RBB Economics

Competitive Dynamics of Generative AI[1]

RBB Economics, 12 June 2025

[1] The preparation of this paper has been supported by the Computer & Communications Industry Association. The opinions expressed are those of RBB Economics and do not necessarily represent the views of CCIA or of its members.

Contents

1 Executive Summary

1.1 Overview of findings

The current GenAI landscape exhibits considerable dynamism, with firms using a variety of strategies to compete

The early evidence we have examined shows a range of GenAI deployment strategies across companies and sectors. Firms deploying GenAI in their businesses are frequently multi-sourcing from different model providers; market-led solutions are emerging to help reduce the cost and difficulty of switching between GenAI providers; and there is significant variety in business models, competing to serve the same end user needs. Partnerships between GenAI firms are helping to support innovation by pooling resources and combining capabilities, without any current evidence of market foreclosure or concentration.

Early concerns about concentration have not materialised, and many of the features we observe – including low switching costs and evidence of multi-sourcing – reduce the likelihood that GenAI markets will "tip" towards a single or few providers, in the way that some other digital markets have. Moreover, there is no evidence of foreclosure resulting from vertically integrated players in the value chain, e.g., firms that integrate their own foundation models into downstream applications. Instead, the evidence shows non-integrated firms competing successfully with these vertically integrated providers for the same downstream applications. As of now, GenAI markets show characteristics consistent with healthy competition and appear to be displaying considerable dynamism with continued high levels of investment, entry, and innovation.

Europe faces some challenges in GenAI deployment, but also has significant opportunities

European markets for GenAI lag behind the US, in large part due to a significant and persistent gap in availability of private capital for funding GenAI ventures, which in turn makes scaling-up harder in Europe and affects talent retention. Market fragmentation, in terms of linguistic and regulatory differences across national borders, presents further challenges to achieving scale, as does the complex regulatory landscape, which can increase compliance costs for European firms compared to their US counterparts.

At the same time, industrial expertise in certain sectors where European firms have a long-standing presence creates significant opportunities for GenAI deployment. Domain knowledge in automotive engineering, pharmaceuticals, and luxury goods for instance should mean European firms are well positioned to identify and develop GenAI solutions for these sectors. Such solutions have potential for significant benefits for European productivity and growth, given the importance of these industries to the region. A culture of cross-border collaboration in R&D may help mitigate some of the challenges Europe faces in taking advantage of these significant opportunities, though a clear role remains for policymakers in promoting and supporting GenAI deployment.

Policymakers in Europe should maintain an active approach to monitoring but intervene only where there is clear evidence of harm

Current evidence shows significant competitive dynamism in the GenAI sector, with competition appearing to work for the benefit of end-users. As such, there appears to be no need for intervention in these nascent markets. Given the market features we observe and the rapid pace at which the technology is evolving, there currently appear to be more risks associated with premature intervention – which could disrupt effective competitive processes or chill innovation during critical development phases – than there are gains from regulation in response to potential concerns. However, monitoring key competitive indicators as markets mature will remain important.

Looking beyond competition policy, there is a role for broader policy initiatives to support deployment adoption. To this end, policymakers could consider ways to facilitate deployment in the region by helping to address or mitigate structural barriers, including capital access and market fragmentation, whilst encouraging GenAI adoption by European firms throughout the economy.

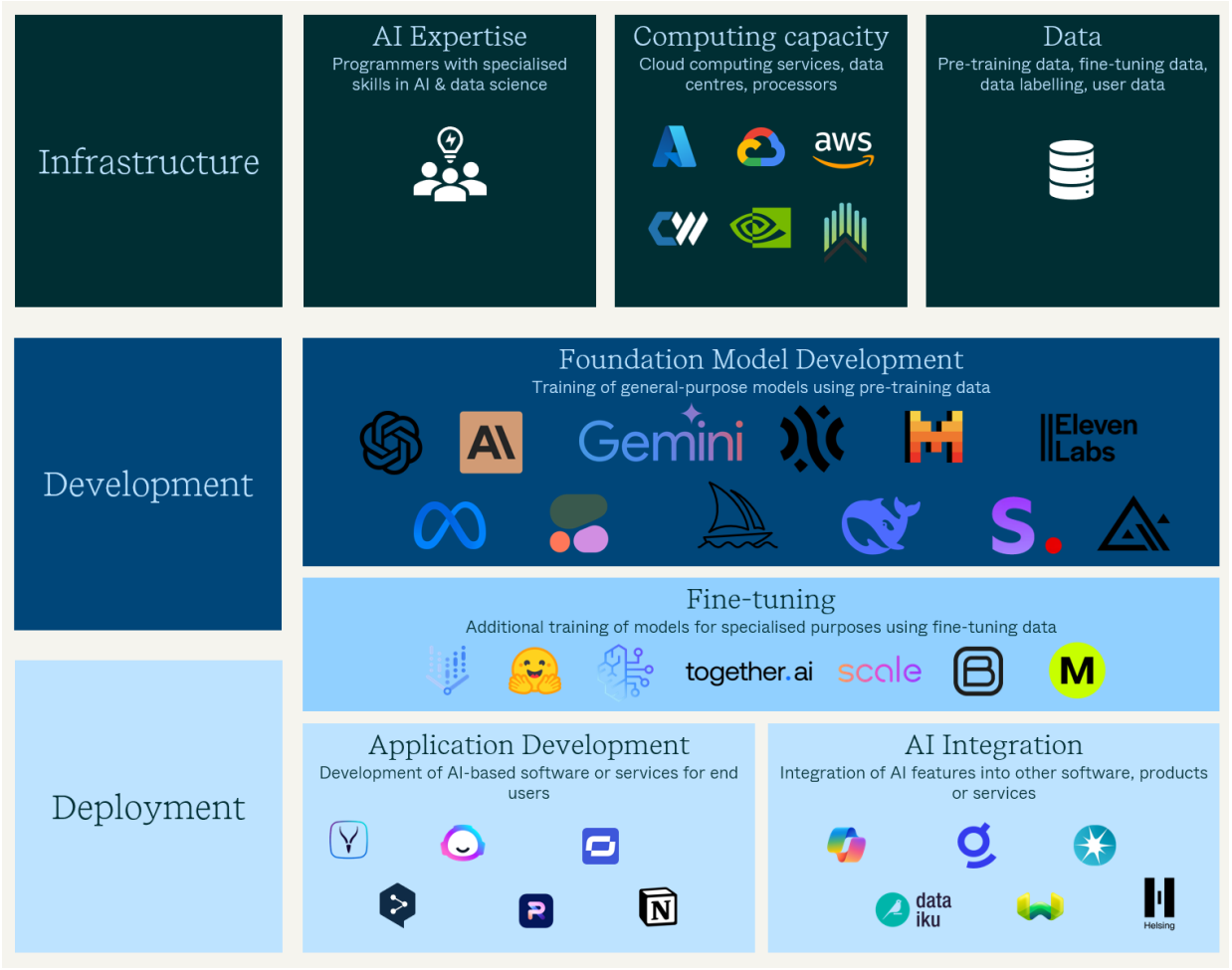
1.2 Global and European GenAI development

The GenAI landscape has undergone significant evolution in its relatively short existence and continues to move at pace. Many of these trends are observed at the global level, although Europe faces specific challenges and opportunities in the sector.

1.2.1 Overview of the global GenAI landscape

The GenAI value chain consists of three broad layers: infrastructure (or compute power), model development, and downstream applications. Figure 1 below provides an overview of this value chain, and some of the key players active at each level.

Figure 1: Overview of the GenAI value chain



Source: *RBB Economics based on Competition & Markets Authority (2023a) and European Economic and Social Committee (2025).*
Note: *Fine-tuning is commonly associated with both the development and deployment stages of GenAI. During development, model creators take a pre-trained FM and fine-tune it on a specific dataset to adapt it for particular use cases, domains, or customer needs. This step enhances performance, relevance, and alignment with desired outputs before the model is deployed for use. Fine-tuning occurs closer to deployment, especially in enterprise settings where models are tailored to proprietary data.*

Opportunities exist across the value chain

The use of GenAI is expected to soar in the upcoming years as the sector moves from experimentation to scaled deployment in downstream use cases. Global GenAI investment exceeded $56 billion in 2024, almost doubling from $29 billion in 2023, and Morgan Stanley anticipates that revenue from GenAI could exceed 1 trillion USD by 2028.[2] This creates significant opportunities for entry and expansion across the value chain.

There is significant differentiation between foundation models

At the foundation model layer, developers compete by bringing diverse capabilities to market. In this respect, it is notable that there is no single best GenAI model that outperforms all others in every scenario. Instead, model performance varies significantly depending on the specific task and user need, reflecting a degree of differentiation between models. Some models excel at certain capabilities, such as reasoning, code generation, or multimodal understanding, while performing less well in other areas. This feature could support long-term diversity in the sector, as it suggests user choice between models will depend on the nature of the application being developed, as well as specific user preferences.

[2] See https://www.spglobal.com/market-intelligence/en/news-insights/articles/2025/1/genai-funding-hits-record-in-2024-boosted-by-infrastructure-interest-87132257; https://www.morganstanley.com/insights/articles/genai-revenue-growth-and-profitability (both accessed in June 2025).

Competitive activity is intensifying within the deployment layer

While the initial focus of competitive activity was foundation model development, competition is now starting to intensify within the application layer, with companies beginning to make use of market-specific data and domain expertise to develop industry-specific GenAI solutions. There is significant untapped potential in deployment, which offers a particularly attractive investment opportunity given the high commercial value of industrial applications. Capital seems to be increasingly flowing toward this layer of the value chain to take advantage of the opportunities it offers. Bloomberg Intelligence projects the industry to grow more than 30-fold from 2022 levels over the next decade.[3] This rapid increase is reported to be driving VC investors' interest in deployment and GenAI application businesses that build specialised software using third-party FMs for consumer or enterprise use. In addition, some major partnerships have focused on deployment and application layer integration, which further signals GenAI's move toward deployment.

Open-source models, APIs and partnerships play an important role in lowering entry barriers

Within this context, open-source models, APIs, and partnerships are all playing an important role in lowering barriers to adopting GenAI by providing access to advanced GenAI capabilities without the high development costs of building proprietary models from scratch. For example, open-source models such as Meta's Llama and AI platforms like Hugging Face help businesses integrate AI functionalities to their existing processes at a lower cost than proprietary models and platforms. Partnerships also help in reducing entry barriers by helping firms distribute costs, obtain funding, and share expertise across the value chain, with collaborations increasingly occurring at the deployment layer as the industry evolves.

Improvements in model efficiency and advances in technical capabilities are also facilitating deployment

Other developments are supporting the shift from development to deployment. First, the costs of deployment are decreasing, as recent improvements in model efficiency mean high-performance models can now be accessed at significantly lower cost than in the recent past. For example, inference costs for GPT-3.5-level systems fell by 99.65%, a more than 280-fold reduction, between November 2022 and October 2024, making AI capabilities that were once expensive now accessible to a much broader range of businesses.[4]

Second, recent developments in technical capabilities have expanded potential industrial use cases. Enhanced reasoning capabilities in models like GPT-4 and Claude 4 enable stronger chain-of-thought reasoning and complex problem-solving, supporting tasks from legal document analysis to financial risk modelling. Multimodal integration allows models to process text, images, audio, and video simultaneously, unlocking applications in healthcare diagnostics and media production. These improvements have enabled the emergence of autonomous agents that can independently accomplish complex multi-step tasks through planning and adaptation, creating opportunities for workflow automation and business operations.

These advances are transforming where and how GenAI can deliver value, enabling new applications to various industrial use cases and, as a result, increasing overall demand for deployment and accelerating take-up.

---

[3] See https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/ (accessed in June 2025).

[4] Stanford Institute for Human-Centered Artificial Intelligence, *The 2025 AI Index Report*, page 28, Highlight 6, available at https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf (accessed in June 2025).

1.2.2 The European landscape

Against this backdrop of rapid global development, Europe is pursuing its own strategy for expanding GenAI capabilities whilst facing distinct opportunities and challenges.

Europe has already launched several infrastructure initiatives to help build its GenAI capabilities

Several targeted initiatives have been launched in Europe across the AI value chain to help build home-grown capabilities, from infrastructure investments to model development. Infrastructure initiatives include the €200 billion InvestAI programme and EuroHPC Joint Undertaking, which aim to expand computational capacity across European research institutions. The development of National AI compute hubs including France's Jean Zay supercomputer, Germany's Jülich centres, and Italy's Leonardo HPC initiative, are also aimed at strengthening regional computing capacity.

There are some early signs that Europe is benefiting from these initiatives, including examples of successful AI innovation within the region. For example, European model developers like Mistral AI and Aleph Alpha show that home-grown companies can innovate successfully and compete in a global playing field, despite relatively constrained resources. Nonetheless, concerns have been raised that Europe may need to do more to achieve the goals of its innovation agenda, which includes specific AI ambitions.

Europe faces structural challenges to scaling

Europe faces specific structural obstacles that affect its global competitiveness in the industry.

- **Persistent funding gap**. Europe faces a significant and persistent funding gap in relation to GenAI, particularly relative to the US. Funding in the US was over 20 times the level of investment in GenAI in Europe in 2024, making it significantly harder for European firms to obtain necessary capital during critical scaling-up phases. Persistent differences in funding have contributed to a drain on critical talent, with a net outflow of AI expertise from Europe to the US that further adds to Europe's challenges.
- **Market fragmentation increases costs**. The European market is fragmented by a multiplicity of jurisdictions, differences in legal and regulatory regimes and linguistic diversity that can create frictions for start-ups and firms seeking to grow. As a result, developers based in Europe can face greater challenges in achieving scale than their counterparts in the US and China, where such barriers do not exist to the same extent.
- **Regulatory complexity creates implementation burdens**. Europe has been early to implement AI regulation, with the GDPR already providing standards for data privacy and governance. Whilst this regulation aims to benefit market participants, the complexity and cost of compliance may create significant challenges for product development and deployment, particularly for startups and SMEs lacking resources for extensive legal and technical compliance programmes.

European firms have significant opportunities to benefit from GenAI

At the same time, we find considerable opportunities in Europe within the GenAI sector.

- **Industrial expertise creates deployment advantages**. Europe's long-standing presence in industries such as manufacturing, automotive engineering, and pharmaceuticals provides a strong foundation for GenAI innovation. European firms with deep knowledge of these sectors can draw on their expertise to identify and develop sector-specific GenAI use cases. Taking these opportunities would benefit productivity in the region significantly, given that the industrial sector has represented over 20% of the EU's GDP in recent years.
- **Cross-border collaboration networks can help accelerate adoption**. Europe has a long-standing history of R&D collaboration, with established R&D partnerships promoting multinational cooperation.

This collaborative model can help institutions in the region pool resources and expertise across national boundaries, which may help to partially overcome some of the challenges that Europe faces in GenAI deployment.

Despite these mitigating factors, significant challenges remain, creating a clear role for policymakers to support the European GenAI sector in taking advantage of its significant opportunities, as discussed further below.

1.3 Different approaches to deployment help support competition in the sector

Our case studies reveal two key patterns that support competition in the GenAI sector. Firstly, many firms deploying GenAI solutions are adopting multi-sourcing strategies, supported by market-based tools that aim to reduce switching costs. Secondly, different business models – including both vertically integrated and non-integrated players – are successfully competing in the same market segments, suggesting that the presence of some firms at multiple layers of the value chain is not currently acting as a barrier to entry and expansion. Our case studies also identify how European firms have used GenAI capabilities to achieve measurable results, underscoring the opportunities for productivity gains in the region if GenAI deployment can be successfully promoted and supported.

1.3.1 Multi-sourcing and low barriers to switching mean GenAI markets are unlikely to tip

The multi-sourcing strategies documented across our case studies mitigates the risk that GenAI markets will "tip" towards a single or few suppliers, at least in the relatively near term. Economic literature on competition in markets with low switching costs and significant multi-homing suggest these market features are likely to contribute to effective competition.

Multi-sourcing of models is a consistent theme across sectors deploying GenAI

Case studies across enterprise applications, legal services, and European industries demonstrate that these early deployers typically use multiple AI providers rather than rely on a single vendor. By way of example, Estée Lauder's distributed strategy uses different providers for specific business functions. Goldman Sachs' integrated platform approach routes tasks between multiple models based on requirements and cost, enabling sophisticated model allocation whilst reducing external vendor dependency. Within its productivity tool, Notion AI also relies on multiple foundation models "under the hood", all the while providing a unified interface to its customers.

Open-source alternatives like Meta's Llama and Mistral AI's open-weight models provide credible competitive pressure, offering firms looking to deploy GenAI models a genuine choice between proprietary and open alternatives.

Market-based tools have emerged to reduce switching costs and support flexible deployment

This demand for multi-sourcing has led to the emergence of market-led solutions aiming to facilitate switching between models to fulfil a commercial need. Unified access platforms like OpenRouter and development frameworks like LangChain help to simplify the technical management of multiple AI providers through standardised interfaces. Analytics tools provide quantitative performance data on specific queries for different foundation models, which helps developers select foundation models based on factual evidence, and to optimise their model portfolio and usage dynamically. Open-source development frameworks like LangChain demonstrate how community-driven solutions can help reduce the risk of vendor lock-in, whilst platforms like Hugging Face show how open-source community models can reach significant scale.

Industry standardisation efforts, including Anthropic's Model Context Protocol (MCP) and Google's Agent-to-Agent standard (A2A), establish common integration methods that reduce implementation complexity and minimise switching costs between providers.

These market developments enhance provider interoperability and help to maintain competitive dynamics by increasing transparency around model performance and reducing technical barriers to changing vendors.

Model-agnostic designs help reduce switching barriers

SpringBok's legal AI application is designed to work regardless of the foundation model, without requiring system rebuilds in the event of a supplier change. Its approach relies on "prompt architecture", a technique that builds complex workflows based on carefully crafted instructions to guide a pre-trained model without changing its underlying capabilities through fine-tuning. This architecture can be transferred between models, by contrast to fine-tuning models. Fine-tuning can create switching costs since customising a model involves retraining it with specific data, and this customisation must be redone in the event of a change of foundation models.

SpringBok's value proposition seems to have resonated with its intended audience: Dentons reported 65% firm-wide adoption within six weeks, and Cleary Gottlieb Stein & Hamilton, a large international law firm, chose to acquire the GenAI startup in 2025.[5]

The competitive implications of these emerging solutions are substantial. By reducing technical switching costs and enabling easier comparison of model performance, these tools and techniques help maintain competitive dynamics in FM markets. When deployers are able to evaluate and respond to differences in performance for specific tasks by easily switching between foundation models thanks to portable architectures, incentives to innovate remain strong across the ecosystem, helping reduce the risk of market concentration in a market where different models excel at different things.

1.3.2 Vertically integrated and non-integrated firms compete against each other in the same GenAI market segments

Alongside these multi-sourcing patterns, we observe competition between different business models, with both vertically integrated and non-integrated firms finding viable strategies to succeed in the same market segments. This market feature also points to a healthy competitive dynamic, as it suggests smaller firms can compete effectively in a particular deployment niche without the need to integrate across the value chain, helping to support entry into the sector and providing a challenge to larger integrated players.

This competition occurs across multiple market segments, from AI platforms that intermediate between model providers and deployers, to productivity software where specialists challenge established incumbents.

AI platforms, which intermediate between model providers and deployers, include both vertically integrated platforms and specialists

AI platforms compete through varied approaches serving different market needs. For example, Amazon Bedrock and Google Vertex AI offer general-purpose solutions with extensive model catalogues and enterprise-grade infrastructure services. Hugging Face's success through open-source community development demonstrates how collaborative models can compete effectively against proprietary cloud-native platforms, reaching a valuation of $4.5 billion in 2023 through community engagement rather than

[5] See https://www.linkedin.com/posts/springbok-ai_dentons-springbok-collaboration-activity-7132676616547131392-byUB; https://www.clearygottlieb.com/news-and-insights/news-listing/cleary-gottlieb-acquires-springbok-ai (both accessed in June 2025).

vertical integration.[6] Dataiku focuses on governance and regulatory compliance, which can be particularly valuable for European enterprises navigating complex regulatory requirements.

Specialised providers compete against vertically integrated incumbents in the productivity segment

There are a number of vertically integrated players bringing GenAI applications to market but, despite the advantages these solutions can offer to end users, early evidence suggests that non-integrated firms are able to provide a successful challenge. For example, both Microsoft and Google embed GenAI capabilities directly into productivity suites. Their vertical integration of GenAI into existing offerings creates efficiencies for users, as it provides AI-enhanced functionality without requiring customers to switch between applications. Despite this, the productivity segment has seen the emergence of new, disruptive competitors powered by GenAI. Notion reached 100 million users in August 2024 and reports being now used by teams in 62% of Fortune 100 companies.[7] This illustrates how GenAI-powered innovation can overcome incumbency advantages when the value proposition is compelling. DeepL achieved 82% adoption among language service providers in a 2024 industry survey through translation expertise, almost double the adoption of Google and Microsoft translation services.[8] The company then expanded into adjacent language workflows whilst maintaining integration with major productivity platforms, including Microsoft 365 and Google Workspace.

Vertical integration across the GenAI value chain offers potential efficiency benefits

The vertically integrated model offers significant potential for cost and integration efficiencies. For example, Microsoft, Amazon, and Google all achieve cost efficiencies by supplying GenAI services to themselves at cost rather than market rates, eliminating external cloud hosting fees and markup costs. As noted, vertical integration can also help enable smoother user workflows by embedding GenAI directly within existing applications rather than requiring switching between separate tools. Having control of both the infrastructure and application layers also means integrated firms can accelerate the deployment of new features through coordinated updates across product suites without waiting for third-party approvals.

This is true of both AWS and Google's AI platforms, and their respective productivity suites. However, in both segments, the vertically integrated players continue supporting third-party GenAI providers: both Google Vertex AI and Amazon Bedrock offer third-party foundation models in addition to their proprietary models, and both Google Gemini for Workspace and Microsoft Copilot offer official add-ins and connector programmes for competing solutions such as DeepL or Notion.

European industries are using GenAI to enhance their existing competitive strengths

Our European case studies provide examples of companies making use of their domain-specific expertise to deploy GenAI solutions in their existing business processes and achieving measurable results in doing so. For example, Stellantis has used its automotive engineering expertise to enable effective GenAI deployment, achieving significant improvements in voice recognition systems while expanding manufacturing efficiency through AI-enhanced predictive maintenance across global facilities. AstraZeneca is a good example of how pharmaceutical companies can combine domain knowledge with AI capabilities, using multiple partnerships to identify promising drug targets that leverage both genetic data and AI analysis for research advancement. LVMH shows how luxury brands can scale GenAI applications whilst preserving brand values, implementing comprehensive employee training programmes that position AI as augmenting rather than replacing human expertise across creative and operational functions.

---

6 See https://originality.ai/blog/huggingface-statistics (accessed in June 2025).
7 See https://www.notion.com/blog/100-million-of-you; https://www.notion.com/enterprise (both accessed in June 2025).
8 See https://www.prnewswire.com/news-releases/deepl-is-2024s-most-used-machine-translation-provider-worldwide-among-language-service-companies-302270449.html; https://www.deepl.com/en/blog/alc-survey-results (both accessed in May 2025)

Across all three industries, AI integration has been used to enhance productivity and performance, and, through this, competitiveness. This suggests competition in downstream markets is likely to drive further demand for deployment, given the competitive advantage that successful GenAI deployments can bring.

1.4 Partnerships have helped enable innovation without any current evidence of competition concerns

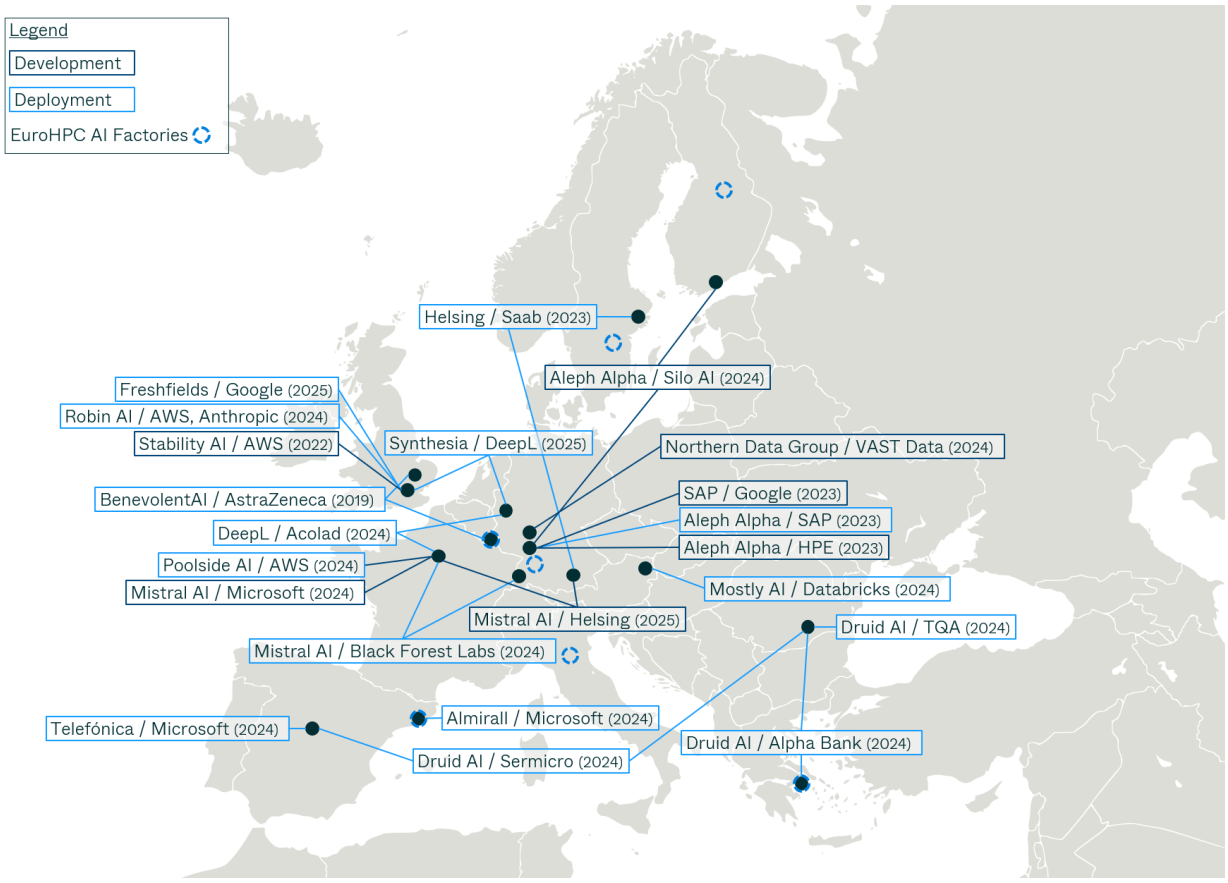
Strategic partnerships are an important feature of the GenAI landscape. They can offer pragmatic solutions to high development costs, technological uncertainties, and the need for specialised expertise across multiple domains. Our analysis of selected partnerships provides examples of partnerships helping to reduce entry barriers and enable innovation, supporting the competitive dynamics observed in the deployment case studies.

GenAI partnerships take diverse forms

The European GenAI partnerships we have observed take varied forms across infrastructure, development, and deployment layers. Development stage partnerships between compute providers and model developers can help innovation by enabling the sharing of resources whilst maintaining competitive independence. Deployment partnerships between model providers and consumer-facing businesses can help accelerate GenAI adoption and integration into existing services.

European cross-border networks demonstrate collaboration across national boundaries and value chain layers, helping to support pan-European GenAI development. Figure 2 below illustrates the dense network of GenAI partnerships spanning European countries and various layers of the GenAI value chain. The map presents a non-exhaustive selection of partnerships and therefore offers a conservative representation of the broader, dense network of GenAI collaborations across European countries.

Figure 2: GenAI partnerships span Europe, connecting diverse players across all layers of the value chain



Source: *RBB desk research based on announcements by the partnering firms. Background map by Wikimedia Commons (accessed in May 2025).*

Note: *Partnerships are classified as either Development or Deployment depending on which value chain layer the partnership supports, as interpreted by RBB Economics. This classification does not necessarily match the layer where the firms usually operate, and some partnerships may cover multiple layers of the value chain, such as supplying both cloud services and a distribution platform. The year in brackets indicates the start of the partnership.*

Partnerships vary in their corporate governance and contractual frameworks, including equity arrangements, control rights, consultation mechanisms, and exclusivity provisions, with many non-exclusive relationships preserving competitive options. For instance, Mistral AI partners with Microsoft, Amazon, and Google simultaneously, ensuring competitive distribution channels remain open. Anthropic maintains separate partnerships with both Google and Amazon on non-exclusive terms, contrasting with Microsoft-OpenAI's exclusive arrangements. European cross-border networks demonstrate collaboration across national boundaries and value chain layers, supporting pan-European GenAI development.

Pro-competitive effects can be observed across partnership types

- **Resource sharing can help reduce entry barriers** where partnerships like Anthropic x Google enable smaller firms to access essential infrastructure. Strategic collaborations help distribute high development costs, combine complementary capabilities, and accelerate innovation across the value chain.
- **Market access acceleration** helps European start-ups to reach enterprise customers through established channels. DRUID AI's partnerships with automation providers across multiple European countries demonstrate how conversational GenAI solutions can be assisted in gaining market traction through the use of established distribution networks.

- **Capability combination** enables firms to focus on their own core expertise whilst also accessing complementary skills, helping to facilitate innovation by reducing duplication of development efforts and technical risks.

Competition authorities have not found evidence of competitive harm to date

In their general publications on GenAI, the European Commission and other authorities have identified pro-competitive effects, including resource sharing, innovation acceleration, and market access facilitation. They have also identified potential theories of harm, but these do not appear to have materialised. In fact, despite ongoing monitoring and reviews of GenAI partnerships by competition authorities, no enforcement actions have been taken to date.

Combined with the evidence on tangible benefits that a number of partnerships have brought to the GenAI sector, current evidence suggests that partnerships are likely to have a pro-competitive effect on the sector in Europe.

1.5 Policy considerations

The evidence we have seen suggests European GenAI markets are currently characterised by significant innovation and competition. Our case studies point towards a variety of deployment strategies, partnership structures, and business models, and provide examples of a number of European companies successfully taking advantage of the opportunities offered by GenAI deployments to improve efficiency in their productive processes. Policy should focus on supporting this momentum while ensuring markets remain open and contestable.

1.5.1 Competition policy: Active monitoring without premature intervention

We see a role for active monitoring as the market continues to evolve to ensure competition concerns do not arise. However, it is important regulators act with caution in relation to any competition interventions, given that the risks of intervening too early currently appear to outweigh the risks of watching to see how the market develops. Our analysis reveals European GenAI markets characterised by competitive variety, multi-sourcing strategies, and relatively low switching costs between providers. These conditions suggest that markets are currently functioning well, with no evidence of the market tipping or foreclosure effects that would justify intervention.

Our competition policy recommendations, therefore, focus on preserving these positive dynamics while avoiding premature intervention that could disrupt effective competitive processes and chill innovation:

- **Maintain active monitoring with clear intervention thresholds based on evidence of actual concerns**. Market features such as multi-sourcing strategies, low switching costs, and business model variety suggest competition in the sector is currently healthy. Evidence shows diverse deployment approaches, sustained entry and expansion across value chain layers, and competitive partnerships that create value without restricting competition. Intervention should address actual rather than theoretical concerns, given the risks of disrupting effective competitive processes during market formation.
- **Continue to recognise the pro-competitive potential of partnerships**. Partnership diversity across exclusive and non-exclusive arrangements enables different strategic approaches whilst maintaining competitive alternatives. Resource sharing arrangements address legitimate business needs, including high development costs, technical complexity, and market access challenges. Restricting beneficial partnerships, particularly those involving non-European investment in European GenAI companies, risks undermining European competitiveness during a critical scaling phase.

- **Exercise caution in extending *ex ante* regulation to GenAI services**. Evidence does not indicate anti-competitive leveraging by existing digital platforms. Effective competition observed between integrated and specialised providers illustrates that distribution advantages can be overcome when value propositions prove compelling. Clear guidance specifying that *ex ante* obligations will apply only where there is concrete evidence of harm would provide necessary regulatory certainty for continued innovation.
- **Monitor interoperability developments and support market-based solutions that preserve flexibility**. The sustainability of current competitive dynamics relies on continued foundation model competition, effective interoperability standards adoption, platform competition, and preservation of deployment flexibility. Competition authorities should monitor these developments and consider intervention only to ensure no proprietary barriers emerge that harm consumers, for instance, by creating substantial foreclosure effects.
- **Avoid creating additional layers of competition-specific obligations**. Current regulatory frameworks, including GDPR and the emerging AI Act, already address many potential concerns about GenAI deployment. Additional competition-specific rules risk creating overlapping obligations that could particularly burden European companies relative to global competitors, and smaller players relative to large organisations.

1.5.2 Broader policy recommendations: Address barriers to scaling

Competition policy alone cannot ensure Europe maximises the opportunities GenAI presents. The following broader policy recommendations address structural challenges and opportunities for European industries to build on their existing strengths through effective GenAI deployment:

- **Address barriers to private capital access for European GenAI companies**. Policy should address regulatory and structural barriers limiting European GenAI companies' access to growth capital from both domestic and international sources. This could include completing the Capital Markets Union initiative, removing barriers to cross-border venture capital investment, and ensuring regulatory frameworks do not inadvertently discourage investment in European GenAI startups.
- **Target public funding toward areas where market failures give rise to under-investment**. Public investment should focus on foundational research with longer-term horizons, open-source initiatives benefiting the European ecosystem broadly, multilingual capabilities serving European market needs, and infrastructure investments supporting the startup ecosystem. Implementation should aim for additionality rather than crowding out private investment in commercially viable applications, targeting areas where projects generate significant social benefits that cannot be fully captured by private investors.
- **Develop policies that support and promote the efficient adoption of GenAI deployment**. Policy could support deployment momentum through sector-specific guidance on regulatory compliance, technical assistance programmes helping SMEs navigate GenAI integration, and initiatives leveraging European strengths in developing GenAI standards for interoperability and trustworthiness.
- **Establish EU-wide coordination mechanisms for consistent GenAI policy implementation**. An EU-wide forum for GenAI policy coordination could help ensure that key stakeholders are actively involved in shaping policy and implementation decisions. This would help avoid unintended consequences from divergent national interpretations whilst preserving single market benefits.
- **Road test remedies and explore softer regulatory approaches where concerns arise**. Experimental regulatory frameworks can allow market participants to trial novel technologies in supervised environments, providing regulators with insights before implementing permanent rules.

Policymakers should consider participatory approaches where industry plays key roles in delivering solutions to identified concerns, which can minimise the risks of unintended consequences from policy interventions in a rapidly evolving industry.

2 An overview of the rapidly evolving GenAI landscape

This section provides an overview of the global Generative Artificial Intelligence (GenAI) landscape, including recent developments in investment patterns, cost dynamics, and technical capabilities. We then consider European-specific trends within this global context, looking at the initiatives Europe has in place to fund and promote GenAI applications, as well as Europe's challenges and opportunities in the GenAI landscape more broadly.[9]

The GenAI value chain consists of three broad layers: infrastructure, model development, and downstream applications.[10] Whilst competitive activity initially focused on model development, competition is now starting to intensify at the application layer, where companies are beginning to make use of market-specific data and domain expertise to develop industry-specific GenAI-powered solutions. Investment data shows capital increasingly flowing toward the development of these specialised applications as technical advancements in model efficiency, reasoning capabilities, and multimodal processing expand use cases across economic sectors.[11]

Against this backdrop, Europe is making investments in infrastructure initiatives and foundational model (FM) development. Europe faces substantial funding gaps compared to the US and China, particularly in areas where market failures are likely to lead to under-investment (i.e., investments with high social value), and European GenAI companies face significant challenges to achieving scale. Policymakers have a role to play in encouraging private finance into the region, as well as ensuring public funds are directed to where they will deliver the greatest social returns. Moreover, whilst some European firms are successfully deploying GenAI in early use cases, there is still considerable scope for Europe to incorporate GenAI into a broader range of applications: GenAI can deliver significant productivity gains, particularly in industries where European firms have a long-standing presence and can take advantage of domain-specific knowledge to identify and develop GenAI solutions. This creates significant opportunities for GenAI deployment within the region, and a further role for policymakers in promoting and supporting the efficient and cost-effective deployment of GenAI across the European economy.

2.1 The global GenAI landscape shows increasing activity at the deployment layer

The GenAI industry is shifting from experimentation to scaled deployment. Competition is intensifying around applications that combine technical capabilities with domain-specific expertise, whether data security for investment banking, product development for car manufacturing, or research tools for legal services.[12] This creates new competitive dynamics as firms move beyond general-purpose models toward specialized solutions.

This shift matters for competition policy because it changes where barriers to entry or expansion may exist and how market power can emerge. As deployment becomes the key battleground, success depends less on massive infrastructure investments and more on understanding specific industry needs and accessing the right combination of models and data.

The section examines:

---

9 In this report, the term "Europe" is used in a broad sense, referring to the continent as a socio-economic area that includes both public and private entities, unless otherwise specified.

10 Infrastructure includes the computational hardware (e.g. GPUs, data centres), cloud platforms, and software frameworks needed to train and deploy GenAI systems. Model development involves creating and refining FMs and other generative systems, including pretraining, fine-tuning, and alignment work. Downstream applications refer to the tools, services, and user-facing products that apply GenAI models to specific tasks or domains (e.g. chatbots, coding assistants, industrial automation). See Figure 3 below for an overview of the GenAI value chain.

11 Multimodal processing refers to a model's ability to integrate multiple types of data, such as text, images, and video. See https://www.ibm.com/think/topics/multimodal-ai (accessed in May 2025).

12 See section 4 for various detailed GenAI deployment case studies.

- **Value chain structure:** how the three-layer GenAI stack creates different competitive dynamics at each level;
- **Market organisation:** the role of vertical integration, specialisation, and partnerships in shaping competition;
- **Investment patterns**: why private capital is flowing toward applications while public funds target infrastructure;
- **Technical cost trends**: how efficiency improvements are democratising access to high-performance capabilities;[13] and
- **Capability expansion**: how advances in reasoning and multimodal processing are creating new deployment opportunities.

2.1.1 Opportunities for entry throughout the GenAI value chain

Competition dynamics differ across the GenAI stack, but opportunities for entry exist at all levels. Infrastructure investments are attracting both public funding and private capital, while model development remains vibrant with new entrants like Mistral AI and DeepSeek successfully competing against more established players. The deployment layer is experiencing particularly rapid growth as firms customise existing models to build specialised applications, creating the most diverse competitive landscape.

