DATA REGULATION AND POLICY IMPLICATIONS OF AI ADOPTION

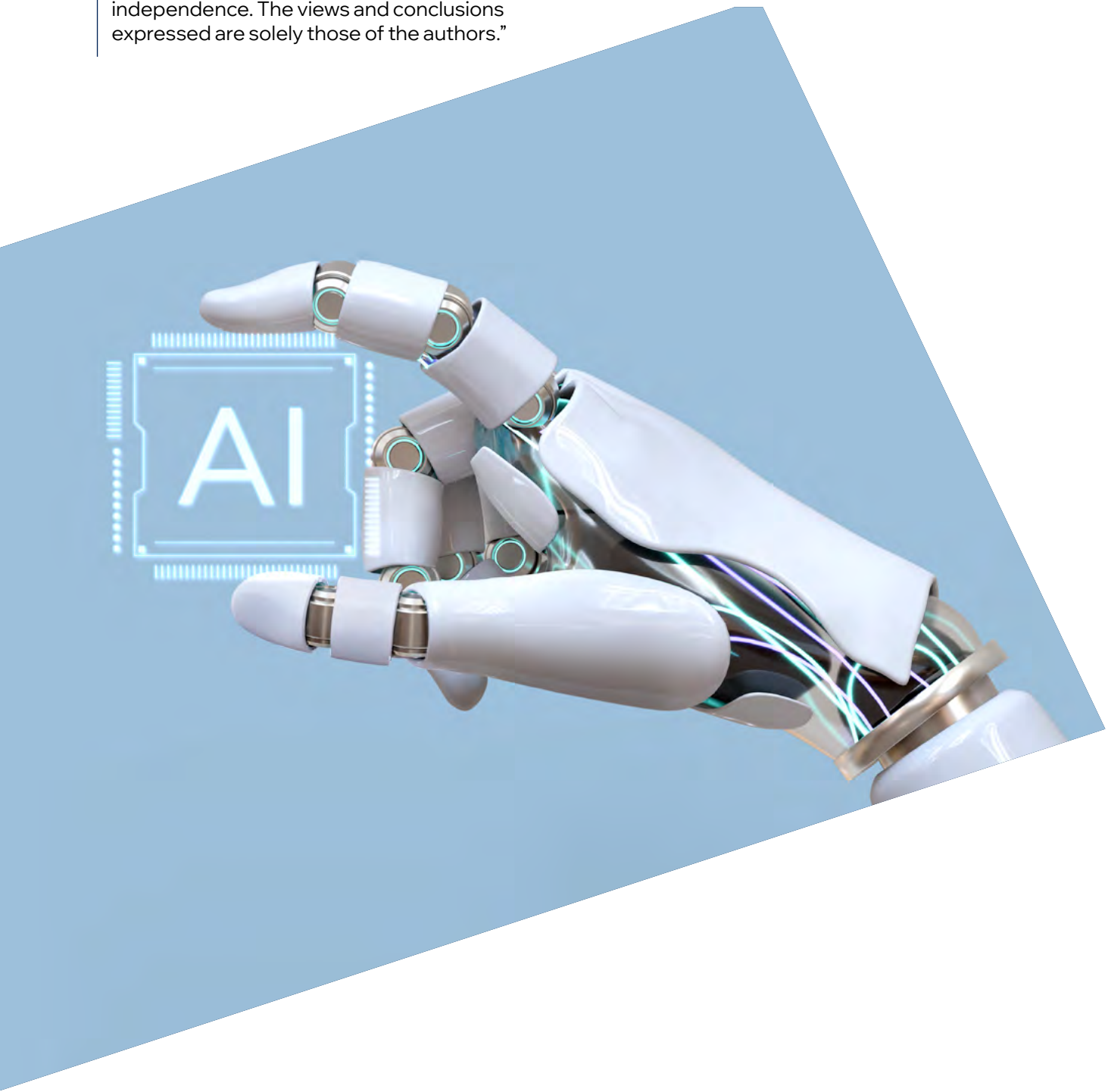
2026

SDA Bocconi
SCHOOL OF MANAGEMENT

DEVO LAB - DIGITAL ENTERPRISE VALUE AND ORGANIZATION
PLATFORM ECONOMY AND REGULATION MONITOR

sdabocconi.it

“This research was sponsored by Google Denmark ApS but conducted with full academic independence. The views and conclusions expressed are solely those of the authors.”



