exposed workers. However, the nature of AI and specifically gen-AI puts into question whether this technology will act as a complement or a substitute for workers in related tasks (see, e.g., Autor (2022)). For this reason, the pre-gen-AI literature on ML has looked to empirical evidence on the relation between AI exposure or adoption measures and various outcomes, while more recent studies produced experimental evidence on gen-AI. In what follows, we first present studies on exposure measures and then turn to the empirical evidence that may be used to interpret these measures, and understand whether AI might differ from previous waves of automation.

2.2 AI Task Exposures

The literature on non-AI automation, focusing primarily on industrial robots, found strong exposure for blue-collar and less educated workers (Acemoglu and Restrepo 2022), which has been compounded by increasing substitutability of low-skilled workers with machines (A. Berg et al. 2024). Under the lenses of potential task exposures, however, AI appears different.

Exposure Estimates. Early studies on AI task exposures focused on ML. Namely, Brynjolfsson, Mitchell, and Rock (2018) find that machine learning can displace tasks for occupations throughout the wage distribution in a rather uniform way, which contrast markedly with the low-skill bias of industrial robots. In a similar vein, Webb (2019) finds that AI is mainly directed at high-skilled tasks and will affect highly educated and older workers. The author also estimates that AI could reduce 90:10 wage inequality by 5 to 10%. This finding is under the assumption that the wage elasticity to AI exposure is the same as the wage elasticity to software and robots' exposure, in line with the task framework of automation of Acemoglu and Restrepo (2018) that sees larger task exposure as leading to lower relative wages. E. Felten, Raj, and Seamans (2021) build a comprehensive measure of exposure to machine learning linking 10 AI applications (e.g., image generation, language modeling, etc.) to 52 human abilities (e.g., oral comprehension and expression, inductive reasoning, etc.) which are in turn mapped to occupations. Highly educated, highly paid, white-collar occupations appear most exposed to machine learning. E. W. Felten, Raj, and Seamans (2023) obtain analogous findings when restricting their measure to ML applications

6

more related to gen-AI such as language models and image generation. Eloundou et al. (2023) estimate that up to 80% of US workforce could have at least 10% of their tasks replaced and 19% of workers stand to lose at least 50% of their tasks to LLMs. This analysis also shows that about 15% of all worker tasks in the US could be completed significantly faster at the same level of quality if LLMs are employed. Once again, this exposure is higher for higher-paid and high-skilled professional occupations. In a similar fashion, Gmyrek, J. Berg, and Bescond (2023) also show that the highest share of tasks exposed to AI technology are clerical jobs, where the majority of tasks fall into medium- to high-level exposure. Additionally, the authors analyze separately automation (labor substitution) and augmentation effects, finding that the latter outweigh the former across both low- and high-income countries. High-income countries see 5.5% of total employment exposed to automation and 13.4% to augmentation from AI, while the same percentages are 0.4% and 10.4% for low-income countries, respectively. Interestingly, the earliest study to incorporate earlier machine learning capabilities into a measure of potential computerization (Frey and Osborne 2017) found a negatively sloped relation between IT (Information Technology)-driven labor substitution exposure and average occupational earnings. Therefore, if we interpret the time of the studies as indicative of the state of machine learning capabilities, it appears that AI has become more biased towards high-skilled occupations in recent years.

Correcting for Complementarity. While some of the studies described above explicitly interpret AI exposure as a measure of potential substitutability, most are agnostic on what exactly task exposure entails. By contrast, Pizzinelli et al. (2023) construct an indicator of AI occupational exposure adjusted for the complementarity of the technology with the exposed occupation (C-AIOE). The resulting measure therefore provides a more direct gauge of the potential displacement from AI. Occupations' substitutability and complementarity are assessed independently from exposure itself based on US O*NET occupation characteristics.[5] The authors first compute an unadjusted exposure index, reporting that AEs have a higher share of high-exposure jobs with both high and low levels of complementarity. When considering demographic characteristics, female and high-skilled workers appear more exposed. Adjusting for complementarity, the C-AIOE measure shows a lower exposure for high-skilled occupations relative to its unadjusted counterparts, leading to a more uniform exposure of AI throughout occupations. Employing this framework in model-based scenarios, Cazzaniga et al. (2024) find that AI would raise wage inequality for sufficiently high complementarity, and wealth inequality regardless of complementarity. Similarly, Huang (2024) classifies tasks that are exposed to AI into groups that allow a better assessment of whether the technology will substitute or complement human labor. Using this classification, the author shows that high-skill workers with a high AI exposure would see their employment share in the economy increase relative to less AI-exposed. This increase would be similar to that experienced by progress in IT,

---

[5]For details on the O*NET data, see National Center for O*NET Development, O*NET OnLine: `https://www.onetonline.org/`

which affects predominantly low-skilled workers, but substitutes them instead of complementing them.[6]

Kogan et al. (2023) provide an estimate of wage and employment effects of AI on occupational groups that explicitly incorporates the potential of AI to both augment and displace tasks. The authors calibrate a model that matches empirical evidence on the impact of labor-saving and labor-augmenting technologies patented since the 1980s. In the model, technology substitutes routine tasks, augments non-routine tasks, and causes skill obsolescence for incumbent workers who cannot efficiently use it. Accordingly, each technology has both a labor-saving and a labor-augmenting exposure. Assuming that workers with the highest labor-saving exposure will lose 10% of their earnings from AI, the model computes the direct effects–that is, excluding reallocation across sectors–of current AI technology by occupational group. On average, workers stand to lose 4% of their earnings. This effect stems from skill obsolescence as, for almost all occupations, the potential of task augmentation undo the negative task displacement effects. In terms of skill content, more routine occupations such as office and admin work, production and transportation stand to lose the most (6-8%), while higher-skilled occupations like management, engineering, legal, education and social service occupations log the lowest losses (around 2%).

Estimates of Aggregate Employment Impacts. Think tank and private sector estimates provide some additional information on the potential extent of AI impacts on the labor market. Pew Research (2023) finds that 19% of all US workers are in occupations most exposed to AI.[7] By the same token, Briggs and Kodnani (2023) find an employment exposure of 25% for the US, 24% for the euro area, and 18% worldwide, with effects concentrated in AEs. Ellingrud et al. (2023) offers somewhat smaller estimates, predicting that 8% of hours currently worked will be automated by 2030 because of gen-AI, ranging from around 16% of all hours in STEM (science, technology, engineering, and math), education, and business and legal professional occupations, to 4% in production work and 3% in agriculture. However, when factoring in other structural factors, these figures translate into slowing—but still increasing—demand for higher skilled occupations like STEM, and decreased demand only for selected customer-facing and production occupations.

Aside from Briggs and Kodnani (2023) and Gmyrek, J. Berg, and Bescond (2023), only Pizzinelli et al. (2023) offer a glimpse into the potential difference of AI impact between AEs and EMs. The authors use worker-level microdata from 2 advanced economies (AEs: US and UK) and 4 emerging markets (EMs: Brazil, Colombia, India, and South Africa). Stronger high-skilled workers' exposures also translate into advanced economies being more susceptible to employment displacement from AI compared to EM counterparts. However, this difference is smaller when considering task complementarity.

Economic Feasibility of Task Replacement. In closing this section, it should be noted that

---

[6]In this article, IT refers to software and computer equipment.

[7]See https://www.pewresearch.org/social-trends/2023/07/26/exposure-of-workers-to-ai/

task exposure measures the *theoretical possibility* that AI may carry out some tasks. This feature makes exposure a potentially poor measure of the employment effects that we might expect from the adoption of the technology. Svanberg et al. (2024) explicitly assess the technical and economic feasibility of automating vision tasks that have a high theoretical AI task exposure. They find that, given current costs, most US businesses would not automate most tasks with high exposure. Only 23% of exposed jobs (8% of total jobs) in U.S. non-farm businesses have at leas one task that is economically attractive to automate. When considering the contribution of automated tasks to compensation, the authors conclude that only 0.4% of non-farm wages would be potentially lost to automated computer vision.

2.3 Empirical and Experimental Studies on AI

In the previous section, we saw that high-skilled and higher-paid workers seem to be more exposed to AI. However, as discussed, the consequences of exposure are less clear-cut than in the case of previous waves of automation. For this reason, we now move to survey empirical studies on the effect of AI and some of the recent experimental literature focusing more narrowly on gen-AI.

Firm-level Studies on ML Impacts. Empirical papers on AI adoption have to rely on data collected before the introduction of more powerful gen-AI tools. Alekseeva et al. (2021) and Acemoglu, Autor, Hazell, et al. (2022) analyze AI-related vacancies in the US. Alekseeva et al. (2021) find a large increase in the demand of AI skills across industries over the period 2010-2019. Job postings listing AI skills command a 5% wage premium compared to other same-firm, same-job postings, and firms that demand more AI skills tend to be larger and more innovative. Acemoglu, Autor, Hazell, et al. (2022) employ similar data for 2010-2018 and find that establishments that increase postings related to ML skills simultaneously reduce vacancies unrelated to ML and change their skill requirements more quickly, which might be consistent with substitution of non-AI work. The authors cannot detect aggregate effects of AI technology, which they ascribe to limited economy-wide adoption. More recently, Babina, Fedyk, A. X. He, et al. (2024) and Babina, Fedyk, A. He, et al. (2024) combine US firm data with employee resumes. In Babina, Fedyk, A. X. He, et al. (2024), the authors show that firms investing in AI tend to transition to a more educated workforce mostly focused on STEM and IT. At the same time, AI adoption tends to flatten firms' hierarchical structures, increasing workers in junior positions and decreasing workers in middle management and senior roles. In terms of firm outcomes, Babina, Fedyk, A. He, et al. (2024) find that investment in AI experience led to increased sales, employment, market valuations and product innovation. Further, AI-fuelled growth is more prevalent among larger firms and in industries with higher market concentration.

Copestake et al. (2023) offer a developing country perspective on AI adoption, using demand for ML skills observed in the text of posted job descriptions in India over the period 2010-2019. Consistent with

Alekseeva et al. (2021), the authors document a rapid increase in 'AI demand' after 2016, particularly in the IT, finance, and professional services industries, as a well as a 13 to 17% salary premium for job postings demanding AI skills. At the same time, similarly to Acemoglu, Autor, Hazell, et al. (2022), they document a fall in non-AI job postings and wage offers. Cornelli, Frost, and Mishra (2023) examine AI-related investments across 86 countries for the same period. Their findings link AI-related investments to a shift from mid-skill jobs to high-skill and managerial positions, accompanied by a decline in the labor share of income, higher total factor productivity (TFP) and increasing inequality at the aggregate level. Alderucci et al. (2021) analyze the outcomes of US firms *inventing* AI technologies, which they identify through the text of patent grants. AI-related innovations are associated with faster employment and revenue growth, higher output per worker, and an increase in within-firm wage inequality. The World Economic Forum (WEF) Future of Jobs Survey (World Economic Forum 2023) predicts that 23% of global jobs will change in the next five years due to industry transformation, including AI-related. The survey confirms previous findings: the fastest-growing roles are technology-related, with AI and ML specialists at the top, while clerical and secretarial roles are the ones declining faster.[8] Finally, Milanez (2023) conducted 100 case studies of AI applications over the period 2021-2022 in Austria, Canada, France, Germany, Ireland, Japan, the UK, and the US. Only 23% of firms stated that job quantities declined for affected workers; in these cases, workers were reallocated to different tasks (existing or new) without adverse impact on overall employment.

Experimental Evidence on AI Complementarity. Some more recent experimental studies focus more closely on gen-AI. Brynjolfsson, Li, and Raymond (2023) study gen-AI adoption in customer service call centers. The authors report productivity gains accrued primarily by less experienced and lower-skilled workers. Their conclusion is that AI can disseminate the knowledge accumulated by more knowledgeable workers, helping newer workers move down the experience curve. Similarly, Noy and W. Zhang (2023) show that experimental exposure to ChatGPT increases productivity and output quality in writing tasks, while reducing the performance differential between low- and high-ability workers. In the context of radiology, Agarwal et al. (2023) study the effectiveness of human-AI collaboration through an experiment with radiology experts comparing the performance of humans to AI as well as humans assisted by AI. Their findings suggests that it is optimal to assign cases either to humans or to AI, but rarely to a human with the assistance provided by AI since human radiologists tend to underweight the information provided by AI when it deviates substantially from their own beliefs.

---

[8]The Future of Jobs Survey covers 803 companies employing more than 11.3 million workers in total. These span across 27 industry clusters and 45 economies.

2.4 Policies to Mitigate Displacement and Ease Labor Reallocation

While the empirical evidence surveyed above is still ambiguous on the extent and characteristics of AI-driven labor displacement, several academic and policy studies suggested policies to deal with automating technologies. Therefore, it is important to note that the policy recommendations below chiefly relate to the labor-saving nature of AI.

Skills and Workforce Adaptation. Analogously to previous waves of technological innovation, AI is likely to displace old skills and bring about new ones (Acemoglu and Restrepo 2018; Autor 2022). OECD (2023b) and in particular the section by Lassebie (2023) identify new AI-related skills that complement basic digital and data science competencies, as well as others that can foster diversity in the AI workforce. Among the former, programming, machine learning, and data science stand out, while the latter feature other cognitive and transversal skills—namely, creative problem solving, critical thinking, and mentoring.[9] This study also points out that AI itself might be used to deliver customized training, thereby fostering inclusion and accessibility. Focusing on AI professionals, Borgonovi et al. (2023) similarly conclude that soft skills will continue to be a key requirement for companies hiring AI professionals. Recent evidence on India's services sector (Copestake et al. (2023)), instead show that establishments looking for AI hiring tend to reduce their demand of analytical and complex communication skills as seen in their job postings. As a result of this changing demand for skills, Lassebie (2023) suggests promoting greater training provision by employers, as well as extending such training beyond currently vulnerable groups to ensure smooth future AI adoption and development.

Taxation and Fiscal Policy Trade-Offs. Acemoglu, Manera, and Restrepo (2020) adopt a different theoretical framework, featuring reduced-form labor market frictions as well as a finite capital supply elasticity. There, they show that setting capital taxes too low relative to labor may promote excessive substitution of workers compared to what is socially optimal. In other words, when taxes on capital are too low relative to taxes on labor, robots become artificially cheaper than workers in executing tasks, which causes larger employment losses from automation than socially optimal. Based on the outcomes from this paper, Acemoglu, Autor, and Johnson (2023) suggest equalizing tax rates on labor and ownership of equipment and algorithms to level the playing field between humans and labor-replacing machines, with the aim of potentially incentivizing human-complementary technology choices.[10] M. A. Berg et al. (2021) see automation as a technology that substitutes unskilled workers and raising the incomes of skilled workers and owners of capital. At the same time, automation has a large productivity effect, therefore robot taxation is not efficient, but simply a means to redistribute the benefits of automation at the detriment of lost output in the long run. A tax on robots that slows the accumulation of automating capital or a higher capital income tax combined with an unskilled wage tax cut deliver

---

[9]See Lassebie (2023), Table 5.1 for a complete list.

[10]`https://shapingwork.mit.edu/wp-content/uploads/2023/09/Pro-Worker-AI-Policy-Memo.pdf`

the largest welfare benefits. When imperfect competition is included in the model, a markup tax is the second-best policy after a robot tax. A. Berg et al. (2024) study the returns to corporate tax cuts, increased education spending, and increased infrastructure investment in a model economy, comparing a case featuring a broad definition of "robot" capital with a scenario without it. When "robot" capital is not present, infrastructure investments and corporate tax cuts produce more benefit than education spending. Conversely, the "robot" capital economy sees education spending as the most efficient policy measure, followed by infrastructure investment and tax cuts. To interpret these results, it should be noted that the authors assume that "robot" capital is modelled as a substitute for low-skilled labor, akin to pre-gen-AI automation.[11]

3 AI's Impact on Productivity, Growth, and Development

Contrary to labor market consequences, the potential AI impact on productivity and growth has received considerably less attention. While lack of data prevents a definitive empirical quantification, estimates based on theory or calibrations also remain scant. For this reason, in this section we discuss theoretical channels for the AI impact on productivity and growth, evidence on productivity for adopting firms, and some private sector estimates of aggregate growth effects. This section also includes a discussion of the estimates of AI diffusion since adoption is crucial to realize productivity gains. Finally, we briefly discuss to role of AI as an aid for development in EMDEs.

Theoretical Impact of AI on Productivity and Growth. Trammell and Korinek (2023) review the theoretical growth literature for clues on how accelerating AI progress may affect growth, the labor share and wages. We direct the reader to that excellent synthesis for detailed references. For the purposes of this review, it suffices to note that the vast majority of scenarios in the growth literature entail at least some substitutability of humans with AI in production, as well as a potential for AI assistance in the generation of ideas. Based on these frameworks, AI could produce permanent shifts in growth rates and even exponentially increasing growth, with consequences on the labor share and wages more dependent on specific modelling assumptions. A potential channel through which AI can lead to sharp increases in growth rates is augmentation of research-related tasks (Korinek 2023b). Baily, Brynjolfsson, and Korinek (2023) relying on the estimates of Noy and W. Zhang (2023) and Brynjolfsson, Li, and Raymond (2023), argue that AI could increase aggregate productivity by 33% over 20 years through its impact on knowledge workers' productivity. In addition, Korinek (2023c), drawing on Korinek (2023a), considers the transition to Artificial General Intelligence (AGI)—"AI that possesses the ability to understand, learn, and perform any intellectual task a human being can perform." The introduction of AGI is modelled as an increase in the mass of automatable tasks and the capital-labor ratio determines the effects of automation on

---

[11]A. Berg et al. (2024) defined "robot" capital to encompass robots, AI, computers, big data, digitalization, networks, sensors and servos.

wages.In this scenario, output will always increase, and wages can decrease or increase depending on capital accumulation.

Evidence from Firm-level Studies. Eisfeldt, Schubert, and M. B. Zhang (2023) construct a novel firm-level measure of workforce exposure to gen-AI for the US. Then study the impact of the release of ChatGPT on equity returns at the firm level with varying exposures to this technology shock. They find that this had a sizable positive effect on the value of firms whose labor forces are more exposed, with vast heterogeneity across and within industries. When considering innovation outcomes, Rammer, Fernández, and Czarnitzki (2022) find that the introduction of AI in German firms increases the probability of introducing a new product or process by about 8%. These results are qualitatively in line with analogous findings in Babina, Fedyk, A. He, et al. (2024), who report significant increases in product patents and trademarks associated with the expansion of AI-related workforce. When it comes to productivity effects, however, Babina, Fedyk, A. He, et al. (2024) find no increase in sales per worker. By contrast, Alderucci et al. (2021) estimate through an event study that firms innovating in AI technologies report an increase of 6.8% in sales per worker, while Czarnitzki, Fernández, and Rammer (2022) put this number at 4.4%.

Aggregate Productivity Impacts from Private-Sector Studies. Company (2023) estimated that the adoption of AI can contribute between 0.1 and 0.6pp of productivity growth annually between 2022 and 2040. Goldman Sachs puts this figure at 1.5pp annually for the 10 years following widespread adoption of the technology, extrapolating estimates on worker-level productivity using AI task and occupational exposure (Briggs and Kodnani 2023). In EMs, the impact is seen smaller of 0.7-1.3 percentage as sectors with low AI exposure like agriculture and construction are more prominent.