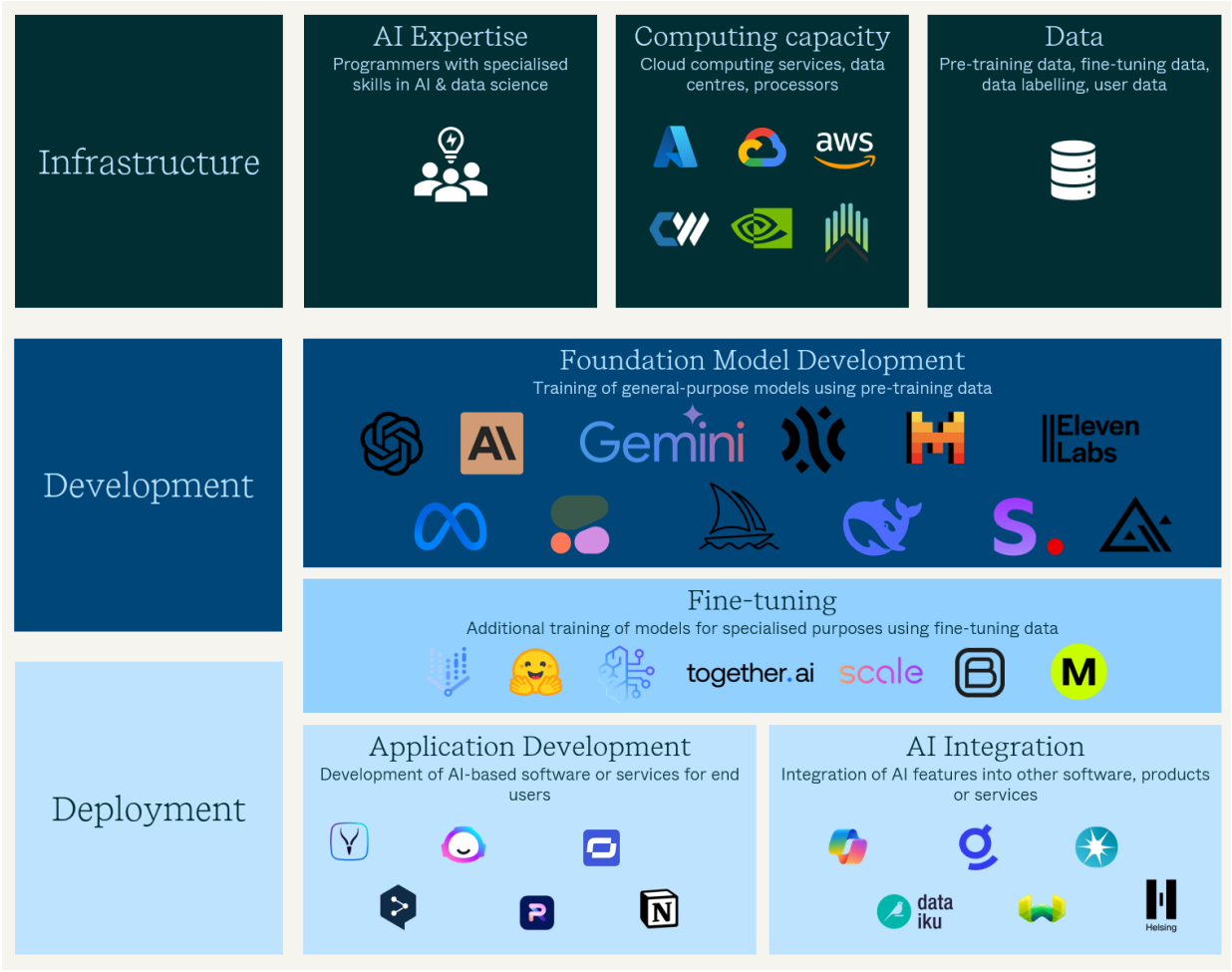
2.1.1.1 Overview of the GenAI value chain

The GenAI value chain consists of three main layers, as illustrated in Figure 3 below: infrastructure (expertise, computing capacity, and data), model development, and application deployment.[14] Understanding this structure is essential for assessing where competitive advantages emerge and how market power can be exercised or constrained.

[13] Edge deployment refers here to AI computing that is performed near the user or the source of inputs, such as on a personal device or a computer in a factory rather than a centralised server. See https://www.ibm.com/think/topics/edge-ai (accessed in May 2025).

[14] See FTC (2025).

Figure 3: Overview of the GenAI value chain



Source: *RBB Economics based on Competition & Markets Authority (2023a) and European Economic and Social Committee (2025).*
Note: *Fine-tuning is commonly associated with both the development and deployment stages of GenAI. During development, model creators take a pre-trained FM and fine-tune it on a specific dataset to adapt it for particular use cases, domains, or customer needs. This step enhances performance, relevance, and alignment with desired outputs before the model is deployed for use. Fine-tuning occurs closer to deployment, especially in enterprise settings where models are tailored to proprietary data. The inputs from the infrastructure layer are utilised at every stage of the value chain.*

- **AI Infrastructure**.
  - *Expertise.*[15] Building and operating AI infrastructure at scale requires deep technical expertise across multiple domains, including chip design, systems engineering, AI research, and cloud orchestration. This specialised talent typically comes from universities and research institutions, then is trained at companies working on large-scale AI systems. Countries and regions must not only develop this expertise through their academic institutions, but also attract talent and retain it against global competition for these relatively scarce skills.
  - *Computing capacity.*[16] Purpose-built AI chips can train GenAI models more quickly and efficiently at scale than general-purpose chips. Given their limited supply, and relatively high cost, most GenAI developers rent access to computing capacity via cloud providers. These chips are housed in large data centres requiring years to build, continual maintenance and sophisticated software for managing their day-to-day operations.[17] This infrastructure remains essential even after development, as

15 See https://www.european-processor-initiative.eu/project/epi/ (accessed in May 2025).
16 See p. 9–11 in FTC (2025).
17 While many GenAI workloads are run on large-scale cloud infrastructure, some developers, particularly leading model firms, invest directly in their own data centres and custom chips to optimise performance and cost. Networking capabilities are especially critical for inference workloads, where latency and data transfer speed affect user experience and system efficiency, though they also play a role in coordinating distributed training. See https://lenovopress.lenovo.com/lp2225-on-premise-vs-cloud-generative-ai-total-cost-of-ownership (accessed in June 2025).

deployed models need computing power for inference, i.e., to generate a response to each user-inputted query. However, this does not necessarily need to be the same infrastructure used for training the models.

- *Data.* GenAI models require training on large-scale datasets. For example, DeepSeek-V3 was pre-trained on 14.8 trillion tokens, roughly corresponding to 11.1 trillion words.[18] After initial training, models are fine-tuned with additional application, industry, or task-specific data.[19] This step provides deep, contextual knowledge to the model, and prepares the model for deployment to end users. Data comes from publicly available sources like Wikipedia and BookCorpus, web-scraped content and proprietary datasets.[20]

- **Development**. Models are trained using statistical techniques that help them learn patterns in data, for example, how words relate to each other in Large Language Models ("LLMs").[21] In the following, we focus our description on these LLMs, but the concepts apply similarly to other non-language-based FMs (audio, video, image, 3D, code, etc).

  Model development can be split into two stages, FM training ("pre-training") and fine-tuning.[22] Pre-training is resource-intensive and sets the model's initial parameters (weights that determine how words link together). This creates general-purpose models can "understand" and produce text, but that are not specialised in a particular task or area. Fine-tuning then reconfigures some weights using application-specific data to improve performance for particular domains. Fine-tuning requires far fewer resources and typically occurs when deploying for specific use cases.

- **Deployment**.[23] Models can reach users in several ways. Developers can create direct consumer interfaces like OpenAI's ChatGPT that offer access to their full model suite. Alternatively, they can provide Application Program Interface ("API") that allow external developers to build GenAI features into their own applications.[24] Developers may also license access through third party AI Platforms that package models with additional tools and services. Open-source models can be downloaded and run locally on a user or developers' own machines or private cloud.

  These deployment strategies aren't mutually exclusive: the same developer might simultaneously offer a public chatbot, provide API access for other companies, and build custom enterprise solutions using the same core model.[25] For instance, OpenAI offers ChatGPT for consumers, APIs for developers, and enterprise versions for businesses. The deployment choice can shape customisation levels: public offerings tend to be generic, while enterprise deployments often involve fine-tuning or prompt engineering tailored to clients' proprietary data and specific requirements.[26] Section 4 of this study provides further details and practical examples of GenAI deployment.

---

18 See https://huggingface.co/deepseek-ai/DeepSeek-V3 (accessed in May 2025). To calculate the number of words, we use the heuristic provided in FTC (2025) that a token is ¾ of a word.

19 See https://www.ibm.com/think/topics/fine-tuning (accessed in May 2025).

20 See p. 13 in FTC (2025).

21 See https://openai.com/index/generative-models/ (accessed in May 2025).

22 See p. 12–13 in FTC (2025).

23 See p. 15–16 in FTC (2025). In addition to accessing proprietary models via APIs or enterprise services, developers can also download and deploy openly licensed GenAI models directly. Models such as Mistral and Falcon are available under permissive or research licenses, allowing users to run them locally, fine-tune them for specific tasks, or integrate them into custom applications without relying on a third-party provider. See https://voicebot.ai/2023/09/28/generative-ai-startup-mistral-releases-free-open-source-7-3b-parameter-llm/; https://www.tftc.io/falcon-180b-leading-the-open-source-ai-revolution-with-advanced-language-modeling/ (both accessed in June 2025).

24 Application Programming Interface (API) - a standardised way for developers to access model capabilities programmatically without hosting the model themselves. Instead of interacting with a model through a chat user-face, developers can send the same user query programmatically, receive the output, and use it in their own applications directly. See https://aws.amazon.com/what-is/api (accessed in May 2025).

25 For example, OpenAI provides public access to ChatGPT through a web interface, offers API access via its developer platform, and supports tailored enterprise deployments through ChatGPT Enterprise. See https://openai.com/api/ (accessed in June 2025).

26 To maximise total customer use of their services, IT providers are incentivised to: (i) allow customers to use the best-in-class GenAI services (e.g., LLMs) developed by third parties on their existing infrastructure services; and (ii) more generally, to win customers' incremental workloads, allow their services to work together with other IT providers' services that a customer might choose to use to run their generative AI solutions. If they do not, customers may choose to use other IT providers' services (e.g., for both their GenAI and other workloads).

2.1.1.2 FM developers compete across multiple dimensions

No single GenAI model outperforms all others across every use case. Models vary widely in their capabilities, with significant product differentiation between developers and no single, standardise measure for a model's performance.[27] Some models excel at reasoning or code generation, while others perform better at multimodal tasks, or offer greater speed, customisation flexibility or cost. This means the best model choice depends on the specific application being developed, and user and preferences.

Key differentiation factors include:

- **Multimodality**[28]**:** Some models process multiple data formats simultaneously (text, image, audio, video), enabling richer analysis. A medical multimodal model might analyse patient scans while listening to doctor consultations to recommend treatments. Others handle only specific formats (e.g. text-to-text) or cross-modal tasks like text-to-video generation.
- **Inference speed**[29]**:** Inference speed measures the time required for a model to generate a response once prompted, affecting real-time applications.
- **Context window**[30]**:** This determines how much information a model can "remember" when responding to a query. Large context windows enable processing lengthy documents without losing continuity, while smaller windows force chunking that can lose important connections.
- **Openness**[31]**:** Open models publish their weights and allow users to adjust these weights (fine-tuning) for their use case. They are often available for free for local deployment. Closed models typically require API access and ongoing payments, and do not always offer fine-tuning options.

This differentiation creates competitive opportunities. Even single developers like OpenAI acknowledge their various models suit different purposes.[32] This differentiation is more pronounced when considering models from different developers, pointing to potential benefits from combining and switching between multiple models. Indeed, as stated by Microsoft: *"Leveraging the strengths of different AI models … can be a great strategy to help* [developers] *meet* [their] *performance objectives. This approach harnesses the power of multiple AI systems to improve accuracy and reliability in complex scenarios."*[33]

Many deployers now use different foundation models for different tasks, for instance, one for summarisation, another for coding. Some even dynamically switch between models for the same task. This drives demand for orchestration tools (see section 4.1.2 below) that assess which model suits each task and manage switching between providers to optimise for performance, cost, or other variables. By reducing technical barriers to using multiple models, these tools enable "multi-sourcing" rather than lock-in. This flexibility enhances competition between model developers, as deployers can more readily switch providers in response to price, performance, or innovation.

2.1.1.3 The GenAI market offers opportunities across various layers of the value chain

Demand for innovative GenAI solutions is growing rapidly across the value chain. The use of GenAI is expected to soar in the upcoming years, with Morgan Stanley anticipating that revenue from GenAI could

---

27 The performance and capabilities of models are measured using benchmarks. There are many benchmarks, and they measure the performance of a model on specific tasks, often by recording the percentage of correct answers on a test. For example, a coding benchmark, such as the HumanEval (developed by OpenAI), measures how a model manages to solve coding questions asked on the internet, such as on Stack Overflow or issues raised on GitHub repositories.

28 See https://huyenchip.com/2023/10/10/multimodal.html (accessed in May 2025).

29 See https://ubiops.com/reducing-inference-costs-for-genai/ (accessed in May 2025).

30 See https://cloud.google.com/transform/the-prompt-what-are-long-context-windows-and-why-do-they-matter (accessed in May 2025).

31 There is a range of degrees of openness of a model. At one end of the range, some models share everything with the public (fully open). At the other end, some models do *not* share anything and cannot be used by third parties (fully closed).

32 See https://help.openai.com/en/articles/11165333-chatgpt-enterprise-models-limits (accessed in May 2025).

33 See https://azure.microsoft.com/en-us/blog/boost-processing-performance-by-combining-ai-models/ (accessed in May 2025).

exceed 1 trillion USD by 2028.[34] This anticipated growth is reflected in GenAI company formation: according to the AI Index Report 2025, newly funded GenAI companies increased from 19 globally in 2019 to 214 in 2024.[35]

While opportunities exist throughout the value chain, deployment offers particularly attractive prospects. Industrial applications using existing models fine-tuned for specific use cases can deliver immediate benefits without requiring massive investments. However, successful new entrants continue emerging in model development and infrastructure as well.[36]

Entry opportunities include:

- *Application layer.* Companies develop industry-specific applications and/or services using existing GenAI models. The demand for industry-specific services is nascent but growing, and many new firms are addressing this demand in industries like manufacturing, legal services, and education, to name a few.[37] Beyond text applications, image, video, and voice generation serve marketing, advertising, and content creation markets.[38]
- *FM development*. Despite high compute and data requirements, new players are still able to successfully enter and compete in a dynamic market. Recent successes include Mistral AI, LightOn, or DeepSeek. Additionally, small language models such as Microsoft's Phi-4 illustrate how computationally efficient models can be high-performing, levelling the playing field.[39]
- *Model fine-tuning*. Specialised firms help businesses with proprietary data but limited technical expertise in-house build their own custom GenAI-powered tools.[40] Such firms bridge the gap between general purpose models and tools and specific business needs.
- *Data management services*. High-quality data preparation is essential for effective GenAI deployment. Companies need document and data to be readily available, in a suitable format, possibly annotated and properly itemised. This creates an opportunity for firms specialising in data preparation and management in anticipation of the deployment of GenAI capabilities.[41]

2.1.2 Open models, APIs and partnerships can all help lower barriers to entry

Open-source models, APIs, and strategic partnerships help smaller firms overcome barriers to entry, particularly resource constraints such as the capital needed for compute power, high-quality data, and specialised talent.

---

34 See https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai; https://www.morganstanley.com/insights/articles/genai-revenue-growth-and-profitability (both accessed in May 2025).

35 See Figure 4.3.5 in Stanford Institute for Human-Centered Artificial Intelligence (2025), "the AI Index Report 2025". Private funding data for the AI Index Report 2025 was provided by Quid. The data is based on Capital IQ and Crunchbase and includes information on early-stage funding rounds and their timings, as well as when the company was founded. See Stanford Institute for Human-Centered Artificial Intelligence (2025); https://www.quid.com/ (both accessed in May 2025).

36 See https://www.mckinsey.com/capabilities/quantumblack/our-insights/exploring-opportunities-in-the-generative-ai-value-chain (accessed in May 2025).

37 Several firms develop generative AI applications tailored to the unique needs of specific sectors. Siemens integrates AI into industrial automation systems to optimise manufacturing processes and predictive maintenance. In legal services, Harvey builds on OpenAI's models to deliver tools for contract analysis and legal research, tailored to the workflows of major law firms. In education, Khan Academy has introduced Khanmigo, a generative AI tutor designed to support personalised learning experiences for students and assist teachers with lesson planning and feedback. See https://press.siemens.com/global/en/pressrelease/siemens-expands-industrial-copilot-new-generative-ai-powered-maintenance-offering (accessed in June 2025).

38 For example, Midjourney, a US-based company, generates images from text; PhotoRoom, a Paris-based company, provides photo-editing services; ElevenLabs, based in the US, generates speech from text.

39 See https://techcommunity.microsoft.com/blog/aiplatformblog/introducing-phi-4-microsoft%E2%80%99s-newest-small-language-model-specializing-in-comple/4357090 (accessed in June 2025).

40 Hugging Face, for instance, offers tools and services that enable developers to fine-tune open-source models on custom datasets for applications ranging from legal document analysis to biomedical research. For more examples of firms that provide similar services, see section 4.2.1.

41 For example, Scale AI provides data annotation and quality assurance services for major AI developers, helping ensure reliable and bias-mitigated training data. Similarly, Snorkel AI focuses on programmatic data labelling, allowing enterprises to rapidly generate domain-specific training sets with minimal manual intervention. These upstream players play a critical role in the AI value chain by improving the quality and efficiency of model development processes.

There is a variety of business models in GenAI deployment. Some firms like Google and Amazon operate across multiple value chain layers –developing FMs (e.g. Gemini and Titan), providing cloud infrastructure, and offering consumer-facing applications. However, specialised firms focusing on single layers currently appear to be competing effectively with their vertically integrated counterparts. Examples include Hugging Face, a GenAI community hub primarily serving as a platform for hosting models and building applications, and JasperAI, a provider of generative AI tools for marketing content creation.

Deployment-focused firms can access third-party FMs through several routes, without developing proprietary models:

- Open-source models are shared publicly on platforms like GitHub and Hugging Face, allowing direct download for on-premises or cloud deployment.
- On-premises deployment lets organisation host proprietary models within their own environment rather than via public cloud. This approach is increasingly supported by providers such as Cohere and Google.[42]
- API access enables model use in the cloud without hosting requirements. This means of accessing foundation models enables rapid prototyping of features like customer service without maintaining full model pipelines.[43]
- Strategic partnerships provide foundation model access without in-house development.

This variety of access methods ensures that firms can choose deployment strategies matching their technical capabilities, security requirements, and business models, lowering barriers for specialised competitors.

2.1.2.1 Open-source models reduce market entry barriers substantially

Open-source models provide access to advanced GenAI capabilities without the high costs of building proprietary models from scratch.[44] Businesses can integrate AI functionalities for free or at lower cost, without requiring deep in-house machine learning expertise or large-scale computing infrastructure. This substantially reduces both implementation time and development cost of GenAI solutions.

Understanding why firms choose to provide models on an open-source basis helps assess the competitive dynamics and the sustainability of open access. Economic literature shows that open-source strategies help firms grow user bases by attracting developer communities that create compelling applications, increasing the value of the platform for users. Incentives for open-source models are particularly strong when companies generate revenue from related products or services, as expanding the user base directly boosts income.[45]

For example, Google's open-source TensorFlow machine learning framework supports Google's cloud computing business by encouraging developers to build applications that run on Google's infrastructure. In the GenAI space, companies like Mistral AI release open-source models alongside premium closed-source versions, using the open models to build developer adoption and showcase capabilities whilst monetising through enterprise services.

---

42 See https://cohere.com/deployment-options; https://cloud.google.com/blog/products/ai-machine-learning/run-gemini-and-ai-on-prem-with-google-distributed-cloud (both accessed in June 2025).

43 Developers of proprietary models often offer APIs for their models, such as Google for Gemini and Anthropic for Claude. See https://ai.google.dev/gemini-api/docs; https://docs.anthropic.com/en/api/overview (both accessed in June 2025).

44 Various open-source initiatives such as the Model Context Protocol (MCP) and Agent2Agent (A2A) are driving greater interoperability and accessibility in the GenAI ecosystem. For more information on such initiatives and tools, see section 4.1.2.2.

45 See Autorité de la concurrence and Competition & Markets Authority (2014).

Some competition authorities have expressed concerns that models may be "open first, closed later", where developers initially offer open-source access to build user bases, then switch to proprietary access for new models once switching costs become significant.[46] However, the risk of consumer harm is likely to be limited in practice in the case of foundation models. Switching costs between models remain relatively low in many use cases, especially with orchestration tools and interoperability standards. In addition, the open-source ecosystem is highly dynamic, with numerous viable alternatives continuously released and improved upon by a distributed global community. This ongoing innovation helps discipline incumbents and preserves user choice.

Even if closure occurred, which has not happened with GenAI models, it would only affect future iterations of the foundation model. Earlier open-source versions would remain publicly available under the original licensing terms. This means users could continue using, modifying, and distributing the previously open versions even if the developer stopped releasing new open-source updates. In practice, many valuable open-source projects like Linux and React have remained open for years, suggesting that maintaining open-source status often proves more commercially advantageous than closure.

Open-source models enhance innovation by enabling startups, researchers, and developers to experiment with and build upon already existing models rather than starting from scratch. This fosters diversity in use cases and problem-solving approaches. Open-source communities accelerate development through global contributions to bug fixes, performance optimisations, and fine-tuned variants for specific domains. Notably, Meta's Llama and Stability AI's Stable Diffusion models demonstrate active community engagement.[47]

2.1.2.2 Partnerships are occurring more frequently at the deployment layer and can help lower barriers to entry

Given the challenges of developing models in-house, including high costs and lead time, partnerships, alongside APIs and open-source models, can play an important role in driving innovation across the value chain.[48] Collaborations between compute providers and model developers can play a pivotal role in overcoming barriers to entry and expansion by distributing costs, securing funding, and sharing expertise. While such collaborations were popular since GenAI early days, their focus and nature has evolved as the market itself matured.[49]

Partnerships now occur more frequently at the deployment layer, reflecting the industry's evolution from building GenAI capabilities to deploying them for specific use cases. Unlike traditional client-customer relationship (such as API access), partnerships involve more formal integration and cooperation structures.

Partnerships come in many shapes and forms. They can vary in structure, and equity stakes, corporate control levels, information sharing arrangements, minimum purchase commitments, and exclusivity rights all differ case by case. No generic partnership template exists, requiring individual and careful examination of each arrangement's details.

However, it is nonetheless worth noting that GenAI partnerships are typically non-exclusive, which reduces the potential for anti-competitive harm since the benefits granted to one party are generally

---

46 See Competition & Markets Authority (2024a).

47 The source files and instructions for operating open-source AI models are usually shared on platforms such as GitHub and Hugging Face. See https://github.com/meta-llama/llama-models; https://github.com/Stability-AI/stablediffusion (both accessed in June 2025).

48 See https://www.mckinsey.com/capabilities/quantumblack/our-insights/strategic-alliances-for-gen-ai-how-to-build-them-and-make-them-work; https://www.cio.com/article/1308158/how-strategic-partnerships-are-the-key-to-ai-driven-innovation.html (both accessed in April 2025).

49 Examples of such partnerships include Google-Anthropic, Microsoft-OpenAI, and Microsoft-Inflection AI. See section 5 below on GenAI partnerships.

accessible to its competitors. This non-exclusive nature helps preserve competitive dynamics while enabling the collaboration benefits that drive innovation and market entry.

Section 5.2 below provides detailed analysis of partnership types, benefits, and potential competitive effects.

2.1.3 Deployment is increasingly attracting private investment

Global GenAI investment exceeded $56 billion in 2024, almost doubling from approximately $29 billion in 2023, according to S&P Global Market Intelligence data.[50] This surge in investment reflects rising investor confidence in the commercial viability of GenAI, where initial investments into GenAI have yielded high returns, as highlighted in IDC's 2024 AI Opportunity Study.[51] Businesses are increasingly recognising GenAI's potential to drive efficiency, unlock new revenue streams, and create competitive advantages.

While governments tend to focus public funding on GenAI infrastructure, private investors are increasingly focusing on applications with clear commercial potential. Investment signals are shifting toward enterprise-grade deployment solutions and vertical applications tailored to specific industry needs.[52] In particular, private equity and venture capital are prioritising application-layer investments and industry-specific AI solutions that can deliver measurable business value. Investment in AI-native applications reached $4.6 billion in 2024, according to Menlo Ventures, an 8x increase from $600 million in 2023, demonstrating this shift toward deployment-focused funding.[53]

Governments worldwide are channelling funds into GenAI infrastructure and capabilities. This is to support various public policy objectives, such as enhancing domestic competitiveness or ensuring access to GenAI resources like compute at a national level. Major public initiatives include Europe's €200 billion *InvestAI* program,[54] the US CHIPS and Science Act's $50 billion semiconductor allocation,[55] and substantial investments by Canada ($2.4 billion), China ($47.5 billion), France (€109 billion), India ($1.25 billion), and Saudi Arabia ($100 billion).[56]

Upstream layers also continue to attract substantial private capital. This is reflected in Figure 4 below, which shows the total private funding of GenAI firms founded in the past three years, separated by region and value chain layer. In both the US and Europe, the development layer received the most private funding, $13 billion and $1.3 billion, respectively, which were approximately 4–6 times more than the downstream deployment layer.[57]

---

50 See https://www.spglobal.com/market-intelligence/en/news-insights/articles/2025/1/genai-funding-hits-record-in-2024-boosted-by-infrastructure-interest-87132257 (accessed in June 2025).

51 See https://blogs.microsoft.com/blog/2024/11/12/idcs-2024-ai-opportunity-study-top-five-ai-trends-to-watch/ (accessed in May 2025).

52 "Vertical applications" refer to software solutions designed specifically for a particular industry or sector. In contrast, general-purpose or "horizontal" tools can be used across many sectors. See https://phoenixnap.com/glossary/vertical-application (accessed in May 2025).

53 See https://www.globenewswire.com/news-release/2024/11/20/2984728/0/en/menlo-ventures-2024-state-of-generative-ai-report-reveals-6x-spike-in-enterprise-spending-as-market-leaders-shift.html; https://magai.co/generative-ai-landscape/ (all accessed in June 2025).

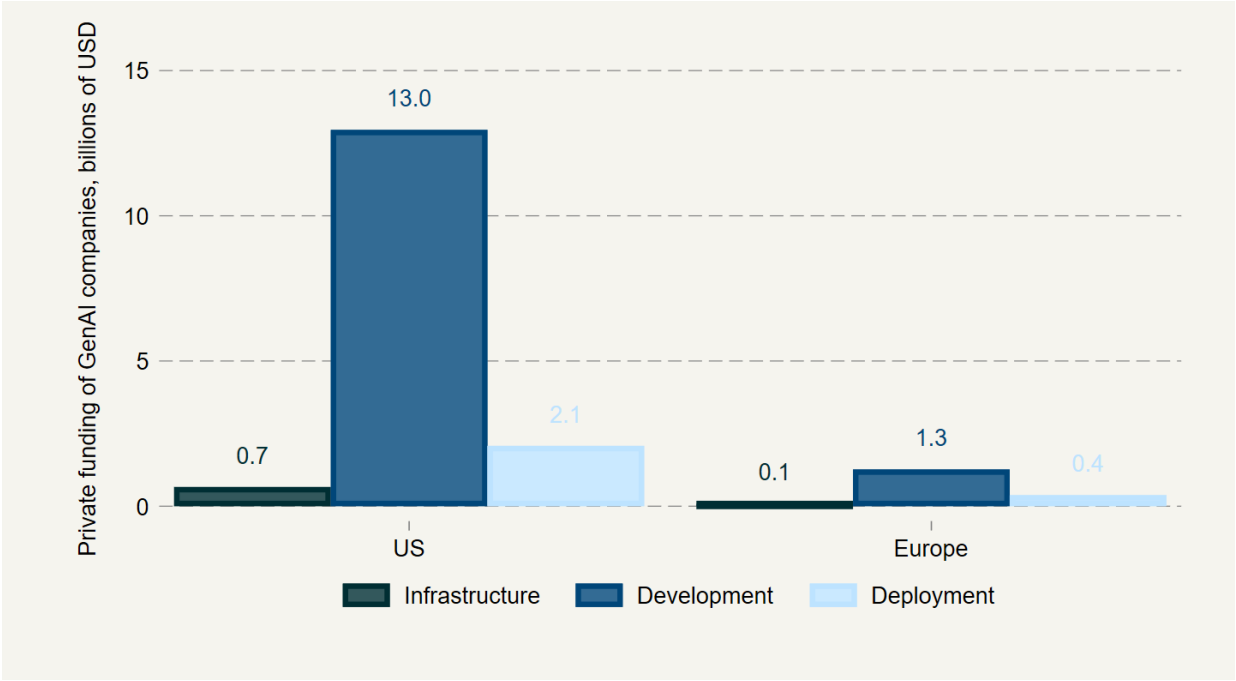
54 See https://digital-strategy.ec.europa.eu/en/news/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence (accessed in April 2025).

55 See https://www.nist.gov/chips (accessed in April 2025).

56 See the AI Index Report 2025, page 327.

57 Figure 4 only includes funding of GenAI companies founded in the past 3 years and therefore doesn't include additional funding acquired by more established companies. Definitions of GenAI companies also vary between sources, which may lead to differences in reported levels of investment.

Figure 4: Private funding of GenAI firms founded between May 2022 and April 2025 by value chain layer



Source: *Data from Crunchbase (accessed in April 2025).*
Note: *GenAI companies here include companies founded between 1 May 2022 and 30 April 2025 with the Generative AI industry tag and available funding data on Crunchbase. The assignment of companies into value chain layers was done by RBB Economics based on industry tags and keyword matching from company descriptions. Europe includes companies with headquarters in the EU or the UK. Companies without a headquarters location in the data, whose funding represents c. 5% of the total in the period, are omitted from the figure.*

While upstream layers continue to attract the most private investment in absolute terms, there are signs that private capital is beginning to move more towards the deployment layer. The growing popularity of consumer-facing GenAI tools like Google's Gemini and OpenAI's ChatGPT has fuelled market expansion, including by raising awareness of GenAI and its capabilities. Bloomberg Intelligence is projecting the industry to grow more than 30-fold from 2022 levels over the next decade.[58] This rapid increase is reported to be driving VC investors' interest in deployment and GenAI application businesses that build specialised software using third-party FMs for consumer or enterprise use.[59]

In addition, some major partnerships have focused on deployment and application layer integration, which further signals GenAI's move toward deployment. Salesforce's collaboration with Hugging Face, for instance, supports the development of GenAI assistants within its CRM ecosystem,[60] while AWS's partnerships with firms like Anthropic aim to accelerate the operationalisation of FMs for enterprise use.[61]

The increasing focus on the deployment layer reflects the considerable commercial opportunities offered by industrial GenAI applications. There is growing enterprise demand for purpose-built or customised GenAI tools.[62] These customised tools offer enhanced performance and easier integration compared to general-purpose tools, creating competitive advantages for specialised providers over incumbent generalists, and driving investment towards the deployment layer.

58 See https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/ (accessed in June 2025).
59 See https://www.lexology.com/library/detail.aspx?g=24084477-088f-41ed-ae7f-a0c8d3f2a6dd (accessed in April 2025).
60 See https://www.salesforce.com/news/stories/agentforce-ai-models-announcement/ (accessed in June 2025).
61 See https://venturebeat.com/ai/exclusive-aws-accenture-and-anthropic-partner-to-accelerate-enterprise-ai-adoption/ (accessed in June 2025).
62 For example, see https://www.kalisa.ai/insights/beyond-generic-ai-the-rise-of-customised-vertical-genai-applications-unpublished; Markets and Markets (2025).

Recent adoption data demonstrates this shift towards specialisation and customisation. According to the AWS GenAI Adoption Index, 58% plan to build custom applications using pre-trained models, while 55% aim to rely on fine-tuned models, and 25% intend to develop solutions entirely from scratch in-house reflecting both model adaptability appeal and in-house expertise limitations.[63] This trend creates opportunities for firms that can bridge the gap between general-purpose models and specific business needs.

Successful targeted tools span multiple sectors with measurable impact:

- Jasper AI focuses on marketing content (raising over $125 million in 2022);[64]
- Harvey AI reports serving 235 law firms across 42 countries legal research tools for law firms, achieving a reported $50 million in annual recurring revenue by 2024 (400% growth year-over-year);[65] or
- GitHub Copilot reached 1.3 million paid subscribers by early 2024, with enterprise adoption accelerating as developers report 30% acceptance rates for AI suggestions and significant productivity gains in surveys.[66]

A growing number of generative AI startups focused on the application layer have even surpassed the $1 billion valuation mark primarily due to strong traction in enterprise deployments and clear product-market fit.[67] Some examples of these so-called "unicorns" in the application layer are:[68]

- Glean (valued at $4.6 billion, December 2024), which offers GenAI-powered enterprise search and knowledge management tools;
- Perplexity AI ($9 billion, December 2024), a conversational GenAI search engine focused on real-time, citation-based question answering; and
- Synthesia, a platform for generating AI-powered video content using synthetic avatars and voiceovers, reached $100 million in annual recurring revenue by 2025, serving over 60,000 customers including 60% of Fortune 100 companies, with a $2.1 billion valuation.[69]

The specialisation trend appears to reduce barriers to entry for new competitors at the deployment layer. Application-focused firms can achieve revenue generation relatively rapidly, as evidenced by the growth trajectories above. The performance of these firms suggests that market value increasingly derives from solving specific use cases rather than developing general-purpose capabilities, though this comes with dependence on foundation model providers for core functionality.

Technology firms are responding by integrating GenAI into existing products rather than competing solely on model capabilities. Microsoft's OpenAI integration across Office 365 and Azure turns FM access into a feature of mainstream enterprise platforms. Adobe's internally developed Firefly image-generation model is now embedded across its creative suite, exemplifying how large firms are building GenAI capabilities into their products to boost competitiveness.

---

63 See https://amazongca.getbynder.com/share/F5DE1F38-F55B-4953-9B66D6D023045E27 (accessed in May 2025).
64 See https://www.jasper.ai/blog/jasper-announces-125m-series-a-funding (accessed in April 2025).
65 See https://fortune.com/2025/02/12/legal-ai-startup-harvey-300-million-series-d-funding-3-billion-valuation-sequoia/ and https://aimresearch.co/market-industry/harvey-ai-came-out-of-nowhere-and-took-over-legal-tech (accessed in June 2025)
66 See https://www.ciodive.com/news/github-copilot-subscriber-count-revenue-growth/706201/, https://github.blog/news-insights/research/research-quantifying-github-copilots-impact-in-the-enterprise-with-accenture/ and https://www.opsera.io/blog/github-copilot-adoption-trends-insights-from-real-data (accessed in June 2025)
67 See https://www.cbinsights.com/research/ai-agent-market-map/ (accessed in April 2025).
68 See https://www.forbes.com/lists/ai50/ (accessed in April 2025).
69 See https://www.glean.com/blog/glean-series-e-prompting-launch; https://www.ft.com/content/d4fb70f9-b971-433b-884c-2f01d1d08968; https://www.ft.com/content/3a35f3ba-7273-41ea-a0a5-77fe46965e63, https://www.synthesia.io/post/synthesia-secures-180m-in-series-d-funding (all accessed in April 2025).

This investment momentum towards deployment is supported by favourable cost trends that make GenAI applications increasingly accessible to a broader range of organisations.

2.1.4 The deployment of high-performance models is becoming cheaper, even though frontier model training costs grow

Two divergent cost trends are shaping GenAI competition. Training costs for frontier models, the most advanced AI systems that push the boundaries of capabilities, have grown at an estimated 2.4x per year since 2016.[70] Recent examples illustrate this trajectory: training GPT-4 cost an estimated $78 million, while Google's Gemini Ultra required $191 million worth of compute, compared to just $930 for the foundational Transformer model in 2017.[71] If current trends continue, the largest training runs will exceed $1 billion by 2027.[72]

However, costs for high-performance (non-frontier) models development and deployment are declining significantly.[73] This creates favourable conditions for GenAI deployment, which typically does not require frontier capabilities. This cost divergence follows a familiar pattern from earlier technological waves: expensive cutting-edge R&D, followed by optimisations that reduce downstream costs and broaden adoption.[74]

Smaller, cost-effective models reduce deployment costs. Architecture optimisations such as knowledge distillation and quantisation have significantly reduced model size and inference costs while preserving core performance.[75] These techniques enable efficient deployment on less expensive hardware and mobile devices.

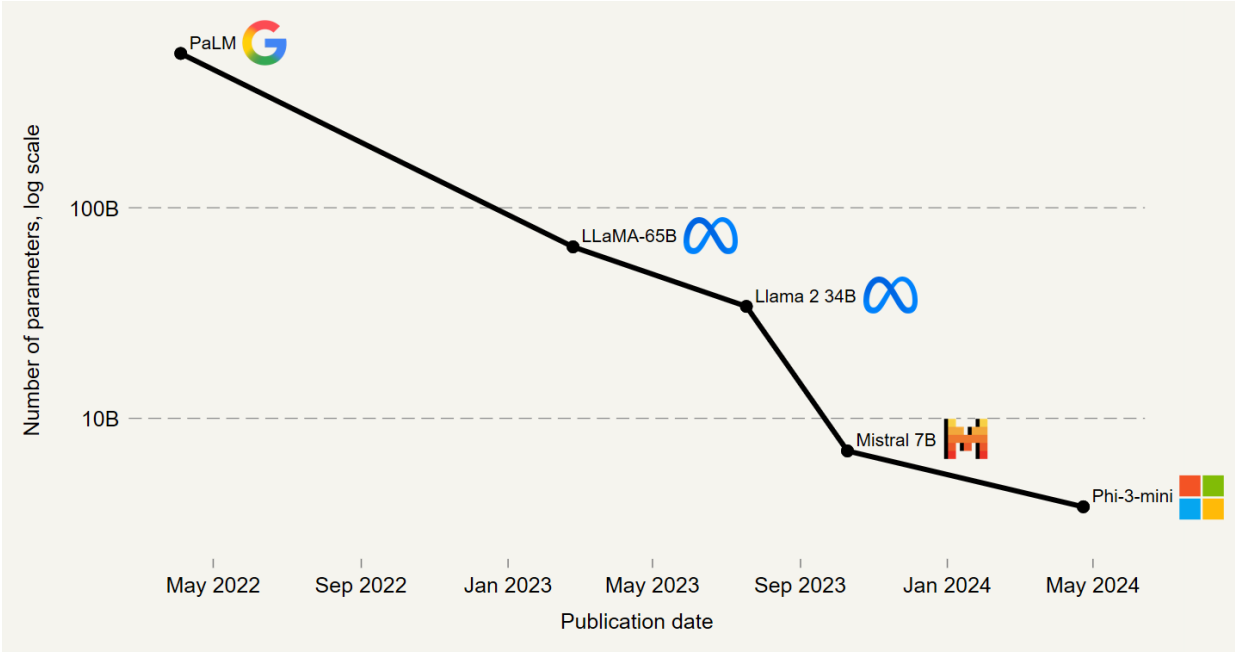
Recent examples demonstrate the potential of these optimisation approaches. DeepSeek's models achieved performance comparable to leading frontier systems while requiring dramatically fewer resources: it was reported that DeepSeek-V3 was trained for approximately $5.6 million compared to over $500 million reportedly spent on Meta's Llama 3, representing roughly an 11x efficiency gain.[76] Similarly, Mistral Medium 3, released in May 2025, delivers state-of-the-art performance at 8x lower cost with radically simplified enterprise deployments.[77]

Systems performing at the level of GPT-3.5 saw inference costs reduce dramatically. According to the AI Index Report 2025, inference costs for these systems reduced more than 280-fold between November 2022 and October 2024, in large part due to these technical advances.[78] Figure 5 below illustrates this trend: the smallest foundation models (i.e. with the fewest parameters) scoring above 60% on the Massive Multitask Language Understanding (MMLU) benchmark, which measures general reasoning abilities across academic and professional subjects, decreased substantially between 2022 and 2024.[79]

---

70 See Cottier, B., et al. (2024).
71 See https://www.voronoiapp.com/technology/The-Training-Costs-of-AI-Models-Over-Time-1334 (accessed in June 2025)
72 See https://epoch.ai/blog/how-much-does-it-cost-to-train-frontier-ai-models (accessed in June 2025)
73 See Cottier, B., et al. (2024).
74 See https://www.britannica.com/technology/computer/Supercomputer (accessed in April 2025).
75 Knowledge distillation is a model compression technique where a smaller "student" model learns to replicate the behaviour of a larger "teacher" model, preserving performance while reducing computational requirements. Quantisation reduces the numerical precision of model weights (e.g., from 32-bit to 8-bit or 4-bit), significantly decreasing memory requirements and inference costs with minimal performance impact. See https://www.ibm.com/think/topics/knowledge-distillation; https://www.qualcomm.com/news/onq/2019/03/heres-why-quantization-matters-ai (both accessed in April 2025).
76 See https://www.maginative.com/article/deepseek-v3-achieves-frontier-ai-performance-at-a-fraction-of-the-cost/, https://www.rdworldonline.com/this-week-in-ai-research-a-0-55-m-token-model-rivals-openais-60-flagship/ and https://epoch.ai/gradient-updates/what-went-into-training-deepseek-r1 (accessed in June 2025).
77 See https://mistral.ai/news/mistral-medium-3 (accessed in May 2025).
78 See the AI Index Report 2025, page 64.
79 Measuring MMLU is a prominent benchmark used for evaluating the general capabilities of LLMs. See https://crfm.stanford.edu/2024/05/01/helm-mmlu.html (accessed in June 2025).