RESEARCH TEAM

Roberta Pisani (SDA Bocconi)
Assistant Professor of Practice
of Digital Transformation
roberta.pisani@unibocconi.it

Carmelo Cennamo (CBS & SDA Bocconi)
Fellow
Digital Transformation
carmelo.cennamo@unibocconi.it

EXECUTIVE SUMMARY

The performance and social value of Artificial intelligence (AI) and generative AI (GenAI) technologies depend critically on access to high-quality, timely, diverse and reusable data. At the same time, AI deployment raises substantial concerns related to privacy, accountability, safety, and fairness. Although much of the European Union's regulatory framework for digital technologies predates the advent of generative AI, the EU has continued to pursue an ambitious regulatory approach that combines horizontal and sector-specific instruments — notably the AI Act, the GDPR, the Data Act, and the Data Governance Act — aimed at enabling trustworthy AI while safeguarding fundamental rights.

This report examines how this layered regulatory system shapes the development and deployment of AI, particularly high-risk AI systems, and identifies the key data-related regulatory and policy bottlenecks—such as barriers to data sharing, portability, and quality that currently limit scalable and effective adoption. While the EU framework is comprehensive and principled, its interaction across legal domains and sectors generates frictions that constrain access to data, increase compliance costs, and introduce uncertainty for organizations seeking to develop and deploy AI at scale.

Three structural challenges emerge. First, **fragmented consent and authorization regimes**, combined with strict purpose-limitation requirements, create legal uncertainty around secondary data use — an essential feature of modern AI development. Second, **limited interoperability and weak data standardization** — both technical and legal — impede data portability, reuse, and cross-border collaboration, reducing the feasibility of large, representative datasets. Third, **governance**, **incentives**, and legal and liability uncertainty discourage data sharing: data holders often face disproportionate legal and reputational risks relative to expected benefits, particularly in highly regulated sectors.

These frictions are most acute in domains such as healthcare, financial services, pharmaceuticals, and critical infrastructure, where "high-risk" AI systems must comply not only with the AI Act and GDPR but also with demanding sectoral regimes (e.g., MDR, EHDS, Solvency II, Clinical Trials Regulation). In these settings, organizations face clear trade-offs between technological ambition and regulatory feasibility. As a result, many AI applications are viable only under tightly controlled conditions, relying on interpretable models, modular data architectures, and advanced privacy-preserving techniques. As a result, **the way in which regulation operationalizes a risk-based approach — particularly the degree of proportionality and differentiation across risk levels —does not merely constrain AI adoption; it also shapes its architecture**.

To address these challenges, the report advances a set of policy recommendations centered on enabling **responsible data access rather than unrestricted data openness**. Key priorities include: clarifying how GDPR principles apply to the collection, secondary use, and reuse of data for AI training purposes; harmonizing and standardizing consent and authorization mechanisms; promoting fiduciary data stewardship through data trusts and certified intermediaries; accelerating interoperable, standards-based data infrastructures (e.g., European Data Spaces); and supporting the deployment of privacy-enhancing technologies through sandboxes, certification, and targeted regulatory and policy incentives.

Overall, the report argues that Europe's AI ambition will be realized not by weakening safeguards, but by **aligning regulation, governance, and technology**. By reducing legal uncertainty, strengthening data governance, and investing in interoperable and privacy-preserving infrastructures, policymakers can enable high-risk AI systems that are both innovative and trustworthy — unlocking economic and societal value while maintaining robust protection of fundamental rights and public interests.

1 INTRODUCTION

Artificial intelligence (AI) and generative AI (GenAI) are transforming economic, social, and technological landscapes across Europe. From healthcare and financial services to energy, manufacturing, and mobility, AI systems increasingly rely on high-quality, timely, and shareable data to deliver accurate, reliable, and innovative solutions. At the same time, AI technologies raise significant questions related to privacy, security, accountability, and fairness. Ensuring the responsible deployment of AI therefore requires navigating a highly complex regulatory system that integrates cross-sectoral and sector-specific legal frameworks.

