ECIPE | EUROPEAN CENTRE FOR INTERNATIONAL POLITICAL ECONOMY

ECIPE Policy Brief – No. 09/2026

The Copyright Trap

By **Oscar Guinea** *(Director),* **Dyuti Pandya** *(Analyst); and* **Vanika Sharma** *(Economist), at ECIPE*

EXECUTIVE SUMMARY

The current debate on EU copyright law and AI training must move beyond the creative sector to include the wider European economy. While the rights of creators are legitimate and must be protected, the rules governing data use for AI models have far-reaching economic consequences. At its core, this is a debate about Europe's economic competitiveness and its ability to reverse the technological stagnation identified in the Draghi report.

The EU's current approach to AI is marked by a profound policy paradox that helps explain why its private AI sector is lagging behind the US and China. While the EU invests billions of euros in public funds to build world-class supercomputing and AI-related infrastructure, its restrictive digital regulations constrain the data needed for its private AI economy to flourish. These regulatory restrictions deprive European start-ups of essential inputs such as data, undermining the development and deployment of AI in the EU.

EU policymakers must carefully weigh the economic trade-offs of making copyright law more restrictive. The European creative industries are undeniably significant, generating

With research assistance from Ismail Abdi, Research Assistant at ECIPE

€202 billion in value added. However, the combined value added of the rest of the economy is far larger. To offset just a 1 per cent reduction in value added across data-intensive sectors such as ICT, financial services, automotive, pharmaceuticals and chemicals the creative industries would need to increase output by more than 9 per cent.

The EU cannot subsidise its way out of a self-imposed data scarcity problem. Instead, policymakers should preserve the current copyright framework as a genuine asset for AI development and work towards making data an abundant economic resource. Such an approach would not only support a thriving domestic AI economy capable of developing the models that will power and transform Europe's industrial and service sectors, but also contribute to achieving the strategic autonomy sought by EU policymakers. This brief sets out six targeted policy recommendations to achieve these objectives:

1. Treat the Text and Data Mining (TDM) exception as a strategic asset: the EU should recognise the TDM exception as central to its technological competitiveness and strategic autonomy. Policymakers should subject any proposals that would narrow this exception to a mandatory economic impact assessment.

2. Contain the opt-out mechanism: expanding it risks rendering the TDM framework unworkable. The EU should resist broadening the categories of content subject to opt-out and maintain the requirement for affirmative, machine-readable reservations of rights.

3. Stabilise the legal framework: to reduce regulatory uncertainty, the EU should avoid further legislative changes to the TDM framework. The EU's digital rulebook is already undergoing significant change. Recently adopted digital regulations should be allowed to reach operational maturity before new compliance obligations are introduced, and efforts to simplify EU digital rules should be implemented before any changes to EU copyright law are considered.

4. Exercise caution on top-down licensing: policymakers should reject statutory mandatory licensing regimes for AI training data, which disproportionately harm start-ups. The EU should allow private markets to develop licensing solutions, intervening only where clear market failures emerge.

5. Reframe the policy problem: the debate surrounding generative AI is fundamentally about market competition and economic displacement, rather than copyright infringement. Policy responses should therefore focus on market adaptation rather than expanding copyright protection.

6. Learn from innovation-first jurisdictions: the EU should draw lessons from countries such as Singapore and Japan which implement broad, technologically neutral exceptions for computational data analysis and actively attract AI investment by treating data as a vital economic endowment.

1. INTRODUCTION

EU copyright rules are coming under increasing pressure. On 10 March 2026, the European Parliament's Committee on Legal Affairs (JURI) adopted a report signalling a push towards tighter restrictions on the use of copyrighted content in AI model training.[1] At the same time, the Court of Justice of the European Union (CJEU) is set to rule in Like Company v Google Ireland (C-250/25),[2] its first major case directly addressing the intersection of generative AI (GenAI) – the type of AI systems capable of producing text, images, code, and other outputs from large-scale training data – and copyright. Meanwhile, the European Commission launched an open consultation on technical protocols for the opt-out reservation of rights under Article 4(3) of the Copyright in the Digital Single Market Directive (CDSMD).[3]

At the heart of the issue is the Text and Data Mining (TDM) exception. Established under the CDSMD, it arguably provides the legal foundation for training AI models, including the collection and analysis of the datasets required for that purpose.[4] This open access is critical for European developers and start-ups, as it lowers transaction costs and barriers to entry, making the EU regulatory framework more innovation-friendly than in other jurisdictions.[5] However, this balance could easily shift towards a more restrictive regime, depending on upcoming court rulings and the policy developments mentioned before.

The push for tighter rules is driven by a central concern. Proponents of tighter restrictions emphasise the need to protect creators' livelihoods, arguing that technology companies should not monetise copyrighted works without fair compensation. However, while these concerns are legitimate, the core issue is one of economic displacement rather than conventional copyright infringement.

Translating these economic anxieties into restrictive copyright law relies on a flawed premise: that AI systems act like traditional databases storing and retrieving protected content. In reality, AI training relies on statistical inference rather than duplication. During training, original works are not preserved as human-readable form; instead, they serve as raw material. AI systems extract unprotected elements, such as factual information, linguistic regularities and abstract relationships, and transform them into mathematical parameters, weights and probabilistic associations. The resulting model generates new outputs based on underlying causal and mathematical structures, rather than reproducing the authorship of its training inputs.

---

1 European Parliament, Committee on Legal Affairs. (2026, February 25). Report on copyright and generative artificial intelligence – opportunities and challenges (Document No. A10-0019/2026). Available at: https://www.europarl.europa.eu/doceo/document/A-10-2026-0019_EN.html

2 Like Company v. Google Ireland, Case C-250/25 (Court of Justice of the European Union 2025). Available at: https://curia.europa.eu/juris/liste.jsf?num=C-250/25

3 European Commission. (2025, December 1). Commission launches consultation on protocols for reserving rights from text and data mining under the AI Act and the GPAI Code of Practice. Shaping Europe's digital future. Available at: https://digital-strategy.ec.europa.eu/en/consultations/commission-launches-consultation-protocols-reserving-rights-text-and-data-mining-under-ai-act-and

4 This understanding is also reflected in the Artificial Intelligence Act, which, as enacted law, assumed that AI training may involve the use of copyright-protected content and refers to the applicability of TDM exceptions, subject to rightsholders' ability to reserve their rights (Recital 105).

5 Most EU countries have open research exceptions. Peukert, C Copyright Atlas. Swiss National Science Foundation Project. Available at: https://www.copyrightatlas.com/

Understanding this process helps clarify the application of two foundational principles of copyright law to this discussion: the distinction between synthesis and duplication, and the idea-expression dichotomy. Copyright law seeks to prevent duplication, such as illegally photocopying a novel or distributing a pirated film, because it preserves the original expressive form and acts as a direct market substitute. By contrast, synthesis extracts generalised statistical relationships without retaining protected expression. The idea-expression dichotomy holds that copyright protects only concrete expressions, not underlying facts or concepts. Because AI training breaks data down into mathematical patterns, rather than preserving or communicating individual expressive works, the resulting model captures the unprotectable elements of its inputs rather than reproducing their protected expressions.

This fundamental distinction between expressive content and statistical structure exposes the flaws in analogies often used in political discourse. For instance, Axel Voss, a Member of the European Parliament and the lead rapporteur of the report mentioned above, argued that "just as a musician is paid when their song is used in a film, creators should be compensated when their work is used to train AI".[6] This comparison is misleading, as it rests on an analogy that conflates statistical model training with the reproduction of protected works, an error that extends copyright concepts beyond their intended scope.[7] Using a song in a film involves the reproduction of a protected expression. By contrast, AI training does not duplicate the work; it analyses the inputted data to learn the underlying rules of language, structure, and style. In this way, it does not reproduce the expressive contents of the work.

