Centrum für Europäische Politik
FREIBURG | BERLIN

cepInput

No 1 | 2026 20 January 2026

Competition at Risk in Generative AI Markets

An Updated Assessment of Competition Dynamics in the Value Chain

Anselm Küsters and Matthias Kullas

Source: Figure created via LLM prompting.

In March 2024, we published an early assessment of the generative AI value chain. Two years later, many of our original concerns have been borne out, while new techno-economic dynamics have emerged that require more consideration by EU competition authorities. The most significant developments since our initial study include the dramatic increase in the use of open-source, efficient AI models, the controversial implementation of the EU AI Act, and the emergence of agentic, personalised AI systems as the next competitive frontier.

- Compute remains the most significant bottleneck, exacerbated by increasing energy demands and pressure on the grid. NVIDIA's dominance increased further, despite the rise of Neo Clouds and Google's TPUs.
- Foundation models have become notably more competitive than we anticipated. The market has evolved from potential "winner-takes-all" concentration toward multi-polar competition among OpenAI, Anthropic, Google, and others. Open-source models (Llama, DeepSeek) have proven viable alternatives.
- Training data concerns have only moderately increased, with exclusive licensing deals between major AI labs and publishers proliferating and the shift to reinforcement learning necessitating more high-quality data. However, collective licensing and the accelerating role of synthetic data provide counterbalancing forces.
- Downstream AI services warrant increased attention. The emergence of agentic, personalised AI systems as critical infrastructure and Big Tech's integration of genAI into existing products create concentration risks.
- Traditional three-to-five-year investigation timelines are ill-suited to an ecosystem where competitive dynamics can reverse multiple times during a single investigation. Maintaining enforcement effectiveness demands institutional reforms such as (partially automated) monthly market assessments by DG COMP, fast-track merger reviews for major AI partnerships, and structured coordination with the AI Office.

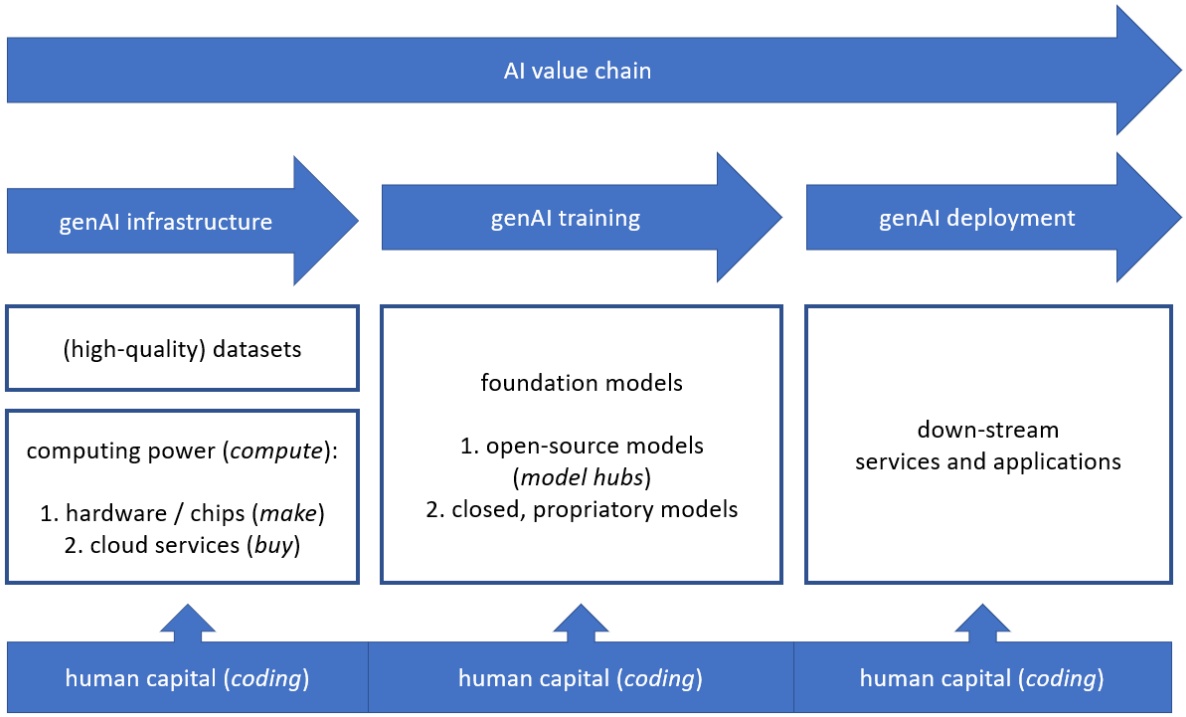
Content

Figures

1 Introduction

When we published our initial analysis of the emerging market for Artificial Intelligence (AI) models and services in March 2024,[1] following the exponential success of OpenAI's ChatGPT bot since its release in late 2022, generative AI (genAI) had already begun to reshape economies and societies across the world. Our study divided the genAI value chain into three interconnected layers, as reprinted in Figure 1: **infrastructure**, encompassing both training data and computing power; **foundation model** training; and **downstream deployment** in services and products. At the time, we warned that the combination of high fixed costs, network effects, switching costs, and vertical integration posed significant risks to competition. We also cautioned that this development could concentrate market power in ways that threaten economic efficiency, democratic values, and EU digital sovereignty.

Fig. 1: Value chain for generative AI as a service



Source: Own illustration.

The subsequent two years have validated several of our predictions while complicating others. The AI market has grown explosively: According to PitchBook data, AI startups attracted USD 192.7 billion through Q3 2025.[2] The European AI ecosystem has shown some dynamism: Mistral AI reached a €11.7 billion valuation in September 2025[3] and the EU AI Act entered its phased enforcement phase, with obligations for providers of General Purpose AI (GPAI) models becoming applicable in August 2025.[4] Yet the fundamental competitive tensions we identified remain. NVIDIA's dominance in AI chips has intensified further, with the company commanding up to 92% of the AI GPU market.[5] The oligopoly of

[1] See: https://www.cep.eu/de/eu-themen/details/competition-in-generative-artificial-intelligence-cepinput.html.
[2] See: https://techcrunch.com/2025/10/04/if-youre-not-an-ai-startup-good-luck-raising-money-from-vcs/.
[3] See: https://www.cnbc.com/2025/09/09/ai-firm-mistral-valued-at-14-billion-as-asml-takes-major-stake.html.
[4] See: https://artificialintelligenceact.eu/implementation-timeline/.
[5] See, although only up to Q1 2025: https://wccftech.com/nvidia-dominates-aib-gpu-market-share-in-q1-2025-amd-intel-drop/#:~:text=NVIDIA%20Leads%202025%20With%20Massive,PC%20market%20have%20been%20published. Other sources cite an „an estimated 80-90% market share in AI training chips": https://us.plus500.com/en/newsandmarketinsights/nvidia-earnings-power-global-tech-rally-as-imf-warns-on-growth.

cloud hyperscalers largely persists.[6] While the development of foundation models has become more competitive, concerns have emerged around "agentic AI" systems and platform lock-in through personalisation, as conversation/context data that emerge from the interaction with an AI bot cannot be exported to other model developers and providers. One potential solution that could be explored in the future would be to demand "AI conversation portability", meaning dominant AI platforms such as ChatGPT would need to provide users with "complete, machine-readable exports of their LLM interactions, behaviors, and algorithms".[7]

Against the backdrop of rapid economic and technological change, **this updated assessment revisits each layer of the genAI value chain** (section 2). Each sub-section begins by briefly summarising our findings from March 2024 and then provides a revised analysis, informed by current market data, technological trends, and regulatory developments. On this basis, we then describe some cross-cutting topics and concerns (section 3). We conclude by deriving institutional recommendations for European competition authorities observing the genAI value chain (section 4).