At the European level, the AI Act (Regulation EU 2024/1689) establishes the first comprehensive AI-specific legal framework, adopting a risk-based approach that prohibits certain practices, imposes stringent obligations on high-risk AI systems, and introduces documentation , safety and security requirements for general-purpose AI models. Complementary regulations, including the GDPR, the Data Act, and the Data Governance Act, govern personal data protection, access, sharing, and governance, shaping the design, deployment, and operationalization of AI across sectors. Together, these instruments create a layered and interconnected regulatory environment that defines when and how AI can be lawfully developed, deployed, and monitored within the European Union.

Despite this structured framework, numerous bottlenecks and frictions persist. Fragmented consent regimes, inconsistent interoperability standards, weak governance structures, commercial restrictions, and liability uncertainties create systemic barriers that limit access to high-quality data, inhibit scalable data exchange, and increase legal and compliance risks. These challenges are particularly acute in sensitive and highly regulated domains such as healthcare, financial services, and pharmaceuticals, where high-risk AI applications must comply not only with the GDPR and AI Act but also with sectoral frameworks such as the Medical Device Regulation, Solvency II, the European Health Data Space, and the EU Clinical Trials Regulation. As a result, organisations face structural trade-offs between technological ambition and regulatory compliance, with many high-risk AI applications viable only under tightly controlled conditions or requiring advanced privacy-preserving and traceable data infrastructures.

Addressing these challenges requires coordinated policy, technological, and organisational solutions. Harmonisation of consent and authorisation mechanisms, interoperable data standards, fiduciary data stewardship through certified intermediaries, and privacy-preserving analytics such as federated learning or synthetic data are critical to enabling secure, auditable, and scalable AI ecosystems. Sectoral case studies — from healthcare diagnostics and hospital data platforms to banking credit scoring and pharmaceutical R&D — highlight the practical implications of these regulatory and operational constraints, illustrating how AI governance, data quality, and ethical considerations intersect with legal obligations.

This report provides a comprehensive overview of the European regulatory landscape for AI, identifies key sectoral constraints and bottlenecks, and proposes actionable policy recommendations to support responsible, efficient, and sustainable AI deployment. Given that the EU AI Act imposes most obligations on high-risk AI systems, integrating regulatory clarity, governance innovation, technical safeguards, and standardized infrastructures would enable policymakers and organizations to unlock the societal and economic potential of high-risk AI while maintaining robust protections for individuals, organizations, and critical public interests.

Next, we first map out the different data-related regulation, at the EU and sector-specific level, that affects the use and deployment of AI technologies. Then, in section 3, we discuss the main bottlenecks and frictions that these regulations can impose on AI deployment. A special focus on GDPR is provided given the far reach of the law, and on few, selective sectoral case exemplars. In section 4 we consider implications for policy and advance some recommendations.

2 EU DATA REGULATORY LANDSCAPE

The primary and direct regulatory reference is the AI Act **(Regulation EU - 2024/1689)**, which for the first time introduces a comprehensive and AI-specific legal framework. The AI Act adopts a risk-based approach: it prohibits certain AI practices deemed incompatible with fundamental EU values, imposes particularly stringent requirements on "high-risk" AI systems, and requires systems with limited risk to have some transparency measures. It also introduces documentation and risk-mitigation obligations for general-purpose AI models. As a result, the AI Act directly affects how AI systems and models must be designed, trained, documented, placed on the market and used.

Alongside the AI Act, other regulations apply that are not AI-specific but nevertheless have a substantial impact on the use of AI, in particular through the regulation of data. The **GDPR (Regulation EU - 2016/679)** applies whenever an AI system processes personal data, thereby affecting the entire AI lifecycle, from training to operational use. It sets limits and conditions for data use, imposes transparency and accountability obligations, and protects the rights of individuals, directly influencing technological and organisational choices related to AI.

The **Data Act** and the **Data Governance Act** further complete the regulatory landscape by governing, respectively, access to and sharing of data, as well as data availability, governance and intermediation mechanisms. These regulations affect AI insofar as they determine which data may be used, by whom and under which conditions, with a direct impact on the ability to train AI models, supply them with up-to-date data and reuse datasets originating from industrial or public sources.