Demanding compensation for this process risks transforming copyright law into a tool for controlling learning itself, moving beyond its traditional function of protecting original expression. Extending protection this far would be akin to a novelist demanding royalties from anyone who learns the rules of grammar or narrative form. Copyright protects the expression of a work, not the underlying linguistic patterns used to create it.

While the copyright debate has historically centred on the creative industries, the stakes are far broader. Because AI functions as a general-purpose technology, the regulatory frameworks governing data impact not only the future of the European creative industries, but also the baseline competitiveness of the entire European economy. AI applications are improving processes across the economy and accelerating the development of new scientific applications, products and services. Any changes to the framework governing data access for AI training will therefore have far-reaching implications for the competitiveness of the European economy.

This policy brief introduces an economic perspective to broaden the copyright debate, reframing data as a critical economic resource that underpins industrial competitiveness. Chapter 2 examines the EU's policy paradox: it subsidises AI development while simultaneously limits the

---

[6] European Parliament. (2026). REPORT on copyright and generative artificial intelligence – opportunities and challenges. 2025/2058(INI). Available at: https://www.europarl.europa.eu/doceo/document/A-10-2026-0019_EN.html; also see: Voss, A. (2026). Post on generative AI and copyright. LinkedIn. Available at: https://www.linkedin.com/feed/update/urn:li:activity:7437134296504766464/

[7] See: Larroyed, A. (2025). The fallacy of the file: How the memorisation metaphor misguides copyright law and stifles AI innovation, where the author demonstrates how metaphor-based comparisons in AI and copyright debates can distort legal reasoning and lead to flawed public policy by conflating technical processes with copyright concepts. Available on SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5782882

growth of its AI industry. Chapter 3 explains how changes to EU copyright law could affect the EU's competitiveness, while Chapter 4 examines the disproportionate impact on small and medium-sized enterprises (SMEs). Finally, Chapter 5 sets out policy recommendations to ensure regulatory balance, protecting rightsholders while ensuring that Europe's industrial and start-up ecosystems have access to the data they need to thrive.

2. THE PARADOX OF EUROPEAN AI POLICY

2.1 The Accelerator: EU Public Investment

There is a structural contradiction at the centre of the EU's AI policy. On the one hand, the EU actively promotes AI development through public investment. On the other hand, its legal framework introduces uncertainty and restrictions that complicate access to data and increase compliance costs for developers. These legal constraints do not merely regulate AI deployment; they shape the feasibility and location of AI development itself. The result is a schizophrenic policy environment in which the EU seeks to promote AI innovation while simultaneously constraining the conditions necessary for its emergence.

Since 2017, the EU has introduced a series of policy initiatives to strengthen its technological competitiveness in AI. Following the Tallinn Digital Summit,[8] the European Council invited the European Commission to develop a coordinated European approach to AI. In response, the Commission's 2018 AI strategy set the objective of ensuring that the EU was at the forefront of AI development and adoption. Subsequent policy initiatives have reinforced this objective. The Commission's 2020 White Paper on AI,[9] the Coordinated Plans on AI,[10] and more recent initiatives such as the AI Innovation Package[11] and the AI Continent Action Plan[12] have made clear how important AI is to the EU's economic development.

These ambitions are backed by more than words. In recent years, the EU has ramped up public spending to support AI infrastructure. In December 2024, the European High Performance Computing Joint Undertaking (EuroHPC) selected seven proposals for establishing AI factories, i.e., public data centres that include supercomputing power, data, and expertise to build and train advanced AI models, at a cost of EUR 1.5 billion. A few months later, at the 2025 AI Action Summit, the EU launched InvestAI, a programme that aimed to mobilise EUR 200 billion, with a EUR 20 billion ring-fenced fund to build four to five 'AI Gigafactories' to house over 100,000 next-generation AI chips each. In October of 2025, as part of its 'Apply AI Strategy', the European Commission announced that it will mobilise EUR 1 billion (drawing from existing funds like Horizon

---

[8] European Commission. (2017, September 14). Tallinn Digital Summit. Shaping Europe's digital future. Available at: https://digital-strategy.ec.europa.eu/en/events/tallinn-digital-summit ; ComplexDiscovery Staff. (2025, October 9). Tallinn's digital convergence: Europe's blueprint for AI security governance. ComplexDiscovery. Available at: https://complexdiscovery.com/tallinns-digital-convergence-europes-blueprint-for-ai-security-governance/

[9] European Commission. (2020, February 19). White paper on artificial intelligence: A European approach to excellence and trust. Available at: https://commission.europa.eu/document/d2ec4039-c5be-423a-81ef-b9e44e79825b_en

[10] European Commission. (2021, April 21). Coordinated plan on artificial intelligence (2021 review). Shaping Europe's digital future. Available at: https://digital-strategy.ec.europa.eu/en/policies/plan-ai

[11] European Commission. (2024, July 25). AI innovation package. Shaping Europe's digital future. Available at: https://digital-strategy.ec.europa.eu/en/factpages/ai-innovation-package

[12] European Commission. (2025, May 7). AI continent action plan. Shaping Europe's digital future. Available at: https://digital-strategy.ec.europa.eu/en/factpages/ai-continent-action-plan

Europe and the Digital Europe Programme) to accelerate the adoption of AI in ten strategic sectors.[13] Lately, on January 15, 2026, the European Commission announced the allocation of EUR 307.3 million to the development of 'trustworthy AI services'.[14]

2.2 The Brake: How Regulation Induces Data Scarcity

The EU's primary legislative instrument governing the collection of data for AI training is the CDSMD (Directive (EU) 2019/790), adopted in 2019. The Directive introduced two mandatory exceptions for text and data mining (TDM), i.e., the process by which automated systems extract patterns, correlations, and information from large bodies of digital content. Article 3 CDSMD grants research organisations and cultural heritage institutions an unconditional right to conduct TDM for scientific research purposes, provided they have lawful access to the relevant content. Article 4 extends a broader exception to any actor, including commercial firms, again subject to lawful access. Although the CDSMD was not drafted with GenAI specifically in mind, its technologically neutral framing was deliberate: the legislature intended the exceptions to remain adaptable to emerging forms of computational analysis.

At first glance, this architecture appears permissive. A broad TDM exception open to commercial actors, with no licensing requirement by default, suggests a framework aimed at supporting, rather than limiting, access to data. That reading is incomplete. The TDM exception does not operate in isolation. It sits within a scaffolding of public and private regulatory instruments, transforming a nominally open framework into one that is far more restrictive than what first meets the eye.

First of all, it is important to recognise that the TDM is not a free-for-all exception. Under Article 4 of the CDSMD, rightsholders may expressly reserve their rights provided that such reservation is made in an appropriate manner, such as through machine-readable means in the case of content made publicly available online. If opt-outs are exercised extensively – a scenario that is legally permissible – the availability of high-quality training data will be reduced. This is an important restriction, because AI models depend on broad, diverse, and statistically representative training datasets.