2 GenAI infrastructure: an updated assessment

2.1 Training data

March 2024 assessment: Our original study found no immediate competition problems in the market for high-quality training data. However, we flagged exclusive licensing agreements between AI developers and media companies as warranting close monitoring. We warned that if data became better protected and exclusively licensed, new entrants might face significant barriers to training competitive models.

Since our initial assessment, the existence of and demand for training data landscape has evolved, though not entirely in the directions we anticipated. Our concern about exclusive licensing has proven partially justified, but the market has also developed mechanisms that may mitigate these risks.

The bilateral deals we observed in early 2024 have proliferated into a **systematic licensing ecosystem**. OpenAI now maintains agreements with major publishers including the Associated Press, Politico, Business Insider, Axel Springer, Financial Times, Vox Media, Time, Le Monde, and others.[8] Google has partnerships with over 2,800 publications across more than 31 countries through News Showcase.[9] The Wall Street Journal reported in July 2025 that Amazon's licensing deal with the New York Times is worth USD 20 million to USD 25 million per year.[10] The deal apparently covers NYT content for Alexa and AI model training. Furthermore, the **structures of these deals have become increasingly sophisticated and complex**. Fixed upfront payments range include USD 10 million in the case of Taylor & Francis from Microsoft.[11] While Meta and Reuters did announce a multi-year AI content licensing deal, the financial terms were not publicly disclosed.[12] News Corp-OpenAI deal was reported at over

6 See: https://www.quantumrun.com/consulting/cloud-market-share/.
7 For this proposal, see: https://www.diplomaticourier.com/posts/reimagining-ai-data-ownership-models-for-prosperity.
8 See: https://openai.com/index/conde-nast/.
9 See: https://blog.google/products/news/news-showcase-is-launching-in-switzerland/.
10 See: https://www.wsj.com/business/media/amazon-to-pay-new-york-times-at-least-20-million-a-year-in-ai-deal-66db8503.
11 See: https://www.thebookseller.com/news/academic-authors-shocked-after-taylor--francis-sells-access-to-their-research-to-microsoft-ai.
12 See: https://www.reuters.com/technology/artificial-intelligence/meta-platforms-use-reuters-news-content-ai-chatbot-2024-10-25/.

USD 250 million over five years, including cash and OpenAI technology credits.[13] These costs remain manageable for frontier AI labs but represent meaningful barriers for smaller entrants. The deals are so complex that they raise questions of "circular financing" and an "AI bubble" (section 3.2).

Licensing deals are a strategic response by AI providers to both regulatory pressure and mounting copyright litigation. For instance, the EU AI Act's transparency requirements for GPAI models compel developers to disclose summaries of training data sources and categories (section 3.1). The USD 1.5 billion Anthropic copyright settlement with authors in September 2025 established a landmark precedent for AI training data compensation at approximately USD 3,000 per copyrighted work.[14] Again, this settlement and similar licensing requirements may paradoxically reinforce incumbent advantages: while USD 3,000 per work appears modest for well-funded companies like Anthropic or OpenAI, systematic licensing costs could prove prohibitive for smaller AI developers and European startups. Meanwhile, the November 2025 Munich Regional Court ruling in *GEMA v. OpenAI* marked the first EU judicial decision on the lawfulness of AI training on copyrighted material, holding that "memorization" of copyrighted works in model parameters constitutes copyright-relevant reproduction and that text-and-data-mining exceptions do not cover such reproductions.[15] The court ordered OpenAI to cease reproducing song lyrics both in its models and in outputs, potentially requiring the company to stop offering ChatGPT in Germany unless it licenses the works or develops new models. This ruling signals a stricter European approach compared to the US, where courts have shown greater deference to fair use arguments. In the future, the **territorial fragmentation of AI copyright regulation could shape competitive dynamics**, particularly if European jurisdictions systematically require licensing while US markets permit broader fair use defences.

A potential counterforce might be **Real Simple Licensing** (RSL), launched in September 2025 and backed by major web publishers including Reddit, Quora, and Yahoo.[16] Co-founded by RSS co-creator Eckart Walther, RSL aims to create machine-readable licensing agreements at internet scale. If taken up by many followers, such a standardized approach could democratize data access. Rather than negotiating bilateral deals, AI companies would then obtain licenses through a collective organization. However, whether frontier AI labs will participate remains uncertain at the time of writing.

Moreover, our March 2024 observation that smaller models with efficient training could reduce data requirements has been vindicated. Over the past two years, **synthetic data has become an important part in AI model training**. For its famous R1 model, the Chinse AI lab DeepSeek used synthetic data, primarily 600,000 reasoning-focused samples, to fine-tune its models, which significantly improved performance for tasks like math and coding.[17] Gartner predicts 80% of AI training data will be synthetic by 2028.[18] Major AI firms including NVIDIA and Databricks have built scalable synthetic data pipelines.[19] This shift reduces, though does not eliminate, the strategic importance of exclusive access to proprietary human-generated data. Critically, the current trend of doing RL-based training

[13] See: https://www.wsj.com/business/media/openai-news-corp-strike-deal-23f186ba.
[14] See: https://www.ropesgray.com/en/insights/alerts/2025/09/anthropics-landmark-copyright-settlement-implications-for-ai-developers-and-enterprise-users.
[15] Linda Kuschel and Darius Rostam, 'If It Looks Like a Duck', *Verfassungsblog: On Matters Constitutional*, ahead of print, Fachinformationsdienst für internationale und interdisziplinäre Rechtsforschung, 19 November 2025, https://doi.org/10.17176/20251120-142049-0.
[16] See: https://techcrunch.com/2025/09/10/rss-co-creator-launches-new-protocol-for-ai-data-licensing/.
[17] See: https://toloka.ai/blog/the-data-behind-deepseek-s-success/.
[18] See: https://www.ibm.com/think/insights/streamline-accelerate-ai-initiatives-synthetic-data.
[19] See: https://www.databricks.com/blog/building-scalable-synthetic-data-generation-pipelines-perception-ai-databricks-and-nvidia.

(reinforcement learning) requires curated task environments rather than internet-scale data, which will reorient AI data strategy from mass web scraping toward precision task construction.[20]

Overall, competition concerns in training data have modestly increased but remain moderate. The concentration risk we initially identified – where exclusive or expensive licensing could crowd out start-ups and competitors – has partially materialized, with major AI labs securing preferential access to premium content. However, counterbalancing factors include the potential for collective licensing through mechanisms like RSL, increasing availability of online datasets for model training, and the accelerating role of synthetic data. For competition authorities, we recommend continued monitoring of exclusive arrangements but view this layer as less concerning than the others.

2.2 Computing power ("compute")

March 2024 assessment: We identified computing power as the most concerning layer for competition. Our analysis highlighted NVIDIA's dominance in AI chips, the oligopolistic structure of cloud services, and the problematic combination of high demand, scarce supply, high switching costs, and growing vertical integration. We warned that this mix posed non-negligible risks to competition, for instance in the form of self-preferencing or discrimination.

In retrospect, our concerns have proven accurate – and in some respects, the situation has even intensified. The **compute layer remains the most significant "chokepoint"**[21] **in the genAI value chain**. This position is now further reinforced by **rapidly rising energy demand** from AI data centres and the resulting strain on electricity grids, which makes access to scarce power and grid connections another strategic bottleneck that only the largest, best-capitalised players can reliably secure.

Clearly, **NVIDIA's market position has solidified to an extraordinary degree**. The company now commands between 80% and 92% of the AI chip market, depending on the segment measured.[22] The company became the first to reach a USD 5 trillion market capitalization in late October 2025.[23] The Blackwell architecture, NVIDIA's latest chip generation, delivers strong performance improvements. However, it remains supply-constrained, which is reflected in global deals: Korea secured 260,000 Blackwell chips,[24] while the UK obtained 120,000[25] and Germany's Deutsche Telekom Industrial AI Cloud project in Munich 10,000,[26] increasing the country's AI capacity significantly. Access to cutting-edge AI compute has become a geopolitical priority. Beyond pure hardware scarcity, Nvidia has consistently tried to own the standards and software layers that define each market it enters, from

[20] See: https://newsletter.semianalysis.com/p/rl-environments-and-rl-for-science.