In addition, the evolving EU product liability framework plays a growing role in shaping the deployment of AI systems. The revised Product Liability Directive and the introduction of the AI Liability regime aim to adapt traditional liability rules to digital products and AI-enabled systems, clarifying responsibility and compensation mechanisms in cases of harm caused by defective AI systems or software. These instruments influence risk management strategies, documentation practices and the allocation of responsibilities along the AI value chain.

Finally, recent policy initiatives such as the Digital Omnibus proposal further contribute to the regulatory environment by addressing cross-cutting issues related to data use, innovation and legal certainty. In particular, the recognition of AI training as a legitimate interest under specific conditions seeks to reduce legal uncertainty surrounding data use for AI development, while maintaining appropriate safeguards for fundamental rights.

Taken together, these regulations establish an integrated regulatory system that defines when and how AI can be lawfully developed and used within the European Union. Table 1 below illustrates, for each regulation, its main purpose, its specific application to the use of AI, the key obligations involved and the overall level of compliance effort required. Its purpose is to provide an integrated view of the EU regulatory framework, highlighting how different regulatory instruments – although pursuing distinct objectives – jointly govern the development, implementation and use of AI in both the private and public sectors.

Table 1 **EU DATA POLICY STACK**

Act	Main Purpose	Application and AI Use Scope	Core Obligations
GDPR	Protect personal data and harmonise EU data protection.	Applies to AI whenever the system uses or processes personal data (e.g., user, customer or employee data) for training, operation, or improvement of models. AI use is allowed only with a valid legal basis and respecting data subject rights.	Organisations must process personal data lawfully, transparently, and for specific purposes only. They must inform data subjects about AI use, respect their rights (access, rectification, erasure, objection), assess risks when AI may significantly impact individuals, and adopt security measures to prevent misuse or data breaches.
DATA ACT	Enable access and use of IoT and industrial data; prevent unfair terms; ensure cloud switching; Business to Government (B2G) access.	Relevant to AI when systems rely on data generated by IoT devices or industrial data. Regulates who can access such data and under which conditions, affecting AI training and operation.	Companies must design products and services to allow authorised users access to data, share data upon request fairly and non-discriminatorily, and avoid unfair contractual clauses. They must also ensure the possibility to switch cloud providers without undue obstacles.
DATA GOVERNANCE ACT	Improve availability and trust in data sharing; regulate data intermediaries; enable reuse of protected public-sector data; encourage data altruism.	Relevant to AI when systems rely on data generated by IoT devices or industrial data. Regulates who can access such data and under which conditions, affecting AI training and operation.	Companies must design products and services to allow authorised users access to data, share data upon request fairly and non-discriminatorily, and avoid unfair contractual clauses. They must also ensure the possibility to switch cloud providers without undue obstacles.
AI ACT	Risk-based framework regulating AI systems; bans unacceptable uses; sets requirements for high-risk AI and obligations for limited risk AI systems; includes obligations for General Purpose AI.	Central regulation for AI use. Applies directly to AI development, market deployment, and use in the EU, with obligations depending on system risk level.	• Providers and users must assess and manage risks, use adequate-quality data, document system operations, ensure human oversight, and guarantee accuracy and robustness. High-risk systems require stringent pre- and post-market controls; limited risk AI systems require some transparency measures; general-purpose AI requires documentation and systemic risk mitigation measures.
PRODUCT LIABILITY DIRECTIVE	Modernize liability rules for defective products, including digital and AI-enabled products	Applies to AI systems and software placed on the market that cause damage due to defects	Strict liability for defective products, evidentiary facilitation, documentation and traceability requirements
DIGITAL OMNIBUS PROPOSAL*	Reduce legal uncertainty and foster innovation in the digital economy	Cross-cutting application to digital technologies, including AI, particularly with respect to data use and training	Clarification of lawful bases (e.g. legitimate interest for AI training), alignment of existing digital rules

*Based on ongoing discussions at the EU-level at the time of the writing of this report