Figure 5: Smallest GenAI models scoring above 60% on MMLU, 2022–2024



Source: Figure 2.1.38 in the AI Index Report 2025.

These efficiency improvements make high-performance GenAI models accessible to organisations with limited resources, expanding the competitive field and scope for deploying GenAI applications.

2.1.4.1 Advancements in edge and hybrid architectures expand use cases

Alongside more efficient models, deployment architectures are evolving beyond centralised clouds or local machines. "Edge AI", which refers to running foundation models directly on end-user devices such as smartphones, IoT devices, and wearables, rather than on remote servers, cuts latency, improves privacy and reduces infrastructure costs. For example, AI startup Hugging Face has released a new app for iOS that uses offline, local AI to describe what's in view for an iPhone's camera.[80] Similarly, Google's Gemini Nano has been developed specifically to run for on-device tasks.[81] Reflecting this trend, OpenAI recently announced the acquisition of the AI hardware startup io Products, co-founded by former Apple design chief Jony Ive, to develop AI-native consumer devices.[82]

Hybrid architectures combine on-device processing for routine tasks with cloud-based support for complex workloads and emerge as a tool for enterprises wishing to deploy GenAI while balancing responsiveness, cost, and control. Apple, for example, runs elements of its AI processing (e.g. Siri and image enhancements) directly on-device, while Microsoft's Azure Percept platform enables hybrid edge-cloud deployments for industrial use cases.[83]

This architectural shift matters for competition. Edge capabilities reduce dependence on infrastructure for deployment and enable new application types. As edge performance improves, more GenAI functions could run independently of large-scale infrastructure, lowering barriers for developers lacking access to expensive computing resources.

80 See https://techcrunch.com/2025/03/19/hugging-faces-new-ios-app-taps-ai-to-describe-what-youre-looking-at/ (accessed in May 2025).
81 See https://deepmind.google/models/gemini/nano/ (accessed in May 2025).
82 See https://openai.com/sam-and-jony/ (accessed in May 2025).
83 See https://machinelearning.apple.com/research/introducing-apple-foundation-models; https://azure.microsoft.com/en-us/blog/azure-percept-edge-intelligence-from-silicon-to-service/ (both accessed in April 2025).

2.1.4.2 The cost of training advanced frontier GenAI models remains high

While deployment costs fall, frontier model development appear to require increasing resources. Due to the business sensitive nature of costs, there is limited publicly available (and reliable) information on the topic. Epoch AI's, a research institute, estimates frontier GenAI models training costs around $40–100 million in 2024, $1 billion for cutting-edge 2024 models.[84] These frontier training costs have increased significantly in the past years and are expected to keep increasing.[85]

These rising costs stem from the empirical relationship between the size of a model and its performance. Larger models, i.e. those trained on more data and with more parameters, have typically been found to perform better on a wide range of general benchmarks compared to smaller models.[86] Most experts consider that there are still performance gains to be achieved through increases in size. Thus, it is likely that the cost of training Frontier GenAI models will continue to increase due to rising computational requirements.[87]

However, this size-performance relationship does not hold universally. Smaller models trained on specialised data may even outperform general purpose larger models on specific tasks.[88] For instance, Microsoft's Phi-4, specialised in complex reasoning, outperformed much larger models on math-related reasoning.[89] This reinforces the emerging trend towards task-specific optimisation rather general-purpose scaling in deployment.

The cost divergence creates distinct competitive dynamics. Frontier development concentrates among well-funded organisations pursuing scientific advancement. Practical deployment increasingly focuses on efficiently matching models to use cases, where specialised approaches can compete effectively without frontier-level resources, at increasingly low deployment and inference costs.

2.1.5 Recent developments in technical capabilities have expanded potential GenAI use cases

Recent technical developments in reasoning capabilities, multimodal integration, and autonomous agents are creating new use cases for GenAI across industries.[90] These advances matter for competitive dynamics as they expand the addressable market, enlarging the scope of where GenAI can bring value, and create opportunities for specialised applications relying on GenAI capabilities that can be deployed at scale.[91]

Each of the key technical advances of the last year has expanded the ways that GenAI can help firms and users. Notably, these advances include:

- enhanced reasoning and multimodal integration enabling complex problem-solving;
- emergence of autonomous agents for complex workflows automation; and
- evolving developer tools that democratise access.

---

84 Concerning the cost of training models: *"There are models in training today that are more like a billion"*. See https://www.tomshardware.com/tech-industry/artificial-intelligence/ai-models-that-cost-dollar1-billion-to-train-are-in-development-dollar100-billion-models-coming-soon-largest-current-models-take-only-dollar100-million-to-train-anthropic-ceo (accessed in May 2025).

85 See Cottier et al. (2024).

86 *"Model performance improves with more parameters and training data"*. See Cottier et al. (2024).

87 *"Improving AI capabilities demand exponential increases in computing power"*; "[regarding the future] *the most extensive publicly available model will cost one billion dollars to train"*. See https://humandrivenai.com/2024/07/16/the-future-of-ai-scaling-laws-and-the-path-forward/ (accessed in May 2025).

88 See https://www.weforum.org/stories/2025/01/ai-small-language-models/ (accessed in June 2025).

89 See https://techcommunity.microsoft.com/blog/aiplatformblog/introducing-phi-4-microsoft%E2%80%99s-newest-small-language-model-specializing-in-comple/4357090 (accessed in June 2025).

90 See Department for Science, Innovation & Technology (2025).

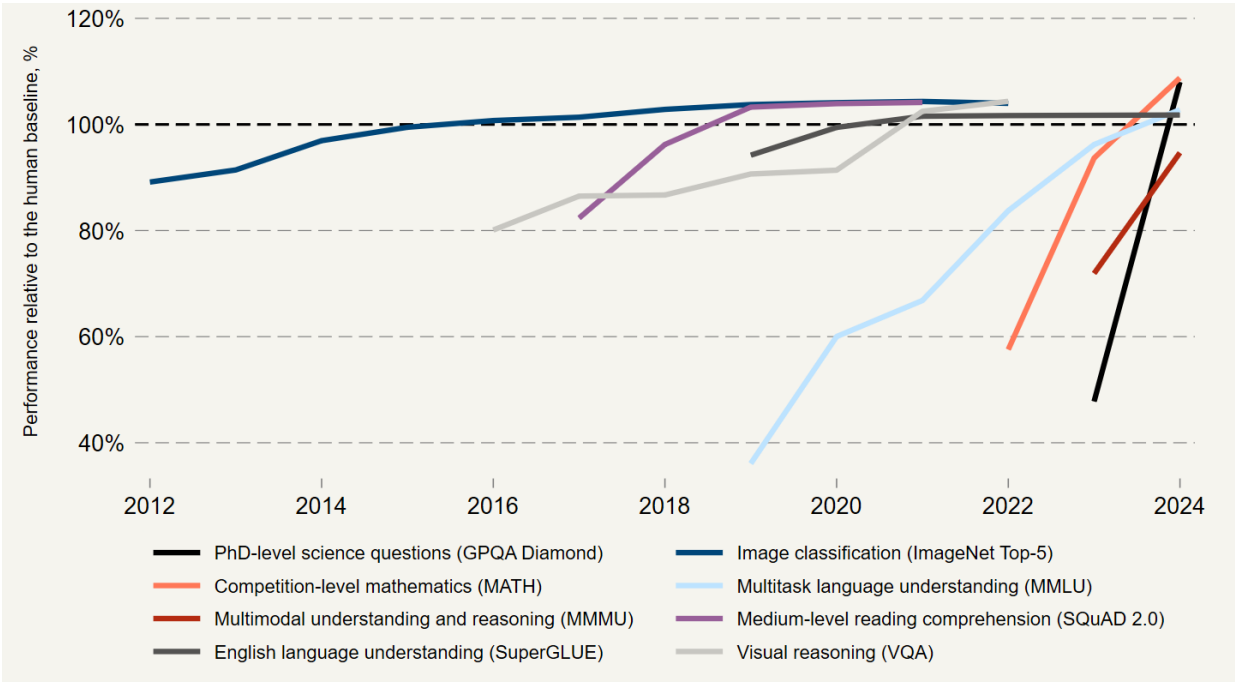
91 See McKinsey & Company (March 2025), *The State of AI: How Organizations Are Rewiring to Capture Value*. See https://www.mckinsey.com/~/media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai/2025/the-state-of-ai-how-organizations-are-rewiring-to-capture-value_final.pdf (accessed in June 2025).

Enhanced reasoning and multimodal integration. Advanced models like GPT-4, Claude 4, and Gemini 1.5 exhibit stronger "chain-of-thought reasoning" and contextual awareness, making them able to handle complex problem-solving. This enables applications beyond simple content generation, such as summarising legal documents, debugging code, or modelling financial risk.

Interestingly, research suggests that these reasoning improvements in LLMs are more efficiently obtained with increased inference time, rather than increased model size alone. Smaller models with additional test-time compute can outperform models 14 times their size on reasoning tasks, suggesting that specialised optimisation may compute effectively with model size scaling.[92]

Meanwhile, multimodal integration has expanded the reach of GenAI systems beyond text processing. Models that simultaneously handle text, image, audio, and video enable richer analysis across media, healthcare, and human-computer interaction. OpenAI's GPT-4 with Vision and Google's Gemini series represent major advances in this direction.[93] Figure 6 below shows the AI model performance benchmarks relative to the human baseline.

Figure 6: AI model performance benchmarks relative to the human baseline (100%)



Source: *Figure 2.1.33 in the AI Index Report 2025.*
Note: *The human baselines are based on the results of human participants in studies that are generally conducted by the research team that initially presented each novel benchmark, and may involve experts, non-experts, or both depending on the benchmark. The human baselines and their measurement therefore differ for each benchmark. The performance of AI models per task is then scaled against the measured human baseline.*

Emergence of autonomous agents for complex workflows. Autonomous AI agents represent a significant development: these are systems that combine FMs with planning capabilities, memory, and tool usage to independently accomplish complex multi-step tasks with minimal human oversight. Agents can act as scheduling assistants, travel planners, or internal workflow automation tools by integrating memory, API calls to other software, and goal-driven behaviour.

Examples include open-source projects like AutoGPT and commercial offerings like Cognosys and Adept's ACT-1, demonstrating how agents are being trialled in areas ranging from customer service to

[92] See Snell et al. (2024).
[93] OpenAI, *GPT-4 Technical Report*. Google DeepMind, *Gemini Technical Overview*. See https://openai.com/research/gpt-4; https://deepmind.google/technologies/gemini (both accessed in April 2025).

enterprise operations.[94] Interoperability standards are also emerging, such as the A2A protocol announced by Google and a number of industry partners, which aims to facilitate collaboration between agents.[95]

Evolving developer tools and democratised GenAI deployment. *Low-code and no-code platforms* such as Cursor or Replit enable non-specialists to build GenAI-powered tools, while *code generation assistants* like GitHub Copilot or Amazon CodeWhisperer streamline and accelerate software development. These tools lower technical barriers and expand the pool of potential GenAI and software developers.

These technical advances increase deployment demand by expanding use cases while reducing implementation barriers. The combination of improved performance and broader accessibility makes GenAI adoption attractive across more sectors and organisations.

This expansion creates particular opportunities for regions with strong domain expertise in specific industries. Where local firms possess deep knowledge of sector-specific challenges and data, they can leverage these technical capabilities to build specialised solutions that compete effectively with general-purpose offerings. Europe's established presence in manufacturing, automotive, financial services, and healthcare positions it well to capitalise on these expanding deployment opportunities.

2.2 The GenAI landscape in Europe

Europe faces distinct challenges and opportunities in GenAI competition. While lagging behind the US and China in absolute investment levels, Europe has built meaningful capabilities across the value chain and possesses competitive advantages in specific sectors, which may have significant impacts on the European economy as a whole. Understanding Europe's position requires examining both its structural constraints and its strategic strengths.

This section analyses:

- European initiatives building GenAI infrastructure and capabilities in both development and deployment;
- Structural barriers limiting European scale, including funding gaps and cultural, linguistic, and regulatory market fragmentation;
- Competitive opportunities arising from Europe's industrial expertise and collaborative research networks.

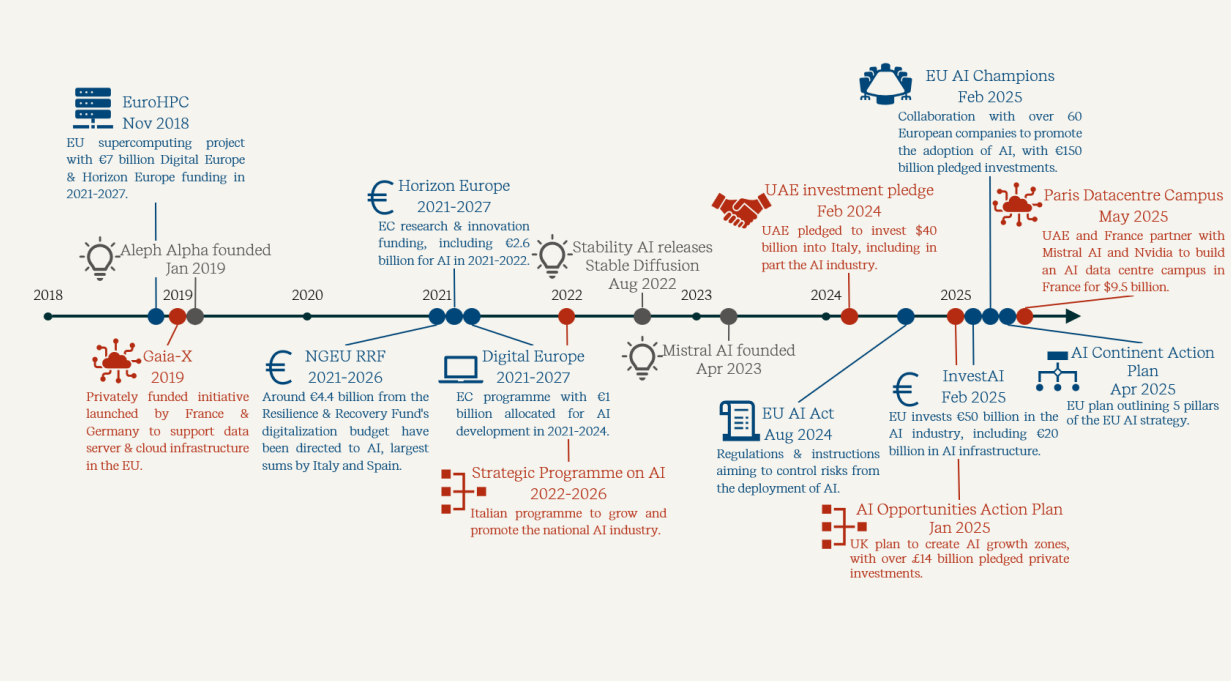
2.2.1 European infrastructure initiatives are focused on funding GenAI capabilities

European institutions have launched targeted initiatives across the GenAI value chain, focusing on areas where coordinated investment can build competitive advantage. These efforts concentrate on infrastructure development and foundation model capabilities.

Figure 7 shows the timeline of major European AI initiatives since 2018, illustrating both the scope and evolution of European strategy.

[94] AutoGPT GitHub repository. Adept ACT-1. See https://github.com/Torantulino/Auto-GPT; https://www.adept.ai/blog/act-1 (both accessed in April 2025).
[95] For more information on protocols, including the A2A, see section 4.1.2.2.

Figure 7: European AI Initiatives Timeline



Source: *https://eurohpc-ju.europa.eu/about/discover-eurohpc-ju_en; https://aleph-alpha.com/; https://gaia-x.eu/; https://www.euronews.com/business/2024/04/29/who-in-europe-is-investing-the-most-in-artificial-intelligence; European Parliament (2024); https://digital-strategy.ec.europa.eu/en/policies/european-ai-research; https://docs.italia.it/italia/mid/programma-strategico-nazionale-per-intelligenza-artificiale-en-docs/en/bozza/the-strategic-programme-on-artificial-intelligence-anchoring-principles-and-goals.html; https://stability.ai/news/stable-diffusion-public-release; https://mistral.ai/about; https://www.reuters.com/world/europe/uae-plan-40-billion-investments-italy-pm-meloni-says-2025-02-24/; https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai; Department for Science, Innovation, & Technology (2025); https://www.investmentweek.co.uk/news/4395743/uk-attracts-gbp14bn-private-investment-ai-days; European Commission (2025c); https://www.euronews.com/next/2025/02/11/heres-what-has-been-announced-at-the-ai-action-summit; https://digital-strategy.ec.europa.eu/en/news/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence; European Commission (2025a) (all accessed in April 2025); https://www.bloomberg.com/news/articles/2025-05-19/nvidia-uae-s-mgx-pledge-to-build-europe-s-largest-data-center (accessed in May 2025)*

2.2.1.1 Expanding AI infrastructure capacity

Key European infrastructure initiatives include:

- *InvestAI* mobilises €200 billion across the European innovation ecosystem, with €20 billion earmarked for the development of AI gigafactories – large-scale computing and model training facilities.[96] This represents Europe's largest coordinated investment in AI capabilities.
- *European High Performance Computing Joint Undertaking (EuroHPC JU)* is a pan-European initiative expanding access to high-performance computing facilities for AI R&D. It consists of AI compute hubs such as the Jean Zay supercomputer and Germany's Jülich AI centre.[97] With €3 billion from the current 2021–2027 EU budget, including €1.9 billion for infrastructure acquisition, deployment, upgrade, and operation, EuroHPC aims to democratise access to computing resources for European researchers and companies.[98]
- *GAIA-X* is a European public-private initiative aimed at creating a framework for secure, transparent data exchange infrastructure. Rather than competing directly with hyperscale cloud providers, GAIA-X focuses on interoperability and data governance.[99]

Despite these investments, implementation faces obstacles. Notably, the European Court of Auditors, in its Special Report on the EU's AI ambitions published in May 2024, highlighted significant shortcomings

96 See https://ec.europa.eu/commission/presscorner/detail/en/ip_25_467 (accessed in April 2025).
97 See http://www.idris.fr/eng/jean-zay/jean-zay-presentation-eng.html; https://www.fz-juelich.de/en/ias/jsc
98 See https://eurohpc-ju.europa.eu/about/discover-eurohpc-ju_en (accessed in April 2025).
99 See https://gaia-x.eu/about/ (accessed in April 2025).

in the EU's innovation agenda to date. More specifically, it points to poor coordination between the Commission and Member States and recommends reassessing investment targets, improving governance, and boosting support for AI innovation in Europe.[100]

Similarly, the Draghi report on European competitiveness published in September 2024 offers a mixed assessment. While praising EuroHPC as "*one-of-a-kind globally*", with three of its supercomputers ranking among the world's top ten, it calls for expanding HPC's computing capacity, developing a "*federated AI model*" based on public-private collaboration, and better support mechanisms for innovative SMEs.[101]

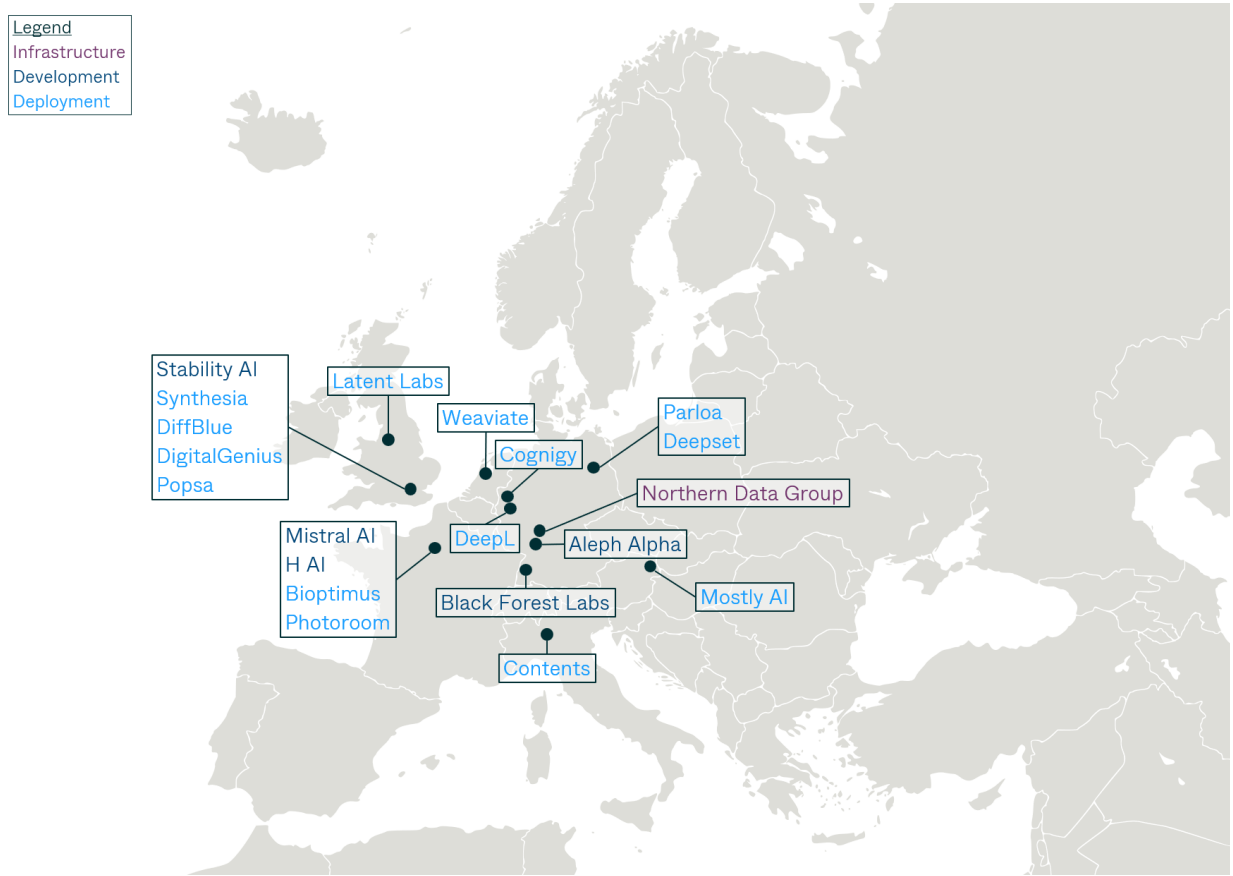
These coordination challenges reflect broader European governance complexities but also highlight the strategic importance policymakers place on building European AI capabilities.

2.2.1.2 Fostering model innovation

European companies have established meaningful positions in foundation model development, often focusing on areas where regulatory compliance and multilingual capabilities create competitive advantages. Mistral AI (France) has gained recognition with models like Mistral 7B and Mixtral that compete effectively in their parameter classes, while companies like Aleph Alpha (Germany) have built enterprise-focused AI solutions.[102] These models claim to focus on (amongst other things) multilingualism, transparency, and energy efficiency.

Figure 8 below presents the 20 highest-funded European companies across infrastructure, development, and deployment layers, illustrating the breadth of European participation in the value chain.

Figure 8: Top 20 European GenAI companies – By private funding



Source: RBB Economics based on Crunchbase (accessed in May 2025). Background map by Wikimedia Commons (accessed in May 2025).

[100] See pages 48–49 in European Court of Auditors (2024).
[101] See Draghi (2024).
[102] See https://mistral.ai/; and https://aleph-alpha.com/ (both accessed in April 2025).

Note: The map includes the 20 highest-funded Generative AI firms headquartered in Europe, according to Crunchbase industry classifications and private funding in USD. Europe here includes the continent. Each firm on the map has a funding of at least $25 million. Classification into Infrastructure, Development, and Deployment by RBB Economics based on our best understanding.

European developers have contributed to foundational research initiatives. LightOn (France) has been active in LLM development since 2020, whilst the BLOOM model, coordinated by the French research initiative BigScience, exemplifies large-scale, multilingual, open-access development capabilities.[103]

Beyond language models, Europe has also seen progress in image and video generation:

- Stability AI, founded in the UK, has established itself in open-source image generation with Stable Diffusion, providing an alternative to closed US systems like DALL-E and Midjourney. The company raised $101 million and reached a $1 billion valuation in October 2022.[104]
- Black Forest Labs, a German GenAI startup founded by former Stability AI researchers, emerged in 2024 with its advanced Flux family of image generation models. The company raised $31 million in seed funding and has established partnerships with companies like Mistral AI, Nvidia and Deutsche Telekom, positioning itself as bringing *"state-of-the-art AI from Europe to the world"* through both open-source and proprietary technologies.[105]
- Synthesia, based in the UK and valued at $2.1 billion, targets enterprise video generation for training and marketing, reaching over 60,000 customers, including 60% of Fortune 100 companies. The company became the first AI video platform to achieve ISO 42001 certification for security compliance in GenAI, setting industry standards for responsible AI deployment.[106]
- Photoroom (France) has built a profitable business model, reaching $65 million in annual recurring revenue (ARR). Unlike many AI startups still seeking sustainable business models, Photoroom broke even just a year after launch by focusing on specific e-commerce imagery needs, particularly background removal and replacement. The company managed to reach $20 million ARR with just $2 million in funding.[107]

Meanwhile, European model developers often integrate privacy and compliance features from the outset, reflecting familiarity with frameworks like GDPR. This gives them potential time-to-market advantages over non-EU developers who must retrofit compliance into existing models. This advantage may help to counterbalance structural disadvantages arising from Europe's smaller scale. Companies like Syntho (Netherlands) and Mostly AI (Austria) develop synthetic data solutions that enable model training without using real personal data, competing against market leaders like Gretel AI (acquired by NVIDIA in March 2025) and Synthesis AI.[108]

EU-funded initiatives such as AI4Europe and the European Network of AI Excellence Centres aim to promote best practices in reproducibility, ethical alignment, and open science in model development, reinforcing Europe's emphasis on responsible AI deployment.[109]

---

103 See https://www.lighton.ai/about-us; https://bigscience.huggingface.co/blog/bloom (both accessed in May 2025).
104 See https://stability.ai/; https://stability.ai/news/stability-ai-announces-101-million-in-funding-for-open-source-artificial-intelligence; https://www.bloomberg.com/news/articles/2022-10-17/digital-media-firm-stability-ai-raises-funds-at-1-billion-value (all accessed in May 2025).
105 See https://blackforestlabs.ai/; https://techcrunch.com/2024/08/14/meet-black-forest-labs-the-startup-powering-elon-musks-unhinged-ai-image-generator/; https://sifted.eu/articles/black-forest-labs; https://bfl.ai/announcements/25-01-03-nvidia (all accessed in May 2025).
106 See https://www.synthesia.io/; https://www.synthesia.io/post/synthesia-world-first-iso-42001-compliant-ai-video-company; https://www.eu-startups.com/2025/01/synthesia-claims-the-crown-as-the-uks-most-valuable-generative-ai-media-company-with-e174-million-in-funding/ (all accessed in May 2025).
107 See https://www.photoroom.com/; https://sacra.com/research/photoroom-background-removal-app/; https://sifted.eu/articles/photoroom-interview-gen-ai-profit (all accessed in May 2025).
108 See https://techcrunch.com/2022/04/28/synthesis-ai-raises-17m-to-generate-synthetic-data-for-computer-vision/; https://techcrunch.com/2025/03/19/nvidia-reportedly-acquires-synthetic-data-startup-gretel/ (both accessed in May 2025).
109 See https://www.ai4europe.eu/ (accessed in May 2025).

2.2.2 Europe's structural challenges in the GenAI race

Despite active participation across the GenAI value chain, Europe faces specific structural obstacles that affect its global competitiveness in the industry. The AI Index Report 2025 documents a substantial funding disparity, with US private GenAI investment in 2024 exceeding European and Chinese investment combined by $25.4 billion (see Figure 9 below). This capital gap, combined with market fragmentation across multiple languages and regulatory systems, creates specific scaling challenges for European AI companies. In particular, three key challenges for Europe in the GenAI industry are:

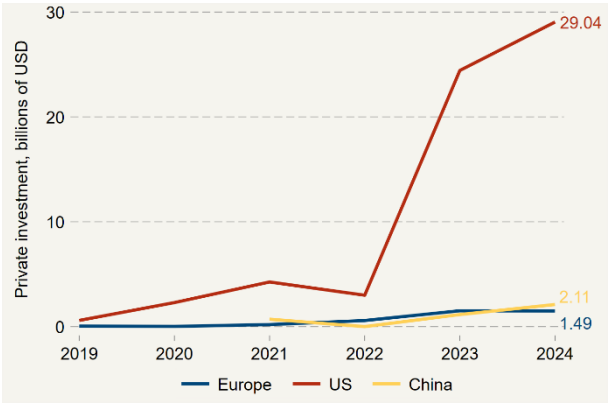
- funding and talent;
- a complex regulatory and cultural landscape; and
- GenAI adoption constraints in the broader economy.

2.2.2.1 Funding and talent

According to the AI Index Report 2025, the disparity in GenAI investment is significant. Figure 9 below compares private investment in GenAI in Europe, China, and the United States from 2019 to 2024. While all regions have increased spending in recent years, the US consistently outpaced Europe and China. In 2023, the US outpaced the combined GenAI investments of Europe and China by $21.8 billion. By 2024, US private GenAI investment exceeded Europe and China combined by over 8x.

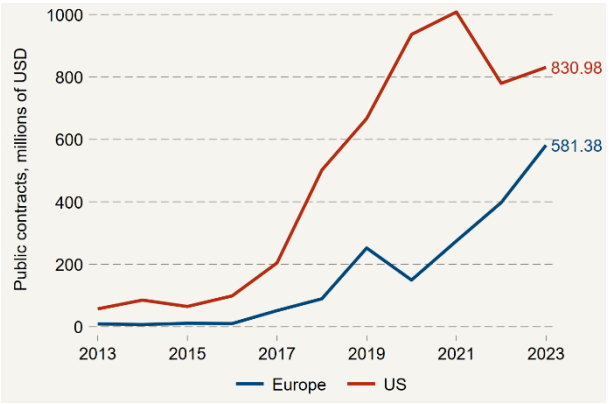
Public investment in the EU also falls behind other regions. As can be seen in Figure 10, the US saw sharp growth starting around 2017, peaking in 2021 with spending of nearly $1 billion. In contrast, Europe's public AI spending remained substantially lower despite growth in 2017 and 2019.

Figure 9: Private investment in GenAI in Europe, US, and China (billions of USD)



Source: *Figure 4.3.11 in the AI Index Report 2025.*
Note: *Europe includes both EU countries and the UK*

Figure 10: Public spending on AI-related contracts in Europe and US (millions of USD)



Source: *Figure 6.3.6 in the AI Index Report 2025.*
Note: *Europe includes both EU countries and the UK. US spending only includes grant-level spending, while European spending includes government contracts in EU member states and the UK.*

This underscores the persistent transatlantic gap in support for AI development, particularly with regards to private investments which was more than twenty times higher in the US than in Europe in 2024. Consistent with trends in private investment, the US led all regions in the number of new AI companies (not limited to GenAI) in 2024, with 1,073, compared to 436 in Europe and 98 in China.[110] Notably,

[110] See Figure 4.3.12 in the AI Index Report 2025. For the purposes of this analysis, Europe includes the UK (116 AI companies), Germany (67), France (59), the Netherlands (24), Switzerland (22), Spain (18), Sweden (16), Denmark (15), Italy (14), Finland (12), Austria (10), Norway (9), Belgium (9), Ireland (9), Lithuania (7), Portugal (5), Poland (5), Estonia (3), Luxembourg (3), Romania (2), Czechia (2), Bulgaria (2), Iceland (2), Cyprus (2), Croatia (1), Slovenia (1), and Hungary (1).

European startups frequently struggle to raise late-stage capital, limiting their ability to scale products across the Single Market.[111]

Prioritising AI expertise has been a recurring theme in recent statements by European policymakers.[112] Talent attraction and retention are ongoing issues: the US stands out as a primary destination for AI talent from Europe, attracted by the its greater funding opportunities, leading tech companies, renowned universities, and extensive research facilities.[113] Research shows that about 85% of all migrating startups move to the US, with European startups raising significantly less capital at home: US startups raised 3x more venture capital than European ones in 2023.[114]

Notable examples include French unicorns like Dataiku (now expanded into GenAI deployment) that have moved its headquarters to the US, often driven by investor requirements and better access to growth capital.[115] Hugging Face also exemplifies this pattern: founded by French entrepreneurs in New York in 2016, it is now a *"French-American company"* with significant operations in both countries but maintains US headquarters for funding and market access.[116]

However, there are emerging counter-trends: US-based Poolside AI relocated to Paris after raising $126M, citing *"high availability and low cost of talent"* and French government AI incentives.[117] This suggests potential for European talent retention with appropriate policy support.

2.2.2.2 Navigating a complex regulatory and linguistic landscape

The EU has been early to implement AI regulation, with the GDPR already providing standards for data privacy and governance. The AI Act, expected to be fully implemented in August 2026, will introduce a tiered framework classifying AI systems by risk and impose new compliance obligations.[118] These rules aim to enhance trust, transparency, and user safety, factors that can encourage adoption over time. Whilst these regulations aim to benefit market participants, the complexity and cost of compliance may significantly slow product development and restrict participation by smaller players, particularly in fast-evolving areas like generative AI. This, in turn, risks delaying users' uptake of innovative solutions across the economy.

More fundamentally, recent strategic reviews, including the Draghi report on European competitiveness, have identified overregulation, fragmented implementation, and a risk-averse innovation culture as core structural challenges for the EU's digital competitiveness. Five specific problems emerge:[119]

1. Compliance costs are likely to burden smaller firms. Startups and micro, small, and medium-sized enterprises (MSMEs) often lack the legal, technical, or financial resources to navigate complex and evolving requirements, potentially restricting their ability to compete in AI development and deployment.

2. High-risk sector restrictions. Strict requirements for healthcare, education, and employment applications may discourage firms from developing or deploying AI in these critical sectors due to potential liability or lengthy approval processes.

3. Legal uncertainty surrounding the practical implementation of rules. Unclear interpretation of provisions, including recent AI Act implementation delays may increase ambiguity for businesses, as it

111 Based on data from Crunchbase (accessed in May 2025), GenAI companies headquartered in the US have raised a total of 1371 angel, pre-seed, seed, or series A funding rounds and 157 rounds in series B and beyond. GenAI companies headquartered in the EU or the UK have raised 420 angel, pre-seed, seed, or series A rounds, but only 21 in series B and beyond.
112 See European Commission (2025b).
113 See Interface (July 2024), *Where is Europe's AI workforce coming from? Immigration, Emigration & Transborder Movement of AI talent*, page 10. See https://www.interface-eu.org/publications/where-is-europes-ai-workforce-coming-from (accessed in April 2025).
114 See Weik, Achleitner and Braun, R. (2024); https://sifted.eu/articles/11x-relocate-silicon-valley (accessed in May 2025).
115 See https://sifted.eu/articles/european-unicorns-relocating-us (accessed in May 2025).
116 See https://sifted.eu/articles/hugging-face-interview-julien-chaumond (accessed in May 2025).
117 See https://sifted.eu/articles/poolside-raises-126m-relocated-france-news; https://www.analyticsinsight.net/artificial-intelligence/ai-and-blockchain-startups-are-flocking-to-europe-heres-why (both accessed in May 2025).
118 See https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (accessed in April 2025).
119 See Draghi (2024).

remains unclear which rules will apply to the AI sector and when they will take effect.[120] This uncertainty may delay launches or deter cross-border experimentation, especially in fast-moving areas like FMs or autonomous agents.

4. National implementation and enforcement could vary significantly. Despite harmonisation goals, enforcement remains nationally driven. Different interpretations may emerge across jurisdictions, especially in the early years of enforcement and in complex cases like autonomous agents or multi-domain foundation models. Such divergences could fragment the internal market and create inconsistent legal exposure, creating barriers to scaling AI solutions across the EU.

5. Global regulatory asymmetry. European rules often exceed requirements in other major markets. Non-EU developers face additional compliance costs when entering the European market, while European firms must meet EU standards then adapt to more permissive frameworks elsewhere, adding costs without first-mover advantages in larger markets. This regulatory asymmetry may put European developers at a competitive disadvantage, especially when time-to-market is critical.

The European market is fragmented not only by regulatory differences but also by cultural diversity, slowing the scaling of AI products and services across borders. Unlike more unified markets such as the US or China, where a single regulatory framework, language, and large domestic customer base can support rapid national scaling, Europe's multiplicity of jurisdictions, legal regimes, and linguistic diversity can create frictions for startups and firms seeking to grow. As a result, developers based in Europe often face duplicated compliance efforts, complex software needs, and limited access to continent-wide data.[121] This can hinder the wider diffusion of AI capabilities and may restrict the growth of a unified, pan-European AI industry.[122]

2.2.2.3 GenAI adoption constraints in the broader economy

Europe's economic structure creates specific obstacles to GenAI deployment that may delay productivity benefits. Three factors stand out: workforce demographics, institutional complexity, and MSME resource constraints.

- **Demographic and institutional factors slow adoption.** Europe's aging population and strong labour protections require careful change management for GenAI implementation. Sectors with older workforces need tailored training approaches, while labour institutions may require engagement strategies that demonstrate AI's potential to enhance rather than replace human work.
- **Traditional sectors face multiple barriers.** Legacy systems, regulatory constraints, and limited digital infrastructure slow AI adoption in established European industries. A study by the European Economic and Social Committee (EESC) identifies key obstacles for MSMEs: insufficient awareness of AI benefits, skills gaps among staff, and inaccessible data.[123]
- **MSMEs lack deployment resources.** Small and medium enterprises form Europe's industrial backbone but often lack capital and expertise for advanced AI deployment. This creates a particular challenge since MSMEs represent the majority of European businesses yet struggle most with AI adoption barriers. The EESC report emphasises the need for targeted support systems, from finance and infrastructure to data availability and interoperability, to enable successful pan-European AI adoption.[124]

---

120 See https://sifted.eu/articles/eu-ai-act-pause-analysis (accessed in June 2025).

121 For example, software may need to account for different character sets, writing directions, and design conventions. For more on the software localisation process, see https://optimational.com/blog/software-localisation-process/; https://eutechloop.com/europes-ai-continent-action-plan-too-lean-on-the-simplification-and-data-part/ (both accessed in May 2025)

122 See Fondation Robert Schuman (28 November 2023), *Artificial intelligence: Europe needs to start dreaming again*, https://www.robert-schuman.eu/en/european-issues/728-artificial-intelligence-europe-needs-to-start-dreaming-again (accessed in April 2025).