[21] For this term, see: Rebecca Giblin and Cory Doctorow, *Chokepoint Capitalism: How Big Tech and Big Content Captured Creative Labor Markets and How We'll Win Them Back* (Beacon Press, 2022).

[22] See, although only up to Q1 2025: https://wccftech.com/nvidia-dominates-aib-gpu-market-share-in-q1-2025-amd-intel-drop/#:~:text=NVIDIA%20Leads%202025%20With%20Massive,PC%20market%20have%20been%20published. Other sources cite an „an estimated 80-90% market share in AI training chips": https://us.plus500.com/en/newsandmarketinsights/nvidia-earnings-power-global-tech-rally-as-imf-warns-on-growth.

[23] See: https://www.reuters.com/business/nvidia-poised-record-5-trillion-market-valuation-2025-10-29/.

[24] See: https://www.reuters.com/business/media-telecom/nvidia-supply-more-than-260000-blackwell-ai-chips-south-korea-2025-10-31/.

[25] See: https://www.investing.com/news/company-news/nvidia-partners-with-uk-firms-to-deploy-120000-blackwell-gpus-93CH-4241320.

[26] See: https://the-decoder.com/10000-nvidia-blackwell-gpus-set-to-increase-germanys-ai-capacity-by-50-percent/.

early graphics APIs to today's CUDA[27] ecosystem.[28] CUDA in particular functions as a private standard: it bundles hundreds of optimized libraries and tools tightly coupled to Nvidia GPUs, around which a huge developer ecosystem has grown. Competing firms are thus not only constrained by foundry capacity and capital requirements; they also face the task of prying developers away from a mature, Nvidia-centric software stack that has become the default interface for AI research and training. In practice, switching away from CUDA entails substantial re-engineering costs.

Similarly, the **cloud oligopoly, which allows to access high-end chips remotely, persists**. As of Q2 2025, Amazon Web Services (AWS) held a 30% share of the global cloud infrastructure market, while Microsoft Azure held a 20% share and Google Cloud held a 13% share.[29] Together, the three companies control 63% of this rapidly growing market. Data from Synergy Research Group shows that enterprise spending on cloud infrastructure services increased to almost USD 99 billion worldwide in Q2 2025, up over USD 20 billion from Q2 2024.[30] Just 15 per cent of the market is made up of European cloud providers combined, such as SAP, Deutsche Telekom, OVHCloud, and Scaleway.[31] However, one must acknowledge that **cloud infrastructure has also democratized access to frontier AI capabilities** for organizations and firms unable to secure their own NVIDIA chips directly. The hyperscalers' capital investments effectively create rental markets that lower barriers for smaller firms. Additionally, "**neo-cloud" providers** like CoreWeave are emerging as specialized AI infrastructure competitors, offering alternatives focused specifically on GPU-intensive workloads and potentially providing more competitive pricing and flexibility than the hyperscalers.

In addition, one must acknowledge the **rapidly escalating electricity demand from AI data centres** equipped with the chips of NVIDIA and others. The extraordinary power density required for AI training and application (known as "inference") strains electricity grids, particularly in regions of rapid data centre expansion. Modern hyperscale AI data centres operate at 100+ megawatts. However, many hyper-scalers have concrete plans to build "gigawatt-scale" facilities in two years or less.[32] Even worse, AI inference workloads exhibit volatile "start-stop" behaviour with power fluctuations of tens to hundreds of megawatts within sub-second intervals, which creates grid stability risks.[33] Consequently, access to reliable, high-capacity electricity and grid connections has emerged as a related strategic bottleneck that only the largest, best-capitalised players can reliably secure through dedicated power purchase agreements. This further entrenches incumbent market concentration and erecting additional barriers to entry for competitive challengers.

Finally, our warning about **vertical integration** has materialized through multiple channels. Microsoft's partnership with OpenAI continues to draw regulatory scrutiny. Under its previous chair Lina Khan, the FTC launched a comprehensive investigation examining whether the partnership creates lock-in effects

---

[27] Compute Unified Device Architecture is a programming interface developed by Nvidia that allows program components to be processed by the graphics processing unit.

[28] This is based on the concise overview article by Megan Kirkwood: https://www.techpolicy.press/examining-the-source-of-nvidias-power-in-the-ai-industry/.

[29] See: https://www.cargoson.com/en/blog/global-cloud-infrastructure-market-share-aws-azure-google.

[30] See: https://www.srgresearch.com/articles/q2-cloud-market-nears-100-billion-milestone-and-its-still-growing-by-25-year-over-year.

[31] See: https://www.srgresearch.com/articles/european-cloud-providers-local-market-share-now-holds-steady-at-15.

[32] See: https://epoch.ai/data-insights/data-centers-buildout-speeds.

[33] See Figure 3 in this paper ("Unlike conventional electromechanical loads, power electronics-based AI data center demand may change by tens to hundreds of megawatts within sub-second intervals, posing substantial stability risks to the power grid"): Xin Chen et al., 'Electricity Demand and Grid Impacts of AI Data Centers: Challenges and Prospects', version 4, preprint, arXiv, 2025, https://doi.org/10.48550/ARXIV.2509.07218.

that deprive start-ups and AI developer partners of key AI inputs.[34] The investigation scrutinizes Microsoft's bundling of productivity software with cloud services, exclusive licensing agreements, and data sharing practices with OpenAI. Significantly, the FTC is also examining whether Microsoft scaled back its own AI research after partnering with OpenAI, potentially reducing competitive pressure in the field. The UK's Competition and Markets Authority dropped its probe into the Microsoft-OpenAI partnership in March 2025, finding Microsoft did not achieve de facto control.[35] However, the restructured partnership announced in late 2025,[36] allowing OpenAI to become a public benefit corporation while preserving Microsoft's technology access, may face renewed scrutiny.

However, there are also some **counterbalancing forces** that have emerged since our analysis from March 2024. In October 2024, NVIDIA competitor AMD raised its AI chip sales forecast to USD 5 billion for 2025.[37] The new platform from Intel, another competitor in this market, may capture some portion of the AI training accelerator market.[38] Most importantly, Google's custom TPUs have garnered increased interest recently as a credible alternative to Blackwell. Google's TPUv7, detailed in a technical analysis by SemiAnalysis,[39] represents a significant architectural advance that targets NVIDIA's dominance through vertical integration of custom silicon with Google's software stack. In our assessment, its specialized design for transformer workloads positions it as a good alternative for firms seeking to reduce dependence on NVIDIA hardware, which might alter the industry in the long run.

Moreover, technical efficiency improvements are reducing compute barriers. The DeepSeek R1 breakthrough in January 2025, achieving GPT-4-competitive performance with much lower training costs, demonstrated that architectural innovation can partially compensate for compute constraints.[40] The Stanford HAI 2025 AI Index reports that inference costs for GPT-3.5-level performance dropped over 280-fold between November 2022 and October 2024.[41] According to data from EpochAI, the rate of decline "varies dramatically depending on the performance milestone, ranging from 9x to 900x per year."[42] In other words, there are efficiency gains across models and benchmarks, which should, at least to some extent, reduce concerns about compute. However, the combination of approaching data exhaustion,[43] rising energy constraints, and fundamental physical limits suggests that efficiency improvements alone cannot indefinitely sustain the scaling trajectory that has driven recent progress in AI.[44]

Overall, competition concerns in computing power remain high and largely unchanged from our March 2024 analysis. NVIDIA's market dominance has strengthened rather than diminished. Vertical integration between cloud providers and foundation model developers continues to raise concerns. While efficiency improvements and the emergence of neo-cloud providers offer some hope for