The opt-out mechanism of the CDSMD is only the first layer of the public and private regulatory enforcement that surrounds data access in the EU. The General Data Protection Regulation (GDPR) – the EU's comprehensive framework governing the processing of personal data – imposes substantial constraints on AI training. Large-scale web datasets routinely contain personal data, even where that data is publicly accessible. Under Articles 6 and 9 GDPR, developers must establish a lawful basis for processing such data. At the same time, Articles 13 and 14 of the GDPR further require that individuals be informed of how their data is used – a

[13] The strategic sectors were: healthcare and pharmaceuticals; mobility, transport and automotive; robotics; manufacturing, engineering and construction; climate and environment; energy; agri-food; defence, security and space; electronic communications; and cultural, creative, and media industries. Reference: European Commission. (2025, October 8). Commission launches two strategies to speed up AI uptake in European industry and science [Press release]. https://digital-strategy.ec.europa.eu/en/news/commission-launches-two-strategies-speed-ai-uptake-european-industry-and-science

[14] European Commission, "EU invests over €307 million into artificial intelligence and related technologies," Shaping Europe's digital future, January 15, 2026, https://digital-strategy.ec.europa.eu/en/news/eu-invests-over-eu307-million-artificial-intelligence-and-related-technologies.

requirement that is practically impossible to fulfil when training datasets contain information relating to millions of individuals.

A second example is the Digital Markets Act (DMA). The Act imposes structural obligations on large digital platforms designated as 'gatekeepers'. Rather than restricting access to data directly, the DMA actively seeks to prevent gatekeeper platforms from combining data assets across their services, on the assumption that such combinations confer anti-competitive advantages. The immediate consequence is a reduction in the usefulness and competitiveness of the services these firms can offer within the EU, an effect already visible in the decision by some platforms to postpone the introduction of new data and AI-driven services in the region.[15]

Finally, the AI Act – which some provisions have been postponed[16] – introduces additional uncertainties with respect to the standards for data collection. For instance, while Article 4 of the CDSMD allows rightsholders to opt out using "any appropriate means," the AI Act (Article 53) demands that developers employ state-of-the-art standards to detect these reservations. However, there is no globally harmonised technical standard for AI opt-outs.[17] Furthermore, the AI Act introduces an extraterritorial mandate, requiring developers to respect EU opt-outs regardless of where the model was trained. This contradicts the traditional territoriality principle of copyright law, which usually ties legal acts to the jurisdiction (in this case) where the training occurred and broadens the scope of the AI Act as it regulates the legality of the data used for AI training.

Compliance with these and other EU regulations is assured by an active regulatory enforcement. Several European data protection authorities have already scrutinised and limited the use of personal data in AI training, including investigations into Meta's use of publicly available user data.[18]

Private regulatory enforcement in the form of litigation constitutes an equally significant, and arguably more unpredictable, constraint on data access. As alluded to in the introductory chapter, the case of Google v. Like Company currently in the ECJ illustrates the point. In that case, a Hungarian publisher challenged Google's use of press content in AI training, raising the question of whether such training falls within the Article 4 TDM exception or instead requires licensing under Article 15 CDSMD – a separate provision granting press publishers a neighbouring right over the digital use of their publications.

---

[15] In July 2024, Meta announced it would withhold its advanced multimodal AI models from the EU market, explicitly citing the "unpredictable nature of the European regulatory environment". Source: https://www.theguardian.com/technology/article/2024/jul/18/meta-release-advanced-ai-multimodal-llama-model-eu-facebook-owner; Apple postponed the EU release of its Apple Intelligence features, including iPhone Mirroring and SharePlay due to "regulatory uncertainties" brought about by the Digital Markets Act (DMA). Source: https://www.reuters.com/technology/artificial-intelligence/apple-delay-launch-ai-powered-features-europe-blames-eu-tech-rules-2024-06-21/

[16] European Parliament. (2026, March 20). Report on the proposal for a regulation of the European Parliament and of the Council amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI) (Document No. A10-0073/2026). https://www.europarl.europa.eu/doceo/document/A-10-2026-0073_EN.html

[17] Downing, K. (2025). The Enforceability of AI Training Opt-Outs. Available at: https://katedowninglaw.com/2025/05/28/the-enforceability-of-ai-training-opt-outs/

[18] Lucchi, N. (2025). Generative AI and copyright: Training, creation, regulation (PE 774.095). European Parliament, Policy Department for Justice, Civil Liberties and Institutional Affairs. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)774095_EN.pdf

The European Commission has indicated that the case may be partly or entirely inadmissible, suggesting that courts should not rule on the general functioning of Google's Gemini without clearly establishing an act of infringement.[19] It remains to be seen how the case will develop, with the Advocate General's opinion expected on 3 September 2026.[20] If AI training is found to fall outside the TDM exception, developers would face licensing obligations across an entire category of content, with corresponding increases in cost and complexity, and the precedent would invite analogous claims across other categories of protected content.

The outermost layer of the scaffolding is also the non-binding General-Purpose AI (GPAI) Code of Practice, published by the European Commission in February 2026 under the auspices of the newly established AI Office.[21] In substance, it functions as a compliance standard. Among other things, signatories are expected to implement internal copyright policies, ensure that training data is drawn only from lawfully accessible sources, identify and comply with rights reservations, and mitigate the risk of infringing outputs. The practical effect is to operationalise the opt-out mechanism under Article 4 CDSMD, transforming what was designed as a rightsholder tool into an ongoing compliance obligation for developers.

2.3 The Consequence: The Private Investment Deficit in European AI

These regulatory choices have significant economic consequences. Crucially, the EU's data governance framework turns data, a core economic resource in the AI value chain, into a binding constraint on AI development in Europe. It raises the cost of acquiring data, reduces the volume of legally accessible content, and creates legal uncertainty that discourages investment.

The economic effects are clear and follow a simple logic. When regulation makes data artificially scarce or costly, the expected returns on investment in data-intensive AI development fall. This weakens incentives for EU firms to participate in data management and the foundational layers of the AI value chain. The result is a European AI ecosystem increasingly concentrated in downstream activities such as application development and deployment, dependent on foundational models developed elsewhere, and structurally unable to capture value at the start of the chain.

Empirical evidence shows that EU regulation is constraining firms' ability to accumulate and use data, while the gap in AI investment between the EU and leading economies continues to grow. For example, upon the GDPR's entry into force, EU firms were compelled to destroy substantial amounts of data.[22] Two years after its implementation, EU firms stored 26 per cent less data on average than their US counterparts and had reduced their computational activities by 15 per

---

[19] This means that a court cannot simply declare an AI model unlawful in general terms. Instead, the claimant must identify a specific instance of alleged infringement and demonstrate that protected material was used without authorisation in a way that constitutes copyright infringement.

[20] Preiss, I. (2026, March 10). Google v Like Company Case analysis. Available at: https://www.linkedin.com/posts/inbar-preiss-5b1b21187_google-told-the-eus-top-court-that-its-gemini-activity-7437180708613267459-pen7?utm_source=share&utm_medium=member_android&rcm=ACoAADTs5lIBu-ZAa2pRzpDzaxm1wosxJjlzMb4

[21] European Commission. (2026, February 2). The General-Purpose AI Code of Practice. Shaping Europe's digital future. Available at: https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai

[22] Erixon, F., & Guinea, O. (2024). From Compliance to Constraint: How Digital Regulation Impacts Productivity and Innovation in Europe. EconPol Forum, 25(6), 18-21.

cent.[23] The consequences extended to market dynamics, with a 50 per cent drop in new app entries[24] and a shift in firm demographics favouring older, larger companies that could better absorb the compliance costs.[25]

These firm-level effects are reflected in global AI investment patterns. The gap between European and American private investment in AI has widened significantly. In 2024, the US private sector invested over €101 billion[26] in AI – equivalent to 73 per cent of all private AI investment globally – bringing the total invested between 2013 and 2024 to €408 billion. By contrast, Europe attracted €18 billion in private AI investment in 2024, with its top member states[27] accumulating a total of €27 billion over the same period. China invested €8 billion in 2024, bringing its cumulative total to €103 billion.[28]

The divergence is most pronounced in GenAI. In 2024, the US attracted €27 billion in private investment in GenAI; the EU attracted €1.3 billion; and China attracted €1.8 billion.[29] The consequences of this investment gap are evident in model development and compute capacity. In 2024, the US produced 40 new AI models. Compared with 15 in China and just three in Europe.