123 See European Economic and Social Committee (2021).

124 See footnote 123 above.

Addressing these constraints requires comprehensive support strategies encompassing education, training, data access, and financial assistance to facilitate digital transformation across Europe's diverse economic base.

2.2.3 There are considerable opportunities in Europe for GenAI deployment

Deployment of GenAI is still at an early stage. While some early adopters within Europe are already using GenAI technology, there are considerable opportunities in Europe within the GenAI sector. European firms have a long-standing presence in key industries where their domain knowledge could be used to identify and develop sector-specific GenAI use cases, to the significant benefit of productivity in the region. Additionally, Europe's history of cross-border R&D collaboration may partially mitigate funding gaps in the region by enabling researchers to pool resources and knowledge across national boundaries.

2.2.3.1 Industrial expertise creates opportunities for application development

Europe's established presence in manufacturing, automotive engineering, precision machinery, and complex financial services provides opportunities to identify new GenAI use cases within these areas.[125] The gains from taking advantage of these opportunities are high: with the industrial sector representing over 20% of the EU's GDP in recent years, productivity gains from custom GenAI deployments in this sector alone could significantly impact economic growth.[126]

Some European companies are already taking advantage of the opportunities GenAI presents, using their domain-specific data and knowledge to develop GenAI solutions tailored for tasks such as automating industrial operations, ensuring regulatory compliance, and improving process control in sectors like manufacturing and energy.[127] For example, Siemens offers GenAI-powered solutions to enhance predictive maintenance, energy efficiency, and product lifecycle management for their customers.[128] Bosch has made AI a core competency across its operations, with AI now featuring in every product or involved in production.[129] The company is actively exploring GenAI applications, from improving automated driving functions through its Microsoft partnership to creating synthetic data for manufacturing optimisation.[130]

Healthcare and life sciences also offer competitive niches, supported by Europe's strong public research base. Europe's renowned academic and research institutions, such as the University of Oxford, ETH Zurich, and the Max Planck Society, provide a steady pipeline of innovation in healthcare-related GenAI applications.[131] From drug discovery to diagnostic imaging and clinical decision support, GenAI is being deployed to address pressing healthcare challenges.

For example, BioNTech, a German biotech firm, drew on machine learning models in its successful development of mRNA vaccines.[132] The French startup Owkin applies federated learning, a technique that allows GenAI models to be trained across multiple hospitals without transferring or pooling sensitive

---

125 See https://www.mckinsey.com/featured-insights/lifting-europes-ambition/leveraging-generative-ai-in-europe-the-opportunities-and-challenges (accessed in June 2025).

126 See https://ec.europa.eu/eurostat/statistics-explained/index.php?title=National_accounts_and_GDP#Gross_value_added_in_the_EU_analysed_by_economic_activity (accessed in June 2025).

127 See Economic Round Table (2025); https://zbrain.ai/generative-ai-for-regulatory-compliance/ (both accessed in June 2025).

128 See https://press.siemens.com/global/en/pressrelease/siemens-expands-industrial-copilot-new-generative-ai-powered-maintenance-offering (accessed in June 2025).

129 See https://www.bosch-ai.com/about-us/about-us/, https://us.bosch-press.com/pressportal/us/en/press-release-26240.html (both accessed in June 2025).

130 See https://www.bosch-ai.com; https://aibusiness.com/verticals/bosch-microsoft-to-use-generative-ai-to-make-roads-safer-bosch-connected-world-2024, https://us.bosch-press.com/pressportal/us/en/press-release-23488.html (all accessed in June 2025).

131 See http://www.xinhuanet.com/english/2018-11/07/c_137587058.htm; https://enspire.ox.ac.uk/event/artificial-intelligence-for-healthcare-challenge; https://ai.ethz.ch/research/core-areas/ai-medicine.html; https://maxplanckneuroscience.org/new-ai-in-medicine-initiative-launched/; https://www.mpi-cbg.de/news-outreach/news-media/article/new-research-division-to-combine-ai-and-biomedicine-in-dresden (all accessed in June 2025).

132 See https://investors.biontech.de/news-releases/news-release-details/biontech-acquire-instadeep-strengthen-pioneering-position-field (accessed in April 2025).

health data in a central location.[133] This adoption of GenAI in the European healthcare sector has been highlighted recently by the European Commission, who pointed to the use of GenAI in clinical trials and public health initiatives.[134]

Examples of other European firms successfully using their domain-specific knowledge to develop GenAI solutions to improve their own performance and productivity are discussed in further detail in section 4.3.

2.2.3.2 Cross-border R&D networks help to facilitate European innovation

Europe has an established framework for multinational research collaboration. This collaborative model helps European institutions efficiently pool resources and expertise across national boundaries, which may help to partially mitigate funding limitations faced by individual countries.

For example, flagship EU programmes, such as *Horizon Europe* (see Figure 7 above), actively promote multinational partnerships between research institutions, startups, corporations, and public bodies to tackle shared technological challenges, including those in GenAI. This programme could facilitate translation of GenAI research into commercial applications.

Alongside these EU-wide efforts, regional ecosystems like AI4IDF in the Île-de-France region exemplify how concentrated academic and industrial partnerships can create AI innovation hubs.[135] Bringing together four major French AI institutes, DATAIA, Hi! PARIS, PRAIRIE, and SCAI, the AI4IDF programme aims to offer a unified vision and interface to industrial and international partners, and leverage skills in mathematics, computer science, and applied research. Such initiatives not only help structure local innovation but also connect to pan-European networks, accelerating the path from research to commercial impact.

In the open-source ecosystem, European researchers and developers have played a significant role in shaping globally relevant tools.[136] For example, Hugging Face has emerged as a leading platform for open-source model development and distribution. European research institutions have also contributed to projects like PyTorch and TensorFlow, and multilingual datasets and benchmarks (e.g. XNLI, OSCAR, and Common Voice), reinforcing Europe's influence in community-driven AI innovation.[137]

Additionally, Europe is increasingly contributing to discussions around the technical rules of the AI industry. Bodies like CEN and CENELEC (the European standards organisations) and European representatives in ISO/IEC working groups are helping define global standards for AI interoperability, safety, and trustworthiness, areas where Europe's emphasis on accountability and user rights is having an impact.[138] The EU's leadership in content provenance and watermarking efforts (e.g. via the *Coalition for Content Provenance and Authenticity*, C2PA) is an example of its "soft power" in shaping AI norms.[139]

On the infrastructure side, initiatives like *GAIA-X* aim to build a federated and interoperable European cloud ecosystem, offering an alternative to hyperscalers.[140] GAIA-X promotes standards for data sharing and identity management across sectors, with applications in health, mobility, and manufacturing. Likewise, the *European Open Science Cloud (EOSC)* is designed to unify fragmented data infrastructures and make scientific data more accessible across national borders, a key enabler for AI research requiring

---

133 See https://www.owkin.com/ (accessed in April 2025).
134 See https://health.ec.europa.eu/ehealth-digital-health-and-care/artificial-intelligence-healthcare_en (accessed in April 2025).
135 See https://ai4idf.fr/aboutus (accessed in May 2025).
136 See EESC (January 2025), *Generative AI and Foundation Models in the EU: Uptake, Opportunities, Challenges, and a Way Forward*. See European Economic and Social Committee (2025).
137 "*When it comes to deep learning frameworks, PyTorch and TensorFlow are two of the most prominent tools in the field. Both have been widely adopted by researchers and developers alike, and while they share many similarities, they also have key differences that make them suitable for different use cases.*" See https://rafay.co/the-kubernetes-current/pytorch-vs-tensorflow-a-comprehensive-comparison/ (accessed in April 2025).
138 See https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/; https://www.unesco.org/en/articles/how-iso-and-iec-are-developing-international-standards-responsible-adoption-ai (both accessed in April 2025).
139 See https://c2pa.org/ (accessed in April 2025).
140 See section 2.2.1 above.

high-quality, labelled datasets.[141] These efforts help to support a more coordinated and resilient foundation for AI development and deployment across the continent in several key aspects.

Europe's tradition of cross-border R&D collaboration is a strategic asset in the global race to develop and deploy GenAI technologies. However, whilst the pooling of resources, talent and infrastructure across national boundaries can help to mitigate some of Europe's structural challenges, a clear role remains for policy-makers in considering ways in which they can further assist firms seeking to develop and deploy GenAI applications in overcoming these challenges: particularly given the potential benefits that widespread GenAI adoption can offer for productivity and growth in the region. These considerations are discussed further in our policy recommendations.

[141] See https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc_en (accessed in April 2025).

3 Economic framework

This section provides an economic framework to guide our assessment of GenAI deployments and partnerships in sections 4 and 5. We use this framework to draw conclusions from our observations on these early deployments and partnerships, and to inform our policy recommendations arising from this study (which we set out in section 6).

This section is structured as followed:

- First, we provide a brief overview of recent exploratory investigations into GenAI markets by competition authorities. We describe the potential theories of harm that have been articulated in these investigations, noting that no authority has made any adverse findings but have stated they are monitoring developments with future-looking concerns in mind.
- Second, we introduce key economic concepts relating to concerns raised by competition authorities. These economic concepts provide a starting point to better understand the theory behind the concern as well as the economic theory behind when concern is mitigated. These considerations can contribute to the assessment of current developments in GenAI and their implications on competition. Specifically, we focus on:
  - tipping, and in particular the role of switching costs as well as the extent to which users single- or multi-home;
  - vertical integration and vertical agreements, including their potential efficiency benefits as well as associated competition concerns; and
  - complementary products and conglomerate firms, covering both their possible pro-competitive benefits and potentially restrictive effects.

3.1 Overview of recent investigations into GenAI markets

GenAI is a nascent market that is in very early phases of development and characterised by high levels of innovation, entry, and expansion, as can be seen from the overview of the global and European landscapes above. Despite this, competition authorities in Europe have been actively exploring the sector, noting the possibility for potential competition concerns to arise and pointing towards the availability of *ex ante* digital regulation to help address any such concerns. It is notable they have not found any evidence of harm or suggested any regulatory interventions to date. This section provides an overview of the potential competition concerns that have been raised, drawing on recent assessments by key authorities, including the UK's Competition & Markets Authority (CMA) and the European Commission's Directorate-General for Competition (DG COMP). This provides relevant context for the discussions that follow on GenAI deployment and partnerships.

Competition authorities across Europe and beyond have been looking at the GenAI value chain and exploring potential future theories of harm. Beginning with the CMA's initial report on FMs in 2023, several competition authorities, including the European Commission, but also national competition authorities such as the French Competition Authority and the Portuguese Competition Authority, as well as the OECD, have conducted investigations to identify potential issues before they materialise.[142] Most recently, in January 2025, the US Federal Trade Commission (FTC) published a dedicated report examining the competitive implications of GenAI partnerships.[143]

[142] See Competition & Markets Authority (2023a); European Commission (2024a); https://www.autoritedelaconcurrence.fr/en/press-release/generative-artificial-intelligence-autorite-issues-its-opinion-competitive; https://www.concorrencia.pt/en/articles/adc-warns-competition-risks-generative-artificial-intelligence-sector; OECD (2024) (all accessed in May 2025).

[143] See FTC (2025).

These efforts reflect growing regulatory interest in a market that remains nascent, rapidly evolving, and structurally fluid. Innovation is ongoing, and players of varying sizes are actively entering and shaping the GenAI value chain, from upstream infrastructure to downstream applications. As such, while authorities have identified several theories of harm associated with GenAI markets, these identified risks are forward-looking and precautionary.

The theories of harm that authorities have identified as potential future concerns include the following:[144]

- **Input foreclosure**. Authorities have noted the risk that dominant firms may restrict or degrade access to essential upstream inputs, such as computing capacity (cloud infrastructure), data, or skilled labour.[145] By denying access or worsening terms, vertically integrated firms may foreclose downstream rivals and shift demand to its own competing downstream subsidiary.
- **Conglomerate leveraging**. Firms with strong positions in adjacent markets may use their existing assets, such as productivity suites, search engines, or operating systems, to favour their own GenAI services and limit user access to alternatives. The concern is that entrenched ecosystem advantages could be extended into GenAI, raising entry barriers and distorting competition.
- **Self-preferencing**. Firms controlling platforms or interfaces may rank, display, or integrate their own GenAI services more favourably than rivals', even when acting as gatekeepers. While often overlapping with vertical or conglomerate effects, self-preferencing raises distinct concerns around user defaults, and biased ranking mechanisms, especially when platforms mediate consumer choice.
- **Margin squeeze**. Vertically integrated firms that both supply key infrastructure and develop GenAI models may adopt pricing strategies that disadvantage independent competitors in downstream markets.
- **Market "tipping" and ecosystem effects**. Due to network effects, data feedback loops, and economies of scale, GenAI markets risk "tipping" in favour of a few players. This is particularly relevant where GenAI functionalities are embedded into dominant platforms or productivity suites, reinforcing user lock-in and weakening contestability.
- **Strategic partnerships and control**. Authorities are scrutinising partnerships between GenAI developers and large digital firms, such as Microsoft/OpenAI or Amazon/Anthropic, where exclusivity or strategic influence could potentially distort competition. Such arrangements could theoretically grant effective control without triggering formal merger control.
- **Killer and reverse killer acquisitions**. There is concern that large incumbents may acquire innovative GenAI startups with the goal of neutralising potential future threats, rather than integrating complementary capabilities, thereby dampening long-term innovation and consumer choice.
- **Horizontal collusion and information sharing**. While not yet observed, authorities note the theoretical risk of coordination between GenAI developers, especially as the market becomes more concentrated. This could involve the exchange of commercially sensitive information or the softening of competitive rivalry.

These potential concerns can be grouped into three broad categories: input-related restrictions (input foreclosure, margin squeeze), platform-mediated advantages (conglomerate leveraging, self-preferencing, market tipping), and agreements between firms (strategic partnerships, acquisitions, horizontal coordination).

144 See Competition & Markets Authority (2024b); European Commission (2024a); and Curry & Hill (2024).
145 See Figure 3 above for an overview of the GenAI value chain.

Current regulatory activity remains exploratory, with a focus on monitoring and early engagement. No authority has found any evidence of harm to date but are actively tracking market developments to ensure they can take early action should any concerns materialise. The overarching goal is to preserve competitive conditions, by ensuring open access, user choice, and continued innovation as the GenAI sector scales, while recognising the high stakes and uncertainty involved in shaping policy for a technology still in formation.

3.2 Economic considerations relevant in addressing competitive dynamics in GenAI

Here, we introduce some economic concepts that we draw on when assessing competitive dynamics in GenAI deployments and partnerships, including tipping, vertical integration, and conglomerate firms.

3.2.1 Tipping, switching costs and multi-homing

Some competition authorities have raised concerns that GenAI-related markets could "tip" towards a single supplier, in the way a number of earlier digital markets have.

Tipping refers to a situation where one firm, often an early mover, gains a significant advantage over its rivals. Under certain circumstances, this advantage can become self-reinforcing, making it increasingly difficult for rivals to challenge the leading provider until the market "tips" to just a single (or few) supplier.[146]

Tipping tends to occur in the presence of strong network effects or scale effects.

Direct network effects arise when the value of the product or service increases as more people use it. For example, as more users join a social network, this makes the social network more attractive to prospective users. As a result, more users will choose to join, which will then make the social network even more attractive to prospective users, and so on.

Indirect networks effects, associated with multi-sided markets (markets that serve at least two distinct customer groups), occur when the value of a product on one side of the market increases as more firms/consumers use it on the other side of the market. For example, as more users join an online marketplace/intermediation platform, this encourages more providers to sell via the marketplace. This in turn leads to more users seeking out the platform, and so on.

If network effects are strong, this can lead to users being drawn to the providers with the highest number of users/suppliers. This creates a virtuous feedback loop between size and growth for the largest firm, until the market eventually "tips" in its favour: leaving other suppliers unable to effectively compete.

Scale effects have also been identified as potentially important in some digital markets. Scale effects, like network effects, can result in market tipping under certain conditions. For example, if the service requires a large amount of user-generated data to reach a sufficient quality to compete, the first firm to reach that scale might be able to offer the highest quality service. Users will potentially be drawn to that high quality offer, potentially implying that competing services cannot reach the minimum scale required to produce a service of equally high quality. Ultimately, scale effects may also lead the market to tip to one (or a small number of) suppliers.

Even where markets are characterised by strong network effects or scale effects, they may not always tip. Other market characteristics can significantly reduce the risk of tipping, including low switching costs and multi-homing behaviour. Markets with low switching costs and high levels of multi-homing are less prone to tipping and more likely to remain competitive over time:

[146] See p. 416 in Bishop and Walker (2010).

- **Switching costs** refer to the expenses customers may incur, or obstacles users have to overcome to change suppliers. High switching costs may lock customers, conferring a first mover advantage to existing providers, potentially allowing early advantages to perpetuate. Conversely, where switching costs are relatively low, customers can move between suppliers more easily, reducing dependency and encouraging continued competition. In such environments, suppliers must compete aggressively on price and quality to retain customers, promoting innovation and economic efficiency.[147]

- **Multi-homing** refers to users engaging with multiple products or services simultaneously, rather than committing to just one. When users split their activity across several providers, no single supplier gains enough exclusive usage to become the dominant market leader. Multi-homing can significantly weaken network effects since the benefits of additional usage are not concentrated with one provider. As customers or users are not locked with a single provider, this maintains competitive opportunities for rivals as well as lowers barriers for new entrants.[148] By contrast, if single-homing is the norm, tipping is more likely, especially in presence of network effects.

3.2.2 Economics of vertical integration and vertical agreements

A vertically integrated firm is one that controls multiple stages of the supply chain. Instead of relying on external suppliers or distributors, the vertically integrated company brings different parts of the value chain within its own organisational structure.[149] Vertical integration can offer significant benefits to the firm, but it may not always be practical or desirable for a firm to bring all activities fully in-house. In this case, a firm may seek instead to achieve some of the benefits of vertical integration through the use of vertical agreements, which can be used to coordinate activities between firms active in different layers of the value chain, without requiring them to be fully integrated.

Potential benefits of vertical integration

We note three main potential benefits of vertical integration:

- **Lower prices through elimination of double marginalisation**: When firms operate separately across the supply chain, both upstream and downstream firms add their own markup to costs independently. This can lead to higher final prices compared to vertical integration where both stages have common ownership and set a single price.[150] In addition, vertical integration can create more efficient outcomes by minimising transaction costs or aligning incentives between the upstream and the downstream firms.[151]

- **Better product quality and compatibility**: Vertical integration can lead to better interoperability between upstream and downstream products. Closer coordination across supply chains can help make sure parts fit together better, since an integrated firm has stronger incentives and greater control to ensure each part fits well with the next.

- **Enhanced innovation coordination**: Vertical integration can improve coordination across the value chain, increasing incentives to invest when significant innovation opportunities exist both upstream and downstream.[152] For example, upstream innovation by cloud providers, such as developing more efficient chips or optimising data centres, can significantly reduce the cost and speed of training large models. Downstream, model developers can tailor FMs to specific applications, like coding assistants or search tools. Vertical integration can better align the timing and direction of these investments, making the overall system more efficient and adaptable.

147 See for example Farrell and Shapiro (1986).
148 See Bedre-Defolie and Nitsche (2020).
149 See p. 189 in Bishop and Walker (2010).
150 See Spengler (1950).
151 See, for example, Grossman & Hart (1986); and Williamson (1981).
152 See, for example, p. 88–120 in Liu (2016).

Potential competitive concerns

However, vertical integration can also give rise to competitive concerns under some conditions. Theories of harm that are potentially relevant in a GenAI context include:

- **Input foreclosure**: A vertically integrated firm may restrict or degrade access to essential upstream inputs for downstream competitors.[153] In GenAI, this could involve a cloud provider limiting access to computing infrastructure, specialised chips, or development tools for rival model developers, or offering these inputs on less favourable terms compared to its own GenAI division.

  A vertically integrated firm may also set prices for upstream inputs at levels that squeeze the margins of downstream competitors while maintaining profitability through its own integrated operations.[154] This could occur when a cloud provider charges high prices for infrastructure services to independent GenAI developers while subsidising these costs for its own GenAI operations.

- **Customer foreclosure**: A vertically integrated firm may restrict access to important sales channels or customers, preventing upstream rivals from reaching end users.[155] For example, a firm controlling both FMs and popular applications (that purchase access to FMs) might favour its own FM within those applications, making it difficult for competing model providers to reach users.

- **Self-preferencing**: Within platforms or ecosystems, the integrated firm may systematically favour its own products over competitors' offerings through better placement, integration, or functionality.[156] In GenAI, this might involve prioritising proprietary models in search results or providing superior API access for internal products.

- **Information advantages**: Vertical integration may provide privileged access to commercially sensitive information about competitors' strategies, customer preferences, or technical developments, creating unfair competitive advantages in related markets.[157]

The assessment of whether vertical integration creates net benefits or competitive harms requires careful case-by-case analysis, weighing efficiency gains against potential foreclosure effects and considering market structure, entry barriers, and the availability of alternative suppliers or distribution channels.[158]

3.2.3 Complementary products and conglomerate firms

Many firms active in GenAI application markets are also present in adjacent markets, creating conglomerates that supply several complementary products. Competition authorities have raised concerns about bundling practices and strategies that involve linking or tying complementary products by firms that supply multiple related products used by the same customers.[159] However, as with vertically integrated firms, there are economic benefits associated with firms offering several adjacent products with complementary features.

Benefits of supplying complementary products

The benefits of supplying several related products under common ownership stem from better utilisation of the connections between products. In GenAI, such benefits arise when applications are used in combination with other products, such as search engines or productivity tools:

153 See Salop and Scheffman (1983); European Commission (2008a), paragraphs 29–46.
154 See Economides (1998).
155 See Hart and Tirole (1990), paragraphs 47–57.
156 See European Commission (2017).
157 See Rey and Tirole (2007).
158 See OECD (2007); European Commission (2008a), paragraphs 19–22.
159 See European Commission (2008b), paragraphs 93–100.

- **Lower costs due to economies of scope**: Supply-side efficiencies occur when a supplier can reduce costs by producing a range of products rather than just one.[160] These efficiencies arise because shared resources, such as technology, data, personnel, or distribution networks, can be used across multiple products, lowering overall costs. Additionally, supplying one product can reduce the cost or risk of supplying a complementary product, enabling conglomerate firms to launch new complementary offerings more easily and quickly by reusing existing capabilities.

- **Enhanced user experience through integration**: Some products exhibit demand-side synergies, such that customers derive greater benefit from consuming the products together than they would separately. In this setting, user experience can be enhanced when several connected products are offered together. For example, customers can derive benefit from not having to search for each component separately, making well-integrated bundles more convenient. Products specifically designed to work together can provide better complementary features, leading to smoother and more seamless integration.[161]

- **Cournot effects and pricing efficiencies**: When firms producing complementary products operate independently, they fail to internalise the positive effect of price reductions on demand for the other product. A merged or integrated firm can internalise these complementarities, potentially leading to lower overall prices for consumers, a phenomenon known as the Cournot effect.[162] This occurs because the integrated firm recognises that reducing the price of one component increases demand for both components.

Potential theories of harm from conglomerate strategies

However, conglomerate firms can also give rise to competitive concerns, including through so-called "ecosystem strategies" that extend beyond traditional bundling. Competition authorities have identified several theories of harm which are potentially relevant:

- **Traditional bundling and tying concerns**: Conglomerate firms may bundle complementary products in ways that foreclose competitors or create barriers to entry.[163] This can occur through pure bundling (products only available together), mixed bundling (discounted bundle plus individual sales), or tying arrangements that make one product conditional on purchasing another.

- **Ecosystem entrenchment**: Unlike traditional conglomerate theories focused on specific product relationships, ecosystem theories examine how firms can leverage a constellation of assets and capabilities across multiple markets.[164] A merger or expansion may allow a firm to entrench its position not just in individual markets, but across an entire ecosystem of interconnected products and services. This can raise entry barriers at the ecosystem level, even where individual markets remain competitive.[165]

- **Ecosystem-level competitive constraints**: Competition occurs not only between individual products but between entire ecosystems offering substitutable complementarities.[166] For example, different productivity ecosystems may compete by offering distinct combinations of applications that work together, even if individual components are not direct substitutes. A firm's acquisition of additional ecosystem components may remove important competitive constraints at this broader level.

---

160 See p. 467, 916 in Bishop, S. and Walker, M. (2010).
161 See Adams & Yellen (1976).
162 See Economides, N. (1996); p. 59–79 in Masson, R.T., Dalkir, S. and Eisenstadt, D. (2014).
163 See Adams & Yellen (1976); McAfee, McMillan & Whinston (1989).
164 See Batra, de Bijl and Klein (2024); and European Commission (2023), which was the first ecosystem theory of harm prohibition.
165 See Caffarra, Crawford and Ryan (2023).
166 See Batra, de Bijl & Klein (2024).

- **Cross-subsidisation**: Conglomerate firms may use profits from one market to subsidise competition in another, potentially engaging in predatory strategies that standalone competitors cannot match.[167] This can be particularly concerning when the subsidising market has high barriers to entry or strong network effects.
- **Data and capability aggregation**: Ecosystem strategies may allow firms to combine data, algorithms, and capabilities across multiple products in ways that create advantages difficult for rivals to replicate.[168] This can include combining datasets that individually might be substitutable but collectively provide unique insights, or leveraging capabilities developed for one product to gain advantages in adjacent markets.

The assessment of whether conglomerate strategies create net benefits or harms requires careful analysis of both traditional product-level effects and broader ecosystem dynamics.[169] This includes evaluating whether competitive constraints exist at the ecosystem level, the degree of complementarity between products, the presence of alternative suppliers or distribution channels, and the potential for innovation benefits to offset any foreclosure concerns. Competition authorities increasingly consider that digital markets may require expanded analytical frameworks that account for multi-product competitive dynamics while maintaining clear limiting principles to avoid overly broad structural presumptions.

[167] See European Commission on Non-Horizontal Mergers (2008), paragraphs 93–100.
[168] See Caffarra, Crawford & Ryan (2023).
[169] See European Commission (2008b), paragraphs 93–100.

4 A focus on GenAI deployment

The economic framework outlined above provides the lens through which we examine actual deployment practices. Our case studies reveal patterns of multi-homing and low switching costs that support the theoretical prediction of sustained competition rather than market tipping.

This section examines how firms approach GenAI deployment through practical case studies across different sectors. We observe considerable variety in deployment strategies, business models, and approaches to FM selection, suggesting few constraints on how firms can successfully deploy the technology.

The analysis is structured in three parts. First, we examine case studies of how firms select and integrate FMs, including evidence of multi-sourcing and tools that support model flexibility. Second, we analyse competition between different business models, from GenAI platform intermediaries to specialised providers competing with vertically integrated incumbents in productivity software. Third, we explore how European companies deploy GenAI across the automotive, pharmaceutical, and luxury goods sectors to enhance existing competitive strengths.

These deployment approaches suggest that firms have considerable options in structuring their FM relationships and competitive strategies, which may support ongoing competition between upstream providers and diverse business models at the deployment layer.

4.1 Case studies of GenAI deployment

This section examines how firms are using FMs to bring GenAI deployments to market, including how they approach FM selection and how they integrate their chosen models in their GenAI deployments. We present three case studies to illustrate different strategies: Estée Lauder uses multiple providers for different business functions, Goldman Sachs built an internal platform that works with multiple models, and SpringBok designed applications that can work with any foundation model provider.

The case studies provide evidence of multi-sourcing across GenAI deployments, with firms using different models for different applications or maintaining flexibility to switch between providers. We also examine market-based tools that reduce technical barriers to using multiple models, including unified access platforms, performance monitoring tools, and industry standardisation protocols that support model flexibility and interoperability.

These deployment approaches suggest that firms have a variety of options in how they structure their foundation model relationships, which may support ongoing competition between upstream model providers.

4.1.1 Early deployments show firms adopting diverse and flexible approaches to FM selection

The multi-sourcing strategies documented across our case studies suggest that GenAI markets are less likely to top, as predicted by the economic literature on competitive dynamics in markets with low switching costs and high multi-homing (section 3.2.1).

This section considers a selection of case studies that illustrate the way in which firms are currently using upstream foundation models to bring deployment scenarios to market. We find examples of different approaches, including, i) multi-model deployment, where applications employ different models for different use functions; ii) model-switching and combination, where applications can change their underlying models as needed and combine these models to improve efficiency; and iii) model-agnostic design, where

applications are designed to be compatible with any FM. While the case studies below do not directly illustrate this possibility, it can be noted that the three approaches above are not mutually exclusive and can be combined by firms when helpful.

Multi-provider strategies can involve considerations such as the following.

Benefits:

- **Functional optimisation**. Each business function or use case can adopt the FM best suited to its specific needs.
- **Risk distribution**. Dependencies are spread across multiple providers rather than concentrated with a single vendor. Reduced dependency on any single vendor provides resilience against technical failures, price changes, or strategic shifts. Issues with one provider only affect specific functions rather than all AI operations, which limits operational risks.
- **Negotiating leverage and competitive dynamics**. Evaluating providers for each new use case creates ongoing competitive pressure rather than defaulting to existing relationships. In particular, existing commercial relationships with multiple providers, already integrated into the workflows, create competitive tension and facilitate switching for each workload if needed.

Potential trade-offs:

- **Economies of scale**. Single-vendor partnerships can offer volume discounts, bundled pricing, and strategic partnership benefits like early access to features or custom model development.
- **Technical complexity and coordination requirements**. Managing multiple APIs, data formats, and security protocols across providers likely requires dedicated infrastructure and specialised expertise.
- **Infrastructure demands and integration overhead**. Creating unified data access for multiple AI providers across may require substantial infrastructure investment. Creating unified data governance and consistent user experiences across disparate systems also requires ongoing investment.

4.1.1.1 Estée Lauder: A multi-provider strategy across different functions

Estée Lauder Companies, a multinational cosmetics company managing brands such as Tom Ford Beauty and Clinique, has embraced GenAI deployment across multiple business functions through a deliberate multi-provider strategy, implementing over 240 custom applications across its operations.[170] This approach demonstrates how enterprises can use different FMs for different purposes, maintaining flexibility while optimising for specific use cases.

Strategic multi-model deployment across operations

The company has integrated three primary platforms, each selected for distinct capabilities.

- **OpenAI ChatGPT Enterprise** powers over 240 custom models for internal operations. When ChatGPT launched in late 2022, Estée Lauder solicited ideas from employees on potential use cases and received over 1,000 submissions, leading to the formation of a cross-functional GPT Lab in April 2024.[171] The model for clinical trials reportedly can determine *"the immediate moisturisation improvement percentage of Advanced Night Repair serum"* from thousands of reports within seconds, a task that previously required hours of manual research.[172] Similarly, the model for fragrance insights

[170] See https://www.ciodive.com/news/Estee-Lauder-ELC-generative-AI-OpenAI-partnership-ChatGPT/733005/ (accessed in April 2025).
[171] See https://www.voguebusiness.com/story/beauty/estee-lauder-companies-forms-ai-innovation-lab; https://luxeplace.com/how-is-estee-lauder-using-chatgpt-from-marketing-and-rd-to-consumer-insights/ (accessed in April 2025).
[172] See https://wwd.com/business-news/technology/estee-lauder-openai-gpt-ai-1236722324/ (accessed in April 2025).

analyses large consumer survey datasets to identify demographic preferences, while the model for copywriting creates on-brand campaign content across multiple platforms.[173]

- **Google Gemini and PaLM 2** enable the company's "Ella" (Estée Lauder Language Assistant) to provide language translation, ad copy generation, meeting summarisation, and document analysis across 40+ countries.[174] The system relies on Google's Vertex AI platform, with PaLM 2 specifically deployed for customer classification and data labelling functions. This enables real-time monitoring of consumer sentiment across social media platforms and call centre operations, allowing a proactive response to emerging concerns or trends.[175]
- **Adobe Firefly** was integrated in March 2025 for visual content creation, responding to industry projections that content demand will quintuple between 2024–2026.[176] The system enables automated resizing, reformatting, and variation generation for digital assets across the company's 1,700+ mobile and e-commerce sites. Marketing teams can create campaign assets that previously required extensive photo shoots and manual editing in a fraction of the time.[177] This capability may be particularly relevant for social commerce initiatives, where speed-to-market is particularly important.

This multi-provider approach reflects deliberate strategic choices. Despite the availability of competing capabilities from single providers, since OpenAI's DALL-E 3 and Sora compete with Adobe Firefly, while GPT-4 could handle translation tasks performed by Google Gemini, Estée Lauder actively chose to source from multiple providers. This contrasts with industry approaches like L'Oréal's deep partnership with IBM for sustainable formulation AI, or Shiseido's integrated VOYAGER platform.[178] Estée Lauder's strategy prioritises flexibility over the deeper integration these single-partner approaches might offer.

According to company statements, the implementation has delivered quantifiable results across multiple metrics. For instance:

- Response times were estimated to have improved by more than 90% across R&D and business teams, with tasks that previously took hours now completed in minutes.[179] The system processes complex queries like "*What are the latest trends for mascara use among Gen Z?*" in seconds, freeing marketing teams from time-consuming data synthesis to focus on strategic work.[180]
- MAC, one of Estée Lauder's brands, uses AI to predict which social media content will perform well before posting. According to their testing, content that the AI flags as likely to succeed achieved 33% higher engagement rates and 30% more video views compared to the brand's average posts.[181] This predictive capability helps creative teams make informed decisions during photo shoots and content creation, allowing them to prioritise elements the AI identifies as more likely to resonate with audiences.
- Supply chain forecasting accuracy improved by 30% after transitioning from spreadsheet-based systems, according to Estée Lauder's head of global supply chain technology, with the AI system optimising inventory levels across distribution centres and predicting demand patterns more accurately.[182]

---

173 See https://www.ciodive.com/news/Estee-Lauder-ELC-generative-AI-OpenAI-partnership-ChatGPT/733005/ (accessed in May 2025).
174 See https://cloud.google.com/transform/101-real-world-generative-ai-use-cases-from-industry-leaders (accessed in May 2025).
175 See https://www.prnewswire.com/news-releases/the-estee-lauder-companies-inc-and-google-cloud-partner-to-transform-the-online-consumer-experience-with-generative-ai-301912131.html (accessed in May 2025).
176 See https://news.adobe.com/news/2025/03/adobe-estee-lauder (accessed in April 2025).
177 See https://www.marketingdive.com/news/estee-lauder-generative-ai-adobe-integration/742129/ (accessed in May 2025).
178 See https://newsroom.ibm.com/2025-01-16-ibm-and-loreal-to-build-first-ai-model-to-advance-the-creation-of-sustainable-cosmetics; https://corp.shiseido.com/en/news/detail.html?n=00000000003893 (accessed in May 2025).
179 See https://www.ciodive.com/news/Estee-Lauder-ELC-generative-AI-OpenAI-partnership-ChatGPT/733005/ (accessed in May 2025).
180 See https://consumergoods.com/how-estee-lauder-cultivating-new-culture-measurement-ai (accessed in May 2025).
181 See https://www.cosmeticsdesign-europe.com/Article/2023/11/30/Estee-Lauder-is-the-poster-child-for-using-AI-to-step-up-its-social-marketing-strategy/ (accessed in May 2025).
182 See https://sloanreview.mit.edu/audio/the-beauty-of-ai-estee-lauders-sowmya-gottipati/ (accessed in May 2025).

The deployment builds on Estée Lauder's existing data infrastructure, leveraging 80 years of consumer data through its ConsumerIQ system, which consolidates information across 25 brands in 150 countries.[183] This combination of historical data assets and modern AI capabilities creates a distinct competitive advantage. As noted by Estée Lauder's executives, *"Beauty startups leap on the latest TikTok trend but they don't have 80 years of market knowledge like Estée does. … And now Estée has the technology to harness it."*[184] The company's decades of consumer data provide depth, understanding which products succeeded in different markets and demographics, while AI provides speed in identifying emerging trends. Without AI, this historical data would be too cumbersome to analyse quickly enough to respond to fast-moving beauty trends. Without the data, AI alone could identify trends but lack the context to develop appropriate products or predict which markets would be most receptive. The strength may lie in the combination: AI enables real-time analysis of historical data, while the data provides the foundation for AI to generate meaningful insights rather than superficial pattern matching.

Trade-offs and competitive implications

While Estée Lauder has made public statements about the benefits of its approach, potential complexities associated with multi-provider strategies or actual costs remain undisclosed. The multi-provider approach likely required significant infrastructure investment to create unified data governance across providers while maintaining security and compliance standards. For example, the creation of the GPT Lab suggests additional organisational overhead to manage provider selection and deployment, as well as to coordinate and prioritise implementation efforts.[185] Beyond technical complexity, Estée Lauder forgoes volume-based pricing advantages and the deeper strategic partnerships that typically come with single-vendor commitments. The company must also manage inconsistent APIs, varying data formats, and different performance characteristics across providers. However, Estée Lauder appears to have determined these costs are outweighed by the operational value of accessing best-in-class capabilities for each function and avoiding dependence on any single provider's technology roadmap or pricing decisions.

Implications for competitive dynamics

This case illustrates how deployment-layer decisions can influence FM markets. When major enterprises actively choose multi-homing strategies, despite added complexity, they help promote competitive dynamics among model providers and intermediaries.

For Estée Lauder, the multi-provider approach allows them to avoid being locked into any single provider's roadmap while benefiting from innovation across the entire AI ecosystem. Whether this complexity delivers net benefits compared to deep single-vendor partnerships will likely depend on factors specific to each organisation, including technical capabilities, use case diversity, and strategic priorities.

4.1.1.2 Goldman Sachs: An internal platform supporting multiple models

Goldman Sachs is a global leader in investment banking. It has partnered with tech companies to develop an in-house centralised GenAI platform ("GS AI Platform") that acts as the single point of entry for all GenAI activities within the company.[186] As of January 2025, the firm has deployed AI tools to approximately 10,000 employees, with plans to reach all knowledge workers by the end of the year.[187]

---

183 See https://news.microsoft.com/source/features/ai/estee-lauder-uses-ai-to-reimagine-trend-forecasting-and-consumer-marketing/ (accessed in May 2025).
184 See https://news.microsoft.com/source/features/digital-transformation/estee-lauder-uses-ai-to-reimagine-trend-forecasting-and-consumer-marketing/ (accessed in May 2025).
185 See https://www.ciodive.com/news/Estee-Lauder-ELC-generative-AI-OpenAI-partnership-ChatGPT/733005/ (accessed in May 2025).
186 See https://www.wsj.com/articles/goldman-sachs-deploys-its-first-generative-ai-tool-across-the-firm-cd94369b; https://eulerpool.com/en/news/ai/goldman-sachs-accelerates-with-centralized-ai (both accessed in April 2025).
187 See https://www.cnbc.com/2025/01/21/goldman-sachs-launches-ai-assistant.html (accessed in May 2025).

