Executive summary

Bertin Martens (bertin.martens@bruegel.org) is a Senior Fellow at Bruegel

European Union policymakers want to close the arti cial intelligence innovation gap with the United States, as a way to accelerate lagging productivity growth. e EU focus is on expanding an existing supercomputer network with more AI hardware and computing infrastructure, with taxpayer support. However, this computing infrastructure is not adapted to AI modelling. e cost of catching up with leading big tech AI computing centres is already prohibit1tinastrg EU btinet o AI moberveprescо(d)ci-15r er,3 (l(delliok,3 (s mit of s,3 ((g cen)7Er)85U mr)15 (d9)4 (collaborate with US big tech rms. Injecting taxpayer subsidies to make up for these missing markets may further distort EU markets. Regulatory compliance costs, including uncertainty about the implementation arrangements for the EU Arti cial Intelligence Act, add to market problems.

The EU should address a wider range of market failures in its policy initiatives. It should strive to increase productivity growth below the AI technology frontier, by facilitating investment and applications of AI-driven services produced by derived and specialised

bruegel

1 Introduction

In the first half of 2024 alone, more than $35 billion was invested globally into artificial intelligence startups[1]. The European Union attracted only 6 percent of that. The EU is doing better in AI patents and in training AI researchers, but the outputs from this tend not to stay in the EU, but rather to flow to the United States (Renda, 2024). Unsurprisingly, this situation has triggered considerable debate in EU policy circles about what can be done at EU level so the EU can catch up with the US and China on AI, in particular by developing its own AI models, fostering more AI startups, accelerating the uptake of AI-based services in the EU economy.

In this context, the European Commission in January 2024 published a package of proposals, decisions and plans to support AI startups[2]. This seeks to capitalise on the European High-Performance Computing (EuroHPC) network of supercomputers – very large, high-performing computers – used primarily for scientific research. The Commission proposed an amendment to the network's governance rules to facilitate collaboration with the private sector – that amendment has since been adopted (Regulation (EU) 2024/1732). The plan is that EuroHPC should be the core of a network of 'AI factories' for the development by EU startups of large-scale general purpose AI models and applications.

This approach recognises that these supercomputers need to be upgraded to AI capabilities, to be financed equally by the EU and the computer-hosting EU countries[3]. But the EuroHPC budget of €7 billion for 2021-2027 remains for now unchanged[4].

The AI computing infrastructure budget could be increased very substantially if the Commission and EU countries listen to former Italian prime minister and European Central Bank governor Mario Draghi. His September 2024 report on the future of European competitiveness, produced to steer EU policy in the next five years (Draghi, 2024), attributed the EU's weak productivity growth to insufficient investment and uptake of digital technologies, including AI.

The consensus in EU policy circles is that catching up on AI requires public sector involvement and subsidies

His proposed remedies include private and public investment in EU-developed general and sectoral AI models, upgrading EuroHPC, creating an AI incubator similar to that of the CERN nuclear and particle physics laboratory, creating EU-wide large data pools for AI model training, facilitating consolidation among EU cloud providers to create hyperscale computing infrastructure and more financial resources for quantum computing. Draghi (2024) also recognised that the EuroHPC computers cannot compete with US-based hyperscale AI firms and proposed to allocate €100 billion for AI infrastructure.

All this suggests a consensus in EU policy circles that catching up on AI requires public sector involvement and subsidies. There has been less analysis, however, of why the EU AI value chain and business ecosystem have ended up falling behind the US and China in terms of AI model development[5] and uptake in services industries, and why this should justify public sector involvement and subsidies. There is even less debate on how these problems could be addressed through structural reforms that could incentivise more private investment in EU AI industries.

This Policy Brief explores why EU AI investment has fallen behind the US and the types of market failure that may have led to that situation. We ask how the EU should position itself

1 Joanna Glasner, 'AI Gobbled A Record Share Of Startup Funding This Year,' [illegible], 4 September 2024, https://news.crunchbase.com/ai/record-share-startup-funding-2024-xai-anthropic/.

2 See European Commission press release of 24 January 2024, 'Commission launches AI innovation package to support Artificial Intelligence startups and SMEs,' https://ec.europa.eu/commission/presscorner/detail/en/ip_24_383 and European Commission (2024).

3 The nine supercomputers are hosted in different countries; see https://eurohpc-ju.europa.eu/supercomputers/our-supercomputers_en.

4 Other initiatives, complementary to EuroHPC and including a European 'CERN for AI' and other moonshot AI initiatives, have been proposed. For an overview, see Renda (2024).