Actors	Sanctions	Compliance Burden
• Data controllers and processors • Companies processing personal data • Public sector bodies	• Fines up to €20M or 4% of global annual turnover (Art. 83). • Corrective measures by supervisory authorities, including warnings, bans on processing, and audits.	High: requires prior risk assessments, detailed documentation, ongoing rights management, and carries significant enforcement risk.
• IoT device manufacturers • Cloud service providers • Companies generating or using industrial data • Data holders	• Penalties defined by Member States (administrative fines must be effective, proportionate, dissuasive). • Contractual invalidity of unfair terms.	Medium-high: requires technical system adjustments, contract revisions, and organisational effort to manage access and sharing requests.
• IoT device manufacturers • Cloud service providers • Companies generating or using industrial data • Data holders	• Penalties defined by Member States (administrative fines must be effective, proportionate, dissuasive). • Contractual invalidity of unfair terms.	Medium-high: requires technical system adjustments, contract revisions, and organisational effort to manage access and sharing requests.
• Providers, deployers, importers, distributors of AI systems • Developers of general-purpose AI • Companies using high-risk AI	• Fines up to €35M or 7% of global turnover (highest in EU regulatory landscape). • Market withdrawal of non-compliant AI systems.	Medium-High: requires significant technical and organisational adjustments, continuous documentation, ongoing monitoring, and high accountability, especially for high-risk and general-purpose AI systems, with a compliance effort that is broadly comparable to, though different in nature from, GDPR requirements.
Manufacturers, software providers, importers	Compensation obligations, civil liability	Medium: indirect compliance burden through risk management, documentation and product design
AI developers, deployers, data controllers	To be defined (legislative proposal stage)	Low to medium: primarily legal and organizational adjustments, with limited direct technical obligations

Data sharing and utilization in **highly regulated sectors** such as Banking, Insurance, and Healthcare are strongly shaped by sector-specific regulations that define both limitations and opportunities. Understanding these rules is essential to ensure compliance, data protection, and responsible technological development.

The following section identifies and summarizes key regulatory frameworks impacting data management in several areas, including PSD2 in finance, MDR in healthcare, and the "high-risk" classification under the AI Act. Table 2 offers a concise overview of these regulations, highlighting their implications for data sharing and usage across the relevant sectors.

Banking and Insurance

The banking and insurance sectors operate under a complex European regulatory framework, where data management, sharing, and protection are strictly governed. **PSD2** establishes open banking, requiring financial institutions to share account data with authorized third parties through secure APIs and regulated consent mechanisms. Its objective is to promote competition and innovation while maintaining high security and consumer protection standards. **Solvency II** governs the insurance industry, imposing structured and frequent reporting obligations to supervisory authorities, with a focus on data accuracy, consistency, and auditability. **The Insurance Distribution Directive (IDD)** enhances transparency toward customers, requiring clear, traceable information flows. This framework balances openness with confidentiality: PSD2 encourages controlled data sharing, while Solvency II increases reporting obligations and the need for integrated data architectures and automated reporting. **GDPR** forms the foundation of personal data protection, complemented by the **Data Act**, **Data Governance Act (DGA)**, and **AI Act**. The latter classifies **certain financial systems — such as credit scoring, underwriting, and automated decision-making — as high-risk, requiring stringent risk management, dataset quality, technical documentation, human oversight, and post-market monitoring. Compliance in banking and insurance involves a high operational, technical, and AI governance burden**, with significant financial and organizational implications.

Healthcare

Healthcare organizations handle highly sensitive data and critical technological systems, including medical devices and AI-driven solutions. The **Medical Device Regulation (MDR**) mandates strict requirements on safety, performance, and traceability, requiring technical documentation, clinical evidence, and continuous post-market surveillance. The **European Health Data Space (EHDS)** regulates primary (direct patient care) and secondary (research, policy, innovation) use of health data, ensuring interoperability, access governance, and structured consent mechanisms. MDR, EHDS, and GDPR jointly impose rigorous obligations: consent management, data traceability, Data Protection Impact Assessments (DPIAs), security measures, and controlled secondary data use. The **Data Act** and **DGA** provide additional tools for data portability, interoperability, and trusted intermediaries for secure data sharing. The **AI Act classifies clinical decision support, diagnostic tools, and surgical robots as high-risk, requiring clinical validation, human oversight, and continuous monitoring. Compliance demands integrated governance across clinical, data protection, and AI domains, generating extremely high operational and regulatory complexity**.