In terms of compute capacity, the US had 212 supercomputers in 2025, compared with 63 in the EU.[30] The ownership structure of these systems also highlights the imbalance between public and private involvement in AI development in the EU, and how different Europe's profile is from that of the US. In the EU, roughly half of all supercomputers are owned by the private sector, with the remainder owned by the public sector. In the US, by contrast, only 18 per cent of supercomputers are publicly owned.

---

[23] Demirer, M., D. J. Jiménez-Hernández, D. Li and S. Peng (2024), 'Data, Privacy Laws and Firm Production: Evidence from the GDPR', NBER Working Paper 32146.

[24] Janßen, R., R. Kesler, M. E. Kummer and J. Waldfogel (2022), "GDPR and the Lost Generation of Innovative Apps", NBER Working Paper 30028.

[25] Chen, C., C. B. Frey and G. Presidente (2022), "Privacy Regulation and Firm Performance: Estimating the GDPR Effect Globally", The Oxford Martin Working Paper Series 22(1).

[26] All values have been converted from US$ to € using the European Central Bank exchange rate averages for the corresponding years. Accessed at: https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/eurofxref-graph-usd.en.html

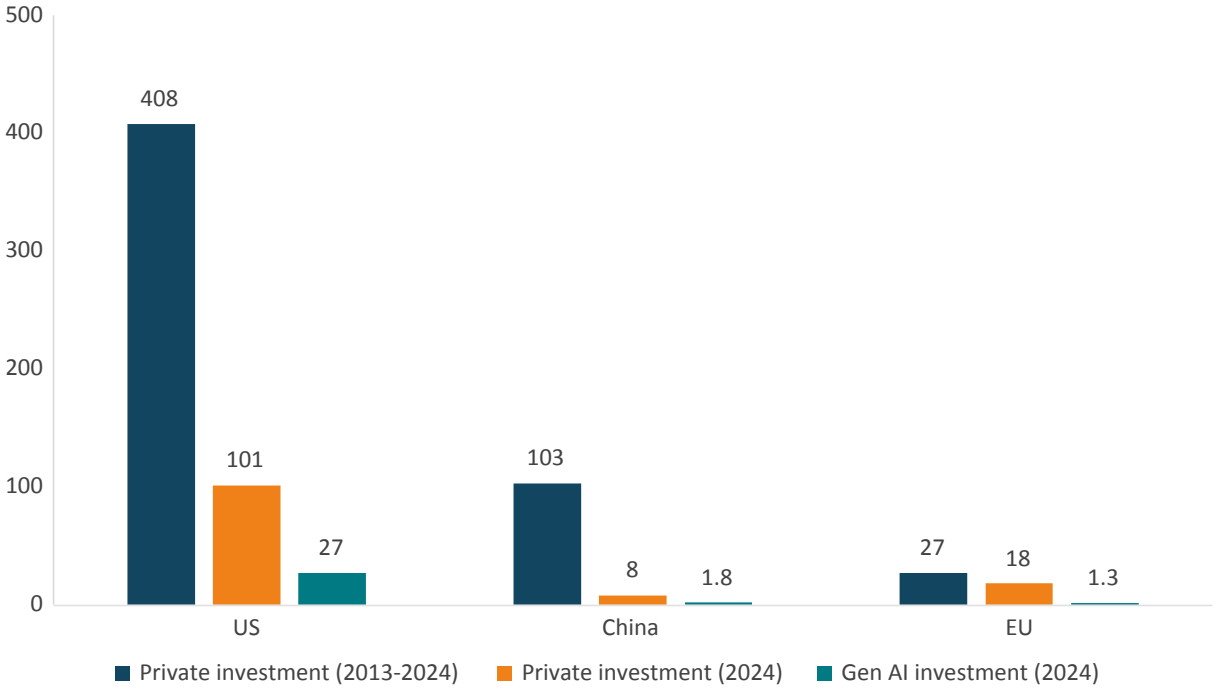
[27] EU member state values were available for Sweden, France, and Germany. In total these countries represent 43 per cent of EU GDP in 2024.

[28] Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., Betts Lotufo, J., Rome, A., Shi, A., & Oak, S. (2025). The AI Index 2025 annual report. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. Available at: https://doi.org/10.48550/arXiv.2504.07139

[29] Ibid.

[30] Epoch AI (2025). GPU Clusters dataset. Accessed at: https://epoch.ai/data/gpu-clusters

FIGURE 1: INVESTMENT IN AI (VALUE, € BILLION)



Source: Author's calculations. Stanford Artificial Intelligence Index Report 2025. Note: EU figures under private investment (2013-2024) combine top-scoring EU member states.[31] EU private investment and Gen AI investment shows private investment in AI by Europe.

The evidence points to a fundamental constraint. Given the scale of resources required to develop AI models, the EU's public investments outlined above will not be sufficient to make a meaningful difference. Without substantial private investment, the European economy will continue to fall behind in the development and deployment of AI. For such investment to materialise, the EU must create a business environment in which data is abundant and therefore a low-cost input. This stands in stark contrast to the EU's current data governance framework, which makes data artificially scarce and as a result more expensive.

3. BEYOND THE CREATIVE INDUSTRY: IMPACT ON THE BROADER ECONOMY

For years, debates on copyright law have been informed by the interests of the creative industries. However, as the economy becomes more digital, the importance of the TDM exception, and access to data more broadly, has become critical for wider economic development. For example, while estimates of AI's impact on economic growth and productivity vary,[32] there is broad

[31] EU member state values were available for Sweden, France, and Germany. In total these countries represent 43 per cent of EU GDP in 2024.

[32] Dario Amodei projects 10-20 per cent annual growth while Daron Acemoglu estimates 0.5 per cent total additional growth over a decade. See: Amodei, D. (2026). The Adolescence of Technology. Accessed at: https://www.darioamodei.com/essay/the-adolescence-of-technology and Acemoglu, D. (2024). The simple macroeconomics of AI. Economic Policy, Volume 40, Issue 121, January 2025, Pages 13–58, https://doi.org/10.1093/epolic/eiae042.

consensus that, in the long run, the net effect of AI development and adoption will be higher growth and productivity.

This reflects the fact that AI is a general-purpose technology capable of transforming a wide range of industries. In Europe's manufacturing and services sectors, the implications are particularly significant. Several globally competitive industries, including automobiles, pharmaceuticals and chemicals, as well as high-value sectors such as financial and professional services, are undergoing rapid AI-driven transformation.