Estimates of AI Adoption. In the US, the Annual Business Survey of 850,000 firms, curated by the Census, shows that the average adoption of AI technologies was just over 18% in 2018, when weighted by employment, varying considerably by industry, and higher in younger more dynamic firms (McElheran et al. (2023)). Few "superstar" cities and emerging hubs were leaders in early adoption of AI. In the EU, the indicator "Artificial Intelligence" of DESI (Digital Economy and Society Index) shows the share of enterprises, with 10 or more persons employed and excluding the financial sector, that use AI technologies.[12] The average in EU in 2021 (pre-ChatGPT) is 7.9% and Big Data used by business stands only at 14%. 21 out of 27 countries do not exceed 10% of AI usage, while Denmark 23.9%, Portugal 17.3%, Finland 15.8%, Netherlands 13.1%, Luxembourg 13% and Slovenia 11.7% are the countries with the highest share.[13] The Path to the Digital Decade target requires that more than 75% of EU companies adopt AI technologies by 2030.[14] As these estimates refer to past data vintages, they

---

[12] This indicator is based on the share of enterprises using at least two AI technologies.

[13] https://ec.europa.eu/commission/presscorner/detail/en/ip_22_4560

[14] https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-desi-2022

do not include the recent spike in gen-AI technologies, and can accordingly be considered lower bounds to current AI adoption. This is particularly relevant in view of the increasing trajectory of AI-related vacancies documented in several studies (Alekseeva et al. 2021; Copestake et al. 2023).

AI and EMDEs. Due to their different sectoral composition, the various studies surveyed above find low potential automation impacts of AI for EMDEs (Pizzinelli et al. 2023; Gmyrek, J. Berg, and Bescond 2023). Here we wish to point out the distinctive benefits that AI can offer for development, following Björkegren (2023). Notably, AI can create new products meeting the specific needs of the developing world, like a financial planner for subsistence farmers to manage the risks in crop choice, or chatbot tutors to alleviate high student-to-teacher ratios. The fast scalability of AI solutions might also allow technologies to diffuse to poorer sectors of the population. Reaping this benefits will require substantial investment in infrastructure and education, as capacity is often lacking in EMDEs.[15] Crucial constraints are also data availability and model training on local languages.[16] The "AI Preparedness Index" developed by Cazzaniga et al. (2024), which incorporates these constraints, quantifies the substantial disparity between AEs and EMDEs in key resources and digital infrastructure. While a potential opportunity, outsourced AI-related work has been related to harsh working conditions, can become a further source of exploitation,and hence might pose its own set of challenges.[17]

Aside from direct effects, EMDEs might also be affected by AI adoption in AEs. As highlighted by Korinek and Stiglitz (2021), labor- and resource-saving AI could result in winner-takes-all dynamics at the detriment of developing countries who are labor- and resource-rich and to the advantage of advanced economies who are better positioned to adopt new technologies. This would lead to income divergence and significant deterioration in developing countries' terms of trade (Alonso et al. (2022)). If instead AI substitutes for more high-skilled workers, as some exposure measures suggest, these results could be overturned and lead to improvements in EMDEs' terms of trade.

4   Regulatory Challenges and Responses

In this section we describe the rationales and ongoing regulatory efforts. We look at regulatory aspects purely from an analytical and research viewpoint. Our objective is to merely report the current state of policy actions, without offering any legal assessment or advice, which is outside the scope of this review. AI-specific regulations here refer to AI models, applications and system: "AI models" are the algorithms trained to perform tasks; "AI applications" are practical uses of AI models; and "AI systems" refer to the complex of AI models and related inputs, process and infrastructure, such as hardware, software, data

---

[15]See the Government AI Readiness Index, `https://oxfordinsights.com/ai-readiness/ai-readiness-index/`.

[16]To compensate for the lack of data on EMDEs new content must be created for the models to train on. Some good examples came recently from crowdsourcing, e.g., WikiAfrica.

[17]`https://www.washingtonpost.com/world/2023/08/28/scale-ai-remotasks-philippines-artificial-intelligence/` and `https://time.com/6247678/openai-chatgpt-kenya-workers/` report instances where workers in developing countries employed to transform raw data into AI source material are underpaid and over-tasked.

processing, and interfaces. We start by describing the reasons for regulation, as described by relevant academic literature or the AI-specific regulations that we consider. In this review, we focus on three main cases: the EU AI Act,[18] the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (US EO hereafter), and the Chinese Interim Measures for the Management of Generative Artificial Intelligence Services. A brief discussion on the theory on AI regulation follows. Then, we summarize AI-specific regulations around the world, with a focus on the experience of the EU, US and China, which we summarize in Table 4.3.4. We focus on the three documents listed above, which explicitly address AI, relegating to Appendix B other legislation in selected countries that may apply to AI, but is not directly designed to this end. We conclude with a description of the state of international cooperation and multilateral actions, and of how AI companies are taking action to promote safety and best practices.

4.1 Reasons for AI regulation

In this section, we discuss the the main reasons for AI regulation. Agrawal, J. Gans, and Goldfarb (2019) and Acemoglu (2024) provide useful overviews of many economic and ethical rationales. In addition to these sources, we consult specialized literature in other fields, as well as the stated aims of proposed and enacted regulations. We choose to focus on six main areas which are prominent in the public discussion or among the intents of regulators. In order, we cover: the effects of AI on competition; privacy; copyright issues; military uses and national security; ethical and bias consideration; and financial stability. We list other issues, among which the impact of AI on electoral systems, in a separate sub-section. Policy initiatives in China, the EU and the US tackling each aspect are briefly mentioned in the relevant sub-section.

4.1.1 Market Competition

Concern over monopoly power in AI-related sectors permeate the public debate, as well as the intent of some regulators, notably in the US and China. In what follows, we discuss three angles that might justify regulation of AI to protect competition: (i) potential sources of market power; (ii) increased concentration in sectors using AI; (iii) unfair competition and price discrimination.

Sources of Market Power. There are different sources of market power that may entail current or future risks of monopolization. While characterized by high barriers to entry, access to computing power and model training represent moderate risks, due to the presence of several countervailing factors. Conversely, the markets for chip and semiconductor design and production, as well as raw materials, are already concentrated, posing a potentially higher risk.

---

[18]We use the latest version of the text as approved by the EU Parliament in plenary session on March 13, 2024. More details on the version and procedure are available at sub-section 4.3.1. We also refer to this version as "provisional EU AI Act".

A first area of concern is potential market power for AI services providers, generated by their preferential access to scarce resources. Following Varian (2019), we can identify the main resources to train a machine learning model as hardware, software, expertise, and data. State-of-the-art generative AI model require specialized chips and vast amount of computing power deployed over long periods. The connected large investments may act as barriers to entry and prevent competition in AI model generation. As noted by Varian (2019), this concern might not be justified when it comes to hardware and software, since cloud computing provides a cheap, variable-cost alternative to large fixed-cost investment. By contrast, the same author notes that access to data is generally harder to come by, especially in vast quantities, requiring specialized infrastructures to collect and maintain and—perhaps more importantly—a source in other commercial activities. This last aspect might give a strong advantage to entrenched players in the tech sector who own other business generating and collecting data. While data access and computing costs represent barriers to entry, it is unclear how effective they will be in deterring competitors to incumbent providers. First, model performance exhibits decreasing returns in data. As a dataset expands, additional informational content pertains mainly to rare occurrences.[19] Accordingly, research shows that representative samples of datasets, so-called "synthetic datasets," can be used for training with little loss, and sometime gains, in performance.[20] However, large amounts of data may be crucial to achieve so called "emerging capabilities," discrete improvements in carrying out tasks that occur after models reach a certain scale.[21] Second, open-source models with lower computational requirements can outperform proprietary counterparts when applied to specialized settings. Through a process known as "low rank adaptation" (LoRA) open-source models can tune a much lower number of parameters than competing alternatives while producing similar results. This is achieved by tailoring learning to specific applications and using high-quality data instead of large quantities thereof. Some examples include Koala at Berkeley and Stanford's Alpaca.[22,23] The emergence of open-source solutions might provide a natural barrier to increasing monopoly power by incumbent firms.

Excessive market power may also occur at different stages in the production of chips and semiconductors, the crucial inputs to build, train, and deploy AI systems. The scarcity and geographic localization of crucial inputs at the very start of the supply chain adds a layer of trade and geopolitical risk to the discussion above. Some highly concentrated markets include raw materials such as gallium, chip-making

---

[19] Diminishing returns to additional data have been documented in different fields. See for instance: Sun et al. (2017) for deep learning in vision; Bajari et al. (2019) in the context of retail sales forecasting; and Varian (2019) for image classification.

[20] For an example on image recognition see: Kim et al. (2022), and press coverage in `https://news.mit.edu/2022/synthetic-data-ai-improvements-1103`.

[21] For a discussion of model scaling and the unpredictability of emerging capabilities in LLMs see Ganguli et al. (2022).

[22] See `https://bair.berkeley.edu/blog/2023/04/03/koala/` and `https://crfm.stanford.edu/2023/03/13/alpaca.html`

[23] The ability of open-source models is further highlighted in a leaked Google's internal memo, which claims that smaller models with LoRA can outperform LLMs in the long run. See `https://www.semianalysis.com/p/google-we-have-no-moat-and-neither` for the full text.

tools, and GPU production and design. [24,25,26] As noted by Agrawal, J. Gans, and Goldfarb (2019), many countries see investments in AI as strategic. As a result, countries may deploy trade restrictions to protect and develop domestic sectors or harm international competitors by denying access to key inputs.

Increased Concentration in Sectors Using AI. As discussed in the previous section, current large tech firms are poised to access large amounts of data as a byproduct of their operations, which might position them to develop powerful models and increase their market dominance. Current digital market leaders' dominant stakes in AI developers are a related matter of concerns for regulators. The EU Commission (DG Competition) in January 2024 launched a call for contributions to all interested stakeholders to gather insights on the level of competition in virtual worlds (virtual reality) and generative AI, as well as the potential role of EU antitrust authorities. The related document cited that Microsoft's investment into OpenAI is under scrutiny as a possible case of unfair competition.[27] Microsoft's role is also being investigated by the UK antitrust authority.[28] In January 2024, the US Federal Trade Commission started an inquiry into gen-AI investments and partnerships.[29]

Unfair Competition and Price Discrimination. Companies may utilize AI to extract surplus from consumers, manipulate users, or relax price competition. First, AI may allow firms to learn more about consumers, thus improving their ability to price discriminate, as discussed in Varian (2019).[30] More perniciously, companies may directly manipulate consumers using subconscious consumers' behavioral biases that algorithms can easily learn but consumers may be unaware of. A specific example, studied in Acemoglu, Makhdoumi, et al. (2023), involves producers tempting consumers into buying products that give instant gratification, but are harmful in the long run. Behavioral manipulation, unlike increased price discrimination, directly reduces social welfare instead of merely redistributing it from consumer to producer. In terms of market competition, Acemoglu (2024) notes that data advantages on specific segments of a market may lead companies to cater to more and more specialized clienteles, which would

---

[24] In 2022, China accounted for 98% of worldwide primary low-purity gallium production. See https://pubs.usgs.gov/periodicals/mcs2023/mcs2023-gallium.pdf.

[25] The Dutch company ASML holds exclusive patents for manufacturer of chip-making tools used by the latest generation of chips and semiconductors. See :https://www.asml.com/-/media/asml/files/investors/why-invest-in-asml/capital-return-and-financing/credit_opinion-asml-holding-nv-09aug2023.pdf?rev=3a50fe8a14164042b31bc790c6108bcc&hash=98D3CB014D8C0C01FBF7497F894B2FFA

[26] Nvidia is a leading company in the design of chips, with a market share estimated at 80-95%, see https://www.cnbc.com/2023/08/08/nvidia-reveals-new-ai-chip-says-cost-of-running-large-language-models-will-drop-significantly-.html

[27] [...] "the European Commission is looking into some of the agreements that have been concluded between large digital market players and generative AI developers and providers. The European Commission is investigating the impact of these partnerships on market dynamics. Finally, the European Commission is checking whether Microsoft's investment in OpenAI might be reviewable under the EU Merger Regulation." See https://ec.europa.eu/commission/presscorner/detail/en/ip_24_85.

[28] See the statement by the Competition and Markets Authority (CMA) of December 2023: https://www.gov.uk/government/news/cma-seeks-views-on-microsofts-partnership-with-openai

[29] The agency issued an order to Alphabet, Inc., Amazon.com, Inc., Anthropic PBC, Microsoft Corp., and OpenAI, Inc. to provide information about participations and investment to assess their impact on competition. See the full press release: https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships.

[30] Cavallo 2018 notes that firms' enhanced ability to optimize on prices does not necessarily lead to more discrimination. Instead, algorithmic pricing increases the uniformity of prices across different geographical locations, even though prices become more volatile due to more frequent adjustments.

reduce economy-wide price competition. Finally, as noted by Agrawal, J. Gans, and Goldfarb (2019), deploying AI to pricing strategies may foster algorithmic collusion as described in Ezrachi and Stucke (2020).

Current Regulations on AI and Competition. American and Chinese regulations both mention the prevention of monopolistic practices and unfair competition among their provisions, while the provisional EU AI Act refers to the Union's competition laws. Chinese regulation and the provisional EU AI Act both prohibit uses of AI to manipulate users. The US Executive Order encourages the development of ethical standards for AI, which we may interpret as incorporating behavioral manipulation.

4.1.2 Privacy and copyright issues

Data and Privacy Protection. Like other data-intensive commercial applications, AI technologies demand and process vast amounts of personal information, which might result in violation and misuse of private data. Unlike other applications, machine learning models pose unique challenges to current data and privacy protection laws due to the profiling and inference capabilities of AI and their distinctive security ramifications. Veale, Binns, and Edwards (2018) discuss the case of the EU General Data Protection Regulation (GDPR), which restricts the use and diffusion of personal data, but does not cover models explicitly. The authors believes this to be a key weakness of the current framework, as models are vulnerable to attacks that can expose their architecture and leak—directly or indirectly— sensitive information. Current vulnerabilities of chat bots are particularly severe, as attacks can take the form of simple instruction prompts that induce them to volunteer desired restricted information.[31] A separate rationale for regulation involves user profiling which may lead to indirect privacy violations or insufficient compensation for users' data. Acemoglu (2024) discusses how the price of individual data is driven to zero even when users obtain positive value of privacy. This because algorithms can learn individual characteristics and data by purchasing other correlated data from agents with a lower value of information, effectively acquiring individual information for free.

Privacy concerns feature in all three contexts that we analyze in detail below. In China, the responsibility of AI providers extends to the protection of personal data. The provisional EU AI Act prohibits any use that violates EU values, which include a right to privacy as part of the European Convention of Human Rights. However, specific applications of the type described above are not explicitly covered. Lastly, the US EO calls for the adoption of privacy-preserving AI techniques and for compliance with relevant privacy laws and regulations.

Copyright. Another key issue in the data protection realm pertains to the use of copyrighted material to train models and the intellectual property rights of gen-AI outputs. When it comes to new

---

[31]See Branch et al. (2022)

regulation, only the US AI Executive Order mentions copyright explicitly. Generally, the decision to apply copyright law to AI inputs and outputs is currently up to court decisions and largely subject to case-by-case evaluations. As discussed in the relevant sections, US and Chinese court decisions seem to disagree on whether copyright laws apply to outputs, while many lawsuits in the US will soon see deliberations on alleged violations in the use of inputs.

4.1.3 Strategic issues and military uses

As discussed in Allen and Chan (2017), AI has the potential to revolutionize national security technology, much like nuclear weapons in the past century, once again thanks to its vast information processing capabilities. Key risks reside in code and model manipulations, hacking, and system errors. As national security and military uses have an intrinsic geopolitical component, the authors suggested a wide collaboration in regulation across countries. A first step in this direction is the "Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy," endorsed by 51 states as of January 2024.[32] Another source of risk may be fast adoption without adequate assessment of the technology's dangers. Indeed, as noted by Horowitz (2018), first movers may accrue key advantages, potentially impacting the international balance of power. This strategic dimension may spur hasty adoption of new weapons and rapidly-evolving cyber warfare technologies.

In a January 2024 report, the UK's National Cyber Security Centre (NCSC) described how AI is already used by many cyber threat actors and is poised to increase security risks.[33] The report notes that AI can improve reconnaissance and make social engineering more effective thanks to its content generation capabilities. At the same time, threat actors will be able to analyze exfiltrated data faster and use it to train models, which will likely contribute to several threats including ransomware. In response to this national security concern, the NCSC and several other national cyber security agencies signed a set of "Guidelines for Secure AI System Development" that we describe more in detail in Section 4.4.

In terms of regulations, the US EO on AI promotes guidelines to ensure the safety of critical infrastructure from cyberattacks as well as in the development of biological material to avoid harmful applications. Further, the EO places specific information obligations on developers of models with advanced capabilities that can be used for both civilian and military purposes ("dual-use foundation models"). Chinese regulation takes as a key objective the safeguard of national security. Finally, the debate is currently ongoing in the EU, with contrasting views across member states (Franke 2019). *Exclusive* military and national security applications are explicitly excluded from the scope of the proposed AI Act, even if

---

[32]See https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy-2/ for the text and https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autono for the list of endorsing states.

[33]See https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development.

bundled in AI models with multiple purposes.[34] Hence, some provisions may still apply to dual-use technologies that involve unacceptable or high risk as classified by the Act and discussed in Section 4.3.1. It is worth noting that the EU has partial or complete competence only on certain areas agreed in the EU Treaties, while other areas—including national security aspects—pertain exclusively to member states.[35]

4.1.4 Ethical and Bias Concerns

The evolution of AI has brought about a speedy diffusion of automated algorithms to support or autonomously carry out decisions. This diffusion comes with opportunities for higher productivity, but also with significant risks. Several scientific studies have highlighted that the use of algorithms can lead to discriminatory and uneven distribution of resources across demographic groups, giving rise to so-called "algorithmic bias." This literature shows that discriminatory outcomes may emerge from features of the algorithm such as its objective function or training data (Klare et al. 2012; Obermeyer, Powers, et al. 2019), or reproduce existing societal biases when deployed for wide use (Caliskan, Bryson, and Narayanan 2017; Zack et al. 2023).

Obermeyer, Powers, et al. (2019) show that black patients received less care than white counterparts when decisions were made following a widespread healthcare algorithm. The reason for this bias turned out to reside in the design of the algorithms, which proxied health needs with expected costs from demographic characteristics. The authors also how correcting this feature lead to unbiased outcomes. In related work, Obermeyer, Nissan, et al. (2021) show that this "label choice bias" is a common source of algorithmic bias in several contexts. Another source of bias that is relatively easy to detect may come from biased training data. For example, Klare et al. (2012) unveil performance differences in face recognition algorithms across demographic groups, which stem from lower representation of minorities in the training sample. The authors propose rebalancing the training data as a solution. The forms of bias described so far have clear sources that can be readily corrected and tackled by regulation that sets standards representative, accurate data and reliable algorithms, as in many of the initiatives put forth so far.