Platform architecture enabling model flexibility

The GS AI Platform evolved from Goldman's existing machine learning infrastructure into a centralised, API-first architecture.[188] This platform design fundamentally differs from deploying separate AI tools: it creates a unified environment where different models can be accessed through consistent interfaces.

The platform integrates multiple FMs from multiple providers:[189]

- **OpenAI**'s GPT-3.5 and GPT-4 variants for general-purpose tasks;
- **Google**'s Gemini models, including versions with 2 million token context windows for processing lengthy documents;
- **Meta**'s Llama open-source models for specific applications, including customer service and document review,[190] and
- **others under evaluation**, such as Anthropic, Mistral AI, and Cohere models.[191]

Goldman selects different models based on task requirements and considers factors like ease of modification and cost. Chief Information Officer Marco Argenti describes this as a "plug-and-play" approach to model integration, where the platform's architecture allows switching between models without modifying applications. Argenti explains that Goldman does not want to rely on just one vendor and is giving the firm *"the flexibility to use a model that may be better for coding, while a rival offering is stronger at reasoning"*.[192]

The platform's flexibility enables diverse use cases across Goldman's business units. In investment banking, AI now generates 95% of IPO prospectus content *"in minutes"* compared to two weeks for a six-person team previously, according to the Bank's CEO.[193] The engineering division, comprising approximately 12,000 developers (one-quarter of Goldman's workforce), uses GitHub Copilot and Gemini Code Assist. CIO reports that in some cases, developers have been able to write as much as 40% of their code automatically using generative AI.[194]

Goldman expects productivity improvements of 30–40% in areas like software development and knowledge extraction.[195] The firm tracks these benefits through multiple metrics, including usage frequency and code acceptance rates, rather than relying on single productivity measures.

Strategic rationale and hybrid AI orchestration

Goldman's platform strategy reflects deliberate choices about vendor relationships and technological flexibility. The centralised approach enables Goldman Sachs to access multiple foundation models through standardised interfaces whilst also customising them with internal data in a secure environment. This ensures compliance with existing data regulatory requirements: for instance, it allows Goldman Sachs to embed control mechanisms to ensure that GenAI models do not present sensitive data to employees without the appropriate access.[196]

---

188 See https://aiexpert.network/goldman-sachs-ai/ (accessed in May 2025).
189 See https://www.cnbc.com/2025/01/21/goldman-sachs-launches-ai-assistant.html (accessed in May 2025).
190 See https://www.bobsguide.com/metas-llama-ai-wins-big-with-goldman-sachs-att-nomura/ (accessed in May 2025).
191 See https://www.cnbc.com/2025/01/21/goldman-sachs-launches-ai-assistant.html (accessed in May 2025).
192 See https://finance.yahoo.com/news/why-goldman-sachs-cio-taking-171310565.html (accessed in May 2025).
193 See https://fortune.com/2025/01/17/goldman-sachs-ceo-david-solomon-ai-tasks-ipo-prospectus-s1-filing-sec/ (accessed in May 2025).
194 See https://finance.yahoo.com/news/why-goldman-sachs-cio-taking-171310565.html; https://fortune.com/2025/03/19/goldman-sachs-cio-ai/ (accessed in May 2025).
195 See https://finance.yahoo.com/news/why-goldman-sachs-cio-taking-171310565.html; https://www.americanbanker.com/news/ai-will-make-workers-become-superhuman-goldman-sachs-cio-marco-argenti (both accessed in May 2025).
196 See https://eulerpool.com/en/news/ai/goldman-sachs-accelerates-with-centralized-ai (accessed in May 2025).

Goldman appears to be implementing what Argenti describes as a *"hybrid AI"* model, where larger models act as *"the brain that interprets the prompt and what the user wants, or the orchestrator that kind of spells out tasks to a number of worker models specialized for a specific task"*.[197] In practice, this means when a user makes a request, a sophisticated model first analyses what's needed, then delegates specific sub-tasks to different specialised models – for example, using one model for financial analysis, another for document summarisation, and a third for code generation. This orchestration approach suggests the platform intelligently allocates tasks based on their specific requirements rather than fixed model assignments, with Goldman factoring in both model capabilities and cost considerations when routing work.[198]

Goldman's approach involves integrating AI with proprietary data through fine-tuning, which involves enhanced training of models with smaller, specialised datasets.[199] This process likely creates models specifically optimised for Goldman's workflows, terminology, and regulatory requirements. However, this extensive internal customisation creates switching costs: if Goldman wanted to change foundation model providers, they would need to repeat the fine-tuning process with their proprietary data on the new models, representing a significant technical and time investment that would be difficult and costly to replicate.

Trade-offs and competitive implications

The platform approach enables sophisticated model allocation whilst creating a different type of dependency – Goldman becomes reliant on its own platform architecture and internally customised model variants rather than being locked into any single external provider. Benefits include task optimisation without requiring users to understand model strengths, cost efficiency through intelligent routing, and reduced external vendor dependency. However, implementation requires building sophisticated orchestration infrastructure, developing internal multi-model expertise, and substantial investment in fine-tuning capabilities.

Goldman's hybrid AI approach represents a more complex form of multi-homing where competitive pressure operates at the platform design level rather than through individual procurement decisions. Rather than users directly choosing between competing models, competitive pressure now operates at two levels: Goldman must still choose which external foundation models to integrate into its platform, and the platform itself must intelligently route tasks to the best-performing model for each function. This means external model providers still compete for inclusion in Goldman's platform, but the competition is filtered through Goldman's orchestration system rather than direct user choice.

However, this approach creates switching costs of a different nature than traditional customer lock-in. Instead of being tied to external providers, Goldman faces significant sunk costs in its platform architecture, internal AI expertise, and fine-tuned model variants. If Goldman wanted to fundamentally change its approach, for example, switching to a different platform provider or rebuilding its orchestration system, it would lose substantial investments in custom infrastructure and internally trained models. Whether this strategy delivers net competitive benefits will depend on Goldman's ability to leverage its internal AI infrastructure for unique competitive advantages that simpler deployment approaches cannot match.

197 See https://www.goldmansachs.com/insights/articles/a-new-generation-of-ai-tools-and-models-is-emerging (accessed in May 2025).
198 See https://www.cnbc.com/2025/01/21/goldman-sachs-launches-ai-assistant.html; https://finance.yahoo.com/news/why-goldman-sachs-cio-taking-171310565.html (accessed in May 2025).
199 See https://www.goldmansachs.com/insights/articles/what-to-expect-from-ai-in-2025-hybrid-workers-robotics-expert-models (accessed in May 2025).

4.1.1.3 Springbok AI: Model-agnostic legal application design

UK-based SpringBok AI demonstrates how specialised European companies can compete in generative AI deployment through technical innovation and domain expertise. The company's March 2025 acquisition by law firm Cleary Gottlieb illustrates the strategic value of model-agnostic architecture in addressing vendor lock-in concerns that constrain legal AI adoption.[200]

Prompt architecting methodology as competitive differentiation enabling model flexibility

SpringLaw employs a model-agnostic architecture that integrates with multiple FM providers including OpenAI, Microsoft Azure AI Services, and emerging alternatives without requiring system rebuilds. The platform achieves this through an abstraction layer that standardises communication across different LLM APIs while maintaining consistent user experience. The platform's three-module structure reflects different legal use cases: a no-code environment for building custom tools (Wizards), document interaction capabilities (Chats), and enterprise-scale document processing (Tables).[201] Each module can utilise different models simultaneously, allowing firms to route routine tasks to cost-effective models while reserving premium capabilities for complex analysis.

SpringBok's key innovation lies in "prompt architecting", i.e. building complex workflows based on carefully designed instructions rather than modifying the underlying foundation models themselves.[202] Instead of treating prompts as simple questions, SpringBok creates sophisticated systems that automatically generate different prompts based on the specific legal task and context, then chain these prompts together to complete complex workflows.

For example, when analysing a contract, SpringBok might automatically generate a sequence of prompts: first asking the foundation model to identify the contract type and key parties, then prompting it to extract specific clauses relevant to that contract type, followed by prompts to check those clauses against standard legal requirements, and finally prompting for a summary of potential risks. Each prompt builds on the previous response, creating a comprehensive analysis through a series of targeted instructions rather than a single complex query.

This approach offers significant advantages over fine-tuning foundation models, which typically requires substantial computational resources, extensive datasets, and technical expertise that many law firms lack. By contrast, prompt engineering works with existing pre-trained models, using carefully crafted instructions to guide the model's behaviour without needing to retrain it.[203] This means SpringBok can achieve specialised legal results through intelligent instruction design rather than expensive model customisation, making advanced AI capabilities accessible to law firms without major technical infrastructure investments.

The practical effectiveness of this approach is demonstrated by SpringLaw Verify, which reportedly achieves 95% accuracy in detecting AI hallucinations through systematic prompt engineering.[204] The system provides line-by-line referencing to source documents and can automatically redraft problematic sections, addressing critical accuracy concerns in legal AI deployment.

Competitive implications of model-agnostic design

SpringBok's model-agnostic approach appears to address real market needs and preferences. Dentons reported 65% firm-wide SpringLaw adoption within six weeks of deployment, with 650+ attorneys

---

200 See https://www.clearygottlieb.com/news-and-insights/news-listing/cleary-gottlieb-acquires-springbok-ai (accessed in June 2025).
201 *"SpringLaw is LLM agnostic and that firms can decide what model they use".* See https://legaltechnology.com/2024/06/11/springbok-ai-announces-general-release-of-no-code-gen-ai-platform-springlaw/ (accessed in June 2025).
202 See https://techcrunch.com/2023/09/18/instead-of-fine-tuning-an-llm-as-a-first-approach-try-prompt-architecting-instead/ (accessed in June 2025).
203 See https://www.ibm.com/think/topics/prompt-engineering (accessed in June 2025).
204 See https://legaltechnology.com/2024/12/05/exclusive-springbok-ai-launches-hallucination-detection-layer-to-verify-genai-output/ (accessed in June 2025).

generating over 21,000 uses in the first month.[205] The platform processed over 7 million words per minute during this period.[206] Charles Russell Speechlys similarly described SpringLaw as their *"most-adopted legal technology"*.[207]

These adoption rates suggest that model flexibility may be valued by legal firms concerned about vendor lock-in and data sovereignty. The ability to deploy on-premises or switch between cloud providers addresses data confidentiality requirements specific to legal practice.

SpringBok's model-agnostic architecture was designed from inception to enable seamless switching between FMs without system rebuilds. This approach delivers several potential advantages: firms can route different tasks to different models based on performance and cost considerations, immediately adopt new models as they emerge, and maintain negotiating leverage with AI providers by avoiding vendor lock-in.

However, the model-agnostic approach involves trade-offs. The abstraction layer that enables flexibility may introduce latency compared to native integrations. As FMs develop increasingly sophisticated capabilities requiring model-specific implementation, maintaining full agnosticism may become more challenging or require sacrificing advanced features that competitors with deeper integrations might access.

From a competition perspective, demand-side innovations like SpringBok's prompt architecting methodology and model-agnostic design can reduce dependency on specific FM providers. By maintaining flexibility over model selection and developing sophisticated prompt engineering capabilities, legal firms can preserve competitive options among upstream suppliers.

The Cleary Gottlieb acquisition in March 2025 represents an unusual move – law firms rarely acquire technology companies outright.[208] This suggests SpringBok's technology and approach were considered sufficiently valuable to warrant bringing in-house rather than licensing externally. The acquisition may indicate that sophisticated legal AI applications require closer integration between legal expertise and technical development than third-party platforms can provide.

4.1.2 Emerging market-led tools reduce switching costs by supporting model flexibility and interoperability

The multi-sourcing strategies observed across the case studies above have created demand for tools that make it easier to work with multiple AI providers. Market-based solutions have emerged in response to this demand, specifically designed to reduce switching costs and enable flexible foundation model usage. These tools address a fundamental challenge: as deployers seek to optimise performance and costs across different use cases, they require technical infrastructure that supports multi-model strategies without requiring separate integration work for each provider. Evidence from deployment patterns suggests significant demand for such flexibility, with market responses including platforms that provide single interfaces for access to multiple foundation models, tools that monitor performance and costs across different providers, and industry standards that make switching between models technically simpler.

The competitive implications of these emerging solutions are substantial. By reducing technical switching costs and enabling easier comparison between model providers, these tools help maintain competitive

[205] Springbok AI LinkedIn: Why Springbok is the right partner for your legal projects. See https://www.linkedin.com/posts/springbok-ai_dentons-springbok-collaboration-activity-7132676616547131392-byUB (accessed in June 2025).
[206] See https://www.artificiallawyer.com/2023/10/02/joe-cohen-on-dentons-fleetai-it-gives-unique-perspectives/ (accessed in May 2025).
[207] See https://www.templebright.com/news/2025/04/temple-bright-advises-on-landmark-legal-sector-ai-deal-with-cleary-gottlieb/ (accessed in May 2025).
[208] See https://www.abajournal.com/news/article/in-rare-move-big-law-firm-acquires-ai-legal-tech-company (accessed in May 2025).

dynamics in foundation model markets. When deployers can efficiently evaluate solutions from different providers and switch between these, market concentration becomes less likely and incentives to innovate remain strong.

4.1.2.1 Unified access platforms facilitate multi-model deployment

Several platforms have emerged to aggregate model access and reduce the technical complexity of managing multiple foundation model providers. These solutions primarily serve software developers, data scientists, and technology teams at companies who want to experiment with or deploy multiple AI models without building separate integrations for each provider. These solutions address the operational overhead that would otherwise discourage multi-sourcing strategies, effectively lowering barriers to competitive model evaluation.

OpenRouter: Standardising model access through unified APIs

OpenRouter has established itself as a leading API gateway for accessing multiple generative AI models through a single interface. Founded in 2023, the platform now provides access to over 300 models from 20+ providers, including OpenAI, Anthropic, Google, Meta, and Mistral AI.[209]

OpenRouter solves practical problems that make using multiple AI providers difficult. Normally, each AI company (OpenAI, Anthropic, Google) has different ways of sending requests and different usage limits. OpenRouter makes them all work the same way, so switching between providers is as simple as changing one setting. If your preferred AI service goes down, it automatically tries another one. This removes the technical complexity that usually stops companies from using multiple providers.

OpenRouter's pricing structure maintains transparency across providers by charging a flat service fee upfront rather than marking up individual model costs. [210] This means users can compare the true costs of different providers without the platform's own pricing creating bias toward particular models. The platform can automatically select models based on the user priorities, for instance choosing the cheapest option available or the fastest-responding model, rather than requiring users to manually compare and select providers for each request.[211]

Early adoption indicators suggest strong market demand for unified access among developers and technology teams. Whilst specific usage statistics are not publicly disclosed, the platform's rapid expansion to 300+ models within two years of founding suggests both provider willingness to integrate and developer adoption of the aggregation approach, with developers using OpenRouter to prototype applications, compare model performance, and avoid lock-in when building AI-powered products.

LangChain: open-source orchestration framework for flexible architectures

LangChain has evolved into a foundational tool for complex AI workflows involving multiple models and systems. Launched in October 2022, the framework has achieved significant adoption with over 107,000 GitHub stars and more than 4,000 contributors as of early 2025.[212]

The framework makes it easy for developers to use different AI providers without having to learn each company's specific requirements. This approach allows complex workflows to combine different models for different steps, for example, using specialised models for reasoning, writing, and sensitive data processing within the same application. LangChain's memory management capabilities keep user interactions consistent when switching between different models mid-conversation. For instance, if a

209 See https://openrouter.ai/docs/overview/models (accessed in May 2025).
210 See https://openrouter.ai/models (accessed in May 2025).
211 See https://openrouter.ai/docs/faq (accessed in May 2025).
212 See https://buttondown.com/agent-k/archive/llm-daily-may-08-2025/; https://www.langchain.com/about (accessed in May 2025).

chatbot switches from one foundation model to another during a customer service conversation, because the conversation moves from checking sensitive data to reasoning, LangChain ensures the new model has access to the conversation history, so the user does not have to repeat themselves.

LangChain's commercial success validates the importance of interoperability tools. The company raised $25 million in Series A funding in February 2024 at a $200 million valuation, with LangSmith achieving over 250,000 user signups and more than 25,000 monthly active teams by February 2025.[213] Enterprise customers including Klarna, Snowflake, and Boston Consulting Group demonstrate adoption amongst organisations with sophisticated AI deployment requirements.[214]

The framework's impact on switching costs appears measurable based on developer adoption patterns. With over 132,000 LLM applications built using LangChain and 28 million monthly downloads, the platform has become integral to many production AI deployments.[215] This widespread use suggests that tools that make switching easier can successfully reduce vendor lock-in at the application level.

Analytics and monitoring tools enable data-driven provider selection

Companies need tools to measure and compare how well different AI providers perform, and how much they cost. Berlin-based Langfuse is one example. It acts like a dashboard that shows a user how fast, accurate, and expensive different AI models are when used for the same tasks. This lets companies make decisions based on actual data rather than guesswork.

Other companies offer similar services: Helicone focuses on making AI requests faster and cheaper; Weights & Biases tracks experiments across many different machine learning projects; and Braintrust specialises in testing AI systems for large companies.[216] All these platforms solve the same basic problem: helping companies compare AI providers fairly so they can choose the best one for their needs.

Whilst specific adoption statistics for individual platforms are not always publicly available, the broader category of AI observability tools has attracted significant investment. The proliferation of competing platforms suggests strong market demand for solutions that enable transparent comparison between AI providers, creating competitive pressure by making provider performance measurable and comparable.

These monitoring tools create competitive pressure by making provider performance transparent. When organisations can quantify differences in accuracy, latency, and cost across models for specific use cases, providers must compete more directly on measurable performance rather than through integration friction or switching costs.

Prompt engineering and management tools address model-specific optimisation

As organisations deploy multiple models, they face the challenge that prompts optimised for one model often perform poorly on others. This technical barrier could create switching costs if not addressed by dedicated tools.

Promptfoo, an open-source testing tool, provides automated evaluation across multiple models to identify which prompts require modification when switching providers.[217] The tool helps organisations understand prompt portability and maintains performance consistency across different models. The tool functions like spell-check, but for ensuring instructions work properly with different AI models. PromptLayer takes a

---

213 See https://github.com/langchain-ai/langchain; https://www.langchain.com/about; https://www.businessinsider.com/sequoia-leads-funding-round-generative-artificial-intelligence-startup-langchain-2023-4; https://research.contrary.com/company/langchain (all accessed in May 2025).

214 See https://research.contrary.com/company/langchain (accessed in May 2025).

215 See https://research.contrary.com/company/langchain (accessed in May 2025).

216 See https://github.com/promptfoo/promptfoo; https://neptune.ai/blog/llm-observability; https://dev.to/lina_lam_9ee459f98b67e9d5/the-complete-guide-to-llm-observability-platforms-in-2025-488n (both accessed in May 2025).

217 See https://github.com/promptfoo/promptfoo (accessed in May 2025).

different approach by keeping track of which prompts work best with which models over time, similar to a notebook that records successful approaches.[218]

These tools make switching between AI providers more predictable by ensuring quality remains consistent. However, they only help with prompts, or the instructions given to AI models. A bigger challenge remains with fine-tuning, where companies train models with their own data to make them work better for specific tasks. Once a model has been fine-tuned with one provider, that customised model only works with that provider. This creates a meaningful switching cost that current tools don't address, though it affects mainly organisations with heavily customised AI applications rather than those using standard prompting approaches.

4.1.2.2 Industry protocols for standardisation and interoperability

Beyond individual tools, two major protocol initiatives have emerged to create industry standards for AI interoperability – essentially common technical "languages" that allow different AI systems and data sources to work together seamlessly: Anthropic's Model Context Protocol (MCP) and Google's Agent 2 Agent (A2A) Protocol.[219]

These protocols establish universal connection methods so that any AI application can access any compatible data source (MCP) or collaborate with other AI agents (A2A) without requiring custom programming for each combination. Whilst positioned as complementary technologies, both protocols aim to reduce technical barriers that favour incumbent platforms and enable smaller providers to compete more effectively.

Model Context Protocol establishes a standard for data integration

Anthropic's MCP, launched in November 2024, functions like a universal adaptor for AI systems. Just as a USB-C port connects any USB-C device to any compatible port without worrying about specific cables or connectors, MCP allows any AI application to connect to any data source without requiring custom programming for each combination. This addresses what the company calls the "M×N problem": if you have 5 AI applications and 10 data sources, you'd need 50 different custom connections. The technical foundation uses standard web communication, with programming libraries available for all major development languages, enabling broad developer adoption.[220]

MCP adoption has accelerated rapidly across major AI providers. OpenAI announced comprehensive MCP support in March 2025, integrating the protocol across ChatGPT desktop, its software development tools, and programming interfaces.[221] Microsoft integrated MCP into Windows 11 as *"a foundational layer for secure, interoperable agentic computing"*, with support across its business software platforms.[222] Google DeepMind has similarly confirmed MCP support in Gemini models.

The protocol's ecosystem growth demonstrates market demand for standardisation. According to public directories, over 5,700 active MCP servers were deployed by May 2025, growing from zero at launch six months prior.[223] Anthropic maintains pre-built MCP servers for common business systems including cloud storage, workplace communication tools, code repositories, and databases, whilst community contributors have developed hundreds of additional integrations.[224] The MCP could be particularly valuable for

---

218 See https://promptlayer.com/ (accessed in May 2025).
219 See https://modelcontextprotocol.io/introduction; https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/ (both accessed in May 2025).
220 See https://modelcontextprotocol.io/introduction (accessed in May 2025).
221 See https://techcrunch.com/2025/03/26/openai-adopts-rival-anthropics-standard-for-connecting-ai-models-to-data/ (accessed in June 2025).
222 See https://blogs.windows.com/windowsexperience/2025/05/19/securing-the-model-context-protocol-building-a-safer-agentic-future-on-windows/ (accessed in May 2025).
223 See https://www.anthropic.com/news/model-context-protocol; https://glama.ai/mcp/servers (accessed in May 2025).
224 See https://github.com/modelcontextprotocol/servers (accessed in June 2025).

companies who want to remain flexible in their GenAI strategy, reducing the cost of adopting newer or better models as they become available.

Google's A2A Protocol Targets Multi-Agent Coordination

Complementing the MCP, Google introduced the A2A in April 2025, developed with over 50 stakeholders including Atlassian, Salesforce, and SAP.[225] The multi-stakeholder development approach helps reduce the risk of the protocol favouring Google's ecosystem, as input from major enterprise players and competitors provides checks against unilateral design decisions that could create dependencies benefiting the originator. Where MCP connects individual agents to tools and data, A2A enables agents from different vendors to discover, communicate, and collaborate on complex workflows.

To use another analogy, if MCP is like giving each AI assistant a universal tool belt that works with any equipment, A2A is like giving them a common language so they can work together as a team. Different AI specialists can coordinate their efforts: one might handle research, another writing, and a third fact-checking, all whilst understanding each other's capabilities and sharing work efficiently.

The protocol's development involved major enterprise software providers including Atlassian, Salesforce, SAP, ServiceNow, and Workday, along with AI companies like Cohere and LangChain. This consortium approach reduces the risk of single-vendor control whilst ensuring enterprise requirements are addressed from protocol design through implementation. A2A's foundation uses standard web communication for agent interactions, but introduces agent-specific concepts like Agent Cards (similar to business cards that tell other agents what each one can do), Tasks (keeping track of work in progress), and Artifacts (completed work output that can be shared).[226] The protocol emphasises enterprise-grade security with standard authentication and access control methods. Currently at version 0.2, A2A targets a production-ready release in 2025.[227] Microsoft has announced support across its cloud AI platforms, potentially reaching hundreds of thousands of organisations through these systems.[228]

These protocols help standardise how models access data (MCP) and agents collaborate (A2A), reducing barriers to entry and supporting long-term competition. Economic literature acknowledges that standardisation involves trade-offs between innovation diversity and coordination benefits, but finds that in network industries, these protocols typically create net competitive benefits and reduce technological barriers that typically favour incumbents.[229] Studies in leading economics journals show that standardisation reduces the high integration costs and technological barriers that prevent smaller firms from entering markets dominated by large incumbents.

4.1.2.3 Market impact and competitive dynamics

These market-led solutions create several competitive effects that sustain diversity in foundation model markets by preventing vendor lock-in and maintaining genuine choice for deployers. This represents a significant market development: innovative solutions have emerged, gained adoption, and actively contribute to keeping the field competitive rather than allowing it to tip towards a few dominant providers.

Unified access platforms like OpenRouter eliminate the technical complexity of working with different AI providers and automatically switch to backup providers if one fails. Performance monitoring tools from companies like Langfuse, Helicone, and Braintrust give organisations clear data on how well different foundation models perform for their specific tasks, allowing them to choose based on actual results rather

225 See https://github.com/google/A2A (accessed in May 2025).
226 See footnote 225 above.
227 See https://medium.com/%40antoniodinoto_10306/bridging-the-gap-understanding-googles-agent2agent-a2a-protocol-1ee3062ec1b8; https://github.com/google-a2a/A2A/releases (both accessed in June 2025).
228 Based on PromptLayer platform capabilities. See https://www.promptlayer.com/; https://blog.promptlayer.com/braintrust-vs-promptlayer/ (both accessed in May 2025).
229 See Matutes & Regibeau (1996); Farrell & Saloner (1985); McAfee, Mialon & Williams (2003); Pehrsson (2009).

than which provider is easiest to work with technically. Tools for managing prompts help optimise performance across different models, though these currently work better for some applications than others.

The standardisation protocols (MCP for data integration and A2A for agent collaboration) reduce barriers to entry and support sustained competition. Academic research confirms that in network industries, these protocols typically create net competitive benefits by reducing the technological barriers that favour incumbents.[230] Studies show that standardisation reduces the high integration costs that prevent smaller firms from entering markets dominated by large incumbents. This proves particularly important for smaller model providers who can focus on specific capabilities without building wide-ranging systems, as standardised protocols eliminate costly custom integrations with each platform.

However, one significant gap remains: organisations that invest heavily in customising foundation models with their own data face meaningful switching costs. When companies fine-tune models using proprietary information, these customisations cannot be transferred to different providers, potentially limiting competitive pressure for enterprises with heavily tailored GenAI deployments. Yet this constraint may prove temporary as the technology evolves. Advanced techniques for crafting sophisticated instructions (rather than modifying the models themselves) could achieve equivalent customisation while keeping that expertise portable across providers. Fine-tuning platforms that standardise the customisation process across different base models could similarly reduce vendor lock-in, whether through low-code interfaces or industry-standard formats for exporting and importing fine-tuned capabilities. The emergence of modular AI architectures, where specialised capabilities are composed rather than baked into monolithic models, would further enhance switching flexibility. These technological shifts suggest that current switching costs from fine-tuning may diminish significantly, potentially restoring competitive pressure even in heavily customised enterprise deployments.

The sustainability of these competitive benefits depends on continued innovation in interoperability tools and the broad adoption of open standards. If switching facilitation tools becomes concentrated or proprietary standards emerge that recreate lock-in at higher levels, the competitive benefits may prove temporary. Early deployment patterns indicate organisations value flexibility enough to invest in additional technical complexity, but this preference may evolve as AI applications mature if performance optimisation becomes more critical than vendor optionality.

These deployment cases demonstrate that firms can avoid vendor lock-in through multi-sourcing strategies and emerging interoperability tools. This pattern of competitive flexibility reduces the likelihood of market tipping and supports continued competition between foundation model providers.

4.2 Vertically integrated firms compete with non-integrated firms

In this section, we examine how companies with different business models compete against one another in the same GenAI markets. These case studies show that some companies are vertically integrated and operate across several steps in the value chain, while others adopt a more focused approach, specialising in deployment and relying on third-party services for other parts of the value chain.

One of these markets, including both vertically integrated players and pure players, is the market for GenAI platforms, which are intermediary firms aiming to connect different layers of the vertically fragmented value chain and assisting third parties in deploying their own GenAI-powered applications. These intermediaries lower the technical requirements for GenAI deployment, therefore helping smaller innovators compete

[230] See Farrell & Saloner (1985); Katz & Shapiro (1985).

effectively with vertically integrated players, supporting competition between business models in GenAI deployment.

4.2.1 AI platforms and intermediaries

AI platforms act as intermediaries between FM providers and application developers, offering different approaches to reducing technical barriers and deployment complexity. These platforms compete by providing distinct value propositions: general-purpose end-to-end services, open-source community models, or specialised governance capabilities, in the aim of facilitating the deployment of GenAI solutions in various contexts where companies lack the in-house capabilities to build from scratch.

Three case studies illustrate how different platform strategies address varied market needs. Amazon Bedrock and Google Vertex AI offer integrated cloud-native solutions with extensive model catalogues and enterprise features. Hugging Face provides community-driven open access to models and development tools. Dataiku focuses on governance and regulatory compliance for European enterprises. Each approach enables firms to deploy GenAI without building complete technical capabilities in-house, though through different models of intermediation.

4.2.1.1 Setting the scene: why firms need to deploy GenAI

AI platforms act as intermediaries between foundation model providers and companies wanting to use AI in their products or for their internal processes. These platforms solve a practical problem: most companies lack the technical expertise and infrastructure to deploy sophisticated AI models on their own.

Companies face a "build versus buy" decision when deploying AI. Building custom solutions offers precise control but requires substantial technical expertise and computational resources. Buying off-the-shelf solutions provides immediate functionality but may not meet specific requirements. Platforms enable a "blend" approach: companies can access pre-trained models through platform APIs, add custom fine-tuning for specific domains, and integrate platform-provided tools for monitoring and scaling.

What AI platforms do

As intermediary market participants, GenAI platforms provide three core services:

- foundation model access through unified APIs, eliminating the need for developers to integrate with multiple model providers separately;
- computational resources for both inference and fine-tuning, allowing developers to customise models without maintaining their own infrastructure; and
- development tools, monitoring capabilities, and workflow management systems that reduce the technical barriers to AI application development.

Different platforms serve different competitive strategies

Platform variety creates competition at two levels. General-purpose platforms like Hugging Face (500,000+ models) and Google Vertex AI (multi-provider access) serve wide-ranging use cases with high flexibility. Others provide specialised services focused on particular use cases, like Roboflow for computer vision applications, or emphasise ease-of-use with more pre-built application templates and simplified interfaces, such as OpenAI's GPT Store or Microsoft Copilot Studio. Beyond pure platforms, model providers like Cohere and enterprise data science platforms like Dataiku, which has pivoted its offering to a governance-focused AI platform, offer alternative deployment paths.

Why this matters for competition

Platforms create competition at two levels.

- First, platforms compete with each other for customers.
- Second, by making AI accessible to more companies, platforms intensify competition in end-user applications. Platforms effectively democratise access to advanced AI capabilities, allowing smaller firms to compete with larger incumbents who might otherwise dominate through superior technical resources.

Platform concentration risks require monitoring

As platforms aggregate access to multiple models and development tools, they may themselves acquire market power. Platform concentration could theoretically recreate competitive bottlenecks, particularly if switching costs between platforms become high or if platforms begin favouring their own models over third-party offerings.

Current evidence suggests these potential concerns may be partially mitigated. Open-source tools like LangChain or OpenRouter, provide standardised interfaces that reduce platform lock-in by allowing developers to switch between providers with minimal code changes, as explained above in section 4.2.1 These tools explicitly promote "vendor optionality" and provide unified APIs across hundreds of models. However, lock-in risks vary significantly across platforms. Hugging Face offers complete model portability, allowing users to download and export fine-tuned models using standard formats.[231] In contrast, it may not be possible to export fine-tuned models from most proprietary platforms.

The key consideration for competition is whether customers have sufficient competitive alternatives and are not harmed by these different approaches to portability, and that competition between platforms remains sufficiently dynamic. Monitoring these dynamics will be important as platform adoption scales.

4.2.1.2 Amazon Bedrock and Google VertexAI: General-purpose cloud-native platforms enable competitive GenAI deployment

Amazon Bedrock and Google Vertex AI represent general-purpose cloud-native GenAI platforms that provide access to advanced capabilities through integrated infrastructure. Both platforms offer extensive model catalogues (Vertex AI includes over 200 foundation models while Amazon Bedrock provides access to over 100) combined with enterprise-grade infrastructure services.[232] These platforms support businesses throughout the application lifecycle, from initial model selection and customisation through application development, deployment, and ongoing maintenance as business requirements evolve. Their adoption patterns suggest how platform aggregation can expand competitive GenAI markets whilst creating both well-documented efficiencies and potential dependencies for enterprise customers.

Platform integration reduces deployment complexity

Both platforms integrate GenAI capabilities within their respective cloud ecosystems, reducing complexity for developers whilst creating operational efficiencies. Organisations using Bedrock can chain GenAI models with AWS Lambda functions or store fine-tuned models in S3 through standardised workflows.[233] Similarly, a company using Vertex AI can automatically feed data from BigQuery into an GenAI model for analysis, then store results back to Cloud Storage, all within a single integrated environment with consistent security policies and access controls.[234]

---

231 See https://huggingface.co/docs/transformers/main_classes/model; https://huggingface.co/docs/hub/en/models-uploading (both accessed in May 2025).
232 See https://cloud.google.com/vertex-ai; https://aws.amazon.com/bedrock/ (both accessed in June 2025).
233 See https://aws.amazon.com/bedrock/ (accessed in May 2025).
234 See https://futurumgroup.com/insights/google-clouds-vertex-ai-leap-into-enterprise-ai-adoption/ (accessed in May 2025).

For enterprise customers, these integrations reduce deployment barriers through reduced engineering overhead, consistent security policies, and simplified troubleshooting across integrated services, potentially accelerating development timelines and reducing operational costs.

However, organisations that become heavily reliant on platform-specific features may face higher switching costs when migrating GenAI applications to alternative platforms. AWS notes that *"switching costs have existed throughout the history of IT"* and frames this as natural technological evolution.[235] These dependencies represent the cost of taking advantage of platform capabilities and can vary significantly based on usage patterns: organisations using basic API access may migrate more easily than those relying on deep platform integration features.

The extent to which these costs meaningfully constrain competitive choice depends on how extensively organisations utilise platform-specific versus standard features, particularly around fine-tuning capabilities, where current technical limitations prevent easy transfer of customised models between platforms.

Multi-model aggregation enables supplier competition

Both Google's Vertex AI and Amazon Bedrock aggregate multiple GenAI providers, potentially creating competitive pressure among model suppliers while offering additional services through platform-specific features. Bedrock provides access to models from Anthropic, AI21 Labs, Cohere, Meta, Mistral AI, and Amazon's own Titan models through a unified API.[236] Vertex AI similarly offers Google's proprietary Gemini models alongside third-party options, including Anthropic's Claude suite and Mistral suite of foundation models.[237]

This aggregation approach allows customers to compare models and switch between providers with minimal changes to the deployment code. Both platforms emphasise model flexibility, with Amazon's unified API allowing developers to *"upgrade to the latest model versions with minimal code changes"* and Google's Model Optimizer providing dynamic routing between models based on cost and quality preferences.[238] Beyond simple model access, both platforms offer additional services that may influence vendor selection decisions. Amazon's Provisioned Throughput provides cost savings through commitment-based pricing, where customers agree upfront to use a certain amount of GenAI processing over a fixed period (such as 6 months or a year) in exchange for significantly lower rates compared to paying per individual request.[239] This works similarly to bulk purchasing arrangements, where larger commitments result in lower per-unit costs. Google takes a different approach with multi-model routing capabilities that provide automated cost optimisation.[240] When customers send requests to use GenAI, Google's system automatically selects which model to use based on requirements for speed, cost, or quality. For simple tasks, it might route to a faster, less expensive model, whilst complex requests use more capable models. Customers do not need to choose manually between different models: the platform optimises automatically and potentially reduces costs.

Enterprise adoption patterns suggest customers value integrated capabilities alongside model choice. Amazon reports that over 10,000 organisations use Bedrock across sectors including travel, creative software, and industrial equipment.[241] Similarly, Google Vertex AI serves notable enterprise customers including General Motors, Mercedes-Benz Group, Citigroup, HCA Healthcare, and Warner Bros

---

235 See https://aws.amazon.com/blogs/enterprise-strategy/switching-costs-and-lock-in/ (accessed in May 2025).
236 See https://aws.amazon.com/bedrock/ (accessed in May 2025).
237 See https://cloud.google.com/vertex-ai (accessed in May 2025).
238 See https://www.cloudforecast.io/blog/aws-bedrock-pricing/; https://aws.amazon.com/blogs/aws-cloud-financial-management/optimizing-cost-for-using-foundational-models-with-amazon-bedrock/ (both accessed in May 2025).
239 See https://cloud.google.com/vertex-ai (accessed in May 2025).
240 See https://cloud.google.com/vertex-ai/generative-ai/docs/model-reference/vertex-ai-model-optimizer (accessed in May 2025).
241 See https://www.techtarget.com/searchenterpriseai/definition/Amazon-Bedrock-AWS-Bedrock (accessed in May 2025).

Discovery, spanning automotive, financial services, healthcare, and media industries.[242] Customer testimonials for both platforms emphasise platform reliability and integration benefits rather than model variety alone.[243] These adoption patterns across diverse industries suggest that general-purpose platforms appeal to organisations seeking managed infrastructure and integrated workflows rather than simply access to multiple models.

Fine-tuning capabilities address customisation needs

Both platforms enable model customisation through fine-tuning, addressing business requirements for domain-specific optimisation. Amazon Bedrock enables organisations to fine-tune models using their own proprietary data.[244] Google Vertex AI provides similar capabilities through various tuning options, including adapter tuning and reinforcement learning from human feedback.[245]

Currently, fine-tuned models remain platform-specific due to different technical architectures, meaning organisations investing heavily in customisation face switching costs if changing platforms requires repeating the fine-tuning process. However, this constraint may prove temporary as advanced prompt engineering techniques could achieve equivalent customisation without platform-specific modifications, and portability standards may arise as the technology matures.

Competitive implications

General-purpose platforms make advanced GenAI accessible to organisations that lack the resources to build equivalent capabilities independently. This accessibility effect expands the competitive GenAI market compared to scenarios where only well-resourced firms could deploy sophisticated models. Platform aggregation also maintains competitive options among model providers by offering access to enterprise customers. Both Bedrock and Vertex AI actively court third-party model providers, suggesting that platforms benefit from diverse model offerings while developing proprietary alternatives.[246]

However, platform concentration could potentially affect competitive options if switching costs become substantial or if platforms favour proprietary models over third-party offerings. Current evidence suggests these concerns remain largely theoretical. Both platforms continue expanding third-party model access, and the emergence of standardisation tools described in section 5.1.2 may provide additional competitive constraints that may preserve switching options.

Hugging Face: Open-source community model provides an alternative competitive approach

Hugging Face, a French American company founded in 2016, is an AI platform and open-source community hub. The platform provides an alternative competitive approach through open-source community models rather than vertically integrated infrastructure. Hosting over one million items (models, datasets, and applications combined), Hugging Face reached a $4.5 billion valuation in 2023 through community engagement rather than cloud integration.[247] The platform's growth illustrates how collaborative development can create competitive positioning whilst addressing different market needs than integrated cloud platforms.