---

34 See: https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-issues-staff-report-ai-partnerships-investments-study.
35 See: https://ecp.crai.com/news/cma-closes-its-merger-inquiry-into-the-microsoft-openai-partnership/.
36 See: https://blogs.microsoft.com/blog/2025/10/28/the-next-chapter-of-the-microsoft-openai-partnership/.
37 See: https://www.scmp.com/tech/tech-trends/article/3284409/amd-raises-ai-chip-sales-forecast-us5-billion-2025.
38 See: https://www.jonpeddie.com/news/intels-gaudi-3-accessible-at-ibm-cloud/.
39 See: https://newsletter.semianalysis.com/p/tpuv7-google-takes-a-swing-at-the.
40 See: https://www.heise.de/en/news/Deepseek-R1-AI-training-costs-less-than-300-000-10661970.html.
41 Nestor Maslej et al., *The AI Index 2025 Annual Report*, Technical Report (AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, 2025), 64.
42 See: https://epoch.ai/data-insights/llm-inference-price-trends.
43 Pablo Villalobos et al., 'Will We Run out of Data? Limits of LLM Scaling Based on Human-Generated Data', version 2, preprint, arXiv, 2022, https://doi.org/10.48550/ARXIV.2211.04325.
44 See: https://epoch.ai/blog/can-ai-scaling-continue-through-2030.

democratized access from a long-run, dynamic competition perspective, access to frontier compute – and increasingly available energy – remains a critical bottleneck, at least in the medium term. We continue to view this layer as warranting the most urgent competition policy attention.

2.3 GenAI training (foundation models)

March 2024 assessment: In our initial study, we characterised the foundation model market as a "winner-takes-most" environment, driven by significant economies of scale, high fixed training costs, emergent capabilities at scale, and scarce human expertise. GPT-4 dominated benchmark leaderboards, though we noted early signs of catching-up from others. We argued that the Microsoft-OpenAI partnership should be seen as vertical integration potentially harmful to competition, while acknowledging that competitive dynamics might ease in the longer term.

In our view, the **foundation model landscape has transformed more dramatically than any other layer in the AI value chain**. In particular, the market for such models has become more competitive, i.e. largely reducing our initial concerns. While we noted that competition restrictions might ease in the longer term, the specifics have surprised us. Our March 2024 assessment showed GPT-4 iterations dominating the Chatbot Arena leaderboard. By November 2025, this dominance has eroded substantially. Multiple frontier models now compete across different dimensions: Anthropic's Claude has emerged as the developer's top choice for code generation, capturing 42% market share in coding tools, i.e. more than double OpenAI's 21%.[45] Claude helped transform GitHub Copilot into a billion-dollar ecosystem. In just six months, Claude Code achieved a staggering USD 1 billion in annual recurring revenue.[46] Meanwhile, Google's Gemini models typically offer up more token context windows and multimodal capabilities than other LLMs, such as ChatGPT. In December 2025, the latest Gemini 3 model broke the barrier on several reasoning benchmarks.[47] And according to SimilarWeb data, Gemini is proving more popular than ChatGPT for chatting.[48] Meanwhile, Meta's Llama series continues as the dominant open-source family, although Llama 4 has disappointed many observers.

DeepSeek R1 represented perhaps the most disruptive development: a Chinese open-source reasoning model achieving GPT-4-competitive performance at just a fraction of the cost of models like ChatGPT, as has been confirmed in a September 2025 Nature publication.[49] The company's earlier V3 chatbot model, which formed the basis for the success of R1, cost "only" USD 5.6 million to train. The well-known AI scientist Yann LeCun characterized DeepSeek's success as demonstrating that "open source models are surpassing proprietary ones".[50] Meanwhile, Mistral AI has solidified its position as Europe's AI champion. In September 2025, Mistral AI reached a €11.7 billion valuation[51] and in December of the same year, it offered "Mistral Large 3", a state-of-the-art open model.[52] Beyond Mistral, notable

[45] See: „Claude quickly became the developer's top choice for code generation, capturing 42% market share, more than double OpenAI's (21%). In just one year, Claude helped transform a single-product space (GitHub Copilot) into a $1.9 billion ecosystem.", https://menlovc.com/perspective/2025-mid-year-llm-market-update/.

[46] See: https://www.anthropic.com/news/anthropic-acquires-bun-as-claude-code-reaches-usd1b-milestone.

[47] See: https://www.vellum.ai/blog/google-gemini-3-benchmarks?utm_source=google&utm_medium=organic.

[48] See: https://www.ft.com/content/8881062d-ff4f-4454-8e9d-d992e8e2c4e3.

[49] Daya Guo et al., 'DeepSeek-R1 Incentivizes Reasoning in LLMs through Reinforcement Learning', *Nature* 645, no. 8081 (2025): 633–38, https://doi.org/10.1038/s41586-025-09422-z. See also: https://www.zdnet.com/article/deepseek-reports-shockingly-low-training-costs-for-r1-in-new-paper/.

[50] Quote taken from: https://www.businessinsider.com/meta-ai-yann-lecun-deepseek-open-source-openai-2025-1.

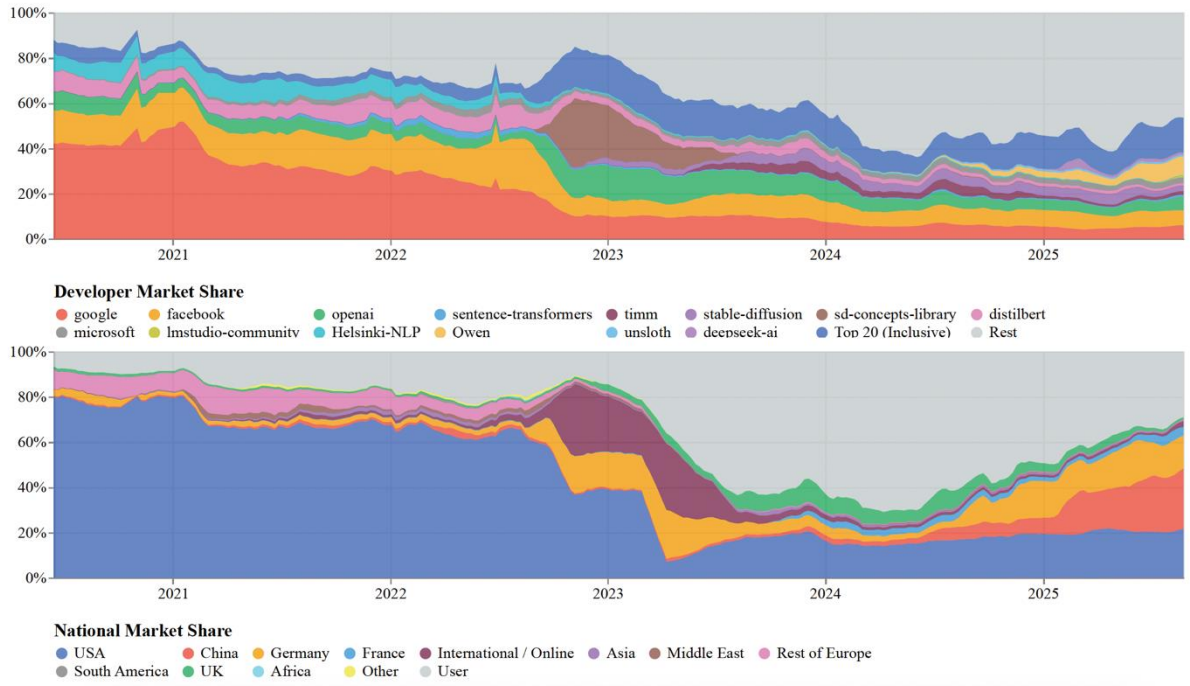
[51] See: https://www.cnbc.com/2025/09/09/ai-firm-mistral-valued-at-14-billion-as-asml-takes-major-stake.html.

[52] See: https://mistral.ai/news/mistral-3.

European successes include Sweden's Lovable, an AI coding platform, and Germany's Black Forest Labs in generative AI.