Case studies from the automotive production sector illustrate how AI can dramatically improve factory performance. Advanced analytics and machine-learning models are now used to optimise assembly-line operations, predict equipment failures, and improve quality control.[33] Beyond the factory floor, the development of autonomous vehicles illustrates the even more acute data requirements of AI in safety-critical applications. Systems designed to handle rare and unpredictable real-world events cannot be trained on simulation alone; they require large-scale access to diverse real-world data, including LiDAR point clouds, camera footage, and annotated edge-case scenarios. European manufacturers such as the VW Group (through MOIA)[34] and Volvo (via Volvo Autonomous Solutions[35] and its partner company Zenseact) are already investing in AI-driven autonomous mobility systems that rely on large-scale real-world data for training and validation, underscoring the growing importance of data access as a key enabler of technological development in Europe.[36]

However, European automotive companies are lagging behind in AI adoption which has led to significant drops in their global market valuations. The top ten European automotive companies have seen a €71 billion drop in valuation since 2015 (a 19 per cent reduction), compared to a quadrupling of US tech over the same period, based on the NASDAQ 100 index.[37] For Europe, where the automotive sector employs 13.8 million people and contributes 7 per cent of EU's GDP[38], AI adoption is critical to maintaining global competitiveness.[39]

Similarly, the pharmaceutical industry, which contributed €311 billion in gross value added and 2.3 million jobs in 2022,[40] is a high-value European sector whose future competitiveness largely depends on its ability to integrate AI into its production processes. AI models can support pharmaceutical companies to accelerate drug discovery by analysing vast biological datasets and

---

33 McKinsey & Company. (2017). Building smarter cars with smarter factories: How automakers can leverage analytics in manufacturing. McKinsey Digital.

34 MOIA. (n.d.). MOIA Turnkey Solution. Available at https://www.moia.io/en/solutions/autonomous-mobility

35 Volvo Autonomous Solutions. (n.d.). Transforming the movement of goods. Available at https://www.volvoautonomous-solutions.com/en-en/

36 As Zenseact noted that the more varied and complex the scenarios on which a vehicle is trained, the better it can generalise across different operational environments. Zenseact. (2023, October 6). Zenseact releases open dataset to advance autonomous driving. Available at: https://zenseact.com/news/zod/

37 McKinsey & Company. (2025). A new 'ERA': An action plan for the European automotive industry

38 European Commission. Automotive Industry. Accessed at: https://single-market-economy.ec.europa.eu/sectors/automotive-industry_en

39 McKinsey & Company. (2023). The State of AI in 2023: Generative AI's breakout year. QuantumBlack, AI by McKinsey.

40 PwC. (2024). Economic footprint of the pharmaceutical industry in Europe. EFPIA. Accessed at: https://efpia.eu/media/3dqjpl3x/economic-footprint-of-the-pharmaceutical-industry-in-europe-report.pdf

identifying potential therapeutic targets.[41] For instance, Google's DeepMind created AlphaFold, which vastly reduced the time required to predict protein structures.[42]

European firms are actively building these capabilities. AstraZeneca has established strong in-house operations and research partnerships to accelerate the screening of potential medicines and computationally predict how well they will fight the disease.[43] In the broader healthcare sector, companies such as the German medical device manufacturer B. Braun[44] are developing diagnostic and predictive models in collaboration with Norwegian[45] and American partners.[46] These applications require large, diverse, and well-annotated datasets including medical imaging, genomic sequencing data, and clinical notes to achieve clinical-grade performance.

Yet, similar to the automotive industry, the European pharmaceutical industry needs to pick up the pace of investment and adoption of digital and AI technologies. In terms of investment in health AI, the US had investments worth €10 billion in 2024, compared to only €7 billion in Europe across all AI sectors.[47] This gap is also reflected in AI adoption. For instance, by 2025, the US authorised nearly 777 AI-enabled radiology devices, compared to approximately only 200 in Europe.[48]

Another interesting case study is the European chemicals industry, accounting for the second largest chemicals market in the world after China. It has a turnover of €635 billion and employs 1.2 million people and is the fourth largest manufacturing sector in Europe.[49] GenAI offers significant opportunities to create value across all business functions in the chemical industry. For instance, GenAI can accelerate material and molecular discovery by two to three times and support optimisation for end-use. European deeptech start-ups such as Dunia Innovations are using AI alongside automated laboratory systems to transform chemical R&D. For example, the company Dunia uses physics-informed machine learning models integrated with chemical robotics in a 'self-driving lab', where AI designs new materials, robotic systems synthesise and test them, and the resulting data is fed back to continuously improve predictions in electrochemical materials discovery.[50]

---

41 Joshi, R. P., & Kumar, N. (2021). Artificial intelligence based autonomous molecular design for medical therapeutics: A perspective. arXiv

42 AlphaFold. Accessed at: https://deepmind.google/science/alphafold/

43 AstraZeneca. (n.d.). Data science and AI. Available at https://www.astrazeneca.com/r-d/data-science-and-ai.html

44 B. Braun. (n.d.). Innovation Hub. Available at https://www.bbraunusa.com/en/about-us/company/innovation-hub.html

45 Cognite. (n.d.). B. Braun partners with Cognite to power its data-first transformation. Available at https://www.cognite.com/en/company/newsroom/b.-braun-partners-with-cognite-to-power-its-data-first-transformation-1

46 IDENTI Medical. (n.d.). IDENTI Medical provide Braun with automated systems for managing surgical sutures. Available at https://identimedical.com/identi-medical-provide-braun-with-automated-systems-for-managing-surgical-sutures/

47 Nawabi, J. & et.al. (2026). Artificial intelligence in radiology: A comparative analysis of reimbursement and regulatory developments in the US and EU. European Journal of Radiology Artificial Intelligence. Accessed at: https://www.sciencedirect.com/science/article/pii/S3050577125000544#:~:text=By%202025%2C%20the%20U.S.%20had,advancing%20regulatory%20oversight.

48 Nawabi, J. & et.al. (2026). Artificial intelligence in radiology: A comparative analysis of reimbursement and regulatory developments in the US and EU. European Journal of Radiology Artificial Intelligence. Accessed at: https://www.sciencedirect.com/science/article/pii/S3050577125000544#:~:text=By%202025%2C%20the%20U.S.%20had,advancing%20regulatory%20oversight.

49 European Commission. (2025). Plan for stronger EU chemical industry. https://commission.europa.eu/news-and-media/news/plan-stronger-eu-chemical-industry-2025-07-08_en

50 Wheatley, M. (2024, October 23). Dunia raises $11.5M to accelerate electrochemical materials discovery. Siliconangle. Available at: https://siliconangle.com/2024/10/23/dunia-raises-11-5m-accelerate-electrochemical-materials-discovery/#:~:text=%E2%80%9CBy%20combining%20machine%20learning%20with%20chemical%20robotics%2C,new%20materials%2C%E2%80%9D%20said%20Dunia%20co%2Dfounder%20Alexander%20Hammer.

The green transition represents another domain where AI data access has direct economic and strategic consequences. In the energy and utilities sector, European companies such as EDF,[51] Engie,[52] and Ørsted[53] use AI to forecast demand and supply, a task that becomes significantly more complex with the integration of variable renewable energy sources. These models require continuous training on historical grid load data, consumption patterns, weather data, and real-time generation outputs. At the same time, next-generation climate models, deployed by firms such as Mitiga Solutions[54] and Ortec Finance,[55] rely on satellite and remote sensing data to improve extreme event prediction, monitor emissions, and detect illegal deforestation. Restricting the data inputs available for these models would not only undermine European AI competitiveness but would also slow the development of tools essential to the EU's own climate objectives.