Community engagement creates network effects through collaborative participation

Hugging Face's competitive positioning derives from network effects generated through community participation. The platform hosts over a quarter million models and tens of thousands of datasets contributed by a global community of developers and researchers, including widely known models such

242 See https://www.appsruntheworld.com/customers-database/products/view/google-cloud-vertex-ai (accessed in May 2025).
243 See https://cloud.google.com/vertex-ai (accessed in May 2025).
244 See https://aws.amazon.com/bedrock/faqs/ (accessed in May 2025).
245 See https://cloud.google.com/vertex-ai/generative-ai/docs/models/tune-models (accessed in May 2025).
246 See https://aws.amazon.com/bedrock/?nc1=h_ls; https://medium.com/google-cloud/vertex-ai-model-garden-all-of-your-favorite-llms-in-one-place-a8940ea333c1 (both accessed in May 2025).
247 See https://originality.ai/blog/huggingface-statistics (accessed in June 2025).

as Mistral, BLOOM, DeepSeek and Llama, thousands of datasets for pre-training and fine-tuning, and many programming packages.[248] This collaborative approach creates value that scales with participation: each additional contributor potentially enhances platform utility for other users by expanding available models, improving documentation, and sharing implementation knowledge.

The community model generates participation cycles where content creators benefit from exposure and feedback when sharing models, whilst users access increasingly diverse resources as the community grows.[249] Educational initiatives, competitions, and collaborative events further support community engagement.[250] CEO Clément Delangue describes these dynamics as creating platform utility that *"becomes more and more useful the more users you have"*.[251]

Developers invest time learning platform conventions, building professional reputations through contributions, and developing networks within the ecosystem. These social investments require sustained engagement to maintain value, potentially supporting platform loyalty through community benefits. However, community-driven development faces coordination challenges that centralised platforms may avoid. Quality control relies on community oversight rather than dedicated engineering teams, potentially creating inconsistencies in model documentation, performance benchmarks, and security practices.[252]

Open-source model access supports flexibility

Hugging Face's emphasis on model portability may enable competitive flexibility by reducing dependencies on specific providers. Models hosted on Hugging Face can typically be downloaded and deployed across different cloud environments or on premises, contrasting with platform-specific implementations that remain within particular cloud environments.[253] This portability potentially enables developers to maintain options across multiple vendors.

The platform's commitment to standardisation extends beyond model hosting to development frameworks. The widely adopted Transformers library has over 115,000 GitHub stars and provides standardised interfaces for accessing models across different architectures.[254] This standardisation may reduce technical barriers that might otherwise favour incumbent platforms through proprietary integration requirements.

Business model balances community access with commercial sustainability

Hugging Face employs an "open core" business model that provides basic services freely, whilst monetising additional features and enterprise capabilities.[255] Core products, including the Transformers library and model hosting, remain open source, whilst the company charges for inference APIs, enterprise security features, and expert support services. This approach enables broad community adoption whilst generating revenue from commercial users requiring additional services. The model reflects value differentiation, as enterprise features address organisational requirements for security, compliance, and professional support that community offerings may not provide.

Revenue has increased rapidly since its foundation, with Hugging Face generating approximately $130 million and 50,000 customers in 2024.[256] The platform's enterprise offerings, including Expert Acceleration

---

248 See https://alexsandu.substack.com/p/how-hugging-face-and-kaggle-bolster; https://huggingface.co/docs/hub/en/index (both accessed in May 2025)
249 See https://www.datacamp.com/tutorial/what-is-hugging-face (accessed in May 2025).
250 See https://alexsandu.substack.com/p/how-hugging-face-and-kaggle-bolster (accessed in May 2025).
251 See https://www.acquired.fm/episodes/building-the-open-source-ai-revolution-with-hugging-face-ceo-clem-delangue (accessed in May 2025).
252 See https://www.datacamp.com/tutorial/what-is-hugging-face (accessed in May 2025).
253 See https://research.contrary.com/company/hugging-face (accessed in May 2025).
254 See https://www.datacamp.com/tutorial/what-is-hugging-face (accessed in May 2025).
255 See https://research.contrary.com/company/hugging-face (accessed in May 2025).
256 See https://getlatka.com/companies/hugging-face#revenue (accessed in May 2025).

Programs and dedicated support, provide pathways for commercial expansion while maintaining community access. This strategy requires patient capital to sustain operations during growth phases.

Competitive positioning

Hugging Face competes by offering model diversity and collaborative development capabilities. The platform serves firms that want to build and deploy their applications collaboratively, benefiting from its large open-source community and contributing back to it. Over 1,000 companies use Hugging Face to develop their GenAI capabilities, including established firms across varied sectors like IBM, Mercedes-Benz, SAP, and Thomson Reuters.[257] Companies choose Hugging Face when they prioritise model flexibility and collaborative development over streamlined deployment and enterprise support from integrated alternatives.

The platform's competitive positioning extends beyond model hosting. In 2024, Hugging Face announced with Meta and French cloud computing company Scaleway the launch of a GenAI accelerator program for European startups.[258] Five startups were selected for the program: Batisia, Kodex AI, Neuralk-AI, Vocal Image and Pruna.[259] The participating businesses received support and services using Hugging Face's platform, reinforcing its role as a collaborative development environment rather than just a model repository.

Hugging Face's competitive sustainability depends on maintaining its position as the preferred platform for organisations valuing model diversity and collaborative development. Major technology companies, including Meta, Google, and Microsoft, contribute models to the platform, suggesting broad industry recognition of this collaborative approach. The evidence indicates this positioning remains viable as long as organisations continue to value model flexibility over fully managed services.

4.2.1.3 Dataiku: European compliance specialist

Dataiku, a European technology company founded in Paris in 2013, has grown into a significant player in the data science and AI platform market, reaching a valuation of $3.7 billion in its most recent funding round.[260] The company provides a broad-based AI development platform with particular specialisation in governance capabilities, a segment where it was recognised by IDC as a leader, and GenAI auditability.[261] Its success illustrates how regulatory complexity creates opportunities for platforms that integrate compliance capabilities throughout their AI development lifecycle rather than adding them as secondary features.

Governance focus as a differentiator among GenAI platforms

Dataiku's competitive strategy centres on GenAI governance frameworks integrated throughout its AI development platform, including structured approval processes, centralised GenAI registries, and compliance templates for regulations like the EU AI Act.[262] Governance frameworks refer to systematic processes that organisations use to oversee, control, and manage their AI systems to ensure they meet regulatory requirements and internal standards.

This governance-integrated approach differentiates Dataiku from platforms that treat compliance as a secondary feature, addressing organisational requirements in regulated industries where governance

257 See https://enlyft.com/tech/products/hugging-face; https://huggingface.co/ibm; https://huggingface.co/Mercedes-Benz; https://aws.amazon.com/blogs/machine-learning/how-thomson-reuters-developed-open-arena-an-enterprise-grade-large-language-model-playground-in-under-6-weeks/; https://community.sap.com/t5/technology-blogs-by-members/solving-renewable-energy-challenges-with-sap-and-hugging-face/ba-p/13572416 (all accessed in May 2025).
258 See https://finance.yahoo.com/news/meta-collaboration-launches-ai-accelerator-151500146.html (accessed in May 2025).
259 See https://about.fb.com/news/2024/10/meta-hugging-face-and-scaleway-unveil-the-5-winners-of-the-european-ai-startup-program-at-station-f/ (accessed in May 2025).
260 See https://sacra.com/c/dataiku/ (accessed in May 2025).
261 See https://blog.dataiku.com/dataiku-leader-idc-marketscape-ai-governance (accessed in May 2025).
262 See https://www.dataiku.com/product/key-capabilities/ai-governance/ (accessed in June 2025).

considerations constrain GenAI adoption. For instance, a leading financial services institution reported a 900% increase in efficiency for deploying compliance models to production and a 90% reduction in time-to-deployment compared to prior desktop solutions.[263] Pharmaceutical companies like Novartis report reduced time-to-insights for GenAI applications whilst maintaining regulatory compliance.[264]

However, governance emphasis introduces explicit operational overhead. Dataiku acknowledges that *"understanding compliance requirements and setting up compliance assessment and reporting can slow things down"*, positioning process requirements as risk management rather than efficiency optimisation.[265] The platform requires structured approval workflows and audit documentation that add deployment steps compared to platforms optimising for rapid iteration, creating a differentiation dimension between platforms: regulated industries often accept additional process steps in exchange for systematic risk management, creating market segmentation between organisations prioritising speed versus governance capabilities.

Pricing and business model reflect Dataiku's specialised positioning

Unlike vertically integrated cloud platforms, Dataiku positions itself as cloud-agnostic and integration-friendly whilst maintaining specialisation in governance capabilities. Cloud-agnostic means the platform can run on different cloud computing providers (like Amazon Web Services, Microsoft Azure, or Google Cloud) rather than being tied to one specific provider. The platform connects with leading enterprise data governance software tools, allowing customers to keep their current systems rather than requiring complete replacement.[266]

Dataiku's pricing strategy reflects its positioning as an enterprise governance solution rather than a broad-market platform. The Discover plan costs approximately $80,000 annually for teams of five users, representing a significant premium compared to typical enterprise AI platforms.[267] This pricing necessarily excludes smaller organisations and individual developers from adoption, and aligns with the platform's focus on large enterprises with explicit compliance requirements and substantial GenAI budgets. Confirmed customers include banking (BNP Paribas), consumer goods (Unilever), and pharmaceuticals (Merck) organisations, who relied on Dataiku to implement GenAI solutions tailored to specific industry compliance requirements.[268] Organisations with significant regulatory requirements may justify premium pricing for platforms that address critical operational needs, whilst price-sensitive users typically choose lower-cost alternatives even when sacrificing governance features.

Dataiku's business model is designed with a focus on "human-in-the-loop" approaches where GenAI augments rather than replaces human expertise.[269] For example, Dataiku's work with European financial institutions involves systems where GenAI generates initial content (e.g., investment reports or customer communications) that human experts then review and refine.[270]

Competitive positioning

Dataiku's growth suggests that specialised governance platforms can compete effectively when regulatory requirements create demand for dedicated compliance capabilities. The governance-focused approach

263 See https://www.dataiku.com/stories/detail/fsi-mlops/ (accessed in June 2025).
264 See https://www.dataiku.com/stories/detail/novartis/ (accessed in May 2025).
265 See https://blog.dataiku.com/ai-governance-perspectives (accessed in May 2025).
266 See https://www.dataiku.com/product/key-capabilities/data-governance-with-dataiku/ (accessed in June 2025).
267 Typical enterprise AI platform subscriptions range from $1,000–$100,000 per month, with most business AI solutions costing $100–$5,000 monthly. At $16,000 per user annually, Dataiku's pricing positions it at the premium end of this market. See https://www.codica.com/blog/how-much-does-ai-cost/; https://www.webfx.com/martech/pricing/ai/ (both accessed in May 2025).
268 See https://www.dataiku.com/stories/detail/bgl-bnp/; https://community.dataiku.com/discussion/29099/unilever-creating-data-driven-product-ideas-based-exclusively-on-consumer-wants-and-needs; https://community.dataiku.com/discussion/36704/merck-a-holistic-approach-for-enterprise-level-data-democratization (all accessed in May 2025).
269 See https://www.dataiku.com/product/key-capabilities/ai-agents/; https://www.dataiku.com/stories/detail/understanding-responsible-ai/ (both accessed in May 2025).
270 See https://blog.dataiku.com/modernizing-finance-teams-with-generative-ai (accessed in May 2025).

indicates that regulatory compliance represents a competitive dimension for some market segments, particularly in highly regulated industries. Platforms can successfully differentiate through specialisation in features where general-purpose platforms may be weaker.

Market segmentation may persist, with governance-focused platforms serving highly regulated industries whilst general-purpose platforms address broader markets with different compliance requirements. This segmentation could enable multiple competitive approaches rather than consolidation around single platform strategies.

The competitive evidence suggests that specialised platforms may compete alongside integrated alternatives when they address organisational needs that generalist solutions cannot easily replicate through feature addition. Dataiku's positioning provides insight into how regulatory requirements can support competitive differentiation in platform markets, though the sustainability of this approach depends on regulatory evolution and competitive responses from broader platforms.

4.2.2 GenAI-powered tools for enhancing productivity

The development of GenAI applications for productivity tools has introduced new features to existing tools such as text editors, email platforms, and task management software. Developers are creating solutions that can automate routine tasks, generate content, analyse data, and offer contextual assistance in real time. These capabilities can reduce cognitive load and potentially enhance creativity and decision-making. As models become more sophisticated and customisable, GenAI-powered productivity tools are evolving as well.

While Microsoft and Google integrate GenAI directly into their productivity suites, specialised providers are competing through focused innovation and functionality in specific domains. Rather than replacing integrated platforms, these competitors seek market penetration by building strong capabilities while maintaining interoperability with existing workflows.

Two case studies examine specialisation strategies against vertically integrated incumbents. Notion has built an extensive GenAI-powered workspace that competes with Office 365 and Google Workspace, achieving enterprise adoption through GenAI features and multi-model optimisation. DeepL used translation expertise to compete against embedded translation tools, then expanded into adjacent language workflows whilst maintaining integration with major productivity platforms.

These cases suggest how specialised providers may overcome the distribution and bundling advantages of integrated platforms when they deliver value in specific use cases.

4.2.2.1 Vertically integrated solutions: Gemini for Workspace and Microsoft Copilot for Office 365

Google and Microsoft, both major players in the productivity software space, are embedding GenAI capabilities directly into their respective product suites.

Google's GenAI-powered productivity assistant, Gemini for Google Workspace (previously Duet AI), operates across Gmail, Docs, Sheets, Slides, and more. Gemini assists users in drafting emails, generating presentations, summarising content, or creating data visualisations or auto-generated images and based on natural language prompts and without leaving their familiar applications.[271] The system draws on Google's Gemini family of multimodal LLMs to streamline workflows and reduce the time spent on repetitive tasks. Google reported that *"more than a million people and tens of thousands of companies*

[271] See https://medium.com/@gearapp/google-workspace-ai-evolves-gemini-vs-duet-ai-a-comprehensive-comparison-092d6cf85339; https://blog.google/feed/google-workspace-generative-ai-features/; https://support.google.com/a/users/answer/14200040?hl=en; https://support.google.com/a/answer/13623623; https://www.ibm.com/think/topics/google-gemini (all accessed in May 2025).

have used generative AI in Workspace" as of September 2024.[272] Users save an average of 105 minutes per week according to Google's recent study of enterprise customers, and 75% of survey respondents report an improvement in the quality of their work.[273]

Similarly, Microsoft's Copilot integrates across Microsoft 365, spanning text editor (Word), email and calendar (Outlook), presentations (PowerPoint), and chat, video call and team collaboration (Teams).[274] This integration creates a unified GenAI assistant experience across Microsoft's larger product offering, while leveraging the company's established distribution channels and large user base. Copilot had an estimated 33 million active users across Windows, app and website as of 2024, and Microsoft reported that nearly 70% of Fortune 500 companies now use Microsoft 365 Copilot as of November 2024.[275] Surveyed users report productivity and quality gains, overall reporting being *"29% faster in a series of tasks like searching, writing, and summarizing"*, with *"the best Copilot users even saved more than 10 hours per month"*.[276] The system takes advantage of Microsoft's partnership with OpenAI while utilising Azure infrastructure for cloud and Dynamic 365 enterprise applications user base, which exemplifies vertical integration from infrastructure to applications for end-users.[277]

While vertical integration can raise concerns related to competition, it can also bring benefits, as discussed above in section 3.2.2. In theory, vertical integration could potentially disadvantage standalone GenAI providers by making it harder for independent companies to reach users who are increasingly served by embedded GenAI within familiar applications. Companies developing specialised GenAI tools for productivity would need to compete against integrated solutions that benefit from existing user bases, established workflows, and cross-subsidisation from other profitable products. There are several benefits from the vertical integration of Microsoft's Copilot and Google's Gemini. We describe some of these benefits below.

Cost efficiencies: Integrated providers supply services to themselves at cost rather than market rates. Microsoft runs Copilot on Azure infrastructure without paying external cloud hosting fees. Google uses Google Cloud Platform for Gemini without third-party charges. This eliminates markups that external cloud suppliers would charge. Internal supply also reduces integration friction costs: Microsoft avoids reconfiguring GenAI models for external cloud environments and the associated data formatting requirements. Google similarly bypasses compatibility issues between different providers.

Smoother user workflows: Vertical integration can reduce end-user friction by embedding GenAI capabilities directly within existing workflows rather than switching between separate tools. In Word, Copilot can generate text using document context without requiring copy-paste content to external interfaces. Similarly, Gemini in Gmail automates tasks using existing email threads and contact information.

Faster product development and deployment: Controlling both infrastructure and applications accelerate features and updates rollout, while maintaining security. Microsoft can update Copilot across Office, Teams, and GitHub simultaneously through Azure. Google coordinates updates across Gmail, Docs, and Sheets through its unified infrastructure. Neither company waits for third-party approvals or

---

272 See https://blog.google/products/workspace/google-gemini-workspace-may-2024-updates (accessed in May 2025).
273 See https://workspace.google.com/blog/product-announcements/google-workspace-extends-gemini-benefits-to-more-customers (accessed in May 2025).
274 See https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/; https://support.microsoft.com/en-us/topic/getting-started-with-copilot-on-windows-1159c61f-86c3-4755-bf83-7fbff7e0982d; https://github.com/features/copilot (all accessed in May 2025).
275 See https://www.businessofapps.com/data/microsoft-copilot-statistics/; https://www.microsoft.com/en-us/microsoft-365/blog/2024/11/19/introducing-copilot-actions-new-agents-and-tools-to-empower-it-teams/ (both accessed in May 2025).
276 See https://www.microsoft.com/en-us/microsoft-365/blog/2024/01/31/the-right-way-to-ai-what-were-learning-from-customers/ (accessed in May 2025).
277 See https://blogs.microsoft.com/blog/2025/01/21/microsoft-and-openai-evolve-partnership-to-drive-the-next-phase-of-ai/ (accessed in May 2025).

integration delays. Vertical integration reduces friction between the software and infrastructure layers leading to faster, more secure upgrades across the product suite.

Better testing and quality control: Integration along the supply chain at the infrastructure level, middleware level (i.e., Graph, APIs: Vertex AI, Google workspace APIs), and the applications level (i.e., Microsoft 365: Google Docs) enables controlled end-to-end testing before wider deployment of GenAI assistant features. For instance, pre-production Azure environments allow quality assurance teams to simulate how Copilot performs using real data and workflows before public release. Additionally, vertical integration means testing can be implemented incrementally through insider programs (Microsoft 365) to catch issues early across the full product suite. Microsoft is able to refine models based on user feedback and performance data stemming from its own applications, without the need to reply on adaptations from third parties.[278] Google is similarly able to test Gemini features across Workspace applications in controlled settings before rolling them out to the public.

In addition to the vertical aspect, competition authorities have raised awareness of potential conglomerate effects. Bundling of GenAI capabilities with established productivity suites could theoretically create ecosystem-level barriers to entry, where competitors face difficulty not just in individual applications but in challenging entire integrated workflows. For example, if Microsoft embeds AI writing assistance into Word, email drafting into Outlook, and meeting summarisation into Teams, a specialist AI writing company would need to compete not just against Microsoft's writing tool, but against the convenience of having all these AI capabilities work seamlessly together within the same productivity ecosystem that users already rely on for their daily work.

However, early evidence suggests these concerns have not yet materialised into foreclosure effects, though the timeline for GenAI integration remains short and market dynamics are still evolving. The historically locked productivity software market has seen some notable new competition since the introduction of GenAI, with companies like Notion and DeepL successfully attracting users by offering differentiated GenAI-powered functionality in specific domains. Notably, these challengers have successfully developed product integrations with the major productivity platforms: Notion offers connectors for Microsoft SharePoint, OneDrive, and Google Workspace, while DeepL provides official add-ins for Microsoft 365 (Word, Outlook, PowerPoint) and Google Workspace applications. These product integration with incumbents in the productivity space which offer their own GenAI-powered tools goes against the potential foreclosure concern and shows that the Google Workspace and Microsoft Office remain open to third-party innovation rather than creating closed ecosystems.[279]

While integrated solutions do bundle GenAI capabilities with existing software subscriptions (Microsoft charges an additional $30 per month per seat for enterprise Copilot features), standalone providers continue to find routes to market. The success of specialised GenAI tools demonstrates that quality and focus can overcome distribution advantages when the value proposition is compelling, as described in further detail in the case studies below.

Rather than barriers to entry, we see examples of competition coming from new challengers. This suggests that while monitoring remains appropriate, the competitive effects will likely depend on whether integrated solutions can maintain quality and innovation pace compared to specialised alternatives that continue to emerge, such as the two challengers described below.

[278] See https://www.ibm.com/think/topics/google-gemini (accessed in May 2025).
[279] However, these integration capabilities may also reduce switching incentives by allowing users to maintain their primary productivity suite while accessing specialised AI tools through connectors, potentially supporting incumbent platforms rather than challenging them.

4.2.2.2 NotionAI: specialised GenAI product competing successfully with vertically integrated incumbents

While Microsoft and Google integrate GenAI into their productivity suites, specialised competitors like Notion demonstrate that focused innovation can successfully compete against these vertically integrated incumbents. Notion's success illustrates how companies can build advanced GenAI experiences in specific domains while maintaining interoperability with the major platforms.

Building a GenAI-powered workspace

Notion is a collaboration and productivity tool that describes itself as a *"single space where you can think, write, and plan"*.[280] The company was founded in 2013 with its first product launching the same year as a no-code application building tool.[281] After struggling for several years with technical issues and poor market fit, the company underwent an in-depth transformation, eventually launching the workspace platform version in 2016, and NotionAI in 2023 (after a private alpha experiment started in November 2022).[282] Since its launch, the company has raised $343.2 million in funding and reached a $10 billion valuation after its latest funding round in 2021.[283]

What started as a failing no-code platform evolved through this pivot into a full-service workspace that directly competes with Microsoft 365 and Google Workspace through enhanced GenAI capabilities and user experience. Notion is reported to be used by teams in over 62% of Fortune 500 companies, and its enterprise adoption spans diverse industries and company types.[284] Notable users include creative companies like Pixar and Figma, tech platforms like Duolingo and Headspace, and business tools companies like Buffer and Typeform, and notably OpenAI itself, who use Notion for internal documentation, project planning, and company-wide knowledge management.[285] The diversity of these use cases demonstrates Notion's platform flexibility beyond traditional productivity software categories.

This enterprise success is reflected in broader market penetration. Notion reached 100 million users in August 2024, 4 million of which are paying customers.[286] A 2024 survey of professionals suggested that 8% of the UK respondents reported a preference for NotionAI, though the sample size was limited and the methodology unclear.[287] That a specialised tool can capture this level of adoption against established platforms with decades of enterprise relationships, bundled pricing, and default installations demonstrates that GenAI-powered innovation can overcome structural and incumbency advantages when the value proposition is compelling.

Notion was one of the first players to gain access to OpenAI's GPT-4, publicly launching its AI Writing Assistant in February 2023. Notion took advantage of this early access to build a substantial competitive advantage in GenAI-powered productivity tools. The company assembled a specialised GenAI team of approximately 20 people focused on indexing, user experience, and model optimisation, areas where specialised companies can often outpace larger, less agile organisations. Following the launch of its AI Writing Assistant, Notion expanded its GenAI capabilities across the platform. The company introduced AI Autofill, automatically populates database fields using intelligent prompts, as well as Q&A, an integrated chatbot that allows users to query their workspace content directly. This extensive GenAI integration

---

280 See https://www.notion.com/help/guides/what-is-notion (accessed in May 2025).
281 See https://research.contrary.com/company/notion (accessed in May 2025).
282 See https://bullet.so/blog/history-of-notion-everything-from-users-funding-and-more/; https://www.notion.com/blog/behind-the-scenes-notion-ai (accessed in May 2025).
283 See https://www.clay.com/dossier/notion-funding; https://taptwicedigital.com/stats/notion; https://getlatka.com/companies/notion (all accessed in May 2025).
284 See https://taptwicedigital.com/stats/notion; https://feather.so/blog/notion-valuation; https://www.notion.com/enterprise (accessed in May 2025).
285 See https://www.klubzero.com/post/notion-case-study-growth-competitors-insights; https://www.notion.com/customers/openai (both accessed in June 2025).
286 See https://www.notion.com/blog/100-million-of-you; https://taptwicedigital.com/stats/notion (both accessed in May 2025).
287 See https://www.techradar.com/pro/exclusive-we-find-out-what-the-worlds-top-productivity-tools-are-and-exactly-why-you-like-them-so-much (accessed in May 2025).

demonstrates how Notion has embedded artificial intelligence throughout its core functionality rather than treating it as a separate add-on feature.

Adopting a multi-model strategy as a competitive advantage

Despite obtaining privileged early access to Open AI's FM, Notion has since expanded its portfolio of FMs and relies on a multi-model strategy that allows it to optimise different capabilities with best-in-class providers rather than relying on a single FM family. The platform now uses both OpenAI's GPT-4 and Anthropic's Claude (with the possibility for the end-user to select their preferred model for certain queries), as well as Cohere's Rerank for its retrieval and search tools (tools that find and organise information from various sources).[288]

This approach provides several advantages over vertically integrated solutions. Anthropic's focus on AI safety and ethical development allows Notion AI to better manage sensitive content, while Cohere's specialised Rerank model delivers improved search accuracy compared to general-purpose models, according to the company.[289] The Cohere implementation also provides cost savings by eliminating expenses associated with converting data and information into a format the AI system can process and store, which would be required with other approaches. [290]

Notably, Notion uses existing powerful foundation models and feeds them relevant information from users' workspaces, rather than building and training their own AI models from scratch. According to Notion, this strategy allows the company to stay at the innovation forefront without extensive training and maintenance overheads that burden larger organisations.[291]

Successful interoperability with popular tools

Despite competing with Microsoft and Google's vertically integrated suites, Notion has successfully developed deep technical connections with both, demonstrating that the major platforms remain open to innovative competitors: Notion offers official connectors for Microsoft SharePoint and OneDrive, as well as Google Drive, allowing users to preview and manage Microsoft and Google workspace files directly within Notion documents.[292]

Technical connections through automation platforms like Zapier further extend Notion's reach, allowing sophisticated workflows between Notion and Microsoft Office. Zapier, for instance, provides triggers that allow specific actions in Notion (creating or updating a page) to initiate corresponding actions in Office 365 (saving files to OneDrive). The most common triggers for integrating Notion with Office 365 include creating or modifying database items, updating content blocks, and changing specific properties within a database.[293] Users can automatically create Notion database entries from Office emails, sync calendar events, and trigger actions across platforms.

Notion also integrates with other popular tools like Slack, GitHub, and Jira. These deep connections benefit users who can seamlessly navigate their workflows across multiple tools, for instance receiving notifications in Slack when there are updates to a Notion data base. By integrating GitHub and Notion, GitHub code repositories can be placed into Notion as databases, which allows for the creation of synced databases.[294] Beyond basic connectivity, Notion AI's Enterprise Search connects with multiple

---

288 See https://www.notion.com/releases/2024-06-18; https://www.anthropic.com/customers/notion; https://cohere.com/customer-stories/notion; https://www.notion.com/integrations (all accessed in May 2025).
289 See https://matthiasfrank.de/notion-features/notion-ai/; https://aws.amazon.com/blogs/machine-learning/improve-rag-performance-using-cohere-rerank/; https://cohere.com/customer-stories/notion (both accessed in May 2025).
290 See https://cohere.com/customer-stories/notion (accessed in May 2025).
291 See https://ainativedev.io/podcast/building-notion-ai-lessons-learned-and-myths-busted-with-simon-last-notion-co-founder-and-cto (accessed in May 2025).
292 See https://www.notion.com/integrations/google_drive; https://www.notion.com/help/notion-ai-connector-for-microsoft-sharepoint-and-onedrive (both accessed in May 2025).
293 See https://zapier.com/apps/notion/integrations/office-365 (accessed in May 2025).
294 See https://www.notion.com/integrations (accessed in May 2025).

productivity platforms including Slack, Microsoft tools, Jira, Google Workspace, and GitHub.[295] This allows users to search across all connected tools from within Notion, creating a unified search experience that pulls information from multiple work apps rather than limiting searches to individual platforms.[296] Additional connectors for Linear, Gmail, Zendesk, Box, and Salesforce are planned, and developers are free to develop their own connectors using the Notion API.[297]

Notion's success demonstrates that specialised GenAI companies can build alternatives to traditional productivity suites rather than accepting niche roles. The company's recent expansion into email illustrates this strategy: Notion launched Notion Mail in April 2025, an GenAI-powered Gmail client featuring auto-labelling and database-like organisation. Early reviews show enthusiasm for specific innovations alongside recognition of current limitations, suggesting that while expansion attempts may not immediately succeed, focused GenAI capabilities can create genuine competitive advantages in individual features.[298]

4.2.2.3 DeepL: focus on multilingual workflows as entry point to competing against generalist incumbents

DeepL provides another example of successful specialisation in GenAI-powered productivity tools, demonstrating how European companies can build competitive advantages through deep focus on specific domains. Founded in Germany in 2017, DeepL has achieved notable success in machine translation, achieving strong adoption among language service providers according to industry surveys by competing directly against the translation capabilities embedded in Microsoft 365 and Google Workspace, reaching a valuation of $2 billion, achieved in 2024, double their previous valuation of $1 billion in 2023.[299]

Proven success through specialised translation focus

DeepL's competitive success in translation is well-documented. A 2024 industry survey revealed that 82% of language service companies now use DeepL, far surpassing Google (46%), Microsoft (32%), and Amazon AWS (17%).[300] The company serves over 100,000 business customers and generated $185.2 million in revenue in 2024, representing 31% growth from 2023.[301] A 2024 Forrester study revealed that "DeepL delivered 345% ROI for global companies, reducing translation time by 90% and driving a 50% in workload reduction".[302]

This success stems from DeepL's strategy and expertise in translation models. Unlike many GenAI applications that rely on and tweak other companies' models, DeepL builds its service from the ground up, with a proprietary LLM optimised specifically for translation that the company claims outperforms GPT-4 and models from Google and Microsoft.[303] Rather than competing across all productivity functions, DeepL focused entirely on language quality and built deep expertise in this domain.

---

295 See https://www.notion.com/blog/notion-ai-for-work (accessed in May 2025).
296 See https://www.notion.com/help/guides/find-answers-and-generate-reports-with-enterprise-search (accessed in May 2025).
297 See https://www.notion.com/blog/notion-ai-for-work; https://www.notion.com/integrations (accessed in May 2025).
298 See for example https://www.tomsguide.com/ai/i-ditched-gmail-for-notion-mail-for-3-weeks-heres-what-its-like-to-use-ai-powered-email; https://www.xda-developers.com/replaced-outlook-with-notion-mail-how-it-went/; https://www.xda-developers.com/ditched-gmail-notion-mail/ (all accessed in May 2025).
299 See https://electroiq.com/stats/deepl-statistics/ (accessed in May 2025).
300 See https://www.prnewswire.com/news-releases/deepl-is-2024s-most-used-machine-translation-provider-worldwide-among-language-service-companies-302270449.html (accessed in May 2025).
301 See https://electroiq.com/stats/deepl-statistics/; https://getlatka.com/companies/deepl.com (accessed in May 2025).
302 See https://www.otpp.com/en-ca/about-us/news-and-insights/2024/deepl-announces-300-million-investment-at-2-billion-valuation-fueled-by-global-demand-for-ai-language-solutions/ (accessed in May 2025).
303 See https://techcrunch.com/2024/11/13/deepl-launches-deepl-voice-real-time-text-based-translations-from-voices-and-videos/ (accessed in May 2025).

Strategic expansion into adjacent language workflows

Building on this translation success, DeepL has begun expanding into adjacent language-related productivity areas, though these efforts remain in early stages. DeepL Write launched in beta in 2023, positioning itself as a writing assistant that goes beyond grammar correction to offer improvements on style, tone, and phrasing.[304] DeepL Write Pro launched in April 2024 as the company's first service powered by its own LLM, targeting business writing with enterprise-grade security features.[305] Most recently, DeepL Voice launched in November 2024, enabling real-time translation for virtual meetings and in-person conversations, with support for multiple spoken languages within the same meeting and instant captioning in over 30 languages.[306]

These products demonstrate DeepL's strategy of leveraging its core translation expertise to address complete multilingual workflows rather than isolated tasks. DeepL Voice for Meetings enables participants to speak in their preferred language while colleagues receive translated captions in real-time, while DeepL Voice for Conversations enables in-person multilingual communication through mobile devices.[307] DeepL Write positions itself as an alternative to Copilot, Gemini for Workspace or Notion use-cases related to drafting documents, with an edge in multilingualism.

While DeepL's expansion demonstrates strategic thinking, concrete adoption data for the newer products remains limited. The company has shared some customer testimonials for both Write Pro and Voice, but has not released comprehensive user numbers or adoption statistics for these newer products. The timeline limitations are significant: DeepL Write has been available for less than two years, Write Pro for eight months, and Voice for only six months. This makes it difficult to assess their competitive success against incumbent productivity platforms or determine whether DeepL's expansion beyond core translation will prove successful.

Despite expanding into areas where Microsoft and Google offer competing features, DeepL has maintained an integration approach. DeepL delivers GenAI writing suggestions within Microsoft 365, Google Workspace, and web browsers, allowing teams to access specialised language capabilities without leaving their primary productivity environment.[308] DeepL Voice integrates directly with Microsoft Teams, enabling real-time translated captions during meetings.[309]

DeepL's case illustrates how specialised GenAI providers can build from proven competitive advantages in specific domains and attempt to expand to address broader workflow needs once they are recognised as a credible player in their niche. However, the ultimate success of this expansion strategy remains to be demonstrated as these newer products mature, and adoption data becomes available.

The coexistence of vertically integrated and specialised providers shows that multiple business models remain viable and compete against one another in the same markets. This diversity of successful business models suggests that vertical integration efficiencies do not reduce the attractiveness of competitive alternatives, reducing concerns about potential market foreclosure.

---

304 See https://www.deepl.com/en/products/write; https://techcrunch.com/2023/01/17/deepl-takes-aim-at-grammarly-with-the-launch-of-write-to-clean-up-your-prose/ (both accessed in May 2025).

305 See https://www.prnewswire.com/apac/news-releases/deepl-launches-new-ai-powered-offering-deepl-write-pro-to-supercharge-business-communication-302128290.html (accessed in June 2025).

306 See https://techcrunch.com/2024/11/13/deepl-launches-deepl-voice-real-time-text-based-translations-from-voices-and-videos/; https://www.prnewswire.com/news-releases/deepl-unveils-next-frontier-for-language-ai-with-voice-translation-solution-deepl-voice-302303610.html (both accessed in May 2025).

307 See https://www.deepl.com/en/products/voice/deepl-voice-for-meetings; https://www.deepl.com/en/products/voice/deepl-voice-for-conversations (both accessed in June 2025).

308 See https://www.deepl.com/en/products/write; https://www.deepl.com/en/teams/customer-service (both accessed in May 2025).

309 See https://www.deepl.com/en/products/voice/deepl-voice-for-meetings; https://support.deepl.com/hc/en-us/articles/17295434995612-Manage-meeting-translation-in-Microsoft-Teams (both accessed in May 2025).

4.3 GenAI deployment in European industries

European companies are deploying GenAI to enhance operations in sectors where Europe leads globally. In this section, we analyse how three European firms, in three European-led industries, are deploying GenAI capabilities to improve their competitiveness: Stellantis in automotive manufacturing, AstraZeneca in pharmaceuticals, and LVMH in luxury goods.

These case studies demonstrate how European companies achieve measurable results while navigating regulatory requirements. Each reflects industry-specific deployment strategies that build on deep sector expertise rather than replacing core competencies.

4.3.1 Stellantis: European automaker's strategic approach to GenAI deployment

Stellantis, a European automotive manufacturer headquartered in the Netherlands, provides an example of how established European companies are strategically deploying GenAI across their operations. It manages numerous car brands including Jeep, Citroen, and Peugeot, and has adopted a wide-ranging approach to GenAI integration that spans manufacturing, vehicle engineering, and customer experience, focusing on enhancing existing products and processes.[310]

Successful GenAI deployment in operational and customer-facing applications

Stellantis has successfully deployed GenAI in consumer-facing applications across multiple brands. Peugeot announced plans in early 2024 to integrate ChatGPT across all its vehicles starting, piloting in five countries (France, Britain, Germany, Italy and Spain).[311] DS Automobiles commenced full production with integrated ChatGPT in March 2024, launching in 13 languages across 18 countries.[312] Citroën followed with its ChatGPT integration launched in July 2024, covering models from passenger cars to small utility vehicles and larger vans, across 17 European countries.[313]

The deployment has achieved measurable results: Stellantis reports that tests of Citroën's ChatGPT system performed in May 2024 showed reduced misunderstandings by 68% and increased voice recognition usage by 70%. The system uses SoundHound AI's voice recognition platform – a company that specialises in voice artificial intelligence – integrated with ChatGPT and activated by "Hello Citroën" commands, and provides educational content, entertainment, and everyday assistance during journeys.[314]

Beyond consumer applications, Stellantis has implemented GenAI as part of broader manufacturing improvements. According to the company's chief manufacturing officer, implementing innovations along with continuous improvement and lean processes has reduced transformation costs by 11%, energy consumption by 23%, and quality issues by 40% since 2021.[315] Among these innovations, the company has deployed a predictive maintenance system using generative AI across 28 locations worldwide, which downloads daily plant problems and provides solutions based on previous similar issues encountered at other facilities.

---

310 See https://www.stellantis.com/en/company; https://mistral.ai/customers/stellantis (both accessed in June 2025).
311 See https://www.reuters.com/business/autos-transportation/stellantis-peugeot-cars-use-chatgpt-talk-drivers-2024-01-30/; https://www.automotivedive.com/news/stellantis-peugeot-adding-chatgpt-ai-voice-assistant/706378/ (both accessed in May 2025).
312 See https://www.automotiveinteriorsworld.com/news/hmi/ai-voice-assistant-with-integrated-chatgpt-launches-in-stellantis-ds-automobiles-in-japan.html (accessed in May 2025).
313 See https://www.media.stellantis.com/em-en/citroen/press/citroen-adopts-chatgpt-to-enhance-the-on-board-comfort-experience (accessed in May 2025).
314 See https://www.media.stellantis.com/em-en/citroen/press/citroen-adopts-chatgpt-to-enhance-the-on-board-comfort-experience (accessed in May 2025).
315 See https://www.automotivemanufacturingsolutions.com/smart-factory/stellantis-turns-to-ai-tools-and-fast-deployment-to-cut-production-costs-and-launch-times/46116.article (accessed in May 2025).

Strategic European GenAI partnerships

Stellantis has been working with French AI company Mistral AI for over a year on projects across vehicle engineering, fleet data analysis, internal car sales and manufacturing.[316] The latest development under this partnership is a GenAI-powered in-car assistant that interacts with users when driving. This partnership focuses on more technical applications compared to the consumer-facing ChatGPT integration. The ChatGPT integration serves as a "travel companion" for general questions and entertainment, while the planned Mistral AI assistant is designed for "more focused support" as a voice-enabled user manual for vehicle-specific queries.[317]

Beyond its in-car assistant feature, Stellantis and Mistral AI have been exploring several other GenAI deployment projects. These include: i) a Stellantis-specific GenAI tool to streamline complex database analysis (Bill of Material "BOM" Data Intelligence); ii) an automated fleet and survey data processing AI tool (Vehicles Feedback Data Analysis); iii) a chatbot designed to assist employees with purchasing company vehicles ("Club Stellantis" Virtual Assistant); and iv) the use of Mistral AI's technology to detect manufacturing errors.[318] As Stellantis Chief Engineering & Technology Officer Ned Curic explained, *"instead of waiting for analysis for weeks, we can do that in minutes and make a decision in the afternoon"*.[319]

Cautious deployment approach with risk mitigation

Stellantis emphasises solutions that can be implemented quickly without major factory rework, rather than those requiring considerable redesign and implementation costs.[320] As part of its strategy for GenAI deployment, the company has also pursued selective acquisitions to maintain control over critical GenAI capabilities in-house. Stellantis has acquired aiMotive, an automotive technology company working on automated driving solutions.[321] Stellantis is using the technology from aiMotive to develop its autonomous driving system, STLA AutoDrive.[322] In 2024, Stellantis acquired CloudMade, a developer specialising in smart automotive GenAI and data-driven solutions.[323] Stellantis cites data privacy risk reduction as one reason for bringing these capabilities in-house.[324]

While Stellantis has achieved concrete deployment success, some ambitious projects remain in development. The Mistral GenAI-powered in-car assistant for vehicle-specific queries is still under development, with unclear launch timelines, representing the company's expansion from general GenAI applications to more specialised automotive use cases.