Other cases of algorithmic bias pose a bigger challenge to regulators as they arise from the reproduction of societal bias in the training process. Caliskan, Bryson, and Narayanan (2017) report how a language algorithm can learn implicit biases from associations in large language corpora, a characteristic

---

[34] (12a) If and insofar AI systems are placed on the market, put into service, or used with or without modification of such systems for military, defence or national security purposes, those should be excluded from the scope of this Regulation regardless of which type of entity is carrying out those activities, such as whether it is a public or private entity.[...]Nonetheless, if an AI system developed, placed on the market, put into service or used for military, defence or national security purposes is used outside those temporarily or permanently for other purposes (for example, civilian or humanitarian purposes, law enforcement or public security purposes), such a system would fall within the scope of this Regulation.[...]In those cases, the fact that an AI system may fall within the scope of this Regulation should not affect the possibility of entities carrying out national security, defence and military activities, regardless of the type of entity carrying out those activities, to use AI systems for national security, military and defence purposes, the use of which is excluded from the scope of this Regulation.

[35] The full list of areas of EU competency is available here: `https://commission.europa.eu/about-european-commission/what-european-commission-does/law/areas-eu-action_en`

that Zack et al. (2023) find in GPT-4. This problem appears particularly severe in view of the lack of transparency of algorithms themselves, which makes human override and corrections potentially difficult. An added challenge is the identification of cases of algorithmic bias. Lambrecht and Tucker (2018) design a gender-neutral ad for STEM education and instruct a social media advertising algorithm to show it to both men and women, and find that the algorithm optimally chooses to display more frequently to men. They then are able to rule out sources of algorithmic bias, and find that this result arose from profit maximization on the part of the social media platform. As advertising space for women was more expensive, the STEM education ad was crowded out by others products and services that paid higher prices.

Ethical and bias concern feature in all the three regulations that we explored in detail, with democracies generally aligned on the main values and principles. Developing measures to combat algorithmic bias is mentioned explicitly in the US EO, in accordance with its stated aim of advancing equity and civil rights. The provisional EU AI Act stresses that AI should conform with EU values, and this regulation aims to protect fundamental rights, democracy, the rule of law and environmental sustainability. In addition, the Act's ex-ante assessment rules should take into account the ethics guidelines for trustworthy AI of the High-Level Expert Group on Artificial Intelligence (AI HLEG), especially for high-risk AI systems.[36] The Act also qualifies many AI applications where bias would allocate critical resources—such as education, employment, training and law enforcement—in discriminatory manners as high-risk. This designation comes with disclosure and testing requirements prior to commercial deployment, as discussed below. The text of the proposal stresses the control of biased AI-assisted decisions as one of the rationales behind the risk classification.[37] Finally, regulation in China requires effective measures to be put in place to avoid algorithmic bias to emerge. All three regulations also explicitly set ethical requirements for AI systems or suggest the development of guidelines towards this end. Lastly, it is important to point out that many governments signed off the 5 OECD AI principles (see sub section 4.4), among which human-centred values and fairness, and transparency and explainability are included.[38]

4.1.5   AI and financial stability

As in other business applications, AI promises to deliver productivity gains to the banking sector as well. For example, the technology may improve forecasting for risk management purposes, and enhance credit scoring or fraud detection. In a recent IMF paper, Shabsigh and Boukherouaa (2023) survey potential risks from the use of gen-AI in the financial sector. A number of risks are not unique to finance and we have covered above the related reasons for regulation and potential remedies, like privacy

---

[36]See https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[37]In the EU, the Council of Europe AI Convention, is also set to be the first international treaty on AI about legal standards in human rights, democracy and the rule of law.

[38]The concept of "explainability" is particularly important in scenarios where AI could be used in sensitive or critical applications, such as service delivery to the citizens by the government (see selection of social program beneficiaries) to help build trust, enables accountability, and allows users to comprehend and validate the AI system's actions.

concerns, cyber threats and embedded bias. Others are more specific to the financial sector and involve outcome opaqueness, performance robustness, and the potential for creating new sources and transmission channels of systemic risks. Bank of England (2022),OECD (2023a) and Danielsson and Uthemann (2024) list several examples. For instance, AI algorithms may end up adopting similar strategies in different firms, or employ sentiment analysis and social media signals as inputs, amplifying prociclicality and herding behavior. Further, the lack of model transparency may challenge the effectiveness of emergency measures in times of stress, and AI may be used for cyberattacks aimed at destabilizing markets. Other concerns involve the use of AI to summarize data and make decisions. For example, a scarce pool of financial data may be used by AI, providing unreliable financial advice, or AI-generated recommendations may be unethical or illegal. Concerns about AI undermining financial markets integrity were also echoed by IMF's First Deputy Managing Director Gita Gopinath (Gopinath (2023)).

To cope with these issues, Danielsson and Uthemann (2024) assess the suitability of AI use by the private sector and public financial regulations, which they base on the answers to six questions: *"1. Does the AI have enough data? 2. Are the rules of the problem domain immutable? 3. Can AI be given clear objectives? 4. Does the authority the AI works for make decisions on its own? 5. Can we attribute responsibility for misbehaviour and mistakes? 6. Are the consequences of mistakes catastrophic?"* The authors look at specific regulatory tasks—from consumer protection to resolution of banks failures and global systemic crises—under the lenses of these criteria. The case of global systemic crises stands out as a domain where AI usage brings critical risks, calling for regulation on the basis of all six listed criteria.

Among regulatory actions, the US EO mentions the financial system among its provisions, in specific mandating the release of best practices for financial institutions managing cybersecurity risks. In the latest version of the EU AI Act, there is an explicit reference to the related EU legislation on financial services which applies also in this context. The Act requires that the competent authorities (within their competencies) should be designated to supervise financial institutions if their services include the use of AI as per the provisions of the Act. At this stage, there are no other provisions to address other financial stability risks described above.

4.1.6 Other Issues

AI's Impact on Political Systems. With its great classification and content generation capabilities, AI threatens to exacerbate political polarization as well manipulate users to influence electoral outcomes. In terms of political polarization, Acemoglu (2024) shows how AI can fuel the creation of social media "echo chambers," whereby users are classified into groups with uniform political opinions and predominantly receive content that confirms their preconceived opinions. Similarly, the ability of AI to easily produce "deep fakes" could easily lead to a proliferation of fake news, with the range of adverse impacts documented in Allcott and Gentzkow (2017). AI misinformation and disinformation recently featured

as the second-highest risk for 2024 in the World Economic Forum's Global Risks 2024. Report.[39]

A potential way to mitigate these impacts consists in mandating watermarking to authenticate genuine content or to label AI-generated content. Guidelines for watermarking are currently being formulated by the US Department of Commerce as part of the provision of the US EO on AI. At the same time, Chinese regulation makes AI service providers directly responsible for the content they diffuse and forbids the use of AI to manipulate information or public opinion.

In addition to negative consequences, AI may improve the functioning of electoral systems and collective decision-making. For instance, Landemore (2023) describes how AI can help deliberating processes and increase diversity and inclusiveness. Other uses, like AI-powered ballot scanners and voter roll monitoring, targeted political advertising and AI-generated summaries of candidates political position, may help electoral processes (Eisen et al. (2023)).

4.2 Theory of regulation

The choice and the setup of a regulation of AI also come with a literature, even if still rather limited. In this section, we review some contributions exploring the welfare consequences of AI regulation, which aim to determine the optimal trade-off between opportunity and–potentially existential–risks from progress in this technology.

Acemoglu (2024) stands out as one of the first papers to provide an organized summary of the risks from unregulated AI, as have been discussed in the previous Sub-Section 4.1. A key claim of the paper is that potential economic and societal risks stem from the *current use and development* of the technology rather than its nature, and downsides could be avoided with appropriate regulation and policies that redirect the course of AI development towards the most productive uses.

Optimal regulation of transformative technologies is at the center of Acemoglu and Lensman (2023) study. The authors build a multi-sector model where risks are revealed *over time* as technology is adopted across sectors. As adoption proceeds, agents in the economy learn about the probability of a "disaster" which incurs losses proportional to the output produced using the new technology. The benefits of the new technology consists in higher productivity as well as higher productivity *growth*. Given this main trade-off, optimal adoption of the technology is gradual and increases its speed over time. It is gradual because this allows learning about the risks of the technology, and its speed increases over time as the disaster probability is assessed to be low enough. A role for taxation and regulation emerges if firms do not fully internalize the potential harms of their technology, which the authors justify with some damages being inherently social in their nature. In this case, adoption is too fast relative to optimal, and the gap between effective and optimal adoption increases with the speed of output growth delivered by the new technology. The more transformative the technology, the larger are potential harms from it, due to the

---

[39]See https://www.weforum.org/publications/global-risks-report-2024/in-full/global-risks-2024-at-a-turning-point.

fact that losses are taken to be a fraction of total output produced with new technologies. Acemoglu and Lensman (2023) propose two potential tools to close this gap: sector-specific taxes that lower the rate of adoption across all sector, or, as a second best, a "regulatory sandbox" that fully restricts the use of new technologies in sectors with high potential losses until a specific time period where risks should be better known.

Guerreiro, Rebelo, and Teles (2023) assess different approaches to AI regulations in a model featuring AI developers that choose the novelty of their algorithm as well as the size of the population to which this algorithm is deployed. Higher novelty delivers higher utility, but also comes with higher variance of damages realized ex-post. The authors consider two alternative setting: full deployment or beta testing targeted at gauging the damages. In the first setting, developers cannot learn about damages until full deployment takes place, while in the second they can test the algorithm on a fraction of the population and measure realized damages before deliberating on full implementation. In both cases, the planner prefers a lower level of novelty than developers since damages manifest as an externality that is only partly internalized by developers, much like in the Acemoglu and Lensman (2023) setting. The authors consider the combination of two policy solutions commonly proposed to bring the economy to the social optimum with or without beta-testing: (i) regulatory approval of release and (ii) developers' liability for realized damages. Without beta testing, it is optimal to set a threshold of novelty equal to the social optimum and forbid algorithms above such thresholds, a policy that the authors compare to the EU Commission's proposal discussed below. Liability for developers delivers the social optimum in this context only if it is full. When beta-testing is available, the optimal policy consists in mandating beta-testing for all levels of novelty and imposing limited liability for damages in case of full deployment. The authors also show that a commonly proposed solution–conditional approval of algorithms that beta-testing reveals to be sufficiently safe–is not optimal since it involves excessive levels of novelty in the testing phase. The authors highlight that this solution might require international cooperation if spillovers are not limited to single countries. In contrast to this approach, Kretschmer et al. (2023) propose a full liability approach to regulation, as discussed more at length in Section 4.3.1.

Without addressing specific policies, Jones (2023) considers when arresting AI development is optimal in the presence of existential risk and different degrees of societal risk aversion. This simple model weighs additional growth against human lives lost, which are modelled as an increased growth rate and a higher death probability, respectively. The planner chooses how long to use AI, realizing growth at the cost of lost lives. AI usage increases with lower risk-aversion, or in an alternative scenario where AI can also deliver increases in life expectancy. A final consideration involves the discount rate attached to future utility. A lower discount rate increases AI if the technology delivers better life expectancy, and lowers its use otherwise.

4.3 Regulation and Governance of AI across Countries

Following Wheeler (2023), we can identify three main challenges for AI regulation: speed, scope, and approach. These challenges are also at the root of major differences in adopted and proposed country-specific frameworks. First, AI's rapid transformation makes rules and standard-setting particularly difficult, as the landscape is rapidly evolving.[40] The speed of AI development motivates some authorities to adopt flexible guidelines that are easier to adapt (e.,g., Japan and UK), while others prefer more rigid regulations to contain risks that might manifest rapidly (e.g., the EU). Second, scope differs widely across regulators, starting from the definition of AI itself and extending to the different areas covered.[41] The Appendix A.1 lists the various definitions contained in EU, US, UK and Chinese source documents.[42] Regulators have so far adopted different strategies ranging from outright temporary bans to voluntary guidelines. The cases of binding regulations have adopted a range of approaches combining ex-ante risk assessments and ex-post liability. The ex-ante assessment approach involves provisions to evaluate the potential risks of AI based on systems' characteristics or criticality areas of application *before* risks manifest. Ex-post liability approaches instead make developers and service providers liable for *actual realizations* of risk events and related externalities. Further, several initiatives have stressed the importance of multilateral organizations, ranging from the EU AI governance architecture (see sub-section 4.3.1) to proposal to establish an authority analogous to the UN's Intergovernmental Panel on Climate Change (IPCC).[43]

Before delving deeper into specific cases, some examples may clarify how scope and approaches vary across the world. In terms of scope, measures proposed and implemented in the EU and US concern broad technologies and their applications. By contrast, China regulates algorithms and their content directly. In terms of approaches, the US decentralized the setting of policies for specific AI applications to relevant departments and agencies, while the EU stands poised to adopt a centralized approach featuring ex-ante risk assessments and a new AI governance structure. Among other examples, Italy stood out for implementing a blanket ban on ChatGPT until adequate data protection issues were clarified at the end of April 2023, while Japan is leaning towards voluntary guidelines.[44]

In the rest of the section, we describe the main regulations developing in selected AEs and EMDEs, including EU, the US, the UK, China, Brazil, and India, with an eye to the current debates and their ramifications. We review the areas that have been covered by each and the regulatory approach followed

---

[40]Due to these rapid developments, the OECD maintains a continuously updated definition of "AI system". See the latest changes at: https://oecd.ai/en/wonk/ai-system-definition-update

[41]See O'Shaughnessy (2022), for Carnegie Endowment for International Peace, and the UK White Paper on AI in paragraph "3.2.1 Defining Artificial Intelligence".

[42]The latest version of the EU AI Act states that *"(12) The notion of 'AI system' in this Regulation should be clearly defined and should be closely aligned with the work of international organisations working on AI to ensure legal certainty, facilitate international convergence and wide acceptance, while providing the flexibility to accommodate the rapid technological developments in this field. [...]"*

[43]https://www.economist.com/business/2023/10/24/the-world-wants-to-regulate-ai-but-does-not-quite-know-how

[44]https://asia.nikkei.com/Business/Technology/Japan-seeks-AI-transparency-with-new-disclosure-guidelines, https://www.japantimes.co.jp/news/2023/11/24/japan/politics/generative-ai-guidelines-no-penalties/, https://www.reuters.com/technology/japan-leaning-toward-softer-ai-rules-than-eu-source-2023-07-03/

by relevant authorities or proposals. None of the countries we consider have focused directly on financial stability, while regulations primarily addressed monopoly, ethical, and privacy risks.[45] US regulation covers explicitly national security and military uses of AI, while the EU has been more ambiguous, as legislation on this matter is decided at the national level. Copyright of AI-generated (or co-generated) material is not explicitly tackled in current frameworks, and in some cases deferred to the discretion of the courts on a case-by-case basis (e.g., in the US and China).

4.3.1 The EU AI Act

The EU AI Act Approval and Implementation. The EU was one of the first movers in AI regulation, with the EU AI Act proposed by the EU Commission in 2021 and set to enter into force by the end of spring 2024.[46] A provisional political agreement was reached on December 8, 2023 in the "Trilogue Agreement" between the European Parliament, the Council of the European Union and the European Commission. On February 2, 2024, the Committee of Permanent Representatives to the European Union ("COREPER") approved the text, which was endorsed on February 13 by the Committee on Internal Market and Consumer Protection (IMCO) and the Committee on Civil Liberties, Justice and Home Affairs (LIBE) of the EU Parliament.[47] The European Parliament approved the Act on March 13, 2024 in a plenary session.[48] We will refer to this version as the "latest text" of the Act or Provisional Act.[49] This is because the *final* approved version of the Act may differ due to the introduction of official rectifications—so-called "corrigenda"—by lawyer-linguists to address translation issues. This edited version will be approved by the European Parliament in the plenary session of April 10-11, 2024, without the need for a vote, which will be formally approved by the Council of the EU in its following session (likely without a vote as well). After the Council of the EU's approval, it will enter into force 20 days after its publication in the official Journal, i.e., expected to enter into force by the end of spring 2024. Following this procedure, it will be fully applicable 24 months after its entry into force, with a gradual implementation and exceptions for specific provisions.[50] In the meantime, on March 8, 2024, the EU Commission established the AI Office as part of their new AI governance as shaped by the

---

[45]We believe that financial stability may have been excluded as macroprudential policies are often the responsibility of independent authorities. For example, in the euro area, this topic falls within the competences of the European Central Bank and other central banks in the European System of Central Banks.

[46]A summary of developments can be found at the website `https://artificialintelligenceact.eu/developments/`, maintained by Future of Life Institute.

[47]A summary is available at:`https://www.europarl.europa.eu/news/en/press-room/20240212IPR17618/artificial-intelligence-act-committees-confirm-landmark-agreement`

[48]The Press Release by the EU Parliament is available at this link:`https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law`

[49]The text of the Provisional Act is available at this link: `https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html`.

[50]In particular, bans on prohibited practices will apply 6 months after the entry into force; codes of practice will apply 9 months after entry into force and general-purpose AI rules– including governance– and provisions on penalties 12 months after entry into force. Moreover, *"(172) [...] While the full effect of those prohibitions follows with the establishment of the governance and enforcement of this Regulation, anticipating the application of the prohibitions is important to take account of unacceptable risks and to have an effect on other procedures, such as in civil law."*

Act.[51]

The EU AI Act: main provisions in the latest text. The main provisions of the regulation are grounded in an ex-ante risk assessment of AI systems, which underpins the regulatory burden imposed on each system. The document supplements this general approach with stricter rules for "general-purposes AI" models and systems–especially those classified as "high risk" or systemic.[52] Finally, the Act establishes a new AI governance.

The Act's provisions apply to all providers (developers) of AI systems to the EU, irrespective of their country of establishment, and to "deployers" (users) of AI systems within the EU.[53] The provisions in the Act should support the view of AI as a "human-centric technology" and in line with EU values. Some provisions are also included to be of support of innovation and SMEs, even if this is not the main scope of the legislation.

The risk-based approach. The provisional EU AI Act adopts a risk-based approach, which consists in classifying AI systems according to an *ex-ante* assessment of their potential risk.[54] The risks stemming from AI systems and the associated provisions are as follows:

- *Unacceptable risk* arises from AI *systems* that contravene EU values, by posing a clear threat to individuals' safety, livelihoods, or otherwise violating fundamental rights. These AI systems are *prohibited.* They include subliminal manipulation resulting in harm, exploitation of vulnerable categories, general purpose social scoring, and remote biometric identification systems (RBIs) with some notable exceptions. The use of RBIs by law enforcement is prohibited, except in exhaustively listed and narrowly defined situations.[55]

- *High risk* arises from explicitly listed AI *systems* that create a *(52) [...]"high-risk if, in the light of their intended purpose, they pose a high risk of harm to the health and safety or the fundamental*

---

[51] In an official statement, it was clarified that the AI Office *"[...] should exercise its tasks, in particular to issue guidance, in a way that does not duplicate activities of relevant bodies, offices and agencies of the Union under sector specific legislation. [...] It is without prejudice to the functions of other Commission departments in their respective areas of responsibility, and of the European External Action Service in the area of Common, Foreign and Security policy."* https://digital-strategy.ec.europa.eu/en/library/commission-decision-establishing-european-ai-office-0

[52] Previous versions of the Act characterized so called "foundation" models as "high risk."