In short, open-source and open-weight models like those from DeepSeek or Mistral have proven more competitive than expected. The **performance gap between open-weight and closed models has narrowed** from 8% to just 1.7% on some benchmarks in a single year.[53] This has implications for our competition assessment. Open research and open-source resources, such as PyTorch and Llama from Meta, allow SMEs to build upon existing work, so everyone can benefit. For example, by making the entire pipeline for training its OLMo models open source, the Allen Institute for AI reduces the barriers for other companies looking to develop their own AI products and services.[54] Recent empirical research confirms this dramatic rebalancing of competitive dynamics in the foundation model ecosystem. Longpre et al.'s analysis of 2.2 billion model downloads across 851,000 models on Hugging Face, which is the primary global platform for open-weight model distribution, documents three distinct competitive eras since 2020 (Figure 2).[55] The "Foundation Embeddings Era" (pre-late 2022) exhibited high concentration, with the top three US companies (Google, Meta, OpenAI) commanding 40-60% of downloads and the USA alone representing over 60% of national market share. This gave way to the "Generative Diffusion Period" from late 2022 to early 2024, when market concentration plummeted as unaffiliated developers and international communities surged following Stable Diffusion's release. Most strikingly, 2025 marks the start of a "Sino-Multimodal Period", where China's share surpassed the US for the first time, driven almost entirely by open-source models from DeepSeek and Qwen.

Fig. 2: Developer/National Download Share over time



Source: Longpre et al. (2025).

[53] Maslej et al., *The AI Index 2025 Annual Report*, 4.

[54] See: https://allenai.org/blog/olmo3.

[55] Shayne Longpre et al., 'Economies of Open Intelligence: Tracing Power & Participation in the Model Ecosystem', version 1, preprint, arXiv, 2025, https://doi.org/10.48550/ARXIV.2512.03073.

However, we emphasise that **technical contestability at the model layer does not guarantee effective competition if non-technical barriers such as brand recognition or ecosystem lock-in systematically channel users toward higher-priced closed models**. There is emerging empirical evidence supporting this warning. Nagle and Yue's analysis of OpenRouter data, representing approximately 1% of the global LLM inference market, found that closed models dominate with 80% of monthly token usage despite averaging 6x the price of open models and offering only modest performance advantages.[56] Most strikingly, users systematically select closed models even when open alternatives are both cheaper and demonstrably superior in performance. They estimate this behaviour costs users USD 24.8 billion in forgone consumer savings in 2025 alone. These findings suggest that the dominance of some models, like OpenAI's models, reflects *at least partly* also non-technical drivers such as switching costs, brand loyalty, or information frictions. In the same vain, Aubakirova et al.'s analysis of over 100 trillion tokens on the same OpenRouter platform documents strong retention patterns: "We analyze long-term retention for the most widely used models, identifying foundational cohorts that define persistent, stickier behaviors. We define this to be a Cinderella 'Glass Slipper' effect, where early alignment between user needs and model characteristics creates a lasting fit that sustains engagement over time."[57] As the authors note later in their analysis: "For the developers or organizations whose workloads finally 'fit,' this alignment creates strong lock-in effects. Their systems, data pipelines, and user experiences become anchored to the model that solved their problem first. As costs decline and reliability increases, the incentive to re-platform diminishes sharply".

Moreover, the empirical research by Longpre et al. mentioned above also identifies a critical new competitive layer: intermediary organizations specializing in quantization and model re-packaging, such as the mlx-community, now effectively function as infrastructure gatekeepers translating frontier models into deployable models.[58] Finally, another factors that must be acknowledged when discussing the model layer is **human capital scarcity, which remains significant**. The concentration of AI talent in specific US metropolitan areas that we flagged in March 2024 persists. However, success stories like Mistral, which was founded by former DeepMind and Meta researchers, demonstrate that European companies can attract world-class talent when adequately funded.

Overall, competition concerns in the foundation model layer have significantly reduced compared to our March 2024 assessment. The market has evolved from potential "winner-takes-most" concentration toward genuine multi-polar competition. OpenAI's dominance has given way to a more fragmented landscape where Claude leads in coding, Gemini excels in multimodal and long-context tasks, open-source models (Llama, DeepSeek) provide accessible alternatives for fine-tuning, and Mistral offers cost-efficient European options. However, we caution that this improved competitive dynamic may prove fragile. The frontier model race continues, and massive capital requirements for next-generation systems (GPT-6 etc.), accompanied by energy demands, could reconcentrate the market. OpenAI's valuation still dwarfs European competitors. Nevertheless, the current trajectory of the model layer is more favourable for competition than we anticipated two years ago.

[56] Frank Nagle and Daniel Yue, 'The Latent Role of Open Models in the AI Economy', preprint, SSRN, 2025, https://doi.org/10.2139/ssrn.5767103.

[57] Malika Aubakirova et al., 'State of AI: An Empirical 100 Trillion Token Study with OpenRouter', December 2025, https://openrouter.ai/reports/state-of-ai.

[58] Longpre et al., 'Economies of Open Intelligence'.

2.4 GenAI deployment: downstream services

March 2024 assessment: We identified emerging risks in the downstream applications and services market, particularly Big Tech's preferential access to advanced foundation models, incumbents leveraging existing market power and customer bases, and the potential spread of "app-store" models creating scenarios similar to Google's Android ecosystem dominance.[59]

Our March 2024 **warning about "app-store"-like platform models has proven prescient**. In October 2025, OpenAI launched "Apps in ChatGPT", positioning ChatGPT as a conversational storefront where services operate natively within dialogue rather than requiring users to navigate to separate applications.[60] Launch partners include Booking.com, Canva, Coursera, Expedia, Figma, Spotify, and Zillow. OpenAI has also introduced checkout options for in-chat purchases,[61] which suggests ambitions toward e-commerce intermediation. Similar ambitions were met with backlash when Amazon launched its "Shop Direct" and "Buy for Me" AI agents in 2025.[62] Online retailers objected to their websites being scraped and their products being listed without consent. Controversies about data usage are likely to increase in 2026 as AI-powered commerce platforms seek to position themselves as intermediaries between consumers and external retailers.

A related development that we did not fully anticipate is the rapid **emergence of "agentic AI"**, i.e. autonomous systems capable of independent decision-making and multi-step task execution. McKinsey's 2025 State of AI survey found widespread experimentation: "Twenty-three percent of respondents report their organizations are scaling an agentic AI system somewhere in their enterprises (that is, expanding the deployment and adoption of the technology within a least one business function), and an additional 39 percent say they have begun experimenting with AI agents".[63] Similarly, MIT Sloan reports that agentic AI has already "reached 35% adoption, with another 44% of organizations planning to deploy it soon".[64] From a competition perspective, agentic AI interacting with LLM APIs or super-platforms creates new dependencies. If enterprises increasingly rely on AI agents embedded in core workflows such as sales or marketing, there might be deepening lock-in effects and switching costs. However, at the moment, most AI agents are prone to errors over longer time horizons, which limits their impact on markets, including the labour market.

The vertical AI market, i.e. industry-specific AI solutions for healthcare, finance, manufacturing, and other sectors, was valued at USD 10.2 billion in 2024 and is growing.[65] If specialized provider succeed in competing with horizontal Big Tech platforms, vertical AI could represent a counterweight to platform concentration. Industry-specific solutions require deep domain expertise that horizontal AI providers may struggle to match, creating opportunities for specialized entrants. At the same time, however, our concern about Big Tech leveraging existing market power has materialized, questioning whether these specialists will succeed. Microsoft has deeply integrated OpenAI's models into its Office 365 suite through Copilot, reaching millions of enterprise users with minimal friction. Google has

[59] In particular, we flagged OpenAI's then-nascent GPT Store as a potential platform concentration mechanism.
[60] See: https://openai.com/de-DE/index/introducing-apps-in-chatgpt/.
[61] See: https://openai.com/de-DE/index/buy-it-in-chatgpt/.
[62] See: https://www.cnbc.com/2026/01/06/amazons-ai-shopping-tool-sparks-backlash-from-some-online-retailers.html.
[63] See: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai.
[64] See: https://sloanreview.mit.edu/projects/the-emerging-agentic-enterprise-how-leaders-must-navigate-a-new-age-of-ai/.
[65] Data taken from: https://www.gminsights.com/industry-analysis/vertical-ai-market.

embedded Gemini across its product ecosystem. These integrations exemplify the downstream leverage we anticipated in our 2024 study.