5 For more details on the global competitive landscape in AI modelling, see Martens (2024b).

in the competition over AI and discuss two possible responses. Should the EU try to catch up with the US, reach the AI technology frontier and develop its own AI capacities, independent of US big tech firms? Or can the EU prosper below the AI technology frontier, in derived AI products and services markets? We also look at the geopolitical context and the risks of EU dependence on US big tech.

2 Building AI models on existing EU supercomputers?

It is clear that the EU is running behind the US in digital technology investment and uptake in general, and in AI specifically. It is less clear whether the response should be to invest taxpayer money in physical infrastructures for AI, as advocated by the Commission policy initiatives and Draghi (2024). It might be possible to resolve some market and regulatory failures in AI-related markets with public money, but many other problems cannot be resolved this way. In this section, we discuss the main considerations that should be factored in to the EU approach to AI.

2.1 Computing hardware issues

The EuroHPC network of nine supercomputers is not up to the task of delivering a state-of-the-art AI computing infrastructure for commercial use. These computers were designed for scientific research, not for training of general-purpose AI models or generative AI models like ChatGPT[6]. Their hardware architecture is not suitable for that purpose. They have no more than a few thousand Nvidia graphics processing units (GPUs) that play a central role in AI model training. This is a tiny capacity compared to Meta's most advanced AI computing centre, which reportedly contains 600,000 Nvidia AI chips[7].

Hobbhahn *et al* (2023) explained how AI hardware differs from classic computing architectures that revolves around central processing units (CPUs). Handling the massive amounts of data in GenAI model training requires GPUs. Nvidia became successful in AI hardware because of its original specialisation in GPUs for gaming applications. Handling data traffic between many thousands of GPUs requires extensive communication bandwidth between GPUs and memory storage, though one way to reduce computational requirements can be to reduce the number of digits behind the decimal point in calculations[8]. AI developers are increasingly designing their own dedicated hardware, including for specific applications such as inference, meaning the making of predictions based on newly supplied data after the model has been trained.

2.2 AI infrastructure costs

Nvidia AI chips each cost more than $30,000. For Meta's most advanced computing centre with 600,000 of these chips, this amounts to $18 billion for the dedicated AI chips alone, excluding other hardware needs. In other words, the cost of chips for a single computing centre is more than twice the current EuroHPC budget.

6 We define Generative AI models as machine learning and neural network models that apply the 'transformer' architecture (Vaswani *et al.*, 2017).

Moreover, technological progress in AI chips is so fast that the latest generation of AI chips will be outdated and written o in less than a year (Hobbhahn *a*, 2023). Spending $162 billion per year (ie nine EuroHPC supercomputers x $18 billion/year) is simply beyond the nancial resources of the EU. Even if the EuroHPC network were to be upgraded to train state-of-the-art AI models, it would still have a hard time running these models on a daily basis to respond to user queries because that requires additional investment in a di erent type of inference accelerator chip, such as NVIDIA's Jetson processors, to reduce the cost of responding to user queries.

e costs of training state-of-the-art generative AI models (ie those that can produce new images, video, audio or text based on prompts) are exploding, running into hundreds of millions of euros (Martens, 2024a). Cottier *a* (2024) estimated that GenAI model training costs are increasing exponentially by a factor 2.4 to 2.6 per year, or around 240 percent per year from 2016 to 2023. Extrapolating the costs of the largest frontier models now to 2030 leads to an estimated training cost for a single GenAI model of $60 billion.

New frontier GenAI models are coming out every week. Cottier *a* (2024) also estimated the cost of AI computing infrastructure at ten times the cost of model training. at infrastructure can be used to train several models but the hardware amortisation rate is estimated at 140 percent per year, or 100 percent depreciation in 8.5 months. By that time, a new generation of AI computing chips will have arrived with superior performance. Infrastructure costs for GPT4 by the end of 2023 may have been as high as $800 million. Extrapolation could push t (t p)-2 (er3.9 4 T3 (d tur)15 (e cosur(e b)-((v)3 (e b)-2t50TJ1.444 end of 2ds t7 (t i)7 i-3 (c)5 (a)7 5 ()85 (us)-2 factories' initiative, and in the AI recommendations in Draghi (2024), is problematic[10]. But the exclusive hardware focus of these plans is not surprising. Eckert (2024) presented an insightful historic overview of EU digital policies over the past 40 years. A recurrent pattern has been the emphasis on telecoms infrastructure and hardware in general, and the almost total absence of digital services markets and business model considerations. Draghi (2024, Part B, Figure 4) showed how the value of telecoms services has become negligible compared to digital services markets. Nevertheless, his recommendations focused on telecoms, cloud and AI hardware, and do not mention digital or AI services markets. More than anything, four decades of path dependency in EU digital policies may have contributed to an ever wider yawning gap between EU and US digital performance – which still continues today.