[53] See Article 3 "Definitions": *[...](3) 'provider' means a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge;(4) 'deployer' means a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity;[...]"* and paragraph *"(21) In order to ensure a level playing field and an effective protection of rights and freedoms of individuals across the Union, the rules established by this Regulation should apply to providers of AI systems in a non-discriminatory manner, irrespective of whether they are established within the Union or in a third country, and to deployers of AI systems established within the Union."*

[54] In previous versions of the Act, the classification involved AI "applications," rather than "systems."

[55] As reported in the press release after the vote at the EU Parliament's Committee in February 2024: *"Real-time" RBI can be deployed only under strict safeguards, e.g. limited in time and geographic scope, with prior judicial or administrative authorisation. Such uses involve, for example, searching for a missing person or preventing a terrorist attack. Using such systems after the fact ("post-remote RBI"), which is considered high-risk, also requires judicial authorisation, and has to be linked to a criminal offence.* See https://www.europarl.europa.eu/news/en/press-room/20240212IPR17618/artificial-intelligence-act-committees-confirm-landmark-agreement

rights of persons." For instance if the systems are used to determine access to education and training, or for job recruitment or to evaluate access to public benefits or resources. [56]

The Act requires high-risk systems to maintain a "risk management system," that is, a process of identification, estimation and evaluation of risks accompanied by appropriate risk management structures. Additionally, these systems should: undergo testing during development and before deployment to the market; satisfy specific data standards in their training sets such as representativeness; keep exhaustive documentation and records of risks; provide information to users on, e.g., their characteristics, limitations and accuracy. They should be also designed to ensure that human supervision is possible and achieve appropriate levels of accuracy, robustness and cybersecurity. Systems in this category must submit to a conformity assessment prior to release, to ensure that requirements are met following specific obligations listed in the Act. Lastly, before placing systems on the market or putting them into service, the provider shall register themselves and their system in a dedicated EU database.

- *Specific transparency risk* arises in AI systems that interact with users and may be subject to manipulation, like chatbots. In this case, specific transparency obligations should be in place to ensure that users are aware that they are interacting with a machine.

- *Minimal risk* arises for all other AI systems. Examples in this category are AI-enabled video games or spam filters. This type of systems is not subject to additional provisions.[57]

General-purpose AI. The provisional EU AI Act defines general-purpose AI as models that display significant generality and are capable to competently perform a wide range of distinct tasks, and that can be integrated into a variety of downstream systems or applications.[58] This definition applies regardless of the way the model is placed on the market.[59] A specific subset of general-purpose AI models is represented by "high-impact" models, defined as those trained on sufficiently large data sets and featuring particularly high complexity, capabilities and performance (high-impact). All general-purpose AI models must: a) provide technical documentation, including training and testing process and evaluation results; b) supply information to downstream providers wishing to integrate the general-purpose AI model into

---

[56]High-risk AI systems include those related to: 1) biometrics, insofar as their use is permitted under relevant Union or national law (RBIs escluding some cases such as biometric verification the sole purpose of which is to confirm that a specific natural person is the person he or she claims to be); 2) critical infrastructure (e.g., utilities); 3) education and vocational training (e.g., evaluate learning outcomes); 4) employment, workers management and access to self-employment (e.g., for recruitment, job posting,...); 5) access to and enjoyment of essential private services and essential public services and benefits (e.g., for triage in emergency healthcare, establishing credit score, risk assessment for insurance,..); 6) Law enforcement, insofar as their use is permitted under relevant Union or national law; 7) migration and border control management, insofar as their use is permitted under relevant Union or national law; 8) administration of justice and democratic processes (these includes AI systems intended to be used for influencing the outcome of an election or referendum). These high risk systems are listed in Annex III of the provisional EU AI Act, referring to Article 6(2).

[57]See "Artificial Intelligence – Questions and Answers" by the EU Commission, from December 2023: `https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683`

[58]A general-purpose AI "system" is an AI system based on a general-purpose AI model that has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems.

[59]We report definitions of general-purpose AI verbatim in Appendix A.1.

their own AI systems; c) comply with the Copyright Directive; and d) publish a sufficiently detailed summary about the content used for training the general-purpose AI model. Free and open-license general-purpose AI models only need to comply with the latter two obligations.[60] The provisional EU AI Act separately tackles *"systemic risk"* general-purpose AI systems, defined as those that can have an impact on the EU internal market, or impinge on fundamental rights. Systemic risk general-purpose AI are subject to additional requirements such as: mitigating possible systemic risks; document and report serious incidents; and ensure cybersecurity protection.[61]

Penalties. Since the Trilogue agreement, the provisional EU AI Act imposes increased penalties for non-compliance relative to the original proposal. The penalties are as follows: 35 million euro or 7% of worldwide annual turnover for the preceding financial year (whichever is higher) for banned AI systems; 15 million euro or 3% of worldwide annual turnover for a failure to notify unacceptable applications; and 7.5 million or 1.5% of turnover for less severe cases (e.g., incorrect information).[62]

Support to innovation. The EU AI act in the last version includes measures to support innovation and SMEs, such as real-world testing, to develop and train innovative AI before placement on the market. In particular, the provisions above do not apply to AI systems and models with *sole* purpose of scientific research as well as to models that are not yet deployed to the market.

The new AI Governance. The agreement introduces a new governance model for AI in the EU, comprised of an AI Office and an AI Board.

The AI Office has been established within the Commission and entered into force in March 2024. Among other goals, the AI Office is tasked with: developing tools, methodologies and benchmarks to evaluate general-purpose AI models' capabilities, with a particular focus on models with systemic risks; monitoring the implementation and application of rules on general-purpose AI and coherent application of the Act; monitoring the emergence of unforeseen risks; investigating possible infringements; ensuring coordination with the Digital Services Act and Digital Markets Act; supporting information exchange and collaboration with national authorities; assisting the EU Commission in setting standards; and fostering international cooperation. A Scientific Panel of independent experts also supports the office to designate and assess general-purpose AI models–including high-impact and systemic cases.[63] Further technical expertise is gathered in an Advisory Forum, representing a balanced selection of stakeholders, including industry, startups and SMEs, academia, think tanks and civil society.[64]

---

[60]More details are in Article 53.

[61]See Article 71 of the provisional EU AI Act for more information.

[62]See Article 99 of the provisional EU AI Act for more details on penalties.

[63]See for more information the decision by the EU Commission on the roles and establishment of the AI Office: `https://ec.europa.eu/newsroom/dae/redirection/document/101625`

[64]For a summary of goals and future of the AI Office see `https://digital-strategy.ec.europa.eu/en/policies/ai-office`. The AI Office *"[...] will enforce the rules for general-purpose AI models. This is underpinned by the powers given to the Commission by the AI Act, including the ability to conduct evaluations of general-purpose AI models, request information and measures from model providers, and apply sanctions. The AI Office also promotes an innovative ecosystem of trustworthy AI, to reap the societal and economic benefits. It will ensure a strategic, coherent and effective*

The AI Board, as established in Article 65, will gather Member States' representatives (one for each Member), serving as a coordination platform and an advisory body to the Commission. An Advisory Forum for stakeholders, such as industry representatives, SMEs, start-ups, civil society, and academia should also be set up to support the AI Board on technical matters.

Academic Discussion on Aspects of the EU AI Act. The Stanford Center for Research on Foundation Models recently released an overview of possible approaches for categorization to categorize AI systems for the purpose of regulating "foundation" models (Bommasani 2023). While recognizing support for the risk categorization in the Act, the article points out the risk from erroneous categorizations, and stresses the importance of categorizing uses of foundation models. At the same time, the ex-ante risk assessment approach typical of the EU AI Act is also a subject of contention. For instance, Kretschmer et al. (2023) considers ex-post liability rules superior. In the view of these authors, liability rules provide the right incentives to improve data quality and AI safety, while encouraging ongoing innovation. They propose a regulatory framework that differentiates between endogenous and exogenous sources of potential harm, which would help to carefully allocate liability between developers of AI technology and providers of related services. By contrast, the ex-ante system puts a high burden on developers only, giving an advantage to larger firms who can pay the legal consequences, and thus potentially limiting technological improvements and innovation.[65] While not at the core of the current AI strategy, the EU is considering liability rules to integrate the AI Act.[66]

4.3.2 US Regulation

On October 30th 2023, US President Biden signed an "Executive Order (EO) on Safe, Secure, and Trustworthy Artificial Intelligence".[67] This EO builds on previous guidance documents such as the AI "Bill of Rights" and the National Institute for Standards in Technology (NIST) AI Risk Management Framework.[68] [69] The order defines AI broadly, which makes its scope wider than gen-AI (See Appendix A.1). The EO lays out eight principles and priorities for the use of AI: 1) AI must be safe and secure; 2) the US should promote responsible innovation, competition and collaboration in AI development; 3) the responsible use and development of AI must come with a commitment to supporting American workers; 4) AI policies must be consistent with the Biden Administration's dedication to improving equity and civil rights; 5) interests of Americans who increasingly use, interact with, or purchase AI and AI-enabled products in their daily lives must be protected; 6) Americans' privacy and civil liberties must

---

European approach on AI at the international level, becoming a global reference point."

[65]See also https://cepr.org/voxeu/columns/global-ai-regulation-race-why-eu-should-focus-data-quality-and-liability-rules

[66]See, e.g.,https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf

[67]The full text is available at: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

[68]The "Blueprint for an AI Bill of Rights," available at https://www.whitehouse.gov/ostp/ai-bill-of-rights/ establishes five principles to guide the development and use of AI: (i) safe and effective systems; (ii) algorithmic discrimination protections; (iii) data privacy; (iv) notices and explanation; and (v) human alternatives, consideration and fallback

[69]The NIST AI Risk Management Framework is available at: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

be protected; 7) the risks from the Federal Government's own use of AI must be adequately managed and the government's capacity to ; 8) the US Federal Government should lead the way and engage with international partners to develop a framework to manage AI risks and unlock potentials.

The EO takes a more decentralized approach compared to EU initiatives, delegating to Agencies or Departments the release of guidelines on specific principles and priorities.[70] Several of the order's provisions task relevant authorities with conducting risk assessments, drafting guidelines and exploring potential legislative action.

Several provisions aim to increase safety in the use of AI. A special focus is given to cyberattacks—especially by foreign entities and when involving large AI models—and the management of critical infrastructure. Notably, critical infrastructure includes the financial system, for which the EO mandates the release of best practices for financial institutions managing cybersecurity risks. Further, the EO tasks competent authorities with setting standards for AI use in the development of biological materials to avoid harmful applications (such as biological weapons). In order to reduce the risks from synthetic content, the order directs the Department of Commerce to produce guidelines for watermarking and clearly identifying AI-produced content.

Further measures direct competent bodies to estimate the technology's potential impacts, to attract AI talent and foster innovation and cross-institutional cooperation. Section 5 is devoted to promoting competition in AI and related technologies, including addressing risks from market concentration of key inputs and consumer protection. Among the measures to promote innovation, the EO instructs the US Copyright Office to issue recommendations on copyright issues including both protection of AI-generated work and the treatment of copyrighted works in training. The order stands out by tackling direct economic impacts that are explored by the academic literature, but that remain largely absent in other regulatory efforts. In particular, Section 6 requires the Council of Economic Advisers to assess the potential labor-market impact of AI. Further, the Secretary of Labor should identify options to support workers potentially displaced by AI, and release principles and best practices for employers to ensure that AI is deployed to advance employees' well-being.

A special mention is given to companies developing large AI models with billions of parameters and a wide range of capabilities ("dual-use foundation models").[71] These companies should notify the Federal Government and share results of safety tests based on "Guidelines, Standards, and Best Practices for AI Safety and Security" to be established within 270 days of the EO. Open-source AI models are also included in these provisions as "dual-use foundation models with widely available weights" and the related risks should be investigated by the Secretary of Commerce.[72]

---

[70]Timelines and responsibilities of the various Agencies and Departments can be found in the EO and are conveniently summarized in the Congressional Research Service's report of November 17th, 2023:https://crsreports.congress.gov/product/pdf/R/R47843

[71]See Appendix A.1 for the full definition of dual-use foundation models.

[72]*4.6. Soliciting Input on Dual-Use Foundation Models with Widely Available Model Weights. When the weights for a dual-use foundation model are widely available — such as when they are publicly posted on the Internet — there can*

Completed actions based on the EO. On January 29, 2024, the White House published a summary of the action items completed by Federal Agencies for which the EO had set 90-day deadlines.[73] These included provisions for developers of the most powerful AI systems to report vital information– like safety test results–to the Department of Commerce, and for US cloud providers to report usage of computing power connected to foreign AI training. The announcement also reported the launch of a pilot of the National AI Research Resource. This initiative, supported by the National Science Foundation (NSF), aims to give access to *"computing power, data, software, [...] open and proprietary AI models, and other AI training resources"* to researchers and students. The 90 days since the EO also saw several other initiatives to support AI education and innovation, as well as the inauguration of a task force in the Department of Health and Human Services. This task force will develop guidelines to address algorithmic bias in healthcare and for the safe use of AI in medical innovation.

Discussions on AI regulation in the US. Shortly after the signing of the EO, a panel of Brookings' fellows commented on it provisions.[74] The experts generally praised the reach and comprehensiveness of the order as well as its treatment of specific aspects, such as privacy and the mobilization of expertise to reach a deeper understanding of potential AI issues. At the same time, commentaries noted the lack of an assessment of AI's impact on climate, of an extensive treatment of financial stability and regulation outside of cyberattacks, and of clear enforcement mechanisms, especially at an international level. Many of the experts noted that the provisions mainly directs competent authorities to issue standards and guidelines, which would not be enforced in the absence of further legislative action. In this sense, several scholars in the Brookings panel highlighted that the most consequential aspect of the act is invoking the Defense Production Act to mandate disclosure and reports by companies developing foundation models.

Other initiatives. In early 2023, the US Copyright Office launched an initiative to examine issues related to copyright. More recently, a D.C. District Court Judge ruled that human beings are an "essential part of a valid copyright claim", which excludes gen-AI from copyright law protection.[75] Several class actions and lawsuits are ongoing centering on the use of copyrighted material in training data and more rulings are expected in 2024. One example is the New York Times' complaint filed on December 27, 2023 against OpenAI and Microsoft.[76] December 2023 also saw the first deal between a publishing house and an AI company to provide real-time breaking news information. Under this deal, signed by OpenAI

---

be substantial benefits to innovation, but also substantial security risks, such as the removal of safeguards within the model. To address the risks and potential benefits of dual-use foundation models with widely available weights, within 270 days of the date of this order, the Secretary of Commerce, acting through the Assistant Secretary of Commerce for Communications and Information, and in consultation with the Secretary of State, shall: [...] start consultations on risks, and start potential mechanisms to limit risks.

[73] The statement is available at this link: https://www.whitehouse.gov/briefing-room/statements-releases/2024/01/29/fact-sheet-biden-harris-administration-announces-key-ai-actions-following-president-bidens-landmark-executive-order/

[74] https://www.brookings.edu/articles/will-the-white-house-ai-executive-order-deliver-on-its-promises/

[75] See Civil Action No.22-1564, Judge Beryl A. Howell's memorandum opinion, available at this link: https://www.documentcloud.org/documents/23919666-thalervperlmutter?responsive=1&title=1

[76] See the full text at: https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf, and a related New York Times' press coverage : https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html.

with Axel Springer, from the first quarter of 2024 the users will be able to receive summaries of stories just published by several brands owned by the publisher as responses to ChatGPT queries. Responses will include links and attribution.[77]

On July 21, 2023, the Biden administration reached a voluntary agreement with seven leading AI companies – Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI. The companies pledged to conduct extensive security testing, sharing information related to security vulnerabilities, develop technical mechanisms to denote when content is AI generated, and prioritize research on societal risks posed by AI—e.g., to avoid bias and protect privacy.

Nine Bipartisan Senate Fora on Artificial Intelligence, chaired by US Senate Majority Leader Chuck Schumer, took place from September to December 2023 to build bipartisan consensus on future legislative initiatives.[78] The themes were varied and included, among others: how to fund development and where to channel resources; how to attract talent and which safeguards should be adopted to protect workers; the role of AI in the financial and health sectors and related risks pertaining to potential bias; the technology's impact on elections and democracy; privacy and liability issues; copyright and algorithm transparency; guardrails in place to prevent or slow Artificial General Intelligence development; leveraging competition in the open source community; international collaboration and competition; and fostering AI innovation for military and security usage.

4.3.3 Regulation in China

China has implemented interim measures to address AI-related risks and impose compliance obligations on entities engaged in AI-related business. In their latest version, provisions only apply to *public usages* of gen-AI by both domestic and foreign individuals and entities involved in AI services in China.

A version of the interim measures was released for comments in April 2023, followed by a substantially revised version dated July 2023 and taking effect on August 2023.[79] In particular, the scope of regulation was narrowed from all uses of gen-AI to *only public usages*, with other cases being less strictly regulated. Similar to the EU AI Act, the measures do not apply to gen-AI technologies used for research purposes

---

[77]See the press releases by OpenAI (https://openai.com/blog/axel-springer-partnership) and Springer (https://www.axelspringer.com/en/ax-press-release/axel-springer-and-openai-partner-to-deepen-beneficial-use-of-ai-in-journalism). See also Reuters press coverage: https://www.reuters.com/business/media-telecom/global-news-publisher-axel-springer-partners-with-openai-landmark-deal-2023-12-13/

[78]Statements are available as part of Senator Schumer's press releases. For example, for the ninth forum: https://www.schumer.senate.gov/newsroom/press-releases/statements-from-the-ninth-bipartisan-senate-forum-on-artificial-intelligence

[79]The official version of the documents is only available in Chinese and for the information and translation to English we followed the US Library of Congress:https://www.loc.gov/item/global-legal-monitor/2023-07-18/china-generative-ai-measures-finalized/. The translation of the April 2023 version is by the Stanford University's DigiChina Project: https://digichina.stanford.edu/work/translation-measures-for-the-management-of-generative-artificial-intelligence-services-draft-for-comment-april-2023/. However, the translation of the July 2023 version cited by the US Library of Congress is not freely available. We had to revert to an alternative source, as per this link: https://www.chinalawtranslate.com/en/generative-ai-interim/. See also press coverage at https://time.com/6314790/china-ai-regulation-us/.

and not deployed to the market.[80] Encouragement for gen-AI development was also added.[81,82,83]

Key points of the July 2023 "Interim Measures for the Management of Generative Artificial Intelligence Services" are:

- Gen-AI services providers with public opinion properties or with the capacity for social mobilization shall carry out security assessments in accordance with relevant state provisions;

- Obligors must comply with laws, social morality, and ethics, and avoid manipulating information or public opinion. During algorithm design, selection of data, model generation and training, and provision of services, effective measures should be employed to prevent biases that result in discrimination. Providers shall fulfill confidentiality obligations towards information input by users and users' usage records in accordance with existing laws.