Stellantis's GenAI deployment reflects broader industry trends where GenAI is becoming a competitive dimension in automotive rivalry. Several other major automakers are similarly implementing GenAI across their operations:

- Audi uses AI in its factories for quality control of spot welds and crack detection in metal parts during production.[325]

316 See https://www.stellantis.com/en/news/press-releases/2025/february/stellantis-and-mistral-ai-strengthen-strategic-partnership-to-enhance-customer-experience-vehicle-development-and-manufacturing (accessed in May 2025).
317 See https://www.stellantis.com/en/news/press-releases/2025/february/stellantis-and-mistral-ai-strengthen-strategic-partnership-to-enhance-customer-experience-vehicle-development-and-manufacturing; https://gettotext.com/transitioning-from-engineering-to-onboard-assistant-the-driving-force-behind-stellantis/ (both accessed in May 2025).
318 See https://www.stellantis.com/en/news/press-releases/2025/february/stellantis-and-mistral-ai-strengthen-strategic-partnership-to-enhance-customer-experience-vehicle-development-and-manufacturing (accessed in May 2025).
319 See https://www.reuters.com/technology/artificial-intelligence/carmaker-stellantis-mistral-ai-expand-their-ai-strategic-partnership-2025-02-07/ (accessed in May 2025).
320 See https://www.automotivemanufacturingsolutions.com/smart-factory/stellantis-turns-to-ai-tools-and-fast-deployment-to-cut-production-costs-and-launch-times/46116.article (accessed in May 2025).
321 See https://www.stellantis.com/en/news/press-releases/2022/december/stellantis-completes-acquisition-of-aimotive-to-accelerate-autonomous-driving-journey (accessed in May 2025).
322 See https://aimotive.com/; https://media.stellantisnorthamerica.com/newsrelease.do (both accessed in May 2025).
323 See https://www.stellantis.com/en/news/press-releases/2024/january/stellantis-to-enhance-personalized-mobility-experience-with-acquisition-of-cloudmade-s-artificial-intelligence-technologies-and-ip (accessed in May 2025).
324 See https://techxplore.com/news/2024-01-stellantis-startup-ai-tech-vehicle.html (accessed in May 2025).
325 See https://www.audi.com/en/innovation/future-technology/artificial-intelligence/ai-in-production/ (accessed in May 2025).

- BMW Group has partnered with AWS to build cloud assistant for IT systems management, offering real-time advice on performance, cost, and security.[326]
- Volvo collaborates with AI startup Waabi to create safer and more reliable self-driving trucks in the US, combining Waabi's virtual driver technology with Volvo's autonomous vehicles.[327]
- Mercedes-Benz is using AI in its MBUX Virtual Assistant for personalised driver experience, including showing real-time traffic and alerting them to nearby pedestrians.[328]

This industry-wide adoption suggests AI capabilities are transitioning from optional innovations to competitive necessities. Stellantis's case illustrates how European manufacturers are successfully deploying GenAI to enhance existing products and processes while managing competitive and operational considerations. The company has achieved measurable operational benefits and successful consumer deployments through a strategic approach that combines multiple GenAI partnerships, gradual implementation, and risk-managed expansion, demonstrating how GenAI is becoming integral to automotive competitiveness.

4.3.2 AstraZeneca: Multi-pathway drug target discovery achieves measurable results

AstraZeneca's approach to GenAI deployment demonstrates how European pharmaceutical companies can adopt AI throughout their R&D operations while managing regulatory complexity. The British-Swedish company has invested in internal GenAI capabilities and created dedicated data science teams whilst forming strategic partnerships, aiming to identify new drug targets, improve clinical success predictions, and enhance disease understanding.[329]

Proven GenAI-generated drug targets deliver portfolio value

AstraZeneca reports that its collaboration with BenevolentAI has identified seven potential "drug targets" across four disease areas since 2019. Drug targets are biological mechanisms in the body (like specific proteins or cellular processes) that scientists believe could be influenced by future medicines to treat disease, essentially identifying where in the body a potential treatment should focus its effects.[330] The partnership identified five such targets by 2023, with two additional targets added in 2024.[331] As of 2023, AstraZeneca was researching four of the five initially identified candidates, with each target selection triggering milestone payments to BenevolentAI.[332]

This approach combines AstraZeneca's genetic and patient data with BenevolentAI's AI and machine learning capabilities to pinpoint these potential treatment opportunities. For idiopathic pulmonary fibrosis, a lung disease with poorly understood causes, the AI system aims to distinguish between what actually causes the disease versus what happens as a consequence of having it, potentially helping researchers design treatments that address root problems rather than just managing symptoms.[333] However, significant uncertainty remains: identifying promising targets is only the first step in a lengthy process.

---

326 See https://aws.amazon.com/blogs/industries/bmw-group-develops-a-genai-assistant-to-accelerate-infrastructure-optimization-on-aws/ (accessed in May 2025).
327 See https://aimagazine.com/articles/how-volvos-autonomous-trucks-are-gen-ai (accessed in May 2025).
328 See https://media.mbusa.com/releases/release-ebe78e1e0abb0f8a2f173a4032054126-mercedes-benz-heralds-a-new-era-for-the-user-interface-with-human-like-virtual-assistant-powered-by-generative-ai (accessed in May 2025).
329 See https://www.astrazeneca.com/media-centre/press-releases/2025/astrazeneca-invests-2-and-half-bn-in-beijing-r-and-d-and-manufacturing.html; https://www.astrazeneca.com/media-centre/press-releases/2024/astrazeneca-invests-3bn-500mn-in-us.html; https://careers.astrazeneca.com/data-science-and-ai (all accessed in May 2025).
330 For a definition of drug target, and where they sit in the drug discovery process, see: https://www.sciencedirect.com/topics/chemistry/drug-target; https://pmc.ncbi.nlm.nih.gov/articles/PMC11334170/ (both accessed in May 2025).
331 See https://www.benevolent.com/news-and-media/press-releases-and-in-media/benevolentai-and-astrazeneca-collaboration-yields-continued-success-further-novel-target-progressed-portfolio/ (accessed in May 2025).
332 See https://www.benevolent.com/news-and-media/press-releases-and-in-media/benevolentai-achieves-further-milestones-ai-enabled-target-identification-collaboration-astrazeneca/ (accessed in May 2025).
333 See https://www.benevolent.com/news-and-media/press-releases-and-in-media/benevolentai-achieves-further-milestones-ai-enabled-target-identification-collaboration-astrazeneca/; https://www.benevolent.com/news-and-media/blog-and-videos/charting-progress-benevolentais-collaboration-astrazeneca/ (both accessed in May 2025).

Developing actual medicines from these targets and bringing them to market typically requires years of additional research with substantial risks of failure.

Multi-provider strategy balances capabilities and dependencies across research functions

AstraZeneca employs multiple GenAI partnerships rather than relying on a single provider, reflecting a deliberate strategy to avoid vendor lock-in whilst accessing best-in-class capabilities for different applications. Beyond BenevolentAI, the company:

- collaborates with Illumina on AI-based genome interpretation tools for drug target discovery,[334]
- uses Amazon SageMaker to fine-tune genomic foundation models,[335] and
- partners with Immunai for oncology applications.[336]

Each partnership addresses different pharmaceutical R&D functions where AstraZeneca selected from competing options. For genomic analysis, alternatives to Illumina include companies like 10x Genomics, Oxford Nanopore Technologies, and Pacific Biosciences.[337] For cloud-based AI infrastructure, AstraZeneca could have chosen Google Cloud's life sciences platform or Microsoft Azure's healthcare solutions instead of Amazon.[338] In drug discovery AI, competitors to BenevolentAI include Exscientia, Recursion Pharmaceuticals, and Atomwise.[339]

This demonstrates that pharmaceutical companies can access specialised GenAI capabilities without being locked into single-vendor ecosystems, suggesting competitive markets exist across different GenAI applications in drug discovery.

AstraZeneca's genomics AI strategy uses multiple approaches to analyse genetic data for drug discovery. The Illumina collaboration provides AI tools that help identify genetic mutations and changes in genetic material that could lead to disease, working alongside AstraZeneca's own analysis systems.[340] Separately, AstraZeneca worked with Amazon to customise HyenaDNA, a specialised AI model that analyses genetic sequences, to better predict whether genetic variations are likely to cause disease. This customised model outperformed existing prediction methods in four out of five comparison tests by an average of 20.9%, demonstrating improved accuracy in identifying potentially harmful genetic changes.[341]

The company applies GenAI across 70% of its small molecule chemistry projects, using computational methods to predict optimal molecular structures and reduce traditional optimisation cycles.[342] AstraZeneca's $18 million collaboration with Immunai demonstrates targeted deployment in oncology, using single-cell genomics and machine learning to optimise cancer immunotherapy trials through dose selection and biomarker identification.[343]

---

334 See https://www.illumina.com/company/news-center/press-releases/press-release-details.html?newsid=610e7fbe-cae9-4f64-a07d-8ebabce60b38; https://pharmaphorum.com/news/astrazeneca-builds-its-ai-capabilities-again-with-illumina-tie-up (both accessed in May 2025).

335 See https://aws.amazon.com/blogs/industries/astrazeneca-fine-tunes-genomics-foundation-models-with-amazon-sagemaker/ (accessed in May 2025).

336 See https://www.reuters.com/technology/artificial-intelligence/astrazeneca-ai-collaboration-with-immunai-inform-cancer-drug-trials-2024-09-26/ (accessed in May 2025).

337 See https://craft.co/oxford-nanopore-technologies/competitors; https://www.genomeweb.com/sequencing/oxford-nanopore-technologies-10x-genomics-partner-single-cell-spatial-rna-seq (both accessed in May 2025).

338 See https://www.microsoft.com/en-us/industry/health/life-sciences; https://cloud.google.com/solutions/healthcare-life-sciences (both accessed in May 2025).

339 See https://ir.recursion.com/news-releases/news-release-details/recursion-and-exscientia-two-leaders-ai-drug-discovery-space; https://www.nature.com/articles/d43747-021-00045-7 (accessed in May 2025).

340 See https://www.genomeweb.com/business-news/illumina-astrazeneca-partner-ai-based-drug-target-discovery (accessed in May 2025).

341 See https://aws.amazon.com/blogs/industries/astrazeneca-fine-tunes-genomics-foundation-models-with-amazon-sagemaker/ (accessed in May 2025).

342 See https://www.astrazeneca.com/r-d/data-science-and-ai.html; https://www.astrazeneca.com/what-science-can-do/topics/clinical-innovation/oncology-partnerships.html (both accessed in May 2025).

343 See https://www.fiercebiotech.com/medtech/astrazeneca-extends-ai-immuno-oncology-rd-pact-immunai (accessed in May 2025).

AstraZeneca combines GenAI with R programming language (a statistical computing tool) to speed up biomarker research, identifying which patients are most likely to respond to specific treatments. The company also uses this combination to evaluate different clinical trial scenarios before actually running them, helping predict potential outcomes and optimise study designs.[344] The company has partnered with Turbine, a London-based biotech firm, to use computer simulations that model drug resistance mechanisms in blood cancers. Turbine's "Simulated Cell" platform runs billions of virtual experiments to understand why cancer treatments stop working in some patients, focusing on blood cancers like leukaemia and lymphoma.[345]

These applications show how GenAI enhances multiple stages of pharmaceutical research and development, though the measurable impact varies significantly across different use cases: some applications like the BenevolentAI collaboration have produced concrete drug targets, while others remain in exploratory phases.

Industry context suggests competitive necessity of deploying AI

AstraZeneca's multi-faceted approach reflects broader pharmaceutical industry trends, with major companies establishing significant GenAI initiatives. Major European and global pharmaceutical companies have established significant GenAI initiatives: Roche has signed over 25 AI partnerships and established a dedicated AI hub, while Novartis has committed to $1.2 billion in potential payments to Google DeepMind's Isomorphic Labs for drug development.[346] GSK partners with multiple AI platforms including Cloud Pharmaceuticals and Insilico Medicine through its Advantage AI program.[347] The global AI in drug discovery market, valued at approximately $1.5–4 billion in 2022–2023, is projected to reach $7–36 billion by 2030 depending on methodology.[348]

This industry-wide adoption suggests AI capabilities are transitioning from optional innovations to competitive necessities. Over 100 deals between pharmaceutical companies and AI drug discovery firms have been completed in the past decade.[349] However, the absence of marketed AI-developed drugs as of 2025 means current deployments remain largely experimental, with actual commercial returns yet to be demonstrated across the industry.[350]

However, industry experts warn that regulatory complexity could disadvantage European companies relative to US competitors facing fewer compliance requirements. The intersection of GDPR data requirements with AI Act transparency obligations creates particular challenges for representative dataset compilation and cross-border research collaboration.[351] Former AstraZeneca digital regulatory strategists note that additional compliance requirements may delay European market entry compared to US competitors. [352]

AstraZeneca's diversified approach contrasts with this broader industry challenge by achieving concrete milestones across multiple AI applications: target identification, genomics analysis, clinical optimisation,

---

344 See https://www.astrazeneca.com/what-science-can-do/topics/data-science-ai/generative-ai-drug-discovery-development.html (accessed in May 2025).

345 See https://turbine.ai/news/turbine-announces-collaboration-to-uncover-biological-mechanisms-of-drug-resistance-in-hematological-cancers/; https://www.businesswire.com/news/home/20240105425138/en/Turbine-Announces-Collaboration-to-Uncover-Biological-Mechanisms-of-Drug-Resistance-in-Hematological-Cancers (both accessed in May 2025).

346 See https://www.swissinfo.ch/eng/multinational-companies/big-pharma-steps-up-race-for-ai-discovered-drugs/79014204

347 See https://www.pharmaceuticalprocessingworld.com/ai-pharma-drug-development-billion-opportunity/

348 See https://www.kingsresearch.com/ai-in-drug-discovery-market-404; https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-drug-discovery-market (both accessed in May 2025).

349 See https://www.swissinfo.ch/eng/multinational-companies/big-pharma-steps-up-race-for-ai-discovered-drugs/79014204 (accessed in May 2025).

350 See https://pmc.ncbi.nlm.nih.gov/articles/PMC11800368/ (accessed in May 2025).

351 See https://www.europeanpharmaceuticalreview.com/article/238250/the-eu-ai-act-will-regulation-drive-life-science-innovation-away-from-europe/ (accessed in May 2025).

352 See https://www.pharmaceutical-technology.com/features/pharmas-ai-prospects-get-nudged-into-the-future-with-eus-ai-act/ (accessed in May 2025).

and internal molecule design. This measured, partnership-based strategy may prove advantageous as regulatory frameworks mature and AI capabilities advance toward clinical applications.

4.3.3 LVMH: Balancing GenAI integration at scale with luxury values

Louis Vuitton Moët Hennessy ("LVMH"), a French multinational holding company specialising in luxury products and services, offers a diverse portfolio of luxury brands, including Louis Vuitton, Dior, and Moët & Chandon.[353] LVMH has adopted a multi-faceted GenAI strategy with multiple partnerships and across multiple operational areas, with the clear aim of introducing GenAI across its luxury portfolio. According to company statements, LVMH reports implementing over 200 GenAI products with plans to train employees from 1,500 in 2023 to potentially 10,000, planning for GenAI adoption across multiple business functions.[354,355]

Chief Omnichannel and Data Officer Gonzague de Pirey emphasises that GenAI tools are designed to *"assist our different people in the organisation and not replace them"*.[356] This "augmentation not replacement" philosophy underpins LVMH's approach to deploying GenAI across Create (product innovation), Move (supply chain optimisation), Show (marketing enhancement), Sell (personalised retail), and Service (customer relationship maintenance) categories.[357]

The strategy combines multiple partnerships: Stanford University's Human-Centered Artificial Intelligence centre for GenAI development, Dataiku for machine learning, and OpenAI's GPT-4 FM for internal applications.[358]

Targeted GenAI-enhanced tools with an adoption focus: LVMH's flagship internal implementation is MaIA, an enterprise chatbot based on OpenAI's GPT-4 FM with privacy-preserving features, designed to assist employees with content translation, mock-up generation, and data analysis tasks across LVMH's operations.[359] The system uses GPT-4's multimodal capabilities to process text, images, and complex business documents, providing predictive insights such as estimating that *"by 2027, 70% of consumer experiences in the beauty industry will be influenced by data and AI"*.[360]

Targeted brand applications demonstrate scalable potential across the portfolio. Louis Vuitton developed a GenAI natural language model for top-tier client advisers to enhance personalised service delivery. Dior Parfums created Dior Astra, a GenAI-powered customer relationship management system that analyses customer data and browsing behaviour to provide tailored recommendations and proactive support.[361] Guerlain's implementation includes automated ingredient selection for fragrance.[362]

LVMH's methodical approach to identifying scalable solutions is exemplified by its annual Innovation Award. The 2024 winner, selected from 1,545 applications across 89 countries, was Chinese startup FancyTech, which created 80 product videos in two weeks for Hublot's Tmall launch, a timeline described as *"almost impossible using traditional the traditional way for the production"*.[363] This represents measurable efficiency gains where traditional video production would require months.

---

353 See https://www.lvmh.com/en/our-group/our-mission (accessed in May 2025).
354 See https://cointelegraph.com/news/lvmh-augmenting-humans-ai-not-replacing; https://wwd.com/business-news/technology/fancytech-lvmh-innovation-award-winner-vivatech-1236394136/ (both accessed in May 2025).
355 However, these ambitious figures reflect corporate intentions rather than verified operational outcomes.
356 See https://wwd.com/business-news/technology/fancytech-lvmh-innovation-award-winner-vivatech-1236394136/ (accessed in May 2025).
357 See https://techinformed.com/lvmh-the-beauty-of-data-transformation-and-ai/ (accessed in May 2025).
358 See https://cointelegraph.com/news/lvmh-augmenting-humans-ai-not-replacing; https://aiexpert.network/case-study-lvmh-embraces-ai/; https://www.dataiku.com/stories/detail/lvmh/ (all accessed in May 2025).
359 See https://cointelegraph.com/news/lvmh-augmenting-humans-ai-not-replacing; https://techinformed.com/lvmh-the-beauty-of-data-transformation-and-ai/ (both accessed in May 2025).
360 See https://techinformed.com/lvmh-the-beauty-of-data-transformation-and-ai/ (accessed in May 2025).
361 See https://www.voguebusiness.com/story/fashion/lvmh-bets-on-generative-ai-with-innovation-award (accessed in May 2025).
362 See https://techinformed.com/lvmh-the-beauty-of-data-transformation-and-ai/; https://digitaldefynd.com/IQ/christian-dior-using-ai-case-study/ (both accessed in May 2025).
363 See https://www.voguebusiness.com/story/fashion/lvmh-bets-on-generative-ai-with-innovation-award (accessed in May 2025).

Adoption strategy centres on extensive training programmes. LVMH established a "Prompt Academy" for staff training and has trained employees across its organisations. The company aims to scale their training from 1,500 in 2023 towards 10,000 employees.[364] This focus addresses a critical sector challenge: whilst operational teams show high GenAI acceptance rates, creative teams demonstrate significantly lower adoption, with only 5% of luxury companies advancing in creative GenAI applications according to Bain & Company analysis.[365]

According to the AIX Expert Network case study, these implementations have *"optimized stock levels and production planning, leading to increased efficiency and reduced environmental impact"* whilst *"smarter distribution strategies have decreased unnecessary product movements, directly benefiting the bottom line"*.[366]

Industry context: Luxury sector GenAI adoption amid reluctance

LVMH's multi-faceted approach operates within a luxury sector showing mixed GenAI adoption patterns. Major competitors are pursuing targeted implementations: Hermès has deployed GenAI-powered personalisation systems analysing customer data and social media activity; Chanel implements predictive analytics for inventory optimisation; whilst brands like Rolex and Omega have achieved specific performance improvements including 20% increased conversion rates through AI-optimised advertising strategies.[367]

However, broader industry analysis reveals general reluctance towards GenAI adoption. Bain & Company research indicates that across the luxury sector, none of 24 identified GenAI use cases has been adopted by more than 30% of brands. According to an industry professional, *"operational efficiency is clearly the number one battle-horse for brands"* with GenAI priorities shaped towards cost management and resource optimisation rather than transformational change.[368]

This context suggests that whilst individual luxury brands are advancing specific GenAI applications, the sector generally maintains cautious approaches to GenAI integration. LVMH's deployment across five operational categories, supported by extensive training programmes and multiple strategic partnerships, represents a more extensive commitment compared to typical luxury sector GenAI adoption patterns.

Enterprise GenAI deployment broadly shows early-stage maturity, with McKinsey research indicating that *"only 1 percent of company executives describe their gen AI rollouts as 'mature'"*.[369] LVMH's structured approach, through employee training, strategic partnerships, and infrastructure investment including MaIA and robust data platforms, positions the company for potential scaling advantages as AI capabilities mature. However, this analysis reflects publicly available company statements and industry data through early 2025, with specific performance metrics awaiting longer-term implementation data and independent assessment of claimed benefits across both LVMH and the broader luxury sector.

These European industry cases demonstrate that GenAI deployment can enhance competitive positioning in sectors where Europe has established strengths. The successful integration of GenAI across automotive, pharmaceutical, and luxury goods industries shows that European companies can apply this technology to build on existing advantages and improve their

364 See https://cointelegraph.com/news/lvmh-augmenting-humans-ai-not-replacing (accessed in May 2025).
365 See https://www.bain.com/globalassets/noindex/2024/bain_report_luxury_and_technology_artificial_intelligence_the_quiet_revolution.pdf.pdf, page 5 (accessed in June 2025).
366 See https://aiexpert.network/case-study-lvmh-embraces-ai/ (accessed in May 2025).
367 See https://digitaldefynd.com/IQ/ai-use-in-luxury-goods-fashion/; https://luxonomy.net/comprehensive-report-on-the-use-of-artificial-intelligence-in-the-luxury-sector-2024-2025/ (both accessed in May 2025).
368 See https://luxus-plus.com/en/artificial-intelligence-luxury-continues-its-transformation-in-2024/ (accessed in May 2025).
369 See https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai (accessed in May 2025).

competitiveness. This suggests significant opportunities for productivity gains and competitive enhancement across Europe's industrial base.

4.4 Conclusion: Early deployment scenarios suggest different approaches can be commercially viable

The case studies examined across enterprise applications, platform intermediation, productivity software, and European industries demonstrate considerable variety in how firms approach GenAI deployment, with different strategies proving commercially viable across markets and use cases.

Three key patterns emerge from this evidence:

- firms commonly adopt multi-sourcing strategies rather than relying on single providers,
- multiple business models appear capable of coexisting competitively (supported in part by open-source alternatives), and
- the current European regulatory framework is designed with the aim of bringing benefits to market participants, although it may create disadvantage for some firms (particularly smaller firms seeking to scale) compared to less-regulated markets.

These patterns have important implications for competitive dynamics and policy considerations.

4.4.1 Companies commonly use multiple AI providers, which may prevent market concentration

The deployment case studies across all sectors examined, from enterprise applications (Estée Lauder, Goldman Sachs) to legal services (SpringBok), platform services (Amazon Bedrock, Hugging Face, Dataiku), productivity software (Notion, DeepL), and European industries (Stellantis, AstraZeneca, LVMH), show firms frequently using multiple AI providers rather than relying on a single vendor.

This pattern appears across different contexts: Estée Lauder uses different providers for specific business functions, Goldman Sachs built an internal system that works with multiple AI models, SpringBok designed applications that can work with any AI provider, AstraZeneca partners with multiple AI companies for different research projects, and Notion uses different AI models for different capabilities. While these represent individual company strategies, the consistency across diverse sectors suggests using multiple providers may be a sustainable approach rather than a temporary trend.

New tools are developing to support this flexibility through unified interfaces (OpenRouter), performance monitoring (Langfuse), and standardisation protocols (MCP, A2A). However, most solutions remain in early stages, and their effectiveness in maintaining switching flexibility requires monitoring as deployment practices become more sophisticated.

Policy relevance: Markets where companies regularly use multiple suppliers (called "multi-homing") and can switch between them relatively easily are typically less likely to become dominated by a single provider. If multi-sourcing strategies remain viable as AI deployment practices mature, this reduces the likelihood that foundation model markets will tip toward monopolistic outcomes. This has implications for whether and when regulatory intervention might be necessary.

4.4.2 Specialised companies can compete with large integrated platforms, with both approaches offering distinct advantages

The case studies show that both vertically integrated GenAI companies (those that control multiple parts of the GenAI value chain) and specialised GenAI companies remain viable across different contexts, each offering different competitive advantages.

The Microsoft Copilot and Google Gemini cases show efficiency benefits from controlling multiple parts of the technology stack, including cost savings (no external fees), smoother user experiences (AI built directly into familiar applications), faster product development (coordinated updates across integrated systems), and better quality control (testing everything together). These advantages help explain why major technology companies pursue integrated strategies.

Despite these integration advantages, specialised providers successfully compete through focused innovation and maintaining compatibility with major platforms. AI platform services demonstrate diverse strategies without requiring full integration: general-purpose cloud solutions (Bedrock, Vertex AI), community-driven open access (Hugging Face), and governance-focused services (Dataiku) serve distinct market needs while enabling AI deployment without extensive in-house capabilities. Open-source models play a significant role in this ecosystem, providing credible alternatives to proprietary solutions and supporting competitive flexibility.

In productivity software, specialised providers compete with integrated incumbents through focused innovation. Notion's multi-model workspace strategy and DeepL's expansion from translation expertise show that distribution and bundling advantages can be overcome when providers deliver compelling value in specific areas. Importantly, both maintain integration with major platforms rather than trying to replace them entirely.

Similarly, in productivity software, specialised providers successfully compete with vertically integrated incumbents through focused innovation and interoperability. Notion's multi-model strategy and advanced workspace, and DeepL's expansion from translation expertise into broader language workflows, demonstrate that distribution and bundling advantages can be overcome when specialised providers deliver compelling value in specific domains. Both cases show specialised providers maintaining integration with major platforms rather than seeking to replace them entirely.

Policy relevance: The coexistence of both approaches suggests that vertical integration efficiencies may not eliminate competitive alternatives, reducing concerns about inevitable market foreclosure through integration strategies.

4.4.3 European regulatory compliance appears compatible with competitive deployment, though risks require monitoring

The European industry case studies provide examples of AI enhancing production processes while navigating regulatory requirements. Stellantis reports operational improvements through partnerships with European providers, AstraZeneca's multi-pathway approach generated drug targets while managing regulatory complexity, and LVMH deployed AI across 200+ applications while maintaining brand positioning.

These cases show European companies adapting deployment strategies to regulatory requirements, particularly in sectors where privacy, safety, and transparency are important. However, the long-term implications remain uncertain: regulatory compliance may create competitive advantages or could impose speed disadvantages compared to less regulated markets.

Policy relevance: Monitoring will be important to ensure regulatory compliance does not inadvertently reduce competition. For example, if compliance requirements become so complex that only large integrated platforms can effectively navigate them, this could favour big companies over smaller specialised providers. Similarly, if regulatory frameworks create technical standards that favour particular providers or if compliance costs disproportionately burden smaller competitors, regulatory requirements could reduce competitive variety despite intentions to preserve it.

4.4.4 Key implications for competitive dynamics

Among the case studies reviewed, **using multiple AI providers is currently common practice** and firms are taking advantage of various market solutions designed to reduce the costs of switching between providers. This has important implications for competitive dynamics in AI markets. As explained in section 3.2.1, markets characterised by multi-homing (using multiple suppliers) and relatively low switching costs are less likely to "tip" toward a single provider, such that it **currently appears unlikely, based on existing competitive dynamics, that the foundation model layer will be dominated by a single supplier**. Moreover, when the risk of tipping is relatively low, there is less cause for concern about any early-mover advantages in AI deployments: **early advantages are less likely to translate into permanent and unmatchable ones**.

This pattern of sustained multi-sourcing has direct policy implications: if deployment flexibility remains viable as the technology matures, the case for immediate regulatory intervention to prevent market concentration becomes weaker, as competitive dynamics may naturally prevent the emergence of dominant positions. Conversely, if the conditions supporting multi-sourcing erode, this would strengthen the case for proactive competition policy measures.

These patterns may not hold across all deployment contexts. Sectors or individual firms requiring deep technical integration or real-time performance optimization may favour single-provider approaches where the costs of switching between models outweigh flexibility benefits. Nevertheless, across the cases examined here, both vertically integrated and specialised approaches appear sustainable, supported by emerging tools that aim to reduce technical barriers and switching costs.

However, these findings reflect early-stage deployment patterns. The durability of current competitive variety will depend on several evolving conditions:

- **Continued competition among foundation model providers**: The multi-sourcing strategies documented across these case studies, from Goldman Sachs's multi-model platform to Estée Lauder's selective sourcing, benefit from having multiple viable AI model providers. Significant concentration among foundation model providers would naturally reduce the value of these flexible deployment approaches.
- **Managing switching costs as deployments mature**: While tools like LangChain and OpenRouter reduce API integration barriers, switching considerations may evolve as firms develop more sophisticated deployments. Accumulating fine-tuning variants, customised AI setups, optimised prompt libraries and instructions, or compliance frameworks tailored to specific providers could gradually increase the complexity of switching suppliers. Moreover, beyond these technical aspects, organisational learning effects, workflow integrations, and compliance investments could create deeper forms of dependency that are harder to quantify but potentially more persistent. The model-agnostic approaches developed by SpringBok for instance will need to adapt as these considerations develop.
- **Adoption of interoperability standards**: The emerging protocols documented here, Anthropic's Model Context Protocol and Google's Agent 2 Agent, aim to standardise how GenAI systems access

data and communicate with each other. Successful adoption of such standards would weaken network effects that favour large integrated platforms and lower technical barriers for new entrants, supporting the competitive variety observed across deployment approaches.

- **Platform competition dynamics**: The diverse platform ecosystem shown here, from Hugging Face's open repositories to specialised services like Dataiku, creates healthy competition among intermediary services. Monitoring concentration trends in this layer will be important, particularly as network effects (where platforms become more valuable as more people use them) around model communities and platform-specific tools develop.
- **Open-source model sustainability**: The competitive flexibility demonstrated across these cases partly relies on credible open alternatives like Meta's Llama and Mistral. Continued innovation in open-source GenAI business models will help maintain this competitive option.
- **Ecosystem-level competition dynamics**: The deployment variety observed depends partly on preventing ecosystem entrenchment where large platforms use GenAI integration to systematically foreclose competitors across multiple markets. This could occur through several mechanisms: bundling AI capabilities with existing dominant platforms in ways that make standalone alternatives unviable, creating proprietary technical standards that lock users into specific ecosystems, or leveraging data advantages from one market to gain decisive advantages in adjacent GenAI applications. The success of specialised providers like Notion and DeepL in maintaining compatibility with major platforms suggests such foreclosure has not occurred, but monitoring whether this competitive coexistence persists will be important for maintaining deployment diversity.

These patterns align with the conditions set out by economic theory for a reduced tipping risk, namely sustained multi-homing and low switching costs (section 3.2.1).

5 A focus on GenAI partnerships

This section focuses on strategic partnerships in the GenAI landscape. Our analysis of partnerships builds on our findings in the previous section on deployment. We examine how collaborative arrangements either support or potentially undermine the competitive flexibility we observed in section 4.

Strategic partnerships are an important feature of the GenAI landscape. They can offer pragmatic solutions to high development costs, technological uncertainties, and need for specialised expertise across multiple domains, as described in sections 2.1.1 and 2.1.2. In this way, strategic collaborations can help companies share financial risks, combine complementary capabilities, and accelerate innovation across the value chain.

The partnership between model developer Anthropic and cloud providers Google and Amazon exemplifies how each party to the agreement contributes with distinct assets – Anthropic's GenAI expertise and the cloud providers' computing infrastructure – to create value more efficiently and effectively than each could on its own.

Over time, the scope of partnerships has broadened across the GenAI value chain, now encompassing user-facing applications. For example, in section 4.1.1.1, this study analysed how consumer-facing companies such as Estée Lauder partnered with GenAI developers Adobe and OpenAI to incorporate GenAI capabilities into their customer experiences and internal processes.

In the below, we describe the role of partnerships in GenAI markets based on publicly available information. First, we describe the variety of partnerships relating to GenAI. We map the landscape of partnerships in Europe and look at the differences of various partnerships. Second, we consider our observations on GenAI partnerships in Europe and how it relates to the likelihood of any competition concerns arising. We conclude that – given there is no current evidence of competition being restricted, and some clear examples of benefits resulting from partnerships – GenAI partnerships are currently contributing positively to the landscape in Europe.

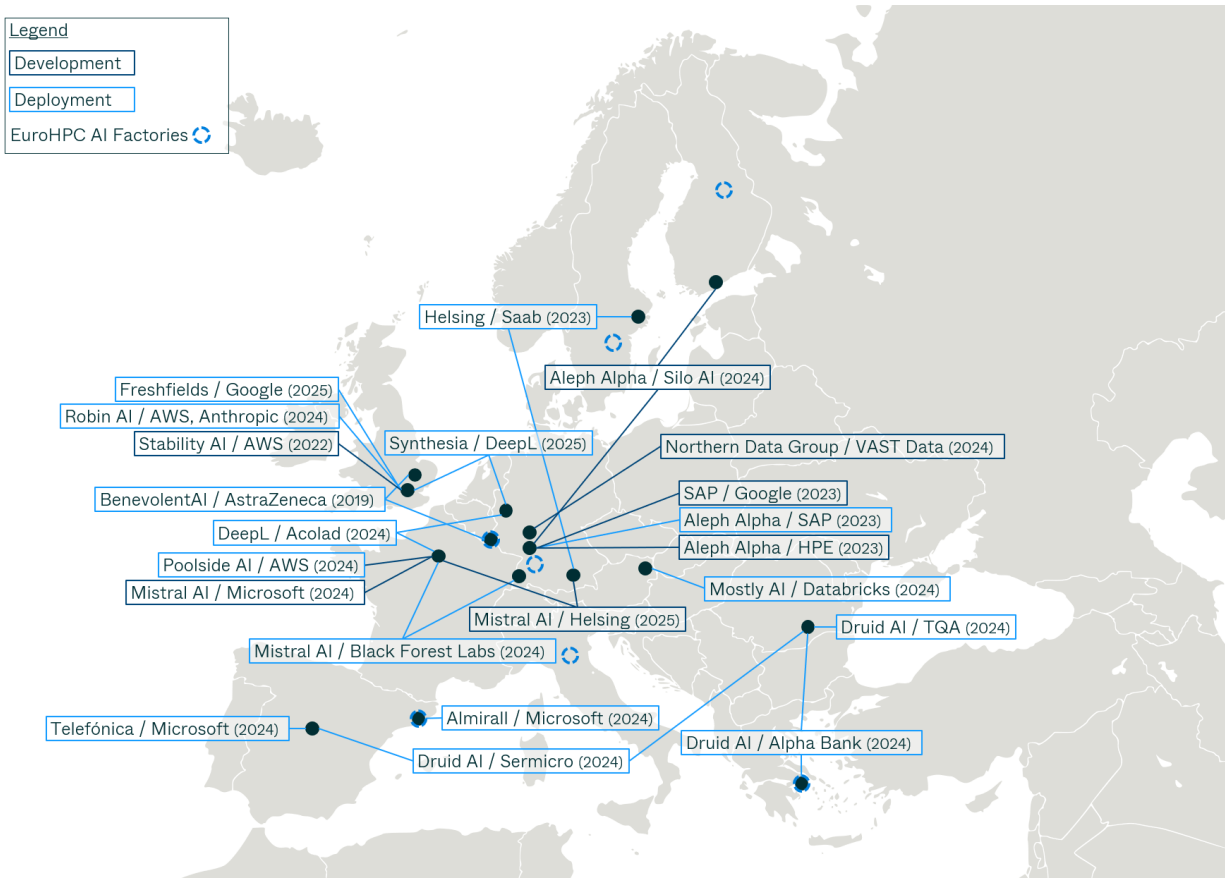
5.1 GenAI partnerships exhibit considerable variety across the value chain

This section begins by examining the landscape of GenAI partnerships in Europe, highlighting the key patterns seen to date, including a high degree of variety (section 5.1.1). Then we look more closely at the different types of partnership, finding that they take diverse forms spanning all layers of the GenAI value chain and differing significantly in their corporate governance frameworks and contractual structures (section 5.1.2).

5.1.1 Mapping the diverse patterns within Europe's GenAI partnership ecosystem

The network of GenAI partnerships across Europe, as shown in Figure 11 below, shows an extensive network of collaborations that has developed rapidly since 2019. This map presents a non-exhaustive selection of significant partnerships, such that the actual collaborative landscape may be even more extensive.

Figure 11: GenAI partnerships span Europe, connecting diverse players across all layers of the value chain



Source: *RBB desk research based on announcements by the partnering firms. Background map by Wikimedia Commons (accessed in May 2025).*

Note: *Partnerships are classified as either Development or Deployment depending on which value chain layer the partnership supports, as interpreted by RBB Economics. This classification does not necessarily match the layer where the firms usually operate, and some partnerships may cover multiple layers of the value chain, such as supplying both cloud services and a distribution platform. The year in brackets indicates the start of the partnership.*

The European GenAI landscape is characterised by diverse partnership models that often transcend national boundaries, including:

- partnerships within individual countries (black dots);
- partnerships across multiple countries (interconnecting lines across European countries); and
- alliances spanning various layers of the GenAI value chain, each addressing distinct and diverse needs (colour-coded boxes).

While innovation hubs in Western Europe show the highest concentration of activity, partnerships exist throughout the region. For example, in Spain, Microsoft has formed collaborations with both Telefónica and Almirall. Similarly, Romania-based GenAI developer Druid AI has partnered with Tquila Automation (UK/Netherlands), Sermicro Digital (Spain) and Alpha Bank (Greece).

Within the European partnership network, we observe the following patterns:

- **Deployment-oriented collaborations**: European partnerships frequently emphasise practical applications over infrastructure alone. For example, companies like DRUID AI (Romania) partner with Tquila Automation (UK/Netherlands) and Sermicro Digital (Spain) to implement conversational GenAI solutions for specific business contexts rather than focusing solely on model development.