Overall, competition concerns in the downstream deployment layer have increased since our initial assessment. The launch of Apps in ChatGPT and Amazon shopping agents validates our prediction of "app-store" and advertising dynamics, as AI firms position themselves as platform intermediaries with potential gatekeeper power. The emergence of agentic AI introduces new dependencies and switching costs as enterprises embed autonomous systems into core workflows. However, the boom in vertical AI and the continued vitality of open-source ecosystems provide competitive alternatives. The market is bifurcating horizontal platform plays (OpenAI, Google, Microsoft) increasingly compete with specialized vertical solutions. EU competition authorities should monitor platform dynamics carefully.

3 Cross-cutting developments

3.1 Regulatory landscape: debates about EU AI Act implementation

The EU AI Act has moved from political agreement (our March 2024 reference point) to phased implementation. The AI Office, now operational, coordinates GPAI Code of Practice development, conducts systemic risk assessments, and oversees leading AI models through its Scientific Panel. Penalties for non-compliance can reach €35 million or 7% of global annual turnover. In our initial study, we suspected that these regulatory requirements would create both constraints and opportunities from a competition perspective. Transparency mandates may level the playing field by forcing disclosure of training data sources, while compliance costs could also reinforce incumbents' advantages. Initially, the planned implementation timeline[66] looked like this:

- 2 February 2025: Bans on AI systems posing unacceptable risks took effect, along with AI literacy requirements
- 2 August 2025: GPAI obligations became applicable, including transparency requirements for foundation model providers, governance rules, and the establishment of the AI Office
- 2 August 2026: Most high-risk AI system obligations will apply
- 2 August 2027: Existing AI models must achieve compliance

Over the past two years, it became clear that particularly smaller players would struggle with documentation and audit requirements, as one of us described in an article for Harvard Business Review.[67] In this context and in an increasingly polarised debate about de-regulation, simplification, and European competitiveness, the European Commission's announced its **Digital Omnibus Package** in November 2025.[68] This package proposes several amendments to the AI Act, including deferring high-risk AI obligations from August 2026 to when supporting standards and guidelines become available. While intended to simplify compliance, the Digital Omnibus has created legal uncertainty for businesses: implementation deadlines are now tied to the availability of yet-to-be-finalized harmonized standards from CEN-CENELEC, and Member States retain discretion to impose additional

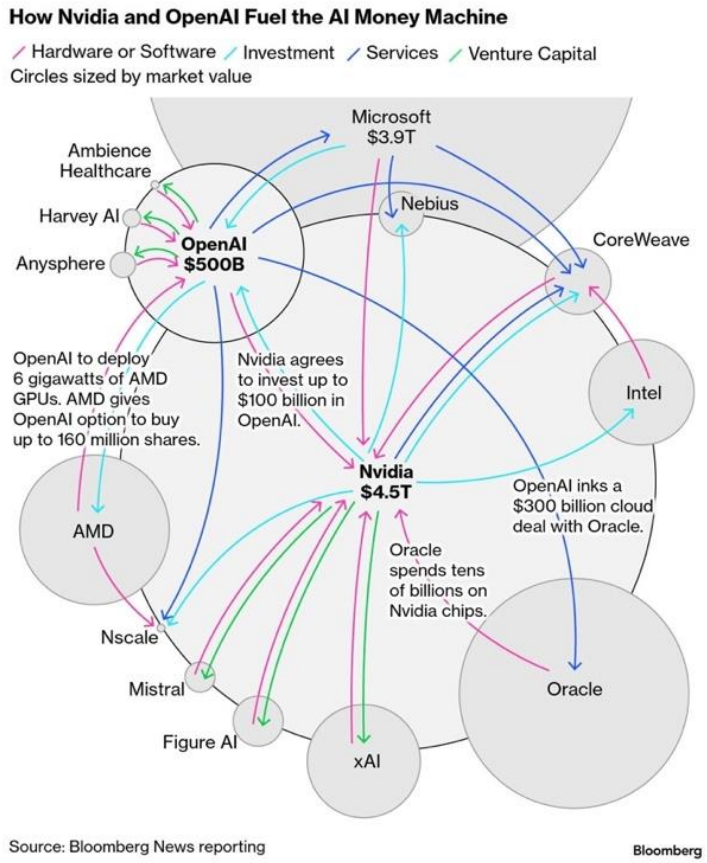
[66] See: https://artificialintelligenceact.eu/implementation-timeline/.
[67] Anselm Küsters and Ben Waber, 'How SMEs Can Prepare for the EU's AI Regulations', Harvard Business Review, 24 September 2025, https://hbr.org/2025/09/how-smes-can-prepare-for-the-eus-ai-regulations.
[68] See: https://digital-strategy.ec.europa.eu/en/faqs/digital-package.

requirements (such as consent for AI development). This regulatory flux and uncertainty might place further strain on SMEs and startups during 2026.

3.2 Industry partnerships and circular financing

In late 2025, **Microsoft and OpenAI signed a memorandum of understanding** that restructured their partnership, clearing the way for OpenAI to convert into a public benefit corporation while preserving Microsoft's privileged access to OpenAI's technology (even beyond any future "AGI threshold" defined in earlier agreements).[69] At the same time, OpenAI entered into a significant long-term cloud agreement with Oracle, reducing its exclusive reliance on Microsoft's Azure infrastructure. Meanwhile, Microsoft began integrating Anthropic's Claude models into its enterprise AI offering – a move that some commentators have described as signalling the end of exclusive, single-vendor "AI partnerships" in the cloud and productivity stack.[70] Despite the start of the Trump II administration, the FTC investigation continues to examine whether the original partnership was structured as a partnership to avoid a merger investigation. In particular, the FTC "wants to determine whether Microsoft's profits from other parts of the business give it an edge over other AI companies. The agency also said it wants details about Microsoft's data centre capacity constraints.[71]

Fig. 3: AI circular financing



Source: Bloomberg.

[69] See: https://blogs.microsoft.com/blog/2025/09/11/a-joint-statement-from-microsoft-and-openai/.
[70] Quote taken from: https://cloudsummit.eu/blog/microsoft-openai-agreement-dawn-multi-vendor-enterprise-ai.
[71] Quote taken from: https://fortune.com/2025/03/12/trump-ftc-microsoft-antitrust-probe-ai-openai-meta-amazon/.

Moreover, the FTC has raised **concerns about "circular spending" practices that characterize many AI partnerships**.[72] Much of Microsoft's initial OpenAI investment came as Azure cloud credits, requiring OpenAI to spend the funding on Microsoft's own services. Similar circular arrangements pervade the sector: NVIDIA invests in AI startups (OpenAI, xAI, CoreWeave, Inflection), which then use those funds to purchase NVIDIA GPUs (see Figure 3). The scale of circular financing has escalated dramatically in 2025. Some firms are even utilizing special purpose vehicles (SPVs) and asset-backed securities structures reminiscent of pre-2008 financial engineering, raising concerns that an AI demand slowdown trigger broader financial instability.[73] Defenders argue these arrangements differ fundamentally from historical bubbles, as most genAI investments are anchored in physical infrastructure such as chips, data centres, electrical equipment, etc., and since most hyper-scalers fund buildouts from their very robust balance sheets.[74] JPMorgan notes that unlike the 1990s dot-com buildup financed by unprofitable companies dependent on credit markets, today's wave benefits from strong free cash flow.[75] Whether circular financing represents strategic necessity or hidden systemic risk ("AI bubble") remains to be seen.