- Content is prohibited if it is against "Core Socialist Values" or if it otherwise endangers national security. Providers are responsible for the legality of content generated and diffused, and for prompt interruption of services and rectification of unlawful content.

- Intellectual property rights should be protected and advantages in algorithms, data, platform and the like must not be used for monopolies or to carry out unfair competition.

- Penalties for non-compliance include warnings, ordering rectifications and corrections, suspension of services, as well as civil and criminal prosecution when relevant.

Other initiatives. In August 2023, The country's internet regulator announced restrictions on the use of facial recognition technology when suitable alternatives are available for identity verification.[84] In late November 2023, the Beijing Court issued a ruling on a copyright case over the use of AI-generated images. The decision established that the images met the requirements of "originality" and reflected a human's original intellectual investment, putting them under the protection of copyright law as works of art .[85]

In terms of international cooperation, the Chinese authorities unveiled a "Global AI Governance Initiative" in October 2023. The initiative calls for international collaboration to develop AI that respects national sovereignty, promotes sustainable development, ensures security, and upholds ethical standards.

---

[80] *"Article 2. [...] These Measures do not apply where industry associations, enterprises, education and research institutions, public cultural bodies, and related professional bodies, etc., research, develop, and use generative AI technology, but have not provided generative AI services to the public."*

[81] *"Article 5: Encourage the innovative application of generative AI technology in each industry and field, generate exceptional content that is positive, healthy, and uplifting, and explore the optimization of usage scenarios in building an application ecosystem.[...]"*

[82] A clear comparison between the version for comments of April 2023 and the "Interim Measures for the Management of Generative Artificial Intelligence Services" version effective in August 2023, can be retrieved at `https://www.chinalawtranslate.com/en/comparison-chart-of-current-vs-draft-rules-for-generative-ai/`. This comparison is based on unofficial translations from Chinese to English.

[83] `https://www.loc.gov/item/global-legal-monitor/2023-07-18/china-generative-ai-measures-finalized/`

[84] `https://www.wsj.com/articles/china-drafts-rules-for-facial-recognition-use-4953506e?mod=world_lead_pos3`

[85] `https://www.natlawreview.com/article/beijing-internet-court-recognizes-copyright-ai-generated-images`

In the connected declaration, authorities also advocated for increased representation of developing countries in global AI governance.[86]

4.3.4 A summary of current regulations in EU, US, and China

We summarize the main characteristics of AI-specific regulations that are in place or under discussion in Table 1. From left to right, the table lists the sources of these regulatory actions, the approaches taken by the relevant authorities, the type of AI regulated according to the definitions reported in Appendix Section A.1 , and the areas subject to regulation defined as in our discussion in Section 4.1. Appendix Section B provides more information on each aspect.

Table 1: **AI-Specific** Regulations in selected AEs and China

Country	AI regulation	Approach	Coverage	Competition	Privacy	Copyright	Military	Ethical	Fin. Stability
EU	Provisional EU AI Act adopted by the EU Parliament (March 13, 2024)	Based on ex-ante risk assessment	AI applications, 'general purpose' AI models (incl. high risk and systemic) also integrated into AI systems.	GN	✓	✓ /GN	X	✓	✓ /GN
US	Executive Order (EO) on Safe, Secure, and Trustworthy Artificial Intelligence (October 2023)	Decentralized based on guidelines, principles and priorities	AI and AI models, incl. "dual use" foundation models	✓	✓	✓	✓	✓	GN (but incl. cybersecurity)
China	Interim Measures for the Management of Generative Artificial Intelligence Services (August 2023)	Ethical model design and content, ex-post liability	Only on public usages of gen-AI	✓	✓	GN	X/GN (but incl. National security)	✓	GN

Note: This table covers AI-specific regulations, while we signal with "GN" (as per "general") when other regulations/laws include these issues in a general context. We do not consider national laws for the EU nor State-specific regulations for the US. The EU AI Act is not officially an EU law (as of March 13, 2024). The Provisional EU AI Act is as approved by COREPER, by EU Parliament's Committees, and in plenary session by the EU Parliament (March 13, 2024). For China, the source is an *unofficial* English translation of the Interim Measures document. For the US, some aspects are also covered in the "Blueprint for an AI Bill of Rights" (October 2022) by the White House Office of Science and Technology Policy (OSTP) aimed to set a roadmap for the responsible use of AI especially on potential human rights.

The "'general purpose' AI model" is defined in the Provisional EU AI Act as *"(63) AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are released on the market; "[...] "(66) 'general-purpose AI system' means an AI system which is based on a general-purpose AI model, that has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems;[...]"*
The "dual-use model" in the US EO is defined as *"AI model that is trained on broad data; generally, uses self-supervision; contains at least tens of billions of parameters; is applicable across a wide range of contexts; and that exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters [...]"*

4.3.5 Regulation in Other AEs: The Case of the UK

In the UK, several authorities are involved in the regulation of AI, and the regulatory approach is characterized by guidelines, best practices and principles. Examples of key initiatives include the Ministry of Defence's "Defence Artificial Intelligence Strategy", UK's Information Commissioner's Office's "Guidance on AI and Data Protection," and the UK's NCSC's "Guidelines for Secure AI System Development." We provide additional details in Appendix B.4.

In March 2023, the UK Government issued a white paper for "A Pro-Innovation Approach to AI Regulation".[87] This paper stresses a regulatory attitude based on specific areas of application of AI

---

[86]https://www.mfa.gov.cn/eng/wjdt_665385/2649_665393/202310/t20231020_11164834.html
[87]The full text of the white paper is available at: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

rather than specific technologies. It suggests that potential future regulation shall focus on principles such as safety, transparency, fairness, accountability, and contestability. The paper excludes immediate regulatory action, delineating instead plans to invest in AI research and development, and to collaborate with international partners to influence global AI governance. A follow up document with Government's responses to the white paper has then been released on February 6, 2024, where it was further clarified the overall UK pro-innovation approach to AI–combining cross-sectoral principles and a context-specific framework with international leadership, collaboration, and voluntary measures on developers.[88] The Department for Science, Innovation, and Technology has since formed a Frontier AI Taskforce gathering experts in AI, national security and related fields to assess risks and enhance AI safety, which became a permanent "AI Safety Institute" in November 2023.[89] [90] The UK Government is currently working on a "Code of Practice" on Copyright Issues and AI.[91] In November 2023, the UK hosted the first global "AI Safety Summit," which we describe below together with other multilateral actions.

4.3.6 Regulation in other EMDEs

As discussed above, EMDEs lag substantially behind AE counterparts in the accumulation of resources needed to reap the advantages of AI. Cazzaniga et al. (2024) identify regulatory frameworks as an important constraint that is likely to hold back AI adoption in these countries. Some multilateral actions focused on helping capacity development in lower-income countries to develop legal frameworks tackling AI-related issues. In June 2023,the EU announced funding support for a UNESCO scheme in this sense. Capacity development will be aimed at furthering UNESCO's 2021 recommendations on the ethics of AI, which include: promotion and protection of human rights; human dignity; and environmental sustainability.[92] November 2023 saw the launch of the "UK Government's AI for Development Programme," which includes funding and provisions to increase AI preparedness in lower-income countries. Among its stated objectives, the program aims to help at least ten African countries create sound regulatory frameworks AI, and making five or more African countries influential in the worldwide conversation on AI. [93]

AI regulation in Brazil and India. The draft law for a "Brazilian AI Framework" delineates an AI strategy for Brazil and sets out a regulatory approach similar to the one adopted by the EU AI

---

[88]The document "Consultation outcome: A pro-innovation approach to AI regulation: government response" is available at this link: `https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response`.

[89]For the first progress report, see: `https://www.gov.uk/government/publications/frontier-ai-taskforce-first-progress-report/frontier-ai-taskforce-first-progress-report`.

[90]For the foundation of the AI Safety Institute, see: `https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute`.

[91]More details and terms of reference available at this link, as per 29th June 2023: `https://www.gov.uk/guidance/the-governments-code-of-practice-on-copyright-and-ai`

[92]On this partnership between UNESCO and the EU see: `https://www.unesco.org/en/articles/artificial-intelligence-partnership-between-unesco-and-eu-speed-implementation-ethical-rules`

[93]See the related Press Release by the UK Government on November 1, 2023: `https://www.gov.uk/government/news/uk-unites-with-global-partners-to-accelerate-development-using-ai`

Act (Belli, Curzi, and Gaspar 2023).[94] The document sets out guidelines to categorizing types of AI based on the risk they pose to society. The law also requires AI developers to conduct and publicize risk assessments before launching a product, particularly for high-risk AI systems, and holds all AI developers accountable for damages caused by their systems. Liability standards are heightened for developers of high-risk products. India has instead emphasized a pro-innovation approach with no targeted legislative initiatives as of January 2024.[95]

4.4 AI regulation, international cooperation, and multilateral actions

As noted by Gopinath (2023), AI's impacts have a global component that calls for international cooperation. We discussed above how AI poses threats to national security and international relations. In addition to strategic considerations, the need for cooperation emerges from externalities inherent to the technology (Acemoglu 2024) that may spill over national borders, or regulation themselves. Agrawal, J. S. Gans, and Goldfarb (2019) see the potential for a "race to the bottom" in regulatory standards in order to attract AI companies. Guerreiro, Rebelo, and Teles (2023) note the need for international cooperation to implement optimal regulatory frameworks. However, the authors note that cooperation may be complicated by contrasting objectives and definitions of social welfare.

We report below the main initiatives relative to international cooperation and multilateral actions relative to AI, ordered from oldest to newest.[96] At the end of this section, we briefly discuss private sector initiatives.

OECD "AI Principles." In May 2019, the OECD adopted a set of "AI Principles" to foster innovative and trustworthy AI that respects human rights and democratic values. The five principles are: 1) inclusive growth, sustainable development and well-being, 2) human-centred values and fairness, 3) transparency and explainability, 4) robustness, security and safety, and 5) accountability.[97] The principles have since been endorsed by 46 countries worldwide and have been embedded in several national and multinational initiatives. OECD (2023c) reports that, as of May 2023, at least 50 countries have envisioned AI strategies, and AI-related policy initiatives have taken shape in 70 countries, half of which are EMDEs. Initiatives are selected for their implementation of one or more of the AI Principles and differ in scope and range. Examples of initiatives include: promoting the use of AI for environmental sustainability; protecting privacy and human rights; disclosing information about use of AI systems; application or new regulations related to potential risks of AI; and having independent oversight bodies

---

[94]See Senate documents [in Brazilian Portuguese] `https://legis.senado.leg.br/comissoes/comissao?codcol=2504`

[95]See, e.g., `https://www.livemint.com/news/indias-ai-regulations-will-allow-room-for-innovation-but-with-eye-on-safety-says-html`.

[96]Other efforts not covered in details have been put in place, for instance in the context of the Asia-Pacific Economic Cooperation (APEC) and the G20. Moreover, on ethics, it is good to signal the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.

[97]For more details on the principles see `https://oecd.ai/en/ai-principles`

to audit the use of algorithms.[98] Lastly, within the OECD umbrella, the OECD AI Policy Observatory has been established with the purpose of sharing data, information, and analyses on AI (more info at sub section 5.2).[99]

NATO AI strategy. In its October 2021 meeting, the NATO Allied Defence Ministers formally adopted an AI strategy for defense and national security, committing to the cooperation and collaboration necessary for its implementation. Signatories committed to "Principles of Responsible Use" for the development and deployment of AI. The six principles are: (i) Lawfulness; (ii) Responsibility and Accountability; (iii) Explainability and Traceability; (iv) Reliability; (v) Governability; and (vi) Bias Mitigation.[100]

World Economic Forum and UN initiatives. In June 2023, The World Economic Forum (WEF) launched the "AI Governance Alliance" to unite industry leaders, governments, academic institutions and civil society organizations around the goal of meaningful AI governance.[101] Ultimately, AI's opportunities and challenges, including governance and regulation, were at the center of the 2024 Davos meetings.[102]

In October 2023, the UN Secretary-General announced the creation of an AI Advisory Body to explore risks, opportunities and the global governance of artificial intelligence.[103] In December 2023, The Advisory Body released the Interim Report "Governing AI for Humanity". The document calls for a closer alignment between international norms and AI's developed and deployment. The AI Advisory Body recommends five guiding principles: *1. AI should be governed inclusively, by and for the benefit of all; 2. AI must be governed in the public interest; 3. AI governance should be built in step with data governance and the promotion of data commons; 4. AI governance must be universal, networked and rooted in adaptive multi-stakeholder collaboration; 5. AI governance should be anchored in the UN Charter, International Human Rights Law, and other agreed international commitments such as the Sustainable Development Goals.* The final recommendations are set to be published in the summer of 2024 after a multi-stakeholder consultation process.[104] In September 2024, the UN Summit of the Future will consider the adoption of a "Global Digital Compact." The Compact should address several aspects of AI, including its governance.[105]

G7: "Hiroshima Process International Code of Conduct for Advanced AI Systems". On

---

[98] Figure 3.1. of OECD (2023c) reports a complete list.

[99] See the OECD AI Policy Observatory website: `https://oecd.ai/en/`.

[100] See: `https://www.nato.int/docu/review/articles/2021/10/25/an-artificial-intelligence-strategy-for-nato/index.html`.

[101] On the WEF "AI Governance Alliance," see: `https://www.weforum.org/agenda/2023/06/we-must-come-together-to-ensure-an-ai-future-that-works-for-all/` and the official website of the initiative: `https://initiatives.weforum.org/ai-governance-alliance/home`

[102] See highlights at `https://www.weforum.org/agenda/2024/01/davos-2024-highlights-ai-growth-climate-security/`

[103] See `https://press.un.org/en/2023/sga2236.doc.htm`.

[104] The full report can be found at:`https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf`

[105] See the UN's Secretary-General May 2023 Policy Brief: "Our Common Agenda Policy Brief 5: A Global Digital Compact — an Open, Free and Secure Digital Future for All," available at `https://www.un.org/techenvoy/global-digital-compact`

October 30, 2023, G7 leaders agreed to the "Hiroshima Process International Code of Conduct for Advanced AI Systems".[106] The code provides "voluntary guidance for actions by organizations developing the most advanced AI systems." The document lays out guidance principles encouraging: risk mitigation in all parts of the AI process; increased transparency during development reporting systems' capabilities and domains of use; sharing of information and reporting of incidents in development; developing governance, increasing security controls and advance the development of international standards; deploying reliable provenance mechanisms such as watermarking; prioritizing research to increase AI's safety and on applications that would help sustainable development goals; implementing data input measures and protection for personal data and intellectual property. Several of these measures were in line with the provisional EU AI Act and the US EO. Italy, assuming its 2024 G7 Presidency, announced that it will focus on AI as one of the key themes, and that it will plan a special session before June's leaders' summit to assess the impact of AI with the involvement of scholars, managers and experts.[107]

The first global "AI Safety Summit". On November 1-2, 2023 the British government hosted the first global "AI Safety Summit," gathered representatives from 28 national governments, including the UK, the US, China and several European countries. The summit also saw the participation of multilateral organizations—such as the European Commission, the Council of Europe, the OECD and the UN—as well as academics, entrepreneurs and civil society representatives. The initiative discussed how to approach and regulate AI technologies.[108]

On this occasion, the participating national governments and the EU signed the "Bletchley Declaration," resolving to further national, multilateral and bilateral action to promote on AI safety research and establish risk-based policy across the respective countries.[109] The declaration acknowledges that specific approaches may differ, while stressing the need for international cooperation. As discussed in Section 4.3.6, the Summit saw the launch of the "UK Government's AI for Development Programme." The initiative is expected to become permanent, with the next two AI Safety Summits to be hosted by South Korea and France in 2024.

Guidelines for Secure AI System Development. In November 2023, the UK National Cyber Security Centre (NCSC), the US Cybersecurity and Infrastructure Security Agency (CISA), and several other national cyber security agencies across AEs and EMDEs released a set of "Guidelines. It also includes inputs from several private companies, AI providers, and universities.[110] It recommends safety guidelines for providers of any system that includes the use of AI (i.e., they apply to all types of AI

---

[106]The full the "Hiroshima Process International Code of Conduct for Advanced AI Systems" is available at this link: https://www.mofa.go.jp/files/100573473.pdf

[107]The official website of the Italian G7 Presidency: https://www.g7italy.it/en/

[108]The official website of the "AI Safety Summit" in the UK is at this link: https://www.aisafetysummit.gov.uk/

[109]See the full text at: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

[110]See report and more info at this link: https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development

systems, not just frontier models) along the main phases of a system's development. These guidelines aim to: a secure design, secure development, secure deployment, and secure operation and maintenance. The report includes suggestions about mitigations to related risks in the view that providers of AI components should take responsibility for the security outcomes of users further down the supply chain.

The Global Partnership on AI and the Ministerial Declaration. The goal of the Global Partnership on AI is to guide the responsible development and use of AI, grounded in human rights, inclusion, diversity, innovation and economic growth. [111] On occasion of the December 2023 general-purpose AI summit in Delhi, Ministers of the 29 member countries signed a join declaration reaffirming their commitment to promote responsible and trustworthy AI, and their dedication to jointly develop regulations, policies, and standards to uphold general-purpose AI's principles. The Ministers also embraced the notion of "collaborative AI," which involves supporting and promoting equitable access to critical resources for AI research and innovation, such as AI computing, high-quality diverse datasets, algorithms, software, test-beds etc.[112]

4.4.1 Selected Private Sector Initiatives on AI Governance

In July 2023, Anthropic, Google, Microsoft and OpenAI founded the "Frontier Model Forum" to promote safety standards and best practices.[113] This industry body is dedicated to ensuring the safe and responsible development of advanced AI models. Its main goals are identifying best practices, sharing knowledge with various stakeholders, and supporting AI applications that address major societal challenges. The Forum will establish an Advisory Board and focus on promoting responsible AI development and safety standards. It also intends to facilitate collaboration with governments and other initiatives in the AI community to benefit society.

Tech companies are also assessing the concept of "Constitutional AI," a way to train AI systems to follow certain sets of rules or "constitutions" (Bai et al. 2022). The proposed implementation would involve human action only to feed AI systems the set of rules. Systems then learn to self-revise based on the constitutional principles via reinforcement learning, effectively self-regulating without human feedback, and potentially becoming instrumental in applying the rules to other AI systems. AI startup Anthropic is at the forefront of the initiative to deploy consistutional AI. The company's AI assistant Claude is trained using the principles listed in Bai et al. (2022), which, among other sources, draw from the UN Declaration of Human Rights.[114] In a similar vein, Google is committing to safety evaluations for their products. The company's gen-AI model "Gemini," to be launched in 2024, was developed for robustness to "real toxicity prompts"[115] Finally, in May 2023, OpenAI launched a grant program *to*

---

[111] More info on the Global Partnership on AI is available on their website: `https://gpai.ai/`

[112] See the entire text of the declaration here:`https://gpai.ai/2023-GPAI-Ministerial-Declaration.pdf`

[113] `https://openai.com/blog/frontier-model-forum`

[114] See: `https://www.anthropic.com/index/claudes-constitution`

[115] `https://blog.google/technology/ai/google-gemini-ai/`

design ideas and tools to collectively govern AI. A report of the findings and the ten winning projects was published in January 2024.[116]

5 Discussion

In this section, we highlight issues left open and possible room for additional research and policy actions, as emerging from the current academic literature and regulations. Subsection 5.3 articulates our key takeaways.