- **Cross-border integration**: European GenAI development typically extends beyond national boundaries, with partnerships frequently connecting companies from different European countries, or connecting companies from Europe with those based outside the continent. For instance, the German GenAI-powered translation company DeepL partners not only with the French content and language services provider Acolad but also with UK-based GenAI video creation company Synthesia, highlighting the pan-European nature of these initiatives. Similarly, enterprise software leader SAP demonstrates a multi-partner strategy through its collaborations with both German Aleph Alpha and US-based Google, enabling it to incorporate diverse GenAI capabilities into its business applications.[370]
- **Public infrastructure initiatives**: Complementing private sector partnerships, Europe has developed significant public AI infrastructure through programs like the EuroHPC Joint Undertaking's AI Factories, as displayed in section 2.2.1. These public initiatives, such as the BSC AI Factory (based in Spain and connected to Portugal, Romania, and Turkey) and LUMI AI Factory (based in Finland and linked to Czech Republic, Estonia, Poland, and others), create foundation infrastructure upon which private initiatives can build.[371]

5.1.2 Variety of partnership structures across the GenAI value chain

As illustrated in section 5.1.1 above, European GenAI partnerships take diverse forms, which we now consider in more detail.

Different partnerships across the GenAI value chain

Partnerships have emerged across all layers of the GenAI value chain:

- **Deployment partnerships** between model providers and consumer-facing businesses aim at accelerating the adoption and integration of GenAI capabilities into existing services. These partnerships, as discussed extensively in section 4, help bring value to end users and create practical applications that enhance products and productivity in specific sectors.
- **Development stage partnerships** (between compute providers and model developers) enable smaller, often cash-constrained firms to share the high costs of model development with their compute suppliers. These collaborations allow model developers to focus on their core areas of expertise, while also securing access to essential AI chips and cloud infrastructure that would otherwise require significant financial resources and practical effort to build independently. For example, Anthropic and Google's partnership provided Anthropic with the necessary infrastructure to continue focusing on developing GenAI models, while Google benefitted from offering the popular suite of Claude models to the developers using its Vertex AI platform to build GenAI tools.[372]

Different contractual structures defining the nature of GenAI partnerships

Partnerships also vary in their corporate governance and contractual frameworks:[373]

- **Equity and/or revenue-sharing arrangements**: Partnerships differ in the extent to which parties acquire equity stakes or establish mechanisms for revenue-sharing. Some partnerships grant significant equity stakes or establish mechanisms for sharing future revenues derived from the

---

370 See https://www.acolad.com/en/news/deepl-partnership-to-advance-ai-solutions.html; https://www.deepl.com/en/blog/synthesia-deepl-partner-video-innovation; https://news.sap.com/2023/07/generative-ai-investments-aleph-alpha-anthropic-cohere/ (all accessed in May 2025).

371 See https://eurohpc-ju.europa.eu/about/discover-eurohpc-ju_en; https://www.bsc.es/join-us/excellence-career-opportunities/bsc-ai-factory; https://lumi-supercomputer.eu/lumi-aif/; European Commission (2024b) (all accessed in May 2025).

372 See https://www.anthropic.com/news/anthropic-partners-with-google-cloud; https://cloud.google.com/products/model-garden/claude?hl=en (both accessed in May 2025).

373 Due to the confidentiality that is typical of any business agreement, publicly available information on the specific terms and clauses of partnerships is limited. Therefore, we mainly rely on public information, such as provided by FTC (2025).

partnership. For instance, Microsoft is reported to hold an equity stake in OpenAI, and to receive a certain percentage of OpenAI's revenues.[374]

- **Control rights**: The degree of influence or "control" partners may exercise over each other varies, whether through voting rights, board representation, and equity holdings. For example, both Amazon and Google have non-voting shares in Anthropic, whereas Microsoft occupied an observer's seat in OpenAI's board until 2024.[375]
- **Consultation rights and information access**: Partnerships vary in how parties consult each other, share information, and access business-sensitive information. This may include regular meetings, temporary staff exchanges, or financial reporting. While specific examples on such type of partnerships are limited, the FTC notes that some partnerships provided partners with *"significant access to assets, IP and research developed by the AI developer partners"*.[376]
- **Other contractual rights and/or commitments**: Partnerships can include specific arrangements such as exclusivity provision, "buy commitments" (a commitment to spending a certain amount on a partner's services), preferential pricing or prioritised access to services. OpenAI and Microsoft's partnership, for example, currently gives Microsoft Azure exclusivity on OpenAI's API.[377] Some partnerships may be non-exclusive by nature. For instance, both Amazon and Google have a non-exclusive partnership with Anthropic, allowing Anthropic's models to be distributed both on Amazon and Google's platforms.[378] Similarly, European model developer Mistral AI also has several non-exclusive partnerships with Microsoft, Amazon, and Google.[379]

These dimensions represent only a subset of variables that differentiate GenAI partnerships. The diversity of these arrangements underscores that there is no uniform model for collaboration: different structures serve different strategic objectives while balancing needs for resources, independence, and market access.[380]

5.2 Assessment of likely competition effects of GenAI partnerships

We examine whether partnerships create the types of dependencies and switching costs that economic theory suggests could lead to market tipping or foreclosure (see section 3.2.2–3.2.3). In this section, we consider what the examples of GenAI partnerships studied above suggest for the potential benefits and concerns competitive authorities have identified could arise in the context of GenAI partnerships. Overall, we observe examples of benefits from a number of partnerships and note that no authority has yet found any evidence of harm. This suggests, at least on current evidence, partnerships are likely to be having a pro-competitive effect on the GenAI landscape in Europe.

5.2.1 Examples of pro-competitive benefits resulting from partnerships

Although the space is nascent and evolving, we find examples of partnerships delivering pro-competitive benefits: including reducing barriers to entry, increasing efficiency, and supporting innovation. For example:

374 See https://www.pymnts.com/artificial-intelligence-2/2025/microsoft-moves-to-protect-its-turf-as-openai-turns-into-rival/; https://techcrunch.com/2025/05/07/openai-expects-to-cut-share-of-revenue-it-pays-microsoft-by-2030/ (both accessed in May 2025).
375 See the p. 18 in FTC (2025).
376 See p. 22 in FTC (2025); see p. 23 for information on relevance of staff exchanges.
377 See https://blogs.microsoft.com/blog/2025/01/21/microsoft-and-openai-evolve-partnership-to-drive-the-next-phase-of-ai/ (accessed in May 2025).
378 See Competition & Markets Authority (2024d).
379 See https://mistral.ai/partners#cloud (accessed in May 2025).
380 For a more comprehensive list and discussion on how partnerships can vary contractually, see FTC (2025).

Lowering barriers to entry

- **Provide a source of capital**: Strategic partnerships between GenAI developers and cloud providers, like Mistral AI's collaboration with Amazon, Google, and Microsoft, might allow smaller, innovative firms to enter the market in situations where they have limited capital. This type of partnership may provide the access to the infrastructure and compute power needed to train and deploy models at scale.
- **Share capabilities**: Partnerships can allow firms to combine complementary expertise rather than requiring each company to develop all necessary capabilities in-house. For example, Anthropic can focus on the quality and safety of their models while relying on Google's computational infrastructure. This could decrease entry barriers by allowing firms to combine and build on their existing capabilities.
- **European public initiatives provide access to resources**: Partnership between public entities augment the private sector's supply of critical GenAI infrastructure components, such as AI chips, which remain in limited supply. Efforts like the BSC AI Factory (Spain, Portugal, Romania, Turkey), Lumi AI Factory (Finland, Czech Republic, Estonia, Poland), the EuroHPC Joint Undertaking's AI Factories aim to broaden access to computational resources in Europe.

Increasing competitive pressure and expanding market reach

- **Create more competitive pressure in markets**: Firms can enhance the quality of their products by incorporating GenAI capabilities by partnering with a GenAI provider, where the collaboration enables tailored solutions specifically designed to meet their unique needs, beyond a simple supply agreement. This can strengthen their competitive position within an industry. For example, Telefónica's partnership with Microsoft to incorporate GenAI capabilities into its Kernel platform allows the telecommunications provider to introduce innovative GenAI-enhanced services, increasing competitive options for consumers in digital services markets.
- **Create multiple distribution channels**: Non-exclusive partnerships, such as Mistral AI's simultaneous arrangements with Microsoft Azure, Google Vertex AI, AWS, HuggingFace and OVHcloud, may ensure that GenAI capabilities reach users through multiple channels. This can contribute to a more dynamic and competitive market, where users benefit from greater choice and easier access to a variety of GenAI products.

Accelerating time-to-market

- **Accelerating adoption through trusted channels**: Deployment partnerships can enable faster market adoption of GenAI capabilities through established, trusted distribution channels. DRUID AI's partnership with Alpha Bank in Greece demonstrates how conversational GenAI can reach financial services customers through an established institution, helping specialised European GenAI solutions gain market traction more quickly.
- **Accelerated time-to-market through joint capabilities**: By combining expertise through partnerships, firms can accelerate innovation and reduce the time-to-market. For instance, DeepL and Synthesia partnered to create GenAI-generated videos in multiple languages by integrating their complementary technologies (video generation for Synthesia, multilingual and translation capabilities in text and audio for DeepL). The product, multilingual GenAI-videos, was arguably brought much faster to the market than had the companies not joined forces.
- **Enabling smaller developers to access large user bases**: Deployment partnerships can allow smaller GenAI developers to integrate their technologies into established platforms, accelerating access to broader customer bases that would otherwise take significantly longer to reach. Aleph Alpha's partnership with SAP enables its GenAI models to reach SAP's extensive enterprise customer base across Europe without having to build its own enterprise distribution network.

These benefits help explain why partnerships have become increasingly important in the developing GenAI ecosystem, particularly in Europe.

5.2.2 Recent competition investigations into GenAI partnerships

Competition authorities across Europe and elsewhere have taken proactive steps to understand potential competition concerns in the fast-developing GenAI market and how partnerships may affect competition. The main theories of harm identified by competition authorities are described in section 3.2. Among authorities and public organisations that have conducted studies to identify potential issues are:[381]

- UK's Competition & Markets Authority;
- the European Commission;
- the OECD;
- the French Competition Authority;
- the Portuguese Competition Authority; and
- the FTC.

In the below, we outline the potential pro-competitive benefits of GenAI partnerships as well as the main concerns raised by policymakers and competition authorities.

It is noteworthy that competition authorities have recognised the potential for pro-competitive benefits from partnerships in the GenAI space, including their capability to:

- **Enhance product quality**: Partnerships can combine complementary skills and assets *"which may result in the issuing of a better or new product or technology that would not otherwise come to light"*.[382] For example, closer collaboration can facilitate improved compatibility between inputs across a vertical value chain and lead to better or new products.
- **Increase innovation**: Partnerships can enable better distribution of *"technological expertise across the market, which may lead to further innovation"*.[383] By close collaboration, teams could share knowledge and expertise which each other and reduce uncertainty, which may promote more and better innovation.
- **Address shortages**: Partnerships can reduce vulnerability to shortages of certain inputs for the party involved. The European Commission has stated that partnerships may reduce *"costs or dependencies when supply of a specific input is limited, which may increase supply and strengthen the internal market"*.[384] Notably, this might be especially relevant for AI Chips, which face supply shortages.
- **Help smaller partners to establish themselves**: Partnerships can help smaller partners to *"boost credibility, provide funding, and compete effectively"*.[385] For instance, by partnering with a large, known organisation, a smaller GenAI firm could become more recognised in the industry, and, as a result, access funding and increase its potential to reach customers.

---

381 See Competition & Markets Authority (2023a); European Commission (2024a); and OECD (2024); https://www.autoritedelaconcurrence.fr/en/press-release/generative-artificial-intelligence-autorite-issues-its-opinion-competitive; https://www.concorrencia.pt/en/articles/adc-warns-competition-risks-generative-artificial-intelligence-sector; https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-issues-staff-report-ai-partnerships-investments-study (all accessed in May 2025).

382 See European Commission (2024a).

383 See footnote 382 above.

384 See footnote 382 above.

385 See https://competition-bureau.canada.ca/en/how-we-foster-competition/education-and-outreach/consultation-artificial-intelligence-and-competition-what-we-heard (accessed in May 2025).

At the same time, authorities have also raised concerns that partnerships could be used to restrict competition in GenAI. The most prominent concerns in this respect are:[386]

- **Leveraging conduct**: Partners may adopt strategies that leverage the market power of one partner in an established market (e.g. in cloud services) to gain advantages in emerging ones, such as development or deployment of GenAI. Such strategies could include restricting access to inputs (as described below in detail), limiting interoperability with competing products, engaging in self-preferencing, bundling, or tying products in a way that strengthen their market position by foreclosing competitors.

- **Input foreclosure**: The GenAI market relies on critical inputs such as data, skilled talent, and computing power that, while resource-intensive, are becoming increasingly accessible. These inputs are essential for developing FMs, which in turn become critical components for deployment into real-world applications. One concern is that a partnership may distort partners' incentives to supply these key inputs to other market participants competing with their partner. For instance, following a partnership, the upstream partner might reduce the provision of these inputs to rivals or offer them under less favourable terms to favour its downstream partner's position in the GenAI market.

- **Switching costs**: Partnerships may increase switching costs that affect a downstream partner's ability or incentive to use the services of competing upstream providers, other than the partner. Such switching costs can materialise through structural and technical means. Structural switching costs may be embedded in the terms and conditions of a partnership, such as "buy commitments" that impose a minimum spend on the upstream partner's services. Technical switching costs may stem from reliance on partner-specific technologies, such as proprietary software or specialised hardware.

- **Information asymmetry and competitive intelligence**: Partnerships may provide established players with privileged access to sensitive technical and business information about emerging technologies and markets, creating information asymmetries that disadvantage other competitors. For example, this information might include GenAI customer usage metrics, model development methodologies, chip design suggestions, and product development recommendations.

So far, no enforcement action has been taken to date against any specific partnership arrangements, which might suggest the absence of materialised harm.[387] The Microsoft-Mistral AI partnership, for example, was reviewed by the CMA, which concluded in May 2024 that it did not qualify for investigation under merger provisions.

It is important to acknowledge that while concerns exist, based on economic theory, they must be carefully assessed on a case-by-case basis and weighed against the potential benefits that partnerships can bring. Moreover, the general trends presented in section 2 as well as observations from deployment in section 4 point towards a market in which competition is currently functioning effectively, with numerous recent examples of successful entry and expansion by firms using a variety of business models. Thus, whilst the GenAI market is still very new and the picture could evolve again going forward, the current evidence suggests that many partnerships are contributing positively to competition.

386 The list has been derived based on RBB's review of relevant authorities and policymakers' studies. See Competition & Markets Authority (2023a); European Commission (2024a); OECD (2024); https://www.autoritedelaconcurrence.fr/en/press-release/generative-artificial-intelligence-autorite-issues-its-opinion-competitive; https://www.concorrencia.pt/en/articles/adc-warns-competition-risks-generative-artificial-intelligence-sector (all accessed in May 2025).

387 Such investigations include, for instance, the UK's CMA assessments of the partnerships Microsoft/OpenAI (2023) and Amazon/Anthropic (2024). See Competition & Markets Authority (2023b); Competition & Markets Authority (2024c).

6 Policy considerations and conclusions

European GenAI markets are currently characterised by significant innovation and competition. Our case studies point towards a variety of deployment strategies, partnership structures, and business models, and provide examples of a number of European companies successfully taking advantage of the opportunities offered by GenAI deployments to improve efficiency in productive processes. Policy should focus on supporting this momentum while ensuring markets remain open and contestable.

In this section, we set out policy recommendations that draw on our analysis of the GenAI market. In doing so, we highlight the role of active monitoring as the market continues to evolve to ensure competition is not hindered. We emphasise the importance of acting with caution in relation to any competition interventions, noting that the risks of intervening too early currently appear to outweigh the risks of letting the market develop. We also consider how policy can support European industries in making the most of this transformative technology and highlight some important considerations for policy design.

6.1 Competition policy recommendations

Our analysis reveals European GenAI markets characterised by a variety of business models, multi-sourcing strategies, and relatively low switching costs between providers. These conditions suggest that markets are currently functioning well, with no evidence of the market tipping or foreclosure effects that would justify immediate intervention. Our competition policy recommendations therefore focus on preserving these positive dynamics while avoiding premature intervention that could disrupt effective competitive processes and chill innovation.

Maintain active monitoring with clear intervention thresholds based on evidence of actual concerns.

Competition authorities are right to monitor GenAI markets given their transformative potential for the European economy. This oversight serves important functions: building expertise in rapidly evolving technologies, understanding competitive dynamics, and positioning authorities to act quickly if genuine concerns emerge.

However, our analysis suggests that risks of premature intervention remain substantial. Regulatory intervention in any nascent market has high potential for unintended consequences, including chilling innovation or shaping the market in a way that is beneficial to incumbents, at the expense of smaller entrants.[388] We find no current evidence of market "tipping" or factors likely to cause it, such that preventive actions may be warranted. Instead, our case studies point to significant multi-homing by deployers and relatively low switching costs between models.

This balance of risk between early and late intervention favours active monitoring by competition authorities of the markets in questions. This means that intervention should be reserved for situations where there is clear evidence of actual concerns: there is no current need to act pre-emptively against potential (hypothetical) concerns.

Continue to recognise the pro-competitive potential of partnerships.

Our analysis of selected European GenAI partnerships provides evidence of significant benefits from such arrangements. Partnerships like Stellantis-Mistral AI demonstrate how European companies can take advantage GenAI capabilities while maintaining strategic autonomy, while the deployment partnerships

[388] See, for instance, p. 120–129 in Thierer (2014).

we have considered show how such arrangements can help smaller European GenAI developers to reach enterprise customers through established channels. At the same time, we understand that no authorities have found any evidence of harm to competition resulting from the partnerships they have reviewed.

Current vertical block exemption regulations and merger control frameworks provide appropriate analytical tools for assessing GenAI partnerships. Authorities should focus on actual rather than theoretical harm, recognising that these collaborations often address legitimate business needs including risk-sharing, capability complementarity, and market access acceleration.

Restricting beneficial partnerships, particularly those involving non-European investment in European GenAI companies, risks undermining European competitiveness during a critical scaling phase.

Exercise caution in extending *ex ante* regulation to GenAI services, unless there is clear evidence of specific harms.

Some competition authorities, including the Commission, have raised concerns that large digital firms may leverage their strengths in existing markets into new GenAI markets. In response, it has been suggested that *ex ante* regulation (including the DMA) could be applied to GenAI products to mitigate against such behaviour.

Our examination of European deployment cases provides no evidence of anti-competitive leveraging by existing digital platforms. Instead, we observe effective competition between integrated and specialised providers. Notion's success against Microsoft and Google's productivity suites, and DeepL's expansion from translation into broader language workflows, illustrate that specialised innovation can overcome distribution advantages when value propositions are compelling.

Given significant efficiency benefits from allowing firms to integrate GenAI into existing products, documented across our case studies from enterprise platforms to consumer applications, authorities should exercise caution before applying *ex ante* obligations to GenAI features. Clear guidance specifying that such obligations will apply only where there is concrete evidence of harm would provide necessary regulatory certainty for continued innovation. If not, there is a risk that designated gatekeepers will be disincentivised from integrating GenAI into their established digital services due to the threat of regulation. This could lead to loss of efficiency benefits from vertical integration and would eliminate a potentially important competitive force in the market.

Monitor interoperability developments and support market-based solutions that preserve competitive flexibility.

Our analysis suggests that the currently observed competitive dynamism in GenAI deployment is driven in part by low switching costs and multi-homing behaviour. The sustainability of current competitive dynamics relies on continued FM competition, effective interoperability standards adoption, platform competition, and preservation of deployment flexibility, as discussed above in section 4.4.4.

The multi-sourcing strategies documented across our case studies, from Goldman Sachs's multi-model platform to Estée Lauder's selective sourcing or NotionAI's use of multiple models at the same time, benefit from having access to multiple viable FM providers as well as low switching costs. If these strategies were no longer available, or harder to implement, this would have clear implications for competitive dynamics.

Similarly, emerging market-based solutions for GenAI interoperability, including protocols like Anthropic's Model Context Protocol and Google's Agent-to-Agent standard (see section 4.1.2.2), help promote

competition by reducing technical barriers that could otherwise favour incumbent platforms. Competition authorities should monitor these developments and consider intervention only if proprietary barriers emerge that create substantial foreclosure effects or systematically prevent competitive alternatives from reaching users.

Where market-based interoperability solutions support competitive variety, as our analysis suggests they currently do, authorities should avoid intervention. However, if technical barriers begin to emerge that prevent effective competition and harm consumers, authorities could consider measures to ensure continued access to the interoperability tools that enable the deployment flexibility documented across our case studies.

Avoid creating additional layers of competition-specific obligations for AI developers and deployers

In the US, the House of Representatives recently passed a 10-year moratorium on state AI laws, with supporters arguing this would prevent *"a confusing patchwork of state AI laws"* and preserve innovation leadership. This reflects broader concerns about regulatory complexity constraining technological development.[389] Recent indications that the Commission is considering pausing certain AI Act provisions reflect growing recognition of implementation challenges and concerns that complex regulatory requirements may burden innovation and competitiveness.[390] Where possible, competition authorities should seek to align their approach to regulation with this simplification effort, and avoid adding additional layers of oversight unless there is clear evidence of harm to address.

The current regulatory framework, including GDPR and emerging AI Act requirements, already addresses many potential concerns about GenAI deployment. Additional competition-specific rules risks creating overlapping obligations that could particularly burden European companies relative to global competitors operating under different regulatory regimes.

6.2 Broader policy recommendations

Our competition policy recommendations focus on preserving the positive competitive dynamics we have reported in this report. However, competition policy alone cannot ensure Europe maximises the opportunities GenAI presents. The following broader policy recommendations address structural challenges and opportunities for European industries to build on their existing strengths through effective GenAI deployment.

Address barriers to private capital access for European GenAI companies.

Our analysis reveals a persistent funding gap constraining European GenAI companies' ability to scale globally (see section 2.2.2). While private GenAI investment reached $56 billion globally in 2024, European companies consistently struggle to raise late-stage capital compared to US counterparts.[391] This affects not only individual company growth but European market positioning during a critical competitive phase.

Policy should address regulatory and structural barriers that limit European GenAI companies' access to growth capital from both domestic and international sources. This includes completing the Capital Markets Union initiative to deepen European capital markets, removing barriers to cross-border venture capital

389 See https://www.techpolicy.press/us-house-passes-10year-moratorium-on-state-ai-laws/ (accessed in June 2025).
390 See https://www.mlex.com/mlex/articles/2344845/eu-commission-eyes-pausing-ai-act-s-entry-into-application; https://sifted.eu/articles/eu-ai-act-pause-analysis (both accessed in June 2025).
391 See S&P Global Market Intelligence data, cited in section 2.1.3, footnote 50.

investment, and ensuring that regulatory frameworks do not inadvertently discourage investment in European GenAI startups.[392]

This is particularly important for the deployment-focused companies that our analysis shows are attracting significant investor interest globally, but which require substantial funding to scale across Europe's fragmented markets. Facilitating access to private capital is essential for European GenAI companies to compete during this critical scaling phase.

Target public funding toward areas where market failures give rise to under-investment

Our findings suggest public funding should aim to complement rather than compete with private investment flows (see section 2.1.3). Private capital naturally targets commercially viable applications and enterprise deployments where returns are clearer. Public investment could focus on areas where private markets alone may under-invest, particularly where projects generate significant social benefits that cannot be fully captured by private investors.

Areas for consideration include foundational research with longer-term horizons; open-source initiatives that benefit the European ecosystem broadly; multilingual and multicultural GenAI capabilities that serve European market needs; and infrastructure investments (computing capacity, data platforms) that support the European startup ecosystem identified in our analysis (see section 2.2.1).

The InvestAI initiative's €200 billion mobilisation represents recognition of these needs, but implementation should ensure additionality rather than crowding out private investment in commercially viable applications.[393]

Develop policies that support and promote the efficient adoption of GenAI deployments

Our case studies of Stellantis, AstraZeneca, and LVMH (see section 4.3) demonstrate how European companies can achieve measurable benefits by applying GenAI to their existing industrial expertise. These deployment successes point toward broader opportunities across European manufacturing, healthcare, automotive, and financial services sectors.

Policy could support this deployment momentum through approaches such as sector-specific guidance on regulatory compliance; technical assistance programs helping SMEs navigate GenAI integration; and initiatives that leverage European strengths in developing GenAI standards for interoperability and trustworthiness that can become competitive advantages globally.

Such initiatives would build on the competitive advantages our analysis identifies: European industrial expertise, established cross-border collaboration networks, and deep sector knowledge (see section 2.2.3), rather than attempting to replicate approaches developed for different market contexts.

Establish EU-wide coordination mechanisms for consistent GenAI policy implementation.

The success of European GenAI companies depends partly on their ability to scale across the single market. Our analysis shows European startups face fragmentation challenges that US and Chinese competitors do not encounter in their domestic markets (see section 2.2.2).

---

[392] See https://ec.europa.eu/info/business-economy-euro/growth-and-investment/capital-markets-union_en; https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/capital-markets-union/capital-markets-union-2020-action-plan_en; https://finance.ec.europa.eu/publications/mid-term-review-capital-markets-union-action-plan_en (all accessed in June 2025).

[393] See https://digital-strategy.ec.europa.eu/en/news/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence (accessed in June 2025).

Because GenAI is such a fast-developing and dynamic area, consistent implementation requires ongoing coordination informed by the latest industry insights. An EU-wide forum for GenAI policy coordination could ensure that key stakeholders contribute to implementation approaches, helping avoid unintended consequences from divergent national interpretations while preserving single market benefits.

Such coordination is particularly important given the Commission's current work on AI Act simplification, where industry insights can help ensure implementation approaches support rather than hinder the competitive dynamics our analysis reveals, while ensuring European companies can develop compliance expertise that becomes a competitive advantage globally.

Road test remedies and/or explore softer regulatory approaches where concerns arise.

In domains where regulation is required, experimental regulatory frameworks can play a role in balancing innovation and oversight. Such frameworks can allow market participants to trial novel technologies and business models in a supervised, low-risk environment, providing regulators with valuable insights before implementing rules with permanent impact.

Policymakers should also consider a more participatory approach to regulation where it finds evidence of concerns requiring intervention, in which industry itself plays a key role in delivering solutions to the concerns in question. This can assist in reducing the regulatory burden and minimising risks of unintended consequences from policy interventions.

Given the rapid evolution of GenAI markets documented in our analysis, such flexible approaches may be more effective than prescriptive rules that risk becoming obsolete or counterproductive as technologies and business models continue to evolve.

These recommendations reflect the competitive dynamics our analysis reveals: a diverse European GenAI ecosystem that benefits from partnerships, deploys varied strategies, and leverages existing industrial strengths. Policy should support this momentum while ensuring markets remain open to continued innovation and competition.

7 Bibliography

Adams, W.J. and Yellen, J.L. (1976), Commodity Bundling and the Burden of Monopoly, *Quarterly Journal of Economics*, Vol. 90, No. 3, p. 475–498, available at: https://academic.oup.com/qje/article-abstract/90/3/475/1852397 (accessed in June 2025)

Autorité de la concurrence and Competition and Markets Authority (2014), The economics of open and closed systems, available at: https://assets.publishing.service.gov.uk/media/5a75c284e5274a545822e01a/The_economics_of_open_and_closed_systems.pdf (accessed in June 2025)

Batra, M., de Bijl, P. and Klein, T. (2024), Ecosystem Theories of Harm in EU Merger Control, Analyzing Competitive Constraints and Entrenchment, *Journal of European Competition Law & Practice*, Vol. 15, No. 6, pp. 357–378, available at: https://academic.oup.com/jeclap/article/15/6/357/7810622 (accessed in June 2025)

Bedre-Defolie, Ö. and Nitsche, R. (2020), When Do Markets Tip? An Overview and Some Insights for Policy, *Journal of European Competition Law & Practice*, Vol. 11, No. 10, pp. 610–621, available at: https://academic.oup.com/jeclap/article/11/10/610/6040005 (accessed in June 2025)

Bishop, S. and Walker, M. (2010), The Economics of EC Competition Law: Concepts, Application and Measurement, 3rd ed., Sweet & Maxwell

Caffarra, C., Crawford, G. and Ryan, J. (2023), 'Ecosystem' theories of harm in digital mergers: New insights from network economics, part 1, *VoxEU*, CEPR, available at: https://cepr.org/voxeu/columns/ecosystem-theories-harm-digital-mergers-new-insights-network-economics-part-1 (accessed in June 2025)

Competition & Markets Authority (2023a), AI Foundation Models: Initial Report, available at: https://assets.publishing.service.gov.uk/media/65081d3aa41cc300145612c0/Full_report_.pdf (accessed in June 2025)

Competition & Markets Authority (2023b), Microsoft/OpenAI partnership merger inquiry, available at: https://www.gov.uk/cma-cases/microsoft-slash-openai-partnership-merger-inquiry (accessed in June 2025)

Competition & Markets Authority (2024a), AI Foundation Models: Technical Update Report, available at: https://assets.publishing.service.gov.uk/media/661e5a4c7469198185bd3d62/AI_Foundation_Models_technical_update_report.pdf (accessed in June 2025)

Competition & Markets Authority (2024b), AI Foundation Models: Update Paper, available at: https://assets.publishing.service.gov.uk/media/661941a6c1d297c6ad1dfeed/Update_Paper__1_.pdf (accessed in June 2025)

Competition & Markets Authority (2024c), Amazon/Anthropic partnership merger inquiry, available at: https://www.gov.uk/cma-cases/amazon-slash-anthropic-partnership-merger-inquiry (accessed in June 2025)

Competition & Markets Authority (2024d), Amazon.com Inc.'s partnership with Anthropic PBC: Decision on relevant merger situation, available at: https://assets.publishing.service.gov.uk/media/676959bae6ff7c8a1fde9d33/Full_text_decision__.pdf (accessed in June 2025)

Cottier, B., et al. (2024), The Rising Costs of Training Frontier AI Models, available at: https://arxiv.org/pdf/2405.21015 (accessed in June 2025)

Curry, K. and Hill, J. (2024), GenAI – Some Reflections on the CMA's Update Paper, available at: https://www.linkedin.com/posts/rbb-economics_genai-some-reflections-on-the-cmas-update-activity-7188090563118649346-lc1p/ (accessed in June 2025)

Department for Science, Innovation, & Technology (2025), AI opportunities action plan, available at: https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan (accessed in June 2025)

Department for Science, Innovation & Technology (2025), International AI Safety Report 2025, available at: https://www.gov.uk/government/publications/international-ai-safety-report-2025/international-ai-safety-report-2025 (accessed in June 2025)

Draghi, M. (2024), The Future of European Competitiveness, European Commission, available at: https://commission.europa.eu/document/download/97e481fd-2dc3-412d-be4c-f152a8232961_en?filename=The%20future%20of%20European%20competitiveness%20_%20A%20competitiveness%20strategy%20for%20Europe.pdf (accessed in June 2025)

Economic Round Table (2025), Harnessing Generative AI for Europe's Industrial Future, available at: https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf (accessed in June 2025)

Economides, N. (1996), Network externalities, complementarities, and invitations to enter, *European Journal of Political Economy*, Vol. 12, No. 2, pp. 211–233, available at: https://www.sciencedirect.com/science/article/pii/0176268095000143 (accessed in June 2025)

Economides, N. (1998), The Incentive for Non-Price Discrimination by an Input Monopolist, *International Journal of Industrial Organization*, 16(3), pp. 271–284, available at: https://www.sciencedirect.com/science/article/pii/S0167718798000058 (accessed in June 2025)

European Economic and Social Committee (2021), Boosting the Use of Artificial Intelligence in Europe's Micro, Small and Medium-sized Enterprises, available at: https://www.eesc.europa.eu/sites/default/files/files/qe-02-21-953-en-n.pdf (accessed in June 2025)

European Economic and Social Committee (2025), Generative AI and Foundation Models in the EU: Uptake, Opportunities, Challenges, and a Way Forward, available at: https://www.eesc.europa.eu/sites/default/files/2025-03/QE-01-25-014-EN-N_0.pdf (accessed in June 2025)

European Commission (2008a), Communication from the Commission – Guidance on the Commission's enforcement priorities in applying Article 82 of the EC Treaty to abusive exclusionary conduct by dominant undertakings, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=oj:JOC_2009_045_R_0007_01 (accessed in June 2025)

European Commission (2008b), Guidelines on the assessment of non-horizontal mergers under the Council Regulation on the control of concentrations between undertakings, paragraphs 93–100, available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=oj:JOC_2008_265_R_0006_01 (accessed in June 2025)

European Commission (2017), Google Search (Shopping), (Case AT.39740), available at: https://competition-cases.ec.europa.eu/cases/AT.39740 (accessed in June 2025)

European Commission (2023), Booking Holdings/eTraveli Group, (Case M.10615), available at: https://competition-cases.ec.europa.eu/cases/M.10615 (accessed in June 2025)

European Commission (2024a), Competition Policy Brief. Competition in Generative AI and Virtual Worlds, available at: https://competition-policy.ec.europa.eu/document/download/c86d461f-062e-4dde-a662-15228d6ca385_en (accessed in June 2025)

European Commission (2024b), Seven consortia selected to establish AI factories which will boost AI innovation in the EU, available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_6302 (accessed in June 2025)

European Commission (2025a), AI Continent Action Plan, available at: https://digital-strategy.ec.europa.eu/en/library/ai-continent-action-plan (accessed in June 2025)

European Commission (2025b), Commission sets course for Europe's AI leadership with an ambitious AI Continent Action Plan, available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_25_1013 (accessed in June 2025)

European Commission (2025c), President von der Leyen launches initiative to mobilise €200 billion for AI investment, available at: https://ec.europa.eu/commission/presscorner/detail/en/ac_25_481 (accessed in June 2025)

European Court of Auditors (2024), EU Artificial intelligence ambition. Stronger governance and increased, more focused investment essential going forward, available at: https://www.eca.europa.eu/ECAPublications/SR-2024-08/SR-2024-08_EN.pdf (accessed in June 2025)

European Parliament (2024), Investment in artificial intelligence in the National Recovery and Resilience Plans, available at: https://www.europarl.europa.eu/thinktank/fi/document/EPRS_BRI(2024)762288 (accessed in June 2025)

Farrell, J. and Saloner, G. (1985), Standardization, Compatibility and Innovation, *The RAND Journal of Economics*, Vol. 16, No. 1 (Spring 1985), pp. 70–83, available at: https://www.jstor.org/stable/2555589 https://ideas.repec.org/a/rje/randje/v16y1985ispringp70-83.html (accessed in June 2025)

Farrell, J. and Shapiro, C. (1986), Dynamic competition with Switching costs, *The RAND Journal of Economics*, Vol. 19, No. 1 (Spring, 1988), pp. 123-137, available at: https://www.jstor.org/stable/2555402 (accessed in June 2025)

Federal Trade Commission (2025), Partnerships Between Cloud Service Providers and AI Developers, available at: https://www.ftc.gov/system/files/ftc_gov/pdf/p246201_aipartnerships6breport_redacted_0.pdf (accessed in June 2025)

Grossman, S.J. and Hart, O.D. (1986), The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration, *Journal of Political Economy*, Vol.94. No.4, 691-719, available at: https://www.journals.uchicago.edu/doi/abs/10.1086/261404 (accessed in June 2025)

Hart, O. and Tirole, J. (1990), Vertical Integration and Market Foreclosure, *Brookings Papers on Economic Activity: Microeconomics*, pp. 205–286, available at: https://www.jstor.org/stable/2534783 (accessed in June 2025)

Katz, M. and Shapiro, C. (1985), Network Externalities, Competition and Compatibility, *American Economic Review*, Vol. 75, No. 3, pp. 424–440, available at: https://www.jstor.org/stable/1814809 (accessed in June 2025)

Liu, Q. (2016), Vertical integration and innovation, *International Journal of Industrial Organization*, Vol. 47, pp. 88–120, available at: https://www.sciencedirect.com/science/article/abs/pii/S0167718716300029 (accessed in June 2025)

Markets and Markets. (2025), Generative AI Market, available at: https://www.marketsandmarkets.com/Market-Reports/generative-ai-market-142870584.html (accessed in June 2025)

Masson, R.T., Dalkir, S. and Eisenstadt, D. (2014), A Note on Complementary Goods Mergers between Oligopolists with Market Power: Cournot Effects, Bundling and Antitrust, *Review of Law & Economics*, Vol. 10, No. 1, pp. 59–79, available at: https://www.degruyterbrill.com/document/doi/10.1515/rle-2013-0014/html (accessed in June 2025)

Matutes, C., and Regibeau, P. (1996), *A selective review of the economics of standardization. Entry deterrence, technological progress and international competition*, *European Journal of Political Economy*, Vol. 12, No. 2, pp. 183–209, available at: https://www.sciencedirect.com/science/article/abs/pii/0176268095000135 (accessed in June 2025)

McAfee, R.P., McMillan, J. and Whinston, M.D. (1989), Multiproduct Monopoly, Commodity Bundling, and Correlation of Values, *Quarterly Journal of Economics*, Vol. 104, No. 2, pp. 371–383, available at: https://academic.oup.com/qje/article-abstract/104/2/371/1854649 (accessed in June 2025)

McAfee, R. P., Mialon, H. M., and Williams, M. A. (2003), Economic and antitrust barriers to entry, *The American Economic Review*, 94, pp.1-26, available at: https://mc4f.ee/Papers/PDF/Barriers2Entry.pdf, (accessed in June 2025)

OECD (2007), Vertical Mergers, DAF/COMP(2007)21, available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2007/11/vertical-mergers_d598244f/b6cf1dcb-en.pdf (accessed in June 2025)

OECD (2024), Artificial Intelligence, Data and Competition, available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf (accessed in June 2025)

Pehrsson, P. (2009), Barriers to entry and market strategy: a literature review and a proposed model, *European Business Review*, Vol.21, No(1), 64-77, available at: https://www.emerald.com/insight/content/doi/10.1108/09555340910925184/full/html (accessed in June 2025)

Rey, P. and Tirole, J. (2007), A Primer on Foreclosure, in Handbook of Industrial Organization, Vol. 3, Ch. 33, pp. 2145–222, available at: https://publications.ut-capitole.fr/id/eprint/1017/1/primer.pdf (accessed in June 2025)

Salop, S.C. and Scheffman, D.T. (1983), Raising Rivals' Costs, *American Economic Review*, 73(2), pp. 267–271, available at: https://www.jstor.org/stable/1816853 (accessed in June 2025)

Snell, J. et al. (2024), Scaling LLM Test-Time Compute Optimally can be More Effective than Scaling Model Parameters, *arXiv preprint*, available at: https://arxiv.org/html/2408.03314v1 (accessed in June 2025)

Spengler, J.J. (1950), Vertical Integration and Antitrust Policy, Journal of Political Economy, Vol. 58, No. 4, pp. 347–352, available at: https://www.journals.uchicago.edu/doi/abs/10.1086/256964 (accessed in June 2025).

Stanford Institute for Human-Centered Artificial Intelligence (2025), Artificial Intelligence Index Report 2025, available at: https://aiindex.stanford.edu/report (accessed in June 2025)

Thierer, A. D. (2014), Permissionless Innovation, The Continuing Case for Comprehensive Technological Freedom, *Mercatus Center at George Mason University*, available at: https://www.mercatus.org/system/files/Thierer-Permissionless-revised.pdf (accessed in June 2025)

Weik, S., Achleitner, A. K., Braun, R. (2024), Venture capital and the international relocation of startups, *Research Policy*, Vol.53, No(7),105031, available at: https://www.sciencedirect.com/science/article/pii/S0048733324000805 (accessed in June 2025)

Williamson, O.E. (1981), The Economics of Organization: The Transaction Cost Approach, *American Journal of Sociology*, Vol.87, No(3), pp.548–577, available at: https://www.journals.uchicago.edu/doi/abs/10.1086/227496 (accessed in June 2025)