Draghi (2024) pointed out that the EU should do more to create its own hyperscale cloud computing infrastructure in support of generative AI model development, and reduce dependence on the US big tech rms that currently dominate the cloud services market in the EU[11].

ere may be competition failures in EU cloud computing services, another important complementary input for AI. A few big tech players can leverage their positions in cloud software- and platforms-as-a-service, rather than just o ering basic infrastructure-as-a-service[12]. is increases entry barriers for smaller EU cloud service providers, leaving them unable to expand their computing infrastructure, which would be suitable for AI model training (Ennis and Evans, 2024; Biglaiser *a* , 2024). rowing taxpayer money at this problem is unlikely to be a good solution, however. Draghi (2024) recommended consolidation among smaller EU cloud players. at does not solve the problem of lack of complementary software and platform services.

2.5 Derived AI model markets are very competitive

ere is no indication of a market failure that would require public policy intervention, let alone taxpayer subsidies, in derived and special applications of AI models. Draghi (2024) recommended that EU AI funds could support European AI startups to develop speci c industry or company application models. at market is already very competitive (Martens, 2024b). While more than a dozen new state-of-the-art GenAI models are released every month, more than a dozen derived models are released per hour[13]. Just as app stores for mobile phones contain millions of special-purpose apps, there are now also millions of industry-, sector- or company-speci c applications of the ChatGPT model in the OpenAI store.

For example, there are ChatGPT applications that help consumers with their shopping questions or nancial decisions. Developers of these applications make them widely available to anyone who can use them. A derived model is created when a company uploads its own proprietary data into ChatGPT for speci c marketing, logistics or industrial process applications within the company. Since they run on proprietary data, these models are of course not made widely available.

2.6 Risk of regulatory failure

Policy intervention may create new market distortions in AI services markets. e amended EuroHPC regulation (Regulation (EU) 2024/1732) now allows collaboration between public and private computing and cloud services providers. Commercial rms can access publicly-owned computers.

10

is raises the question of how scarce computing capacity will be allocated between users, at what price and under what conditions. e amended EuroHPC regulation does not explain this. Will authorities use auctions for commercial applications and sell capacity at market prices? Or will there be a subsidy component in pricing, thereby opening the door to unfair competition with private providers? How will capacity be allocated between paid commercial and presumably unpaid non-commercial use, for instance for scienti c projects, which the EuroHPC network was involved in from the start? More importantly for AI start-ups, what happens after the training of their AI model has been completed? Will they have guaranteed access for inference, daily running of their models? Can they easily scale up capacity when their startup rapidly expands? Computers may be provided by the public sector but they are not non-rival non-excludable public goods. ey are rival and easily excludable.

e European Commission and Draghi (2024) claim that these AI policy initiatives can capitalise on EU regulation, including the EU AI Act (Regulation (EU) 2024/1689; see Box 1), the general data protection regulation (GDPR, Regulation (EU) 2016/679) and other data regulations. e claim is that these EU regulations attract investment because they give users con dence and regulatory certainty.

e available empirical evidence, however, does not support that view. ere is considerable evidence that the GDPR has reduced investment in consumer-oriented online services in the EU (Demirer *a*, 2024; Goldberg *a*, 2023; Jia *a*, 2023; Peukert *a*, 2024). Consumers may be better o without some of these privacy-infringing services, though that may not be the case for all.

ere is also evidence that EU regulation is limiting EU access to AI services. At the request of the Irish Data Protection Commission, Meta held back the roll-out of its most advanced AI models in the EU[14]. European data regulators have doubts that the legitimate interest clause in the GDPR (Article 6(1)(f)) constitutes a su cient legal basis for Meta to use publicly posted messages on its Facebook and Instagram social media platforms as inputs for AI model training. Other US AI developers, including Apple, Google and OpenAI, face similar EU uncertainty about the use of personal data for model training[15]. e Irish Data Protection Commission launched an enquiry into Google's AI services[16]. Social media text has become an important source of AI model training data when other sources are insu cient to meet the training requirements of very large AI models, especially in relation to less-widely spoken languages, for which the available volume of human text data is limited. Regulatory uncertainty about this alternative sourc.9 (1 1 Tf[, 2024; G)-11.9 (,th)7.1 [,lt (io)-0.9(um)7 (ot)1 (e)-(ur)15.1ices)17.1

not only high compliance costs for model developers and deployers, but also considerable regulatory uncertainty regarding the specific implementation rules for copyright and privacy protection. The finalisation of the AI Act in mid-2024 was only the start of a regulatory process that will take several years to complete dozens of implementation guidelines and enforcement standards, including on copyright and data privacy.