5.1 Open questions and gaps in the literature

The emerging literature on AI firmly established the pervasive and potentially transformative role of new technologies based on machine learning, highlighting that all occupations will be affected. However, the limit to our knowledge on its effects remain substantial. In this section we discuss our knowledge gaps progressively broadening the perspective from the labor market, to economy-wide impacts, to development and cross-country implications.

Lack of Consensus on AI Automation Effects. The consequences of AI for the labor market depend on its automating and complementing potential. First, let us consider the automating character of AI. Theory suggests that the effects of labor-substituting automation technology are ambiguous, but they can be determined through the knowledge of some key parameters and data moments. For example, Acemoglu and Restrepo (2022) derive formulas to obtain the net impact on wages and labor demand for different demographic groups. In that framework, wage effects can be directly computed once two quantities are known for occupation and sector: the share of tasks substituted by AI, and an estimate of the resulting cost savings. In this review, we covered several studies trying to provide an estimate of the former through task exposure measures. Although the correlation between different measures is usually positive (see, e.g., Cazzaniga et al. (2024) for a discussion), there is substantial disagreement on whether to interpret exposures as an indicator of potential task displacement. The comparability of different indicators is also in question, as some are indices and can only be interpreted as capturing *relative* task exposure, while others directly report a share of tasks that would be directly affected by the introduction of AI. Getting the indicators to speak to each other and producing related sufficient statistics for employment and productivity effects would be a necessary first step to determine the overall impact of AI on wages and employment. For example, instead of reporting "task exposures," studies could report bounds on the share of employment–at current demand conditions and wages–that could be displaced by AI in each occupation as well as in the overall economy (as in, e.g., Briggs and Kodnani (2023) and Company (2023)). In this context, assumptions on complementarity between AI and tasks carried

---

out by workers should be clearly stated. Ideally, the lower and upper bound of estimated employment effects should be interpretable as arising from the maximum and minimum degree of complementarity, respectively. These numbers could then be used to reach a general equilibrium employment estimates accounting for input-output linkages and worker mobility, as in Acemoglu and Restrepo (2022).

The Need for Task-Level and Sector-Level Productivity Data. The lack of precise productivity estimates further complicates the analysis of employment effects. As discussed in Section 3, available estimates cover specific technologies and contexts and cannot be easily generalized. In the narrow field of gen-AI, research has explored the productivity impact of text-generating AI, while it is silent on image generation, event though the latter has a significant effect on labor market outcomes of affected workers (Hui, Reshef, and Zhou 2023). The debate is still ongoing on how much AI will augment workers. The evidence in this paper points to important areas of complementarity and in this case, exposure measures are not sufficient to assess relative wage and employment effects. Instead, we would need an assessment of the measure of tasks that are complementary or substituting. This could take the form of a two-dimensional exposure measure–that is, measuring at the same time the share of augmented and substituted tasks–similar to Kogan et al. (2023). A similar indicator has however substantially higher data requirement than a simple exposure measure, since it would need to provide information on *how much* tasks are augmented by AI in addition to *which* tasks are affected. Just as noted by Restrepo (2023) for the broader automation literature, the key data gap appears to be the lack of *task-level data,* which would offer information on which tasks are carried out to produce each good and service in establishments with different AI technologies. Efforts to collect this data at scale have a notable historical precedent: the 1899 "Hand and Machine Labor" study, commissioned by the US congress to assess the impact of mechanization on labor (Atack, Margo, and Rhode 2019). The above limitations pose a severe constraint to any study trying to assess the economy-wide consequences of AI progress and adoption. In addition to productivity and task-level data, we would need to further assess the implications of the technology for different economic sectors. The data on occupational exposure only gives us a limited idea of supply-side impacts of the technology. However, we are still uncertain on which sectors' products or markets are likely to benefit the most, and which industries stand to expand or contract. This aspect would be key to assess capacity constraints and skills gap to be filled, as well as retraining needs for potentially displaced workers.

Knowledge Gaps on Growth and Development Issues. The lack of productivity estimates is even more blatant when it comes to general ML and several crucial applications that have now been in use for years around the world. AI-powered precision agriculture stands out as a field for which tightly identified estimates of *economic outcomes*–such as value added or employment–would be highly

beneficial.[117] Such estimates would be key to evaluate the technology's impact on EMDEs, where disemployment effects are expected to be limited. In addition to sectoral implications, we remain unaware of the actual impact of AI on health and education. These points deserve a separate mention for two reason. First, both fields could further augment growth rates, through their impact on demographics and human capital, respectively. Second, these areas represent crucial gaps in the attainment of sustainable development goals (SDGs) in EMDEs.[118] In this sense, experimental findings on the potential contribution of AI to reducing the needed pupil to teacher or patient to doctor ratio would be very valuable. On the theme of EMDEs, a major knowledge gap pertains to the international consequences of AI adoption. Korinek and Stiglitz (2021) and Alonso et al. (2022) study this question theoretically, but we have little empirical guidance on this topic. While we still lack data to evaluate this question, a first assessment may come from bridging theory with empirical evaluation of previous waves of technology that were simultaneously complementing and substituting. A promising candidate for this exercise could be the spillover of past IT adoption in advanced economies on EMDEs and its consequences on capital flows and factor allocations.

5.2 Policy and Regulation

In this part, we list potential issues facing AI regulators: the length of approval and implementation lags compared to the speed of technological advancements; the effects of regulation on sectors and firms; the seeming disconnect between policy and research; the need for data to inform the debate; and the potential role of unions and labor standards.

Regulation and the Speed of Technological Advancements. Several national and supranational authorities are currently either considering or structuring their AI regulation acts. In Section 4, we have highlighted how these initiatives vary widely both in coverage and scope. In addition, differences arise in the timing and speed of adoption. The EU legislative process started in mid-2021. A final act may follow in early 2024, and its complete application would then occur over the next two to three years. The US instead opted for a decentralized setup with involves Agencies and Departments with deadlines by mid-2024. These processes paint a stark contrast between the time needed to reach an agreement compared and the much faster speed of AI technologies. By the time they are finally implemented, regulations might have already become obsolete, unless they allow for a degree of flexibility in their definitions and coverage. Initiatives of self-regulation by AI companies may fill part of this gap.[119]

Effects of Regulation on Firms and Sectors. Several knowledge gaps persist on the effects of

---

[117]There is of course an agricultural literature studying the topic, but it focuses on outcomes expressed in terms–e.,g., yields, water or fertilizer savings for specific crops–that cannot readily translate to overall productivity. In addition, estimates rarely result from random or quasi-random variation.

[118]See Senhadji et al. (2021) for a recent discussion on financing gaps for SDGs.

[119]See Wheeler (2023)

regulations on firm location choices and sectors. Indeed, regulation may push firms to locate where the regulations are less strict, with major repercussions on a country's ability to compete on the international markets and effects on its internal labor market as well. It is also theoretically unclear whether regulations should include sector-specific clauses or exceptions. These may relate to the strategic nature of specific sectors in supply chains or input-output networks, shielding of "national champions," or worker protection concerns. In general, regulation might reduce adoption of AI technologies (Lee et al. 2019) and innovation overall (Aghion, Bergeaud, and Van Reenen 2021). Analyses of firm-level and sectoral effects of new AI regulations appear necessary to guide the research community and policymakers on these aspects.

Policy and Research. Another aspect to consider is the potential disconnect between policy and regulations with research. The main regulations currently under consideration already envision at least some degree of collaboration with research. However, considering the rapid pace of the AI technologies, more exchanges between policy and research may be necessary to understand and measure the potential risks and impacts of AI adoption, and to foster innovation in this field.

Within the new EU AI governance emerging from the Trilogue agreement and the Provisional Agreement, the AI Office (of the EU Commission) will include a scientific panel of independent experts with the goal to advise about general-purpose AI models and to monitor their possible material safety risks. An Advisory Forum—including industry representatives, small and medium enterprises, start-ups, civil society, as well as academia—is also envisioned to provide technical expertise to the AI Board. The same version of the EU AI Act also states that the provisions of the Act would not apply to AI systems used for the *sole* purpose of research and innovation. In the US, promoting innovation and research is one of the eight principles guiding the EO. Practical provisions include expedited visas for noncitizens who seek to travel to the US to work on, study, or conduct research in AI. Further, the EO envisions a program to identify and attract top talent in AI at universities, research institutions, and the private sector, and to establish and increase connections overseas. The EO is also set to establish at least four new National AI Research Institutes and to produce a publicly available report on the potential role of AI in research aimed at tackling major societal and global challenges. The US Senate AI fora, which included experts in different fields (see Section 4.3.2) have been an important step to bridge policy and research to understand risks and opportunities of AI.. Finally, the Chinese regulation explicitly support the coordination between private sectors, education and research institutions, and public bodies in areas such as innovation in gen-AI technology, data resources, applications, and risk prevention.

Data Availability. The scarcity of *updated* data on the last generation of AI technologies stands out as a key hindrance to both regulators and researchers. Surveys and data on AI, its adoption, and progress appear highly critical. In addition to more data, it would be important to establish priorities on the *types* of data needed for policy and research. For example, additional insights would be useful on the market

structure of chip producing sectors, trade linkages, and the concentration of exports and raw materials. Further, we have only scant data on the increasing energy requirements of AI data centers, which may impact energy markets through prices and investments in additional power sources like nuclear plants.[120]

The OECD is currently trying to fill some gaps in the data, providing (or planning to cover) a wide set of AI-related information across countries and over time. Some example of datasets included are: AI news around the world and AI search trends; characteristics of AI developers and AI courses offered worldwide; trends in the demand and supply of AI talents, scientific publications, and patents; R&D funding and private equity investments in AI start-ups; AI models, software developments, and datasets from open-source platforms.[121] The OECD also keeps track of regulations in the world based on their principles for the responsible stewardship of trustworthy AI released in 2021, with the latest report issued in May 2023.[122,123]

Additionally, the project "Our World in Data" (Giattino et al. 2023) now includes aspects related to AI in their interactive charts, such as the language and image recognition capabilities of AI, AI-related patents and new companies, or market share of AI hardware production by countries. The project also collects data from various external sources. Most datasets are updated to 2021.

On surveys, Van Noorden and Perkel (2023) look at researchers on their relationship with AI. They show that researchers are already adopting gen-AI, particularly for writing code, research brainstorming, writing manuscripts and conducting literature reviews.[124] More surveys could bring a better sense of possible impacts of AI, for example covering firms and their adoption of AI and future plans. Looking at small and medium enterprises in a wider set of countries and regions may prove particularly fruitful.

Labor Markets and Trade Unions. The potential role of unions and labor standards in labor market regulations is another area that stands out as needing further investigation. This is especially true in light of concrete actions taken by affected workers' bodies. The United Auto Workers and the Writers Guild of America (WGA) included AI in their labor deal, giving them a say in how they use AI, and establishing a precedent for AI to be the subject of bargaining.[125] In particular, the agreements did not aim to ban AI, but rather demanded a share of the productivity gains from AI.[126] Among regulations, the US EO includes this aspect in its eight principle on supporting the American workforce: *"As AI creates new jobs and industries, all workers need a seat at the table, including through collective bargaining, to ensure that they benefit from these opportunities."* At the same time, the EO does not

---

[120]See https://www.wsj.com/articles/rising-data-center-costs-linked-to-ai-demands-fc6adc0e

[121]Data and charts available at: https://oecd.ai/en/trends-and-data

[122]These principles are: inclusive growth, sustainable development and well-being, human-centred values and fairness, transparency and explainability, robustness, security and safety, and accountability.

[123]https://www.oecd.org/publications/the-state-of-implementation-of-the-oecd-ai-principles-four-years-on-835641c9-en.htm

[124]Data and methodology are publicly available. The survey covers several fields: Computer science, Physical sciences, Life sciences, Health and medicine, Social sciences (economics, econometrics and finance; social sciences; arts and humanities; business, management and accounting).

[125]The WGA memorandum of agreement of September 2023: https://www.wgacontract2023.org/wgacontract/files/memorandum-of-agreement-for-the-2023-wga-theatrical-and-television-basic-agreement.pdf

[126]See for instance Sheffi (2023)

provide binding measures or guidelines. Instead, it delegates the Department of Labor to issue guidance to employers that deploy AI to monitor or augment employees' work on how to comply with existing legal requirements.[127]

5.3 Key Takeaways

Takeaway 1: AI Effects Are Uncertain. Task exposure measures and the theoretical growth literature agree that the impact of AI will be extensive. However, the empirical evidence on employment effects and productivity provides wide estimates of the effects, and remains ambiguous on their direction. We believe that a key reason for this uncertainty resides in the limited comparability across measures, as well as the lack of crucial task-level, sector-level, and productivity data. Uncertainty is complicated by the scarcity of data on latest applications, such as gen-AI.

Takeaway 2: Policy and Research Are Partly Disconnected. At the date of writing, the academic literature has focused primarily on employment and productivity effects. By contrast, *binding* policy actions around the world have been mostly more concerned with ethical and bias concerns, privacy, and safety issues, which, with few notable exceptions, fall outside the scope of economics. The reluctance of regulators to tackle economic issues may be partly motivated by the uncertainty surrounding estimates of AI effects. At the same time, to our knowledge, the academic literature has not provided empirical investigations of key economic issues that are of interest to regulators. In this respect, more research is needed to understand the impact of AI on market concentration—from the competitive landscape of developers and providers to the effects on competition in adopting sectors—, trade networks and linkages in the AI supply chain, and intellectual property. Namely, the effects of AI on incentives to content creation remain unexplored. As discussed above, there are several ongoing initiatives that may contribute to bridging the gap between policy and research.

Takeaway 3: Regulations Differ Widely and Face Difficult Trade-Offs. We have seen how regulators across the world have so far adopted different—or outright contrasting—approaches and covered different areas. Some authorities preferred an ex-ante risk-based approach, other focused on ex-post liability, while some countries preferred a "pro-innovation," regulation-free stance focused on guidelines and best practices. Similarly, there is substantial disagreement on what should be regulated. Efforts for international coordination have not yet changed this landscape. We believe that persistent differences across countries stem from difficult trade-offs that countries face when choosing to regulate AI. Many countries see the development of vibrant domestic AI sectors as strategic, which comes with two concerns. First, governments are reluctant to burden developers with regulation that may stifle innovation, as many countries acknowledge that first-mover advantages in the technology may be large.

---

[127]In the provisional EU AI Act, there is also a guideline to lay down an information/consultation requirement for employers in case of use of AI services (see (92)).

Second, regulators may face a trade-off between safety and attracting AI investments, as companies may choose to localize where policies are looser. Closer multilateral cooperation in development and regulation would go a long way toward addressing these issues.

6   Conclusion

In this paper, we reviewed the academic literature on AI's impact on the economy and the regulatory actions that have taken shape over the last few years. We found that economic impact of AI, while theoretically extensive, remains ambiguous in scope and direction. Perhaps in view of this uncertainty, economic research is only very partially incorporated in regulators' consideration. At the same time, key areas of policy interest remain unexplored or under-investigated. When it comes to actual regulatory actions, our review revealed that countries have taken different approaches with contrasting objectives, which, in our view, motivates increased multilateral cooperation. To counteract these shortcomings, future research should aim to collect more granular, accurate and updated data on AI applications, and strengthen its focus to area of policy interest.

The future of the technology is also highly debated. On the one hand, rapidly-expanding data processing capacity and "emerging capabilities"—the potential for AI systems to learn new, unpredictable, abilities—lead some researchers and commentators to envision human-like Artificial General Intelligence (AGI). In this vein, "Agentic AI systems" would be able to pursue complex goals with limited direct supervision and require specific, constantly-adapting regulatory frameworks (Shavit et al. 2024). On the other hand, some scholars have highlighted the limitations of current systems as well as potential pitfalls that may stifle rapid development. One such example is the possibility that AI models, feeding on excessive AI-generated information, degenerate in self-consuming loops that ultimately compromise precision and recall characteristics (Alemohammad et al. 2023). In the face of such ambiguity, close vigilance will be key to cope with—and potentially steer—the direction of AI technology.

In conclusion, the landscape that emerges from our review features a wide range of estimated economic effects arising from scarcity of data, accompanied by diverging and uncoordinated regulatory actions. The disparity in the pace and scope of regulatory measures partly arises from the variety of political systems, each characterized by unique legislative competences and deliberative processes. At the same time, several actors are pursuing deliberately lax frameworks, motivated by a desire to attract benefits from AI innovation and reap first-mover advantages, while seemingly discounting potential risks. If this gamble proves correct, AI risks will remain contained and localized, and policy responses may adapt in time. However, if more sweeping, cross-border perils materialize, there will be no time left for belated international cooperation.

References

Acemoglu, Daron (2024). "Harms of AI". In: *The Oxford Handbook of AI Governance*. Ed. by Justin B. Bullock, Yu-Che Chen, Johannes Himmelreich, Valerie M. Hudson, Anton Korinek, Matthew M. Young, and Baobao Zhang. Oxford University Press.

Acemoglu, Daron, Gary W. Anderson, David N. Beede, Catherine Buffington, Eric E. Childress, Emin Dinlersoz, Lucia S. Foster, Nathan Goldschlag, John C. Haltiwanger, Zachary Kroff, Pascual Restrepo, and Nikolas Zolas (2022). "Automation and the Workforce: A Firm-Level View from the 2019 Annual Business Survey". In: *Technology, Productivity, and Economic Growth*. Ed. by Susantu Basu, Lucy Eldridge, John C. Haltiwanger, and Erich Strassner. Studies in Income and Wealth. University of Chicago Press.

Acemoglu, Daron and David Autor (2011). "Chapter 12 - Skills, Tasks and Technologies: Implications for Employment and Earnings". In: *Handbook of Labor Economics*. Ed. by David Card and Orley Ashenfelter. Vol. 4B. Elsevier, pp. 1043–1171.

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo (2022). "Artificial Intelligence and Jobs: Evidence from Online Vacancies". *Journal of Labor Economics* 40(S1), S293–S340.

Acemoglu, Daron, David Autor, and Simon Johnson (2023). *Can We Have Pro-Worker AI? Choosing a Path of Machines in Service of Minds*. Policy Memo. MIT Shaping the Future of Work Initiative.

Acemoglu, Daron and Todd Lensman (2023). *Regulating Transformative Technologies*. Working Paper 31461. National Bureau of Economic Research.

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar (2023). *A Model of Behavioral Manipulation*. Working Paper 31872. National Bureau of Economic Research.

Acemoglu, Daron, Andrea Manera, and Pascual Restrepo (2020). "Does the US Tax Code Favor Automation?" *Brookings Papers on Economic Activity*, pp. 231–285.

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), pp. 1488–1542.