Hospitality

Hospitality operators manage extensive personal and behavioral data across the guest journey, including bookings, payments, loyalty programs, Wi-Fi and IoT data, reviews, and information exchanged with OTAs and PMS providers. Data flows often cross borders, increasing compliance complexity. **GDPR** governs guest data, with special attention to profiling, marketing, and consent management. The **Data Act** applies to connected devices, ensuring access and portability of device-generated data, while **DGA** regulates intermediated sharing of aggregated data. **Most AI applications (chatbots, recommendation systems, dynamic pricing) are low-risk, but high-risk AI includes automated recruitment or biometric identification, requiring DPIAs, human oversight, and documentation. Compliance is medium-to-high, necessitating centralized consent management, contractual control of third parties, and AI risk assessments.**

Leisure & Travel

The Leisure & Travel sector processes personal, operational, and biometric data on an international scale, with complex flows among airlines, global distribution systems, travel agencies, and border authorities. **GDPR** protects passenger data, with enhanced requirements for biometric and health-related information. The **Data Act** and **DGA** regulate device-generated data access, portability, and pooled data sharing. **Many AI systems are high-risk, including safety-critical infrastructure components, biometric identification, and automated border control. These systems require risk management, high-quality datasets, human oversight, and continuous monitoring. Compliance demands integrated governance across safety, data protection, and AI frameworks, making the regulatory burden high**.

Manufacturing

Manufacturing relies heavily on IoT telemetry, predictive maintenance, supply-chain data, and worker safety information. **GDPR** applies to personal data, while the **Data Act** promotes access and interoperability of machine-generated data and limits restrictive contractual clauses. **DGA** governs participation in industrial data spaces. **High-risk AI includes robotics and industrial control systems, triggering obligations for risk management, safety testing, documentation, and monitoring. Compliance requires integrated AI governance, contractual oversight, and data security in complex ecosystems. The regulatory burden is high**.

Pharmaceutical

Pharmaceutical companies process clinical trial data, R&D datasets, pharmacovigilance data, and supply-chain records. **GDPR** regulates sensitive health and genetic data, while specific legislation (Clinical Trials Regulation, EMA requirements) ensures data integrity, traceability, and auditability. **DGA** and **EHDS** provide controlled pathways for secondary research use. **High-risk AI includes diagnostics, clinical decision support, drug safety signal detection, and quality control. Requirements encompass risk management, high-quality datasets, documentation, human oversight, and post-market monitoring. Compliance is very high, requiring coordination across data protection, clinical governance, regulatory affairs, and AI risk management**.

Energy & Utilities

Energy and utilities manage critical infrastructures with real-time IoT and OT data. GDPR protects customer and employee data. The Data Act ensures access and portability of smart-meter and device-generated data, while DGA governs pooled energy and environmental data in trusted intermediaries or data spaces. High-risk AI systems—such as those used for grid control, load balancing, and infrastructure management—are subject not only to AI-specific requirements on risk management, human oversight, traceability, and continuous monitoring, but also to stringent sector-specific energy and critical-infrastructure regulations.

Across all sectors, organizations face increasing regulatory complexity, combining data protection (GDPR), structured data sharing (Data Act, DGA), and high-risk AI governance (AI Act), and evolving product liability rules, alongside sector-specific frameworks. Compliance levels range from medium-high in hospitality to very high in healthcare, pharmaceuticals, and energy, requiring ad-hoc advanced governance, risk management, and technological infrastructure.