Box 1: The EU AI Act

The Artificial Intelligence Act (Regulation (EU) 2024/1689), finalised in mid-2024, is intended to regulate AI in the EU by banning certain applications that impinge in citizens' rights and creating a category of high-risk systems and uses, for which risk assessments and measures to offset risks will be required. Decisions taken by high-risk systems should in principle be explainable and appealable. The law also contains transparency requirements, such as labelling obligations for AI-generated images, audio or video, and obliges compliance with EU copyright rules. Parts of the law are being phased in, but it will apply in full from August 2026.

The AI Act also created an AI Office, which was established in May 2024[17], as a monitoring, supervisory and enforcement body in relation to general purpose AI models and systems. Among its responsibilities will be development of specific implementation rules, including on AI and copyright and privacy protection[18].

The text of the AI Act is available at http://data.europa.eu/eli/reg/2024/1689/oj.

3 Elements of an EU AI strategy

In summary, the EU's current approach to AI is based on catching up on AI hardware and infrastructure, while omitting the complementary business model components and not addressing high regulatory uncertainty and compliance costs. Such an approach is unlikely to solve the fundamental AI competitiveness problem because of the shortcomings set out in the previous section. To address these shortcomings, the EU strategy should include the elements we set out here. Overall, it would be a mistake for the EU to try to play the US at its own game on AI – to reach the AI technology frontier and develop its own AI capacities. Instead, the EU can thrive with smaller models to help firms implement AI-driven services. It does not need to reach the AI technology frontier to accelerate AI-driven productivity growth.

3.1 Facilitate collaboration agreements

Complementary inputs and business ecosystems cannot be created by regulation or public money. They need to grow organically. Competition authorities are taking a close look at collaboration agreements between startups and big tech firms, sometimes rightly so because they may contain exclusivity clauses that distort competition. At the same time, these collaboration agreements and even mergers are necessary to provide the complementary inputs that AI start-ups require. Short of exclusionary contractual clauses, such agreements and mergers should be allowed to go through. Rather than cutting off startups from the complementary

17 See European Commission press release of 29 May 2024, 'Commission establishes AI Office to strengthen EU leadership in safe and trustworthy Artificial Intelligence', https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2982.

18 See https://artificialintelligenceact.eu/ai-act-implementation-next-steps/.

inputs they need, EU regulators should focus on solving the missing market failure in private equity markets (as advocated by Draghi, 2024).

3.2 Pro-innovation implementation of the AI Act

The AI Office within the European Commission is in charge of implementing the AI Act, including by designing implementation guidelines and standards (see Box 1). The office should have a razor-sharp focus on pro-innovation implementation and enforcement of the AI Act, minimising compliance costs and navigating the potential pitfalls of strict enforcement. Strict enforcement of existing EU copyright and privacy law is likely to create significant obstacles for AI industries in the EU. The AI Office will have to define an appropriate trade-off between private rights, including the protection of copyright and privacy, and the need to support the development and use of AI services for the benefit of society as a whole.

3.3 Productivity growth below the AI technology frontier

Apart from trying overcoming these market and regulatory failures through regulatory reform, rather than subsidies, what can the EU do set up a pro-active and pro-competitive AI strategy? Would the EU be better off trying to reach the AI technology frontier, or can it prosper below the frontier?

Because of delays in AI productivity uptake (Brynjolfsson *et al.*, 2020), most productivity growth will take place below the frontier of the latest generation of GenAI models. Much of the roll-out of AI as a general-purpose technology across the economy will come from derived, smaller and more specialised AI models that can be trained and run at far lower computing costs[19]. AI applications that can retrieve data in real-time from various sources to respond to user queries will become an important workhorse for industrial applications (Lewis *et al.*, 2024). The CEO of SAP, one of Europe's leading AI applications companies, has argued in favour of smaller AI models[20]. The focus should be on specialised models deed 6-2 (e tr)11a-11 (ew) markets for complementary inputs by focusing on smaller, derived and specialised AI models that do not require hyperscale infrastructure and business ecosystems and could still earn a decent rate of return for patents and skilled researchers. Smaller venture capital funds and private equity could gradually move into that market. Medium-sized EU cloud service providers could expand their infrastructures and services to accommodate smaller AI models and inference operations.

3.5 Geopolitical dependency

In the current geopolitical security setting, can the EU and US be considered as a single and trustworthy AI market, or are they two separate markets?

Admittedly, pursuing economic e ciency below the AI frontier would come at the risk of leaving EU AI industrie to some extent dependent on GenAI frontier models developed and/