3.3 Institutional dimension: Why traditional timelines no longer fit

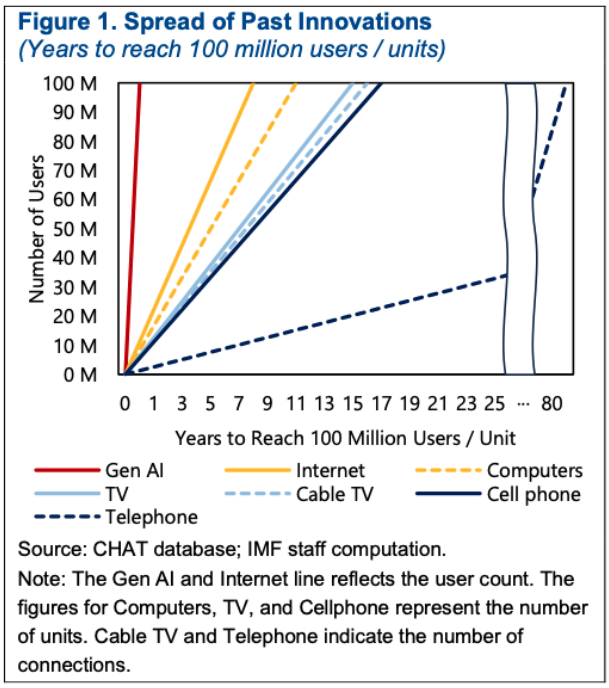
Traditional antitrust enforcement operates on a three-to-five-year timeline: several months for information gathering, about eighteen months for core investigation, and another two years for settlement or trial preparation. For complex matters, total duration often exceeds even five years. With respect to EU merger control, for instance, the latest data show that "[t]he average duration of Phase II investigations skyrocketed to more than 22 months in 2024, while the average duration of Phase I remedy cases settled at 11.2 months, still 20 percent above the 2019-2023 average".[76] In this timeframe within genAI markets, computing efficiency improves, inference costs drop dramatically, new competitors appear, and regulatory frameworks such as the AI Act might be fundamentally adapted. To give an example from our discussion above: Our March 2024 study predicted OpenAI's dominance would largely persist; by early 2026, several serious competitors, ranging from Anthropic and Google to DeepSeek, had successfully emerged. OpenAI evolved from a model developer to a public benefit corporation with a new focus on traditional app and software markets.

In addition, adoption of genAI technology has exploded, even when compared to other general-purpose technologies (Figure 4),[77] further underlining that competition authorities might lack real-time visibility into emerging problems. By the time investigations and cases conclude, the competitive conditions that motivated them no longer exist. This **mismatch between institutional framework and real-world market activities creates a dangerous enforcement gap** where authorities either waste resources on obsolete dynamics or abandon investigations prematurely. Note that this acceleration is not temporary; it will not necessarily disappear once the "AI bubble" has burst and the stock market moves more rationally again. Rather, the rapid pace of technological development reflects the fundamental properties of deep learning itself: continuous improvement through scaling; rapid feedback loops from deployed systems; and exponential capability gains from architectural

[72] See: https://www.mobileworldlive.com/microsoft/ftc-flags-microsoft-openai-antitrust-risks/.
[73] See (in German): https://www.trendingtopics.eu/ai-infrastruktur-anleihen-kredite/.
[74] See: https://www.trustnet.com/news/13464320/circular-financing-in-ai-strategic-necessity-or-hidden-risk.
[75] See: https://am.jpmorgan.com/us/en/asset-management/adv/insights/market-insights/market-updates/on-the-minds-of-investors/does-circularity-in-ai-deals-warn-of-a-bubble/.
[76] See: https://www.dechert.com/knowledge/publication/2025/1/damitt-2024-annual-report.html.
[77] Based on: Florian Misch et al., 'AI and Productivity in Europe', *IMF Working Papers* 2025, no. 067 (2025): 1, https://doi.org/10.5089/9798229006057.001.

breakthroughs. The best example of this is DeepSeek launch of R1 in January 2025, which we already referenced several times above. Despite all the investments of BigTech into genAI, this example shows that rapid innovation can still disrupt the status quo at any time, and in particular faster than any enforcement can respond.

Fig. 4: Spread of AI vs. past innovations



Source: Misch et al. (2025), "AI and Productivity in Europe".

In this context, **maintaining EU competition enforcement effectiveness might require institutional reforms**. Devising such reforms goes beyond the scope of this paper, but we aim to briefly sketch some potential solutions. First, DG COMP should establish monthly market assessments, which could be – at least to a large extent – automated. The recent release by Longpre et al. of an interactive public dashboard for real-time monitoring of concentration dynamics in open model markets, e.g. tracking market shares and national concentration on an ongoing basis, demonstrates that automated monthly market assessments are technically feasible and could be adapted for regulatory oversight purposes.[78] Next, one might explore the idea of fast-tracking merger reviews specifically for AI partnerships (potentially even with some automatic thresholds, such as treating certain expensive arrangements as presumptive concentrations). Finally, the EU AI Act could be amended (for example, through future implementing acts or the Digital Omnibus) to require annual market concentration assessments for providers of GPAI models with systemic risk, with findings automatically triggering DG COMP authority for sector inquiries when competitive concerns exceed material thresholds. This would integrate competition considerations into the AI Office's existing GPAI systemic risk evaluation framework. Alternatively, more institutional coordination between the AI Office and DG COMP should be formalized through an annual joint AI market dynamics report assessing both safety risks and competitive structure across the AI value chain. In general, more research is needed into mechanisms that enable competition enforcement in rapidly evolving digital markets, where traditional timelines have become structurally inadequate.

[78] Longpre et al., 'Economies of Open Intelligence'.

In late 2025, EU competition enforcement around AI finally began to move at something closer to the speed of the markets it targets. Announced on 4 December 2025, the European Commission's investigation into Meta's WhatsApp AI policy signals that authorities are starting to adapt to the urgency of AI markets. The case concerns Meta's October 2025 policy prohibiting competing AI providers from using WhatsApp Business solutions while keeping Meta's own AI assistant accessible, which amounts to a classic textbook self-preferencing concern. Notably, Executive Vice-President Teresa Ribera explicitly framed the inquiry in terms of preventing "irreparable harm to competition in the AI space" and acting "quickly".[79] There is thus some institutional recognition that traditional enforcement timelines risk rendering intervention obsolete. Just five days later, on 9 December 2025, the Commission opened a second AI-focused antitrust investigation, this time examining whether Google imposed unfair terms on web publishers and YouTube content creators by using their content to train AI models and power AI Overviews (generative summaries in search results) without appropriate compensation or opt-out rights, while simultaneously blocking rival AI developers from accessing YouTube content for their own model training.[80] The Google case targets dual concerns: extraction of training data from publishers, and leveraging YouTube's content monopoly to advantage Google's own AI models. Like the Meta investigation, the Commission committed to proceeding "as a matter of priority". A recent analysis by Kena Zheng situates both probes within a burst of five AI-related cases opened in just 22 days, including three DMA market investigations into cloud computing as essential AI infrastructure.[81] If the Commission can translate this prioritization into expedited proceedings (and perhaps suitable interim measures) in the coming months, these cases could serve as **templates for enforcement in fast-moving AI markets**. However, as Zhen has noted, gaps in the DMA's coverage of AI services mean ex ante rules cannot yet "protect the front door" of AI competition on their own, which in practice pushes **classic antitrust enforcement back to centre stage** as the main tool for keeping AI markets fair and contestable.