AI has huge applications in services too. For instance, the European financial services firms are increasingly embedding AI and GenAI technologies into their operations such as automated knowledge management, fraud detection and investment research.[56] To maintain accuracy, these systems require continuous retraining on diverse datasets, including transaction histories, credit records, macroeconomic indicators, and alternative behavioural signals, to adapt to shifting economic conditions and evolving fraud patterns.[57] Yet, in the EU, only 9 per cent of financial firms consider themselves ahead of the curve in AI adoption.[58] Moreover, despite 28 per cent of firms in the European financial services sector accelerating AI adoption over the past year, the majority remain in the early, experimental stages.[59]

The retail and e-commerce sectors offer a further illustration of the cross-economy importance of data access. Companies such as Ahold Delhaize,[60] the Otto Group,[61] and Decathlon[62] are increasingly deploying AI for recommendation systems, dynamic pricing, and demand forecasting. The performance of these systems improves substantially with the integration of multimodal inputs that combine text, product images, and consumer behavioural data, and with access to

---

51 EDF. (n.d.). Artificial Intelligence. Available at https://www.edf.fr/en/the-edf-group/inventing-the-future-of-energy/rd-global-expertise/rd-experience/data-science-ai-world/artificial-intelligence

52 Artefact. (n.d.). ENGIE: Creating an AI Factory to accelerate digital services. Available at https://www.artefact.com/cases/engie-creating-an-ai-factory-to-accelerate-digital-services/

53 Ørsted. (n.d.). Democratise AI. Available at https://orsted.com/en/offshore-wind/innovation/democratise-ai

54 Mitiga Solutions. (n.d.). Home. Available at https://www.mitigasolutions.com/

55 Ortec Finance. (n.d.). Climate MAPS. Available at https://www.ortecfinance.com/en/insights/product/climate-maps

56 Chlouverakis, K. (2024). How artificial intelligence is reshaping the financial services industry. Ernst and Young. https://www.ey.com/en_gr/insights/financial-services/how-artificial-intelligence-is-reshaping-the-financial-services-industry

57 Several major European financial institutions are actively pursuing these strategies. BNP Paribas and Santander have developed strong in-house AI initiatives. with Santander having established a partnership with OpenAI, while Deutsche Bank and UniCredit are integrating AI through long-term strategic operations partnerships. See: BNP Paribas. (n.d.). Data & Artificial Intelligence. Available at https://group.bnpparibas/en/our-commitments/innovation/data-artificial-intelligence; Deutsche Bank. (n.d.). Artificial Intelligence. Available at https://corporates.db.com/files/documents/in-focus/focus-topics/artificial-intelligence/Deutsche-Bank-AG-Corporate-Bank-Artificial-Intelligence-Jan-2024.pdf; Santander. (n.d.). Santander data & AI-first strategy accelerates through OpenAI collaboration. Available at https://www.santander.com/en/stories/santander-data-ai-first-strategy-accelerates-through-openai-collaboration; UniCredit. (n.d.). Digital and Data. Available at https://www.unicreditgroup.eu/en/business/digital-and-data.html

58 Moscetti, L. & Bali, A. (2024). AI Adoption in European Financial Services: Progress, Challenges, and Future Directions. Ernst & Young. Accessed at: https://www.ey.com/en_lu/insights/ai/ai-adoption-in-european-financial-services-progress-challenges-and-future-directions

59 Ibid.

60 Ahold Delhaize. (n.d.). Data & AI. Available at https://aholddelhaize.com/about/technology-innovation/data-ai/

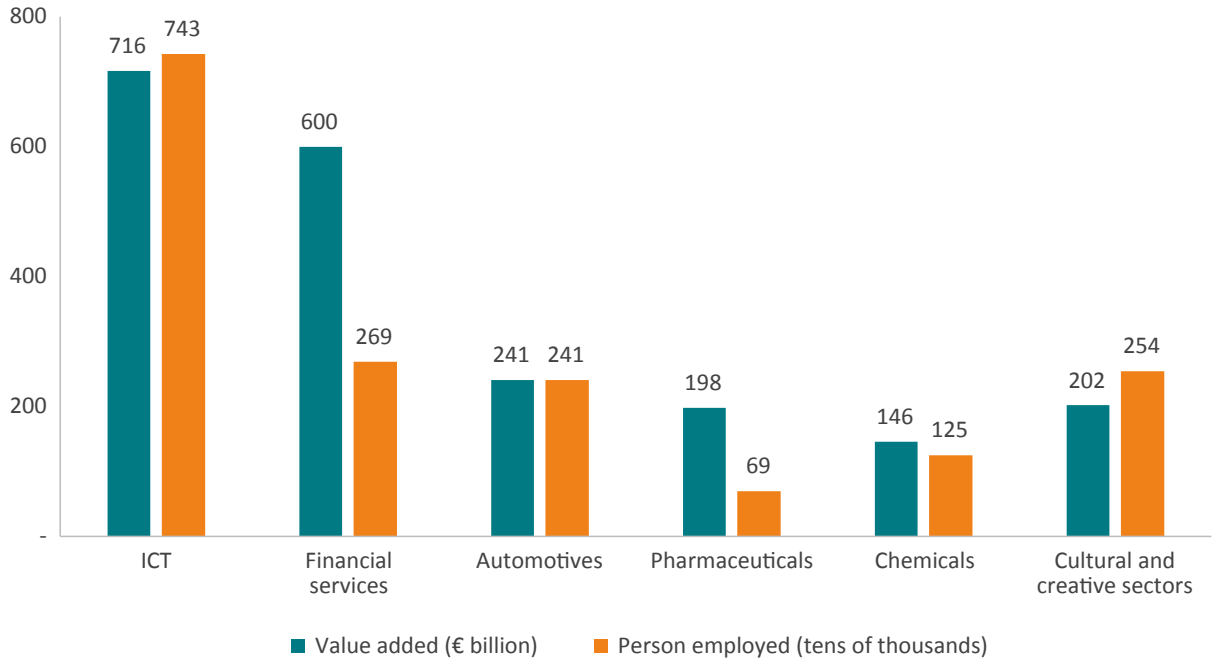
61 Otto Group. (n.d.). AI in action: How Artificial Intelligence enhances customer experience and e-commerce at the Otto Group. Available at https://www.ottogroup.com/en/stories/story/ai-in-action-how-artificial-intelligence-enhances-customer-experience-and-e-commerce-at-the-otto-group.php

62 Corinium Intelligence. (n.d.). AI Transformation in the Retail Industry. Available at https://www.coriniumintelligence.com/content/ai-transformation-retail-industry

large-scale purchase histories and clickstream data. As with other sectors, the competitive advantage of European retailers in the digital economy depends directly on their ability to train and continuously update AI models on rich, diverse datasets.

The European creative industry is also a significant sector for the EU economy. In 2023, the EU's creative industries generated €202 billion in gross value added and employed over 2.5 million workers.[63] However, these data points must be seen in context, given that the policies under discussion affect the broader economy. Figure 2 compares the economic contribution of Europe's creative sector – measured in terms of value added and employment – with a selection of sectors for which data access policies, including changes to the EU copyright framework, are likely to be highly consequential. Europe's ICT, financial services, automotive, pharmaceutical and chemical industries all rely heavily on data and the deployment of digital technologies, including AI, to maintain their competitiveness. The combined value added of these five sectors is more than nine times that of the EU's creative sector, and their total workforce is almost six times larger.

FIGURE 2: VALUE ADDED (2023) AND EMPLOYMENT (2024) BY SECTOR TO THE EU ECONOMY



Source: Authors' calculations based on Eurostat. Data for the Cultural and Creative Sector for 2023.

Comparing the economic contribution of Europe's creative industries with other sectors is not to suggest that the creative industries do not matter, nor that their rights should be sacrificed for the benefit of other sectors' growth. Rather, the aim is to recognise that copyright law has far-reaching economic implications that extend well beyond the creators and rights holders it traditionally governs. Moreover, companies in the creative industry are increasingly sophisticated users of AI, meaning their productivity and future innovation stand to benefit significantly from a robust, data-rich technological ecosystem.