— (2022). "Tasks, Automation, and the Rise in U.S. Wage Inequality". *Econometrica* 90(5), pp. 1973–2016.

Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz (2023). *Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology*. Working Paper 31422. National Bureau of Economic Research.

Aghion, Philippe, Antonin Bergeaud, and John Van Reenen (2021). *The Impact of Regulation on Innovation*. Working Paper 28381. National Bureau of Economic Research.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb (2019). "Economic Policy for Artificial Intelligence". In: *Innovation Policy and the Economy*. Ed. by Josh Lerner and Scott Stern. Vol. 19. The University of Chicago Press, pp. 139–159.

Agrawal, Ajay, Joshua S. Gans, and Avi Goldfarb (2019). "Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction". *Journal of Economic Perspectives* 33(2), pp. 31–50.

Albanesi, Stefania, António Dias Da Silva, Juan F. Jimeno, Ana Lamo, and Alena Wabitsch (2023). *New Technologies and Jobs in Europe*. Working Paper 2831. European Central Bank.

Alderucci, Dean, Lee Branstetter, Heinz College, Eduard Hovy, and Andrew Runge (2021). *Quantifying the Impact of AI on Productivity and Labor Demand: Evidence from U.S. Census Microdata*. Working Paper.

Alekseeva, Liudmila, José Azar, Mireia Giné, Sampsa Samila, and Bledi Taska (2021). "The Demand for AI Skills in the Labor Market". *Labour Economics* 71.

Alemohammad, Sina, Josue Casco-Rodriguez, Lorenzo Luzi, Ahmed Imtiaz Humayun, Hossein Babaei, Daniel LeJeune, Ali Siahkoohi, and Richard G. Baraniuk (2023). *Self-Consuming Generative Models Go MAD*. Preprint. arXiv: 2307.01850.

Allcott, Hunt and Matthew Gentzkow (2017). "Social Media and Fake News in the 2016 Election". *Journal of Economic Perspectives* 31(2), pp. 211–236.

Allen, Greg and Taniel Chan (2017). *Artificial Intelligence and National Security*. Working Paper. Belfer Center for Science and International Affairs, Harvard Kennedy School.

Alonso, Cristian, Andrew Berg, Siddharth Kothari, Chris Papageorgiou, and Sidra Rehman (2022). "Will the AI Revolution Cause a Great Divergence?" *Journal of Monetary Economics* 127, pp. 18–37.

Atack, Jeremy, Robert A. Margo, and Paul W. Rhode (2019). "'Automation' of Manufacturing in the Late Nineteenth Century: The Hand and Machine Labor Study". *Journal of Economic Perspectives* 33(2), pp. 51–70.

Autor, David (2022). *The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty*. Working Paper 30074. National Bureau of Economic Research.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson (2024). "Artificial Intelligence, Firm Growth, and Product Innovation". *Journal of Financial Economics* 151.

Babina, Tania, Anastassia Fedyk, Alex X. He, and James Hodson (2024). "Firm Investments in Artificial Intelligence Technologies and Changes in Workforce Composition". In: *Technology, Productivity, and Economic Growth*. Ed. by Susantu Basu, Lucy Eldridge, John C. Haltiwanger, and Erich Strassner. National Bureau of Economic Research.

Bai, Yuntao et al. (2022). *Constitutional AI: Harmlessness from AI Feedback*. Preprint. arXiv. arXiv: 2212.08073.

Baily, Martin Neil, Erik Brynjolfsson, and Anton Korinek (2023). *Machines of Mind: The Case for an AI-powered Productivity Boom*. Brookings. URL: `https://www.brookings.edu/articles/machines-of-mind-the-case-for-an-ai-powered-productivity-boom/`.

Bajari, Patrick, Victor Chernozhukov, Ali Hortaçsu, and Junichi Suzuki (2019). "The Impact of Big Data on Firm Performance: An Empirical Investigation". *AEA Papers and Proceedings* 109, pp. 33–37.

Bank of England (2022). *Artificial Intelligence and Machine Learning*. Discussion Paper DP5/22.

Belli, Luca, Yasmin Curzi, and Walter B. Gaspar (2023). "AI Regulation in Brazil: Advancements, Flows, and Need to Learn from the Data Protection Experience". *Computer Law & Security Review* 48.

Berg, Andrew, Edward F. Buffie, Mariarosaria Comunale, Chris Papageorgiou, and Luis-Felipe Zanna (2024). *Searching for Wage Growth: Policy Responses to the "New Machine Age"*. Working Paper 2024/03. International Monetary Fund.

Berg, Mr Andrew, Lahcen Bounader, Nikolay Gueorguiev, Hiroaki Miyamoto, Mr Kenji Moriyama, Ryota Nakatani, and Luis-Felipe Zanna (2021). *For the Benefit of All: Fiscal Policies and Equity-Efficiency Trade-offs in the Age of Automation*. Working Paper 2021/187. International Monetary Fund.

Björkegren, Daniel (2023). *Artificial Intelligence for the Poor: How to Harness the Power of AI in the Developing World*. Foreign Affairs. URL: `https://www.foreignaffairs.com/world/artificial-intelligence-poor`.

Bommasani, Rishi (2023). *Drawing Lines: Tiers for Foundation Models*. Stanford Center for Research on Foundation Models. URL: `https://crfm.stanford.edu/2023/11/18/tiers.html`.

Borgonovi, Francesca, Flavio Calvino, Chiara Criscuolo, Lea Samek, Helke Seitz, Julia Nania, Julia Nitschke, and Layla O'Kane (2023). *Emerging Trends in AI Skill Demand Across 14 Oecd Countries*. OECD Artificial Intelligence Papers 2. OECD.

Branch, Hezekiah J., Jonathan Rodriguez Cefalu, Jeremy McHugh, Leyla Hujer, Aditya Bahl, Daniel del Castillo Iglesias, Ron Heichman, and Ramesh Darwishi (2022). *Evaluating the Susceptibility of Pre-Trained Language Models via Handcrafted Adversarial Examples*. Preprint. arXiv. arXiv: `2209.02128`.

Briggs, Joseph and Devesh Kodnani (2023). *The Potentially Large Effects of Artificial Intelligence on Economic Growth*. Global Economic Analyst. Goldman Sachs.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond (2023). *Generative AI at Work*. Preprint. arXiv. arXiv: `2304.11771`.

Brynjolfsson, Erik, Tom Mitchell, and Daniel Rock (2018). "What Can Machines Learn, and What Does It Mean for Occupations and the Economy?" *AEA Papers and Proceedings* 108, pp. 43–47.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan (2017). "Semantics Derived Automatically from Language Corpora Contain Human-Like Biases". *Science* 356(6334), pp. 183–186.

Cavallo, Alberto (2018). *More Amazon Effects: Online Competition and Pricing Behaviors*. Working Paper 25138. National Bureau of Economic Research.

Cazzaniga, Mauro, Florence Jaumotte, Longji Li, Giovanni Melina, Augustus J. Panton, Carlo Pizzinelli, Emma Rockall, and Marina Mendes Tavares (2024). *Gen-AI: Artificial Intelligence and the Future of Work*. Staff Discussion Note 2024/001. International Monetary Fund.

Company, McKinsey & (2023). *The Economic Potential of Generative AI: The Next Productivity Frontier*. Report.

Copestake, Alexander, Max Marczinek, Ashley Pople, and Katherine Stapleton (2023). *AI and Services-Led Growth: Evidence from Indian Job Adverts*. STEG Working Paper WP060. Structural Transformation and Economic Growth.

Cornelli, Giulio, Jon Frost, and Saurabh Mishra (2023). *Artificial Intelligence, Services Globalisation, and Income Inequality*. Working Paper 1135. Bank for International Settlements.

Czarnitzki, Dirk, Gastón P. Fernández, and Christian Rammer (2022). *Artificial Intelligence and Firm-Level Productivity*. Discussion Paper 22-005. ZEW – Liebniz Centre for European Economic Research.

Danielsson, Jon and Andreas Uthemann (2024). *On the Use of Artificial Intelligence in Financial Regulations and the Impact on Financial Stability*. Working Paper 4604628. SSRN.

Eisen, Norman, Nicol Turner Lee, Colby Galliher, and Jonathan Katz (2023). *AI Can Strengthen U.s. Democracy—and Weaken It*. Brookings. URL: https://www.brookings.edu/articles/ai-can-strengthen-u-s-democracy-and-weaken-it/.

Eisfeldt, Andrea L., Gregor Schubert, and Miao Ben Zhang (2023). *Generative AI and Firm Values*. Working Paper 31222. National Bureau of Economic Research.

Ellingrud, Kweilin, Saurabh Sanghvi, Gurneet Singh Dandona, Anu Madgavkar, Michael Chui, Olivia White, and Paige Hasebe (2023). *Generative AI and the Future of Work in America*. McKinsey Global Institute.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock (2023). *GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*. Preprint. arXiv. arXiv: 2303.10130.

Ezrachi, Ariel and Maurice Stucke (2020). "Sustainable and Unchallenged Algorithmic Tacit Collusion". *Northwestern Journal of Technology and Intellectual Property* 17(2), p. 217.

Felten, Edward, Manav Raj, and Robert Seamans (2021). "Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses". *Strategic Management Journal* 42(12), pp. 2195–2217.

Felten, Edward W., Manav Raj, and Robert Seamans (2023). *Occupational Heterogeneity in Exposure to Generative AI*. Working Paper 4414065. SSRN.

Franke, Ulrike Esther (2019). *Not Smart Enough: The Poverty of European Military Thinking on Artificial Intelligence*. Policy Brief. European Council on Foreign Relations.

Frey, Carl Benedikt and Michael A. Osborne (2017). "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114, pp. 254–280.

Ganguli, Deep et al. (2022). "Predictability and Surprise in Large Generative Models". In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, pp. 1747–1764.

Giattino, Charlie, Edouard Mathieu, Veronika Samborska, and Max Roser (2023). *Artificial Intelligence*. Our World in Data. URL: https://ourworldindata.org/artificial-intelligence.

Gmyrek, Pawel, Janine Berg, and David Bescond (2023). *Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality*. Working paper 96. Geneva, Switzerland: International Labour Organization.

Gopinath, Gita (2023). "Harnessing AI for Global Good". *Finance & Development* 60(4).

Guerreiro, Joao, Sergio Rebelo, and Pedro Teles (2023). *Regulating Artificial Intelligence*. Working Paper 31921. National Bureau of Economic Research.

Horowitz, Michael C. (2018). "Artificial Intelligence, International Competition, and the Balance of Power". *Texas National Security Review* 1(3), pp. 36–57.

Huang, Yueling (2024). "Is the Impact of AI Different from That of IT?" *New Technologies, Digitalization, and AI: The Future Is Here*. IMF Research Perspectives 26.

Hui, Xiang, Oren Reshef, and Luofeng Zhou (2023). *The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market*. Working Paper 4527336. SSRN.

Jones, Charles I. (2023). *The A.I. Dilemma: Growth versus Existential Risk*. Working Paper 31837. National Bureau of Economic Research.

Kim, Yo-whan, Samarth Mishra, SouYoung Jin, Rameswar Panda, Hilde Kuehne, Leonid Karlinsky, Venkatesh Saligrama, Kate Saenko, Aude Oliva, and Rogerio Feris (2022). "How Transferable Are Video Representations Based on Synthetic Data?" In: *Proceedings of the 36th Conference on Neural Information Processing Systems*.

Klare, Brendan F., Mark J. Burge, Joshua C. Klontz, Richard W. Vorder Bruegge, and Anil K. Jain (2012). "Face Recognition Performance: Role of Demographic Information". *IEEE Transactions on Information Forensics and Security* 7(6), pp. 1789–1801.

Kogan, Leonid, Dimitris Papanikolaou, Lawrence D.W. Schmidt, and Bryan Seegmiller (2023). *Technology and Labor Displacement: Evidence from Linking Patents with Worker-Level Data*. Working Paper 31846. National Bureau of Economic Research.

Korinek, Anton (2023a). "Generative AI for Economic Research: Use Cases and Implications for Economists". *Journal of Economic Literature* 61(4), pp. 1281–1317.

— (2023b). *Language Models and Cognitive Automation for Economic Research*. Working Paper 30957. National Bureau of Economic Research.

Korinek, Anton (2023c). "Scenario Planning for an AGI Future". *Finance & Development* 60(4).

Korinek, Anton and Joseph E Stiglitz (2021). *Artificial Intelligence, Globalization, and Strategies for Economic Development*. Working Paper 28453. National Bureau of Economic Research.

Kretschmer, M, T Kretschmer, A Peukert, and C Peukert (2023). *The Risks of Risk-Based AI Regulation: Taking Liability Seriously*. Discussion Paper 18517. CEPR.

Lambrecht, Anja and Catherine E. Tucker (2018). *Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads*. Working Paper 2852260. SSRN.

Landemore, Hélène (2023). "Fostering More Inclusive Democracy with AI". *Finance & Development* 60(4).

Lassebie, Julie (2023). "Skill Needs and Policies in the Age of Artificial Intelligence". In: *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market*. OECD Publishing.

Lee, Yong Suk, Benjamin Larsen, Michael Webb, and Mariano-Florentino Cuéllar (2019). *How Would AI Regulation Change Firms' Behavior? Evidence from Thousands of Managers*. Working Paper 19-031. Stanford Institute for Economic Policy Research, Stanford University.

McCarthy, John (2007). *What Is Artificial Intelligence?* Mimeo. Stanford University.

McElheran, Kristina, J. Frank Li, Erik Brynjolfsson, Zachary Kroff, Emin Dinlersoz, Lucia S Foster, and Nikolas Zolas (2023). *AI Adoption in America: Who, What, and Where*. Working Paper 31788. National Bureau of Economic Research.

Milanez, Anna (2023). *The Impact of AI on the Workplace: Evidence from Oecd Case Studies of AI Implementation*. OECD Social, Employment and Migration Working Papers No. 289. OECD.

Noy, Shakked and Whitney Zhang (2023). "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence". *Science* 381(6654), pp. 187–192.

O'Shaughnessy, Matt (2022). *One of the Biggest Problems in Regulating AI Is Agreeing on a Definition*. Carnegie Endowment. URL: https://carnegieendowment.org/2022/10/06/one-of-biggest-problems-in-regulating-ai-is-agreeing-on-definition-pub-88100.

Obermeyer, Ziad, Rebecca Nissan, Michael Stern, Stephanie Eaneff, Emily Joy Bembeneck, and Sendhil Mullainathan (2021). *Algorithmic Bias Playbook*. The Center for Applied Artificial Intelligence at Chicago Booth.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan (2019). "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations". *Science* 366(6464), pp. 447–453.

OECD (2023a). *Generative Artificial Intelligence in Finance*. OECD Artificial Intelligence Papers No. 9. OECD.

— (2023b). *OECD Employment Outlook 2023*.

— (2023c). *The State of Implementation of the OECD AI Principles Four Years On*. OECD Artificial Intelligence Papers No. 3. OECD.

Pizzinelli, Carlo, Augustus J. Panton, Ms Marina Mendes Tavares, Mauro Cazzaniga, and Longji Li (2023). *Labor Market Exposure to AI: Cross-country Differences and Distributional Implications*. Working Paper 2023/216. International Monetary Fund.

Rammer, Christian, Gastón P. Fernández, and Dirk Czarnitzki (2022). "Artificial Intelligence and Industrial Innovation: Evidence from German Firm-Level Data". *Research Policy* 51(7).

Restrepo, Pascual (2023). *Automation: Theory, Evidence, and Outlook*. Working Paper 31910. National Bureau of Economic Research.

Senhadji, Abdelhak, Alexander Tieman, Edward Gemayel, and Dora Benedek (2021). *A Post-Pandemic Assessment of the Sustainable Development Goals*. Staff Discussion Note 2021/003. International Monetary Fund.

Shabsigh, Ghiath and El Bachir Boukherouaa (2023). *Generative Artificial Intelligence in Finance: Risk Considerations*. Fintech Note 2023/006. International Monetary Fund.

Shavit, Yonadav et al. (2024). *Practices for Governing Agentic AI Systems*. White Paper. OpenAI.

Sheffi, Yossi (2023). *The UAW and Other Unions Must Focus More on AI and Automation in Their Negotiations*. Harvard Business Review. URL: `https://hbr.org/2023/09/the-uaw-and-other-unions-must-focus-more-on-ai-and-automation-in-their-negotiations`.

Sun, Chen, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta (2017). *Revisiting Unreasonable Effectiveness of Data in Deep Learning Era*. Preprint. arXiv. arXiv: `1707.02968`.

Svanberg, Maja S, Wensu Li, Martin Fleming, Brian C Goehring, and Neil C Thompson (2024). *Beyond AI Exposure: Which Tasks Are Cost-Effective to Automate with Computer Vision?* Working Paper 4700751. SSRN.

Trammell, Philip and Anton Korinek (2023). *Economic Growth under Transformative AI*. Working Paper 31815. National Bureau of Economic Research.

Van Noorden, Richard and Jeffrey M. Perkel (2023). "AI and Science: What 1,600 Researchers Think". *Nature* 621, pp. 672–675.

Varian, Hal (2019). "Artificial Intelligence, Economics, and Industrial Organization". In: *The Economics of Artificial Intelligence: An Agenda*. Ed. by Ajay Agrawal, Joshua Gans, and Avi Goldfarb. University of Chicago Press.

Veale, M, R Binns, and L Edwards (2018). "Algorithms That Remember: Model Inversion Attacks and Data Protection Law". *Philosophical Transactions of the Royal Society A* 376.

Webb, Michael (2019). *The Impact of Artificial Intelligence on the Labor Market*. Working Paper 3482150. SSRN.

Wheeler, Tom (2023). *The Three Challenges of AI Regulation*. Brookings. URL: `https://www.brookings.edu/articles/the-three-challenges-of-ai-regulation/`.

World Economic Forum (2023). *The Future of Jobs Report 2023*. World Economic Forum.

Zack, Travis, Eric Lehman, Mirac Suzgun, Jorge A. Rodriguez, Leo Anthony Celi, Judy Gichoya, Dan Ju-
rafsky, Peter Szolovits, David W. Bates, Raja-Elie E. Abdulnour, Atul J. Butte, and Emily Alsentzer
(2023). *Coding Inequity: Assessing GPT-4's Potential for Perpetuating Racial and Gender Biases in
Healthcare*. Preprint. medRxiv.

Zeira, Joseph (1998). "Workers, Machines, and Economic Growth". *The Quarterly Journal of Economics*
113(4), pp. 1091–1117.

Zhuk, A (2023). "Navigating the Legal Landscape of AI Copyright: A Comparative Analysis of EU, US,
and Chinese Approaches". *AI Ethics*.