An important objection to these proposals deserves consideration: if genAI markets remain as dynamic as our analysis suggests, with new competitors like DeepSeek successfully challenging incumbents, does this very dynamism render accelerated enforcement unnecessary? This critique can be linked to a Hayekian understanding of markets as „discovery procedures": competitive processes generate information and innovation that regulators cannot predict ex ante, making premature intervention potentially very harmful.[82] Hayek's warning against the "pretence of knowledge", i.e. the assumption that regulators possess sufficient information to outperform spontaneous market ordering, remains particularly relevant in rapidly evolving technological domains.[83] While this point is indeed crucial for preventing overly fast market-correcting measures, we think that in the case of genAI, it is analytically helpful to distinguish two dimensions: competition at the model layer and consolidation at the infrastructure layer. Current model-layer dynamism, which is visible in Claude's capture of the coding market and DeepSeek's architectural breakthroughs, exists precisely because most infrastructure inputs such as computing power (through cloud access), training data, and technical talent have remained somewhat contestable among several well-funded firms and start-ups. Yet consolidation at

[79] See: https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2896.
[80] See: https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2964.
[81] See: https://scidaproject.com/2025/12/12/ex-ante-but-insufficient-the-commissions-five-ai-investigations-launched-in-just-22-days/.
[82] Reprinted in English in: Friedrich A. Hayek, 'Competition as a Discovery Procedure', trans. Marcellus S. Snow, *The Quarterly Journal of Austrian Economics* 5, no. 3 (2002): 9–23.
[83] F. A. von Hayek, 'Die Anmaßung von Wissen', *ORDO: Jahrbuch für die Ordnung von Wirtschaft und Gesellschaft* 26 (1975): 12–21, JSTOR.

the infrastructure layer is accelerating rapidly: NVIDIA's GPU dominance persists and might get more important, as the next generation of frontier models will require access to gigawatt-scale compute and energy arrangements that smaller players cannot replicate or afford to buy; exclusive data licensing deals with premium publishers create permanent cost asymmetries for new entrants; and geographic concentration of AI talent might further intensify. In other words, surface-level dynamism might mask deepening structural bottlenecks.

The institutional reforms sketched above should not replace competition but ensure enforcement capacity during a phase when intervention can still preserve contestability, rather than waiting until consolidation is complete and remedies become structurally ineffective. Hayekian market discovery requires certain structural preconditions such as freedom of entry, absence of insurmountable barriers, and competitive access to essential inputs. In line with this, the emerging empirical evidence on model competition that we discussed in Section 2.3 above[84] suggests that technical contestability at the model layer does not guarantee effective competition if brand recognition, ecosystem lock-in, default choices, or information asymmetries systematically channel users toward dominant, proprietary models from OpenAI. This implies that competition authorities should intervene at an early stage in order to prevent durable market concentration and lock-in effects in downstream GenAI markets, too. Overall, **regulatory scrutiny should thus be directed primarily towards disclosure requirements (ensuring users can meaningfully compare model performance and costs), interoperability standards (reducing switching costs), and vertical integration practices (preventing closed model providers from leveraging adjacent market power to foreclose open alternatives)**.

Recent theoretical work by Susan Athey and Fiona Scott Morton provides a general equilibrium framework that sharpens these concerns by modelling AI as an imported factor of production subject to market power.[85] Their analysis demonstrates that when AI providers can charge both per-unit usage fees and fixed access fees, market power creates distinctive harms that standard partial equilibrium competition analysis overlooks: displaced workers experience falling wages without corresponding price decreases and downstream industry concentration increases as fixed access fees limit entry and variety. Most strikingly, they identify conditions under which workers face what they call "double harm", first displaced by cheap AI that depresses wages through labour market crowding, then further injured when monopoly pricing raises consumer prices without restoring employment (since adoption decisions are discrete and a marginal increase in the AI input price worsens the price index without undoing displacement). This general equilibrium perspective suggests that even if AI adoption appears beneficial in a traditional cost-reduction sense, productivity improvements from AI do not necessarily result in improvements in economy-wide welfare when market power prevents pass-through of efficiency gains to consumers and workers. For our research question, this theoretical work implies that **competition enforcement in the AI value chain should go beyond static allocative efficiency and also aim to ensure that transformative productivity gains translate into broadly shared welfare improvements rather than concentrated rents**. In short, the institutional reforms sketched above might soon become economically essential for realizing genAI's potential social benefits.

---

[84] The two key studies are: Nagle and Yue, 'The Latent Role of Open Models in the AI Economy'; Aubakirova et al., 'State of AI: An Empirical 100 Trillion Token Study with OpenRouter'.

[85] Susan Athey and Fiona Scott Morton, *Artificial Intelligence, Competition, and Welfare*, no. w34444 (National Bureau of Economic Research, 2025), w34444, https://doi.org/10.3386/w34444. For a non-technical summary, see here: https://www.digitalistpapers.com/vol2/atheyscottmorton#footnote2.

4 Conclusion

Roughly two years after our March 2024 assessment of the generative AI value chain, the market presents a mixed competition picture:

- **Computing power remains the most significant bottleneck**. NVIDIA's dominance has strengthened, despite the rise of Google's TPU chips and Neo-Cloud providers, and vertical integration concerns have intensified. The FTC investigation into Microsoft-OpenAI represents appropriate regulatory attention, though outcomes remain uncertain.
- **Training data concerns have moderately increased**, with exclusive licensing deals proliferating. However, collective licensing infrastructure and the rising importance of synthetic data provide counterbalancing forces.
- **Foundation models have become notably more competitive** than we initially anticipated. The market has evolved from potential "winner-takes-most" concentration toward genuine multi-polar competition among OpenAI, Anthropic, Google, Meta's Llama ecosystem, DeepSeek, Mistral, and others. Open-source models have proven more viable than expected.
- **Downstream services warrant increased attention**. OpenAI's platform ambitions through Apps in ChatGPT, the emergence of agentic AI dependencies, and Big Tech integration of genAI into existing products create new concentration risks even as vertical AI specialists provide certain alternatives.

Our original recommendation, namely that **EU competition enforcers take a holistic view to understand vertical integration across the genAI stack**, remains as relevant as ever. The Microsoft-OpenAI-Azure nexus exemplifies how a single undertaking can command strong gatekeeper positions across infrastructure, foundation models, and downstream applications. Similar dynamics apply to Google (Cloud, Gemini, Search/Workspace integration). However, while current AI markets exhibit surface-level dynamism, with new competitors successfully challenging incumbents, this competitive vitality depends on **infrastructure-layer contestability** that **is rapidly eroding**. In our view, this might create a mismatch where traditional enforcement timelines risk arriving after structural bottlenecks have already foreclosed competitive entry. As we argued in March 2024, competition policy must work together with other digital regulations in this fast-paced, dynamic domain. The EU AI Act's implementation provides a new regulatory tool that complements traditional competition enforcement. The institutional reforms we propose aim not to replace market discovery processes but to preserve their structural preconditions. From a European perspective, robust competition in generative AI markets is necessary but not sufficient for achieving technological sovereignty and avoiding strategic lock-in. Given genAI's exponential, cross-sectoral impact, and significant spillover effects, Europe requires complementary industrial policy measures, adapted public procurement guidelines, coordinated media policy, and strategic trade policies. While these broader policy instruments lie beyond this paper's scope, their systematic importance for European technological competitiveness cannot be overstated. Competition enforcement alone cannot substitute for a comprehensive strategy that addresses both market structure and strategic capabilities. In short, the challenge for European policymakers is ensuring that enforcement action is timely enough to prevent harmful concentration without stifling the innovation that has made today's competitive foundation model landscape possible.

Centrum für Europäische Politik
FREIBURG | BERLIN

Authors:

Dr. Anselm Küsters, LL.M., Head of Division Digitalisation & New Technologies
kuesters@cep.eu

Dr. Matthias Kullas, Head of Division Internal Market & Competition
kullas@cep.eu

Centrum für Europäische Politik FREIBURG | BERLIN
Kaiser-Joseph-Straße 266 | D-79098 Freiburg
Schiffbauerdamm 40 Raum 4315 | D-10117 Berlin
Tel. + 49 761 38693-0

The **Centrum für Europäische Politik** FREIBURG | BERLIN, the **Centre de Politique Européenne** PARIS, and the **Centro Politiche Europee** ROMA form the **Centres for European Policy Network** FREIBURG | BERLIN | PARIS | ROMA.

Free of vested interests and party-politically neutral, the Centres for European Policy Network provides analysis and evaluation of European Union policy, aimed at supporting European integration and upholding the principles of a free-market economic system.