[63] Eurostat. Enterprises by detailed NACE Rev. 2 activity and special aggregates. Sector: Cultural and creative sectors – total. Available at: https://doi.org/10.2908/SBS_OVW_ACT

4. THE DISPROPORTIONATE IMPACT OF RESTRICTIVE COPYRIGHT LAW ON SMEs

The EU economy relies heavily on SMEs and start-ups to develop AI applications. These firms develop smaller, more efficient and tailored-made AI models to cater for specific industrial needs.[64] For example, French biotechnology firms have developed AI models for biological applications, while Italian start-ups use machine learning to quantify climate hazards from space.[65]

This reliance on SMEs makes the impact of regulation on firms all the more critical. Regulatory frameworks tend to impose fixed compliance costs. The EU's digital rules mentioned earlier, such as the AI Act and the GDPR, impose substantial regulatory burdens. While large firms have in-house legal teams and the financial resources to absorb these costs, SMEs often lack the staff needed to manage regulatory compliance. Moreover, legal uncertainty associated with some of these regulations has a chilling effect on innovation. When rules governing data collection are ambiguous, AI developers must factor in potential legal risks during the training and development process. As noted earlier, this leads to more conservative design choices, constrained innovation, or relocation to jurisdictions with clearer legal frameworks.

The economics of data exacerbate this asymmetry between firms of different sizes. Data exhibits strong network effects and returns to scale: larger datasets enable more capable models, which attract more users and generate further data, reinforcing existing advantages. Large incumbents benefit from in-house data reserves that sustain this cycle. By contrast, SMEs must rely on the open market to acquire training data in order to build competitive systems. Under restrictive data frameworks, this reliance becomes a significant financial barrier. In practice, high transaction costs, complex licensing negotiations and fragmented data access act as barriers to entry, limiting SMEs' access to essential data.

The pressure on start-ups is evident: nearly half of European start-ups identify data quality and availability as a primary constraint.[66] As a result, access to broad datasets is restricted and SMEs are forced to train models on narrower, lower-quality data, which creates a market structure that favours established incumbents. The consequence of this dynamic is the entrenchment of market power and the degradation of European AI capabilities.

---

[64] The legal sector provides a further illustration of the need for domain-specific training data. Models trained on general text corpora often struggle with citation accuracy and legal reasoning; performance in tasks such as case summarisation, contract analysis, and legal search improves substantially when models are trained on domain-specific materials, including case law, statutes, and regulatory texts. European law firms such as Fidal (France), Loyens & Loeff (Netherlands/Belgium), and Garrigues (Spain) are among those actively fostering AI adoption across their practices. See: Sopra Steria. (n.d.). How law firm Fidal is redefining legal practice with AI. Available at https://www.soprasteria.com/newsroom/client-stories/details/how-law-firm-fidal-is-redefining-legal-practice-with-ai; Loyens & Loeff. (n.d.). How AI is shaping the future of legal and tax at Loyens & Loeff. Available at https://www.loyensloeff.com/insights/news--events/news/how-ai-is-shaping-the-future-of-legal-and-tax-at-loyens--loeff/; Garrigues. (n.d.). Garrigues and Microsoft sign strategic partnership agreement to drive innovation and the use of AI. Available at https://www.garrigues.com/en_GB/new/garrigues-and-microsoft-sign-strategic-partnership-agreement-drive-innovation-and-use-ai

[65] European Commission (2025). A European Strategy for Artificial Intelligence in Science Paving the way for the Resource for AI Science in Europe (RAISE). Communication from the Commission to the EU Parliament and the Council.

[66] Implement Consulting Group (2024). The Economic Opportunity of AI in the EU.

5. POLICY RECOMMENDATIONS

As described in Chapter 2, the EU's regulatory framework for AI is pulling in two directions at once. On one hand, the Union has committed billions of euros to AI infrastructure, computing capacity, and research programmes. On the other, it has constructed a dense scaffolding of regulations that makes the essential input for AI – data – increasingly constrained and legally uncertain to use.

EU policymakers must ensure that data functions as an abundant and accessible economic endowment for European firms. This is not a call to abandon copyright protection. It is a call to stop treating data restriction as a default policy choice without cost. Every regulation that limits access to training data is, in effect, a levy on Europe's AI development, making the EU economy less competitive.

Recommendation 1: Treat the Text and Data Mining Exception as a Strategic Asset, Not a Concession

The EU should formally recognise the TDM exception under Articles 3 and 4 of the Copyright in the CDSMD as a core component of the EU's AI competitiveness strategy. Proposals to narrow, condition, or restrict the scope of the TDM exception, whether through legislative amendment or through the expansion of opt-out mechanisms, should be subject to a mandatory competitiveness impact assessment before advancing.

The TDM exception is one of the few instruments in the EU's regulatory portfolio that moves in the direction of making data more accessible rather than less. All other relevant frameworks, e.g., the GDPR, the DMA, the AI Act, impose restrictions on data access. The TDM exception is the counterweight. Weakening it does not merely reduce a legal permission; it tilts the balance of EU data policy further towards restriction.

The economic logic is straightforward. When the TDM exception is weakened, the cost of AI development rises, the pool of legally accessible training data shrinks, and the incentive to develop AI models within the EU falls. The emergence of EU-based model developers such as Mistral AI (France), Minerva AI (Italy), PhariaAI (Germany), VelvetAI (Italy) and EuroLLM 9B have been directly supported by the legal certainty provided by the TDM exception. Eroding that certainty would not protect rightsholders in any meaningful economic sense; it would simply push AI development to jurisdictions where no such restriction applies.

If any policy is required, the European Commission should issue formal interpretative guidance clarifying that the TDM exception, as currently framed, covers the large-scale reproduction of works for AI training, provided lawful access conditions are met. This guidance should address the unresolved question of whether reproductions made in the context of AI training fall within the scope of the exception – a question that the CDSMD's recitals support. It should also clarify that the exception applies across the full AI development pipeline, including data collection, pre-processing, and model training, as intended by the law.

Recommendation 2: Contain the Opt-Out Mechanism Before It Renders the TDM Exception Unworkable

The EU should resist any expansion of the opt-out mechanism under Article 4(3) of the CDSMD. In particular, policymakers should reject proposals to broaden the categories of content subject to opt-out, lower the technical threshold for a valid rights reservation (i.e. a mechanism for rights holders to opt out of the use of their content), or make opt-out the default for new categories of works. The current framework, under which opt-out requires an affirmative and machine-readable reservation of rights, should be preserved.

Extending the opt-out mechanism represents the most significant vulnerability in the EU's TDM framework. As the preceding analysis has shown, the CDSMD's architecture is already conditional: Article 4 grants a general permission to conduct TDM, but that permission can be withdrawn by rightsholders. If opt-outs are exercised widely, the scope of the exception is significantly reduced. As a result, what was designed as a general permission could become a permission-based regime requiring individual authorisation from potentially millions of rightsholders.

The consequences of this potential policy change are uneven. Large AI developers with substantial legal and financial resources can navigate a fragmented opt-out landscape by negotiating bilateral licensing agreements with major content providers. Smaller European AI developers and start-ups cannot. The result could be a regulatory structure that, in practice, entrenches the market position of large incumbent firms while limiting the development of a competitive European AI ecosystem. This outcome is precisely the opposite of what many EU industrial and digital policymakers want to achieve.

Recommendation 3: Stabilise the Legal Framework and Resist the Temptation to Legislate Under Uncertainty

The EU should impose a moratorium on legislative changes to the TDM framework. The immediate priority should be to allow the Digital Omnibus package to complete its legislative journey through the European Parliament and the Council, and to give the recently adopted EU digital regulations time to reach operational maturity before any new obligations are layered on top.