A   Appendix

A.1   Definition of AI in regulation

Table 2: AI definition in regulation

Country	Definition
EU	[EU AI Act 2021 proposal] "It proposes a single future-proof definition of AI." [Provisional Act March 13, 2024]"A key characteristic of AI systems is their capability to infer." "AI models [...] form part of AI systems."
US	[US Bill of Rights] AI as an "automated system". [EO] AI is "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments."
China	[Interim measures] "Art. 2: These measures apply to the use of generative AI technologies to provide services to the public in the [mainland] PRC for the generation of text, images, audio, video, or other content (hereinafter generative AI services)]."
UK	[White paper] "3.2.1 Defining Artificial Intelligence" "[...] AI by reference to the 2 characteristics that generate the need for a bespoke regulatory response [...]" [adaptivity and autonomy]

Note: this table covers AI-specific regulations. We do not cover *national* laws for the EU.

[**EU - The EU AI Act Proposal 2021**] *"It proposes a single future-proof definition of AI."* [...] *"5.2. Detailed explanation of the specific provisions of the proposal 5.2.1. SCOPE AND DEFINITIONS (TITLE I) Title I defines the subject matter of the regulation and the scope of application of the new rules that cover the placing on the market, putting into service and use of AI systems. It also sets out the definitions used throughout the instrument. The definition of AI system in the legal framework aims to be as technology neutral and future proof as possible, taking into account the fast technological and market developments related to AI. [. . .]" "(6) The notion of AI system should be clearly defined to ensure legal certainty, while providing the flexibility to accommodate future technological developments. The definition should be based on the key functional characteristics of the software, in particular the ability, for a given set of human-defined objectives, to generate outputs such as content, predictions, recommendations, or decisions which influence the environment with which the system interacts, be it in a physical or digital dimension. AI systems can be designed to operate with varying levels of autonomy and be used on a stand-alone basis or as a component of a product, irrespective of whether the system is physically integrated into the product (embedded) or serve the functionality of the product without being integrated therein (non-embedded). The definition of AI system should be complemented by a list of specific techniques and approaches used for its development, which should be kept up-to–date in the light of market and technological developments through the adoption of delegated acts by the Commission to amend that list."* (The EU AI Act (Proposal by EU Commission))

[**EU - Trilogue agreement December 8, 2023**] *General-purpose AI systems and foundation mod-*

els. New provisions have been added to take into account situations where AI systems can be used for many different purposes (general-purpose AI), and where general-purpose AI technology is subsequently integrated into another high-risk system. The provisional agreement also addresses the specific cases of general-purpose AI systems. Specific rules have been also agreed for foundation models, large systems capable to competently perform a wide range of distinctive tasks, such as generating video, text, images, conversing in lateral language, computing, or generating computer code. The provisional agreement provides that foundation models must comply with specific transparency obligations before they are placed in the market. A stricter regime was introduced for 'high impact' foundation models. These are foundation models trained with large amount of data and with advanced complexity, capabilities, and performance well above the average, which can disseminate systemic risks along the value chain. (Council of the EU - Press release)

[**EU - Provisional Act, March 13, 2024**] *"(12) [...](12) The notion of 'AI system' in this Regulation should be clearly defined and should be closely aligned with the work of international organisations working on AI to ensure legal certainty, facilitate international convergence and wide acceptance, while providing the flexibility to accommodate the rapid technological developments in this field. Moreover, it should be based on key characteristics of AI systems that distinguish it from simpler traditional software systems or programming approaches and should not cover systems that are based on the rules defined solely by natural persons to automatically execute operations. A key characteristic of AI systems is their capability to infer. This capability to infer refers to the process of obtaining the outputs, such as predictions, content, recommendations, or decisions, which can influence physical and virtual environments, and to a capability of AI systems to derive models or algorithms from inputs or data. The techniques that enable inference while building an AI system include machine learning approaches that learn from data how to achieve certain objectives, and logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved. The capacity of an AI system to infer transcends basic data processing, enables learning, reasoning or modelling. The term 'machine-based' refers to the fact that AI systems run on machines."(97)[...]Although AI models are essential components of AI systems, they do not constitute AI systems on their own. AI models require the addition of further components, such as for example a user interface, to become AI systems. AI models are typically integrated into and form part of AI systems. This Regulation provides specific rules for general-purpose AI models and for general-purpose AI models that pose systemic risks, which should apply also when these models are integrated or form part of an AI system. It should be understood that the obligations for the providers of general-purpose AI models should apply once the general-purpose AI models are placed on the market. (63) 'general-purpose AI model' means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the*

market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are released on the market; (64) 'high-impact capabilities' means capabilities that match or exceed the capabilities recorded in the most advanced general-purpose AI models; (65) 'systemic risk' means a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain; (66) 'general-purpose AI system' means an AI system which is based on a general-purpose AI model, that has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems; [...] Article 3 "Definitions" [...](1) 'AI system' means a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments;[...](Provisional Act full text)

[**US - Bill of Rights**] *An "automated system" is any system, software, or process that uses computation as whole or part of a system to determine outcomes, make or aid decisions, inform policy implementation, collect data or observations, or otherwise interact with individuals and/or communities. Automated systems include, but are not limited to, systems derived from machine learning, statistics, or other data processing or artificial intelligence techniques, and exclude passive computing infrastructure. "Passive computing infrastructure" is any intermediary technology that does not influence or determine the outcome of decision, make or aid in decisions, inform policy implementation, or collect data or observations, including web hosting, domain registration, networking, caching, data storage, or cybersecurity. Throughout this framework, automated systems that are considered in scope are only those that have the potential to meaningfully impact individuals' or communities' rights, opportunities, or access.*(US Bill of Rights)

[**US - "National AI act of 2020"**] *"The term 'artificial intelligence' means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments."* (State AI - National AI act)

[**US - Executive Order (EO) on Safe, Secure, and Trustworthy Artificial Intelligence**] *Sec. 3. Definitions. For purposes of this order: [...](b) The term "artificial intelligence" or "AI" has the meaning set forth in 15 USC. 9401(3): a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action. (c) The term "AI model" means*

a component of an information system that implements AI technology and uses computational, statistical, or machine-learning techniques to produce outputs from a given set of inputs.

In the same EO, *"(k) The term "dual-use foundation model" means an AI model that is trained on broad data; generally uses self-supervision; contains at least tens of billions of parameters; is applicable across a wide range of contexts; and that exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters, such as by:"*(EO)

[**China - Interim Measures for the Management of Generative Artificial Intelligence Services Current version August 2023**] *"Article 2: These measures apply to the use of generative AI technologies to provide services to the public in the [mainland] PRC for the generation of text, images, audio, video, or other content (hereinafter generative AI services)."* [Official version in Chinese] [Unofficial English translation] (Interim Measures for the Management of Generative Artificial Intelligence Services)

[**China - Measures on the Administration of Generative Artificial Intelligence Services (Draft for Solicitation of Comments) Old version April 2023**] *"Article 2: These measures apply to the research and development and the utilization of generative AI products, and to the provision of services to the public in the [mainland] PRC. "Generative artificial intelligence" as used in these Measures refers to technology that generates text, pictures, audio, video, code, or other content based on algorithms, models, and rules."*(Measures for the Management of Generative Artificial Intelligence Services (Draft for Comment)

[**UK - A pro-innovation approach to AI regulation, White paper**] *"3.2.1 Defining Artificial Intelligence. 39. To regulate AI effectively, and to support the clarity of our proposed framework, we need a common understanding of what is meant by 'artificial intelligence'. There is no general definition of AI that enjoys widespread consensus [One of the biggest problems in regulating AI is agreeing on a definition, Carnegie Endowment for International Peace, 2022] That is why we have defined AI by reference to the 2 characteristics that generate the need for a bespoke regulatory response. The 'adaptivity' of AI can make it difficult to explain the intent or logic of the system's outcomes: AI systems are 'trained' – once or continually – and operate by inferring patterns and connections in data which are often not easily discernible to humans. Through such training, AI systems often develop the ability to perform new forms of inference not directly envisioned by their human programmers. The 'autonomy' of AI can make it difficult to assign responsibility for outcomes: Some AI systems can make decisions without the express intent or ongoing control of a human."* (White paper)

B Details on regulation in AEs and China

B.1 EU regulation and reasons

General EU regulation:[128]

- **Market Competition**: The EU has several laws and regulations aimed at addressing monopolistic practices, including the EU Competition Law. The primary legislation in this area is the Treaty on the Functioning of the European Union (TFEU), specifically Articles 101 and 102, which prohibit anti-competitive agreements and abuse of dominant market positions (European Commission).

- **Privacy**: The General Data Protection Regulation (GDPR) is the key legislation in the EU that governs privacy and data protection. It sets out the rights and obligations regarding the processing of personal data and provides individuals with control over their personal information (European Commission).

- **Copyright**: The EU copyright law consists of 13 directives and 2 regulations, harmonising the essential rights of authors, performers, producers and broadcasters.(EU Digital Strategy). The European Commission has recently proposed a new legal framework for the intellectual property rights of AI-generated works, including copyright (see Zhuk (2023) for a comparison of regulatory frameworks).

- **Military and Security**: The EU does not have specific AI regulations focused solely on military and security. However, the EU Cybersecurity Act establishes a framework for the certification of cybersecurity products, services, and processes, which can be relevant to AI systems used in military and security contexts (European Commission).

- **Ethical and Bias**: The EU has proposed the Ethics Guidelines for Trustworthy AI developed by the High-Level Expert Group on AI. While not legally binding, these guidelines provide a framework for the ethical development and deployment of AI systems in the EU (European Commission).

- **Financial Stability**: The EU has various regulations and directives aimed at ensuring financial stability, such as the Markets in Financial Instruments Directive (MiFID II) and the Capital Requirements Directive (CRD IV). While these regulations do not specifically address AI, they provide a regulatory framework for financial institutions to ensure stability and risk management (European Commission).

---

[128] This section has been structured with the help of gen-AI tools.

Based on the EU AI Act (Provisional Act approved on March 13, 2024):

- **Market Competition**: The EU AI Act refers to the EU Competition laws.

- **Privacy**: The EU AI Act incorporates privacy and data protection principles, building upon the existing General Data Protection Regulation (GDPR). It will likely require AI systems to handle personal data in a transparent and secure manner, ensuring individuals' privacy rights are respected.

- **Copyright**: Provisions are included for providers of general-purpose AI models to comply to Union or national legislation on copyright.[129]

- **Military and Security**: The EU AI Act does not include provisions related to AI systems used *exclusively* in military and security contexts.[130]

- **Ethical and Bias**: The EU AI Act emphasizes the ethical development and deployment of AI systems in line with EU values and also refers to the Ethics Guidelines for Trustworthy AI.[131]

- **Financial Stability**: While the EU AI Act may not have specific provisions related to financial stability, it mention competent authorities for implementation of related provisions.[132]

---

[129] *"[...](106) Providers that place general-purpose AI models on the Union market should ensure compliance with the relevant obligations in this Regulation. To that end, providers of general-purpose AI models should put in place a policy to comply with Union law on copyright and related rights, [...]"*

[130] *Article 2 "[...] This Regulation does not apply to AI systems where and in so far they are placed on the market, put into service, or used with or without modification exclusively for military, defence or national security purposes, regardless of the type of entity carrying out those activities.*

[131] *"(27) While the risk-based approach is the basis for a proportionate and effective set of binding rules, it is important to recall the 2019 Ethics guidelines for trustworthy AI developed by the independent AI HLEG appointed by the Commission. In those guidelines, the AI HLEG developed seven non-binding ethical principles for AI which are intended to help ensure that AI is trustworthy and ethically sound.The seven principles include human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being and accountability. Without prejudice to the legally binding requirements of this Regulation and any other applicable Union law, those guidelines contribute to the design of a coherent, trustworthy and human-centric AI, in line with the Charter and with the values on which the Union is founded. According to the guidelines of the AI HLEG, human agency and oversight means that AI systems are developed and used as a tool that serves people, respects human dignity and personal autonomy, and that is functioning in a way that can be appropriately controlled and overseen by humans.*

[132] *"(158) Union financial services law includes internal governance and risk-management rules and requirements which are applicable to regulated financial institutions in the course of provision of those services, including when they make use of AI systems. In order to ensure coherent application and enforcement of the obligations under this Regulation and relevant rules and requirements of the Union financial services legal acts, the competent authorities for the supervision and enforcement of those legal acts, in particular competent authorities as defined [...] should be designated, within their respective competences, as competent authorities for the purpose of supervising the implementation of this Regulation, including for market surveillance activities, as regards AI systems provided or used by regulated and supervised financial institutions unless Member States decide to designate another authority to fulfil these market surveillance tasks.."*

B.2 US regulation and reasons

- **Market Competition**: In the US, antitrust laws aim to prevent monopolistic practices and promote fair competition. The main pieces of legislation are the Sherman Antitrust Act and the Clayton Act, enforced by the Federal Trade Commission (FTC) and the Department of Justice (DOJ) (Acts)

- **Privacy**: Privacy regulations in the US are primarily governed by sector-specific laws.

- **Copyright**: Copyright in the US is specifically covered in different laws by items. (Copyright Law of the US)

- **Military and Security**: The US has various regulations and policies related to AI in military and security contexts. The Department of Defense (DoD) has issued an AI adoption strategy. Additionally, the National Security Commission on Artificial Intelligence (NSCAI) has provided recommendations on AI for national security and defence (DoD AI Adoption Strategy, NSCAI Report).

- **Ethical and Bias**: While there is no comprehensive federal law specifically addressing the ethical development and deployment of AI, there are ongoing discussions and initiatives to address AI bias and ethics. Organizations such as the AI Now Institute and the Partnership on AI have developed guidelines and recommendations for ethical AI practices (AI Now Institute, Partnership on AI).

- **Financial Stability**: The US has various financial regulations and oversight bodies to ensure financial stability, such as the Securities and Exchange Commission (SEC) and the Federal Reserve. While these regulations do not specifically target AI, they provide a regulatory framework for financial institutions to manage risks associated with AI applications in the financial sector (SEC, Federal Reserve).

Based on the "Executive Order (EO) on Safe, Secure, and Trustworthy Artificial Intelligence":

- **Market Competition**: The EO does specifically address monopolistic practices in the AI sector.

- **Privacy**: The EO emphasizes the protection of privacy and civil liberties in the development and use of AI systems. It calls for the adoption of privacy-preserving AI techniques and compliance with relevant privacy laws and regulations.

- **Copyright**: The executive order recommends further executive actions on copyright and AI with the US Copyright Office responsible to publish a study on AI in copyright issues.

---

[133]This section has been structured with the help of gen-AI tools.

- **Military and Security**: The EO highlights the importance of AI in national security and defense. It emphasizes the need for the development and deployment of AI systems that are safe, secure, and resilient, particularly in critical infrastructure and defense applications.

- **Ethical and Bias**: The EO emphasizes the importance of promoting AI systems that are trustworthy, transparent, and accountable. It calls for the avoidance of unfair bias and discrimination in AI systems and encourages the development of standards and best practices for AI ethics.

- **Financial Stability**: The EO does not specifically address financial stability in relation to AI. However, the EO mandates the release of best practices for financial institutions managing cybersecurity risks.

B.3 China regulation and reasons

General Chinese Regulation:

- **Market Competition**: China has proposed new regulations aimed at curbing the power of its biggest internet companies, potentially signaling a new era of scrutiny over the country's tech giants. (China steps up tech scrutiny with rules over unfair competition, critical data — Reuters).

- **Privacy**: China's Personal Information Protection Law (PIPL) and Data Security Law (DSL) are two major regulations that have been proposed to protect personal data and regulate data-related activities. These laws aim to protect the rights and interests of individuals, and safeguard national security and public interest (Why China's New Data Security Law Is a Warning for the Future of Data Governance – Foreign Policy) (PBOC head urges fintech to secure data (www.gov.cn))

- **Copyright**: China has a law on copyright, but it is not clear if applicable to AI-generated material as well. AI as an entity cannot be protected by the law.(Copyright Law of the People's Republic of China, 2010 amendment)

- **Military and Security**: China's AI strategy includes a focus on military applications. The country has proposed to strengthen the security review and oversight of AI, and promote the application of AI in defense technology and military operations. (The PLA's Strategic Support Force and AI Innovation — Brookings)

- **Ethical and Bias**: China has proposed guidelines for AI development, which include the principle of fairness and non-discrimination. (Global AI Governance Initiative)

- **Financial Stability**: This aspect is included in the general regulation, together with FinTech (Will China's new financial regulatory reform be enough to meet the challenges? (bruegel.org)).

Based on the "Interim Measures for the Management of Generative Artificial Intelligence Services":

- **Market Competition**: Intellectual property rights should be protected and advantages must not be used for monopolies or to carry out unfair competition.

- **Privacy**: Providers are required to assume responsibility as a producer of online information content and fulfill online information security obligations. The measures proscribe not only unlawfully retaining input information that can identify a user or providing users' input information to others, but also "collecting unnecessary personal information".

- **Copyright**: There are no direct provisions in the interim act.

- **Military and Security**: The regulatory objective of China concerning generative AI includes safeguarding national security and social public interests.[134]

- **Ethical and Bias**: The measures require service providers to *"employ effective measures to improve the quality of training data and to enhance the data's veracity, accuracy, objectivity, and diversity."* This could be interpreted as a measure to prevent bias in AI (Art. 7).

- **Financial Stability:** The text does not provide specific information on Financial Stability.

---

[134] *"Article 4: The provision and use of generative AI services shall comply with the requirements of laws and administrative regulations, respect social mores, ethics, and morality, and obey the following provisions: (1) Uphold the Core Socialist Values; content that is prohibited by laws and administrative regulations such as that inciting subversion of national sovereignty or the overturn of the socialist system, endangering national security and interests or harming the nation's image, inciting separatism or undermining national unity and social stability, advocating terrorism or extremism, promoting ethnic hatred and ethnic discrimination, violence and obscenity, as well as fake and harmful information; [...]"*

B.4   UK regulation and reasons

- **Market Competition**: The UK's Competition and Markets Authority (CMA) is responsible for preventing and reducing anti-competitive activities, including those related to AI.[136]

- **Privacy**: The UK's Information Commissioner's Office (ICO) is responsible for upholding data protection and privacy rights in the context of AI. The ICO enforces these laws and provides guidance on compliance, including the use of AI in processing personal data. (Guidance on AI and data protection)

- **Copyright**: The Government is working on a "Code of Practice" on copyright and AI with users and rights holders.(The government's code of practice on copyright and AI)

- **Military and security:** The UK's Ministry of Defence (MOD) has issued guidelines and policies for the ethical development and use of AI in military and security contexts. It also provides principles and guidance for responsible AI use in defense operations. (Defence Artificial Intelligence Strategy - GOV.UK (www.gov.uk))

- **Ethical and Bias:** The UK government recognizes the importance of ethical and unbiased AI. The Centre for Data Ethics and Innovation (CDEI) provides guidance and recommendations on the ethical use of AI. They have published reports on topics such as bias in algorithmic decision-making and AI in the criminal justice system. Other organizations, such as the AI Council, also advise the government on AI policy and strategy. (CDEI), (AI Council).

- **Financial Stability:** The UK has various financial regulations and oversight bodies to ensure financial stability, although specific regulations targeting AI are still developing.

---

[135]This section has been structured with the help of gen-AI tools.
[136]https://www.morganlewis.com/pubs/2023/10/uk-competition-and-markets-authority-releases-initial-ai-foundation-models-rep

PUBLICATIONS