The EU's digital regulatory landscape is, by any measure, exceptionally dense. Between 2022 and 2025, the Union adopted a series of major instruments that have a direct impact on AI development: the AI Act, the Data Act, the Data Governance Act, the Digital Services Act, the DMA, the Cyber Resilience Act and the NIS2 Directive, alongside the continued application of the GDPR. Each of these instruments imposes distinct compliance obligations on AI developers. Taken together, they create a cumulative burden that the European Parliament's Committee on Industry, Research and Energy has described as producing "duplicative, inconsistent, or unclear requirements that deter uptake and delay time to market of new digital and AI-based products."[67]

[67] European Parliament, Committee on Industry, Research and Energy. (2025, October). Interplay between the AI Act and the EU digital legislative framework (Study PE 778.575). European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/STUD/2025/778575/ECTI_STU(2025)778575_EN.pdf

The Digital Omnibus, published in November 2025, reflects the European Commission's acknowledgement that the EU's digital rulebook has become overly complex.[68] The package proposes targeted amendments to EU digital rules, including the AI Act.[69] In July 2025, 56 EU-based AI companies signed a public letter urging the European Commission to pause and simplify parts of the AI Act, warning that compliance costs risked stifling innovation before the industry had a chance to develop.[70] As mentioned, the European Parliament has also voted in favour of postponing the application of certain provisions of the AI Act.

Against this shifting regulatory environment, EU policymakers should refrain from reopening the established TDM exception. The current TDM policy remains one of the few stable and effective pillars of Europe's AI regulatory architecture. Overhauling a system that is already working as intended would introduce further legislative uncertainty, slowing the development and deployment of AI across the EU economy.

Recommendation 4: Exercise Caution on Top-Down Licensing and Allow the Market to Develop Solutions

The EU should not introduce mandatory licensing regimes for AI training data, whether through direct legislative intervention or through the expansion of collective licensing frameworks. Proposals for a centralised licensing registry or a statutory collective licensing scheme for AI training should be assessed against the full range of transaction costs they would impose and the disproportionate burden they would place on smaller companies. The EU should instead adopt a 'wait and observe' approach, monitoring the development of private licensing markets and proposing a policy only where clear market failures can be demonstrated.

A statutory licensing scheme would require the EU to solve, by administrative fiat, pricing problems that markets have not yet fully solved, and for good reason. Lump-sum fees bear no relationship to the actual value generated by AI systems. For example, a news article that helped train a model to understand sentence structure contributes something entirely different from a scientific paper that shaped its reasoning about chemistry and yet both would attract the same fee under any flat-rate scheme. Revenue-sharing is no easier either. A large AI company trains a model, integrates that model into a product, sells that product to businesses, and those businesses use it to generate their own revenue. At each step, value is created. By the time revenue reaches the top of this chain, it is very difficult to trace how much of it originated from any specific piece of training data.

The good news is that the private sector is developing market-based solutions. Major publishers, including Axel Springer, News Corp, the Financial Times, and others, have moved beyond litigation

---

[68] European Commission. (2025, November 19). Simpler EU digital rules and new digital wallets to save billions for businesses and boost innovation [Press release IP/25/2718]. https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2718

[69] Moreno, N., & Melton, C. (2026, January 26). The 2025 European Commission EU Digital Omnibus package: A practical guide and explainer. Kennedys Law. https://www.kennedyslaw.com/en/thought-leadership/article/2026/the-2025-european-commission-eu-digital-omnibus-package-a-practical-guide-and-explainer/

[70] Besliu, R. (2025, November 13). What's driving the EU's AI Act shake-up? Tech Policy Press. https://techpolicy.press/whats-driving-the-eus-ai-act-shakeup/

to sign direct licensing deals with foundation model developers such as OpenAI and Mistral.[71] These developments demonstrate that when the market provides a path to legal certainty, AI companies are willing to pay for it. The appropriate policy response is to ensure that this market can develop without being pre-empted by a mandatory regime that would artificially shape current market structures and potentially foreclose efficient solutions.

Recommendation 5: Reframe the Policy Problem: It is a Market Competition Issue, Not a Copyright Issue

The primary economic concern raised by GenAI is a market-structure problem, not a copyright infringement problem. If there is a market failure, any policy responses should be designed accordingly, focusing on the market and job displacements rather than on expanding copyright protection in ways that restrict data access and harm the broader EU economy.

The creative sector's concern about GenAI is legitimate and deserves a serious policy response. But the nature of that concern has been systematically mischaracterised in the policy debate, with significant consequences for regulatory design. The core issue is not that AI systems copy protected works in the conventional sense. AI training involves the transformation of inputs into abstract statistical representations – a process that extracts unprotected elements such as factual information, linguistic regularities, and structural patterns, rather than reproducing protected expression. In this context, the idea-expression distinction, which is foundational to copyright doctrine, provides significant protection against outputs that are substantially similar to specific works.

The real concern is different: AI systems can generate outputs that are functionally similar to human-created works, potentially competing in the same markets and reducing demand for human-produced content. This is an economic displacement problem. It is analogous to the impact of television on theatre, or of digital streaming on physical music sales. These transitions were disruptive, but they were not resolved by restricting access to the enabling technology. They were managed through market adaptation, new business models, and, where necessary, targeted interventions such as public subsidies for the arts.[72]

Treating AI-driven displacement as a copyright issue leads to the wrong policy response. One in which the government action is focussed towards expanding copyright protection, which restricts AI training and effectively imposes a tax on AI development across the EU.

---

[71] Thornhill, J. (2025, February 27). Help is coming in the AI copyright wars. Financial Times. Available at: https://www.ft.com/content/b98979ba-6ae7-4490-97a9-127381440b1f

[72] This economic tension explains the intervention by the French Competition Authority against Alphabet/Google. In that case, Google was fined EUR 250 million on grounds relating to the effectiveness of rights and the practical possibility for creators to 'opt-out' of LLM training. The authority found that Google had utilised content from press agencies and publishers to train its foundation model (Bard/Gemini) without sufficient notification or fair negotiation. This adjudication suggests that when AI models utilise structured work to achieve re-retention and generalisation, the dispute shifts from one of technical infringement to one of market dominance. In this view, the problem is not the machine learning from facts or ideas, which copyright does not protect, but rather the use of that knowledge to compete directly with the original data providers in a way that disrupts existing market structures. see: Autorite de la concurrence. (2024, March 20). Related rights: the Autorité fines Google €250 million. Available at: https://www.autoritedelaconcurrence.fr/en/article/related-rights-autorite-fines-google-eu250-million

Recommendation 6: Learn from Countries that Adopt an Innovation-First Framework

Jurisdictions that adopt broad, technologically neutral exceptions for AI training enable the accumulation of modern economic endowments such as data and consequently are able to attract AI investment and developed more competitive domestic AI ecosystems. Singapore, Japan or the US are examples of countries that have features from which the EU can learn.

Singapore uses a "computational data analysis" (CDA) exception that permits the use of copyrighted works for data analysis in both commercial and non-commercial contexts. Crucially, this law explicitly prevents contractual overrides. By legally nullifying any contracts that attempt to restrict data analysis, it provides AI developers with robust legal certainty and guaranteed access to training data.

Under Article 30-4 of its Copyright Act, Japan allows the use of copyrighted works for 'information analysis,' including AI training, regardless of whether the purpose is commercial. The primary condition is that the use must not involve the 'enjoyment' of the work's expressive content. This broad interpretation has fostered a highly flexible legal environment for machine learning, directly contributing to the country's AI development.

In this context, the EU's TDM exception is not exceptional. Moreover, as mentioned previously, when it comes to data, EU's TDM own guardrails and EU digital legislation make the EU a comparatively restrictive jurisdiction for AI training. European policymakers must acknowledge these existing limitations and draw lessons from the more flexible, innovation-friendly approaches of its global competitors.