Table 2 **SECTORAL DATA RESTRICTIONS SUMMARY**

Sector	Key Regulatory Drivers	Main Data Constraints	AI Act Impact (Risk Level)	Application and AI Use Scope	Overall Compliance Burden
BANKING & INSURANCE	PSD2, Solvency II, IDD, GDPR, Data Act, DGA	Mandatory API data sharing (PSD2); strict reporting & data lineage (Solvency II); transparency & consent traceability	High-risk (credit scoring, underwriting, AML models)	Secure APIs, consent management, automated reporting, AI governance integration	High
HEALTHCARE	MDR, EHDS, GDPR, DGA, Data Act	Special category data; strict provenance; clinical validation; controlled secondary-use governance	Very High-risk (diagnostics, clinical decision-making)	Clinical validation, post-market monitoring, EHDS governance, DPIAs	Very High
HOSPITALITY	GDPR, Data Act, DGA	Guest data, cross-border transfers, IoT/smart rooms, vendor-heavy ecosystem	Low–Medium-risk (chatbots, dynamic pricing); High-risk if biometrics used	Centralized consent management, vendor contracts, IoT compliance, biometric DPIAs	Medium–High
LEISURE & TRAVEL	GDPR, Data Act, DGA, aviation & border rules	Sensitive data (biometric, travel history), heavy cross-border flows, mobility telemetry	High-risk (border control, biometric gates, safety components)	DPIAs on biometrics/identity, SCCs for transfers, IoT/mobility compliance	High
MANUFACTURING	GDPR, Data Act, DGA	Industrial IoT flows; telemetry access rights; employee monitoring DPIAs; shared data spaces	High-risk (robotics, process-control systems)	Inventory of machine data, AI safety controls, industrial data-space governance	High
PHARMACEUTICAL	GDPR, Clinical Trials Regulation, EMA rules, EHDS, DGA	Clinical trial data, genetic data; strict traceability; long-term retention; cross-border research	Very High-risk (diagnostics, pharmacovigilance AI)	Consent/legal-basis management, clinical-grade AI governance, secure secondary-use pathways	Very High
ENERGY & UTILITIES	GDPR, Data Act, DGA, critical infrastructure rules	Smart meter data; OT/SCADA telemetry; multi-party grid data sharing	High-risk (critical grid control, safety systems)	AI safety programs, Data Act access rights, DGA intermediaries, OT/IT integration	High–Very High

3 BOTTLENECKS AND FRICTIONS FOR AI

The literature consistently identifies a set of regulatory and policy frictions that significantly hinder the adoption of AI and GenAI by restricting access to high-quality and diverse data, impeding scalable data exchange, and increasing legal and compliance uncertainty. These **bottlenecks are widely documented across academic research, European policy analyses, think tank publications and industry reports, with particular emphasis on fragmented consent regimes, interoperability gaps, weak governance structures and misaligned incentives** (Van Noordt et al., 2023; Şahin & Karayel, 2024; Taeihagh, 2025).

Importantly, these frictions do not operate in isolation. Rather, **they form interconnected and mutually reinforcing systemic barriers that undermine the core foundations of AI and GenAI development, namely access to trustworthy, shareable and portable data**. Fragmented consent frameworks, limited interoperability, deficient data governance, commercial rigidity and incentive misalignment together create a structural bottleneck that constrains scalable AI deployment across sectors and geographical contexts (Van Noordt et al., 2023; Jørgensen et al., 2025; Taeihagh, 2025).

Overcoming these challenges requires coordinated and coherent policy interventions, including the harmonization of consent models, the development of interoperable technical standards, the strengthening of data governance mechanisms and the establishment of clearer liability and compensation frameworks, including in relation to evolving EU product liability rules. Such measures are essential to enable secure, ethical and sustainable data-sharing ecosystems that can effectively support the responsible scaling of AI and GenAI technologies. Over-reliance on FRAND-based access models and weak enforcement of interoperability obligations reduce the effectiveness of data portability rights and undermine the pro-competitive intent of EU digital policy (Jørgensen et al., 2025). As the European Commission (2020) notes, “High contracting and interface costs and fragmented markets reduce data sharing incentives.” These commercial barriers restrict the circulation of data necessary for scalable AI innovation.

Building on a broad review of academic, industry, and policy literature, the following list highlights the top regulatory and data-related bottlenecks most frequently cited as barriers to AI and GenAI adoption across sectors.

Consent Fragmentation and Legal Uncertainty

Fragmented consent regimes and inconsistent legal interpretations across sectors and Member States remain one of the most frequently cited barriers to AI and GenAI adoption. Traditional notice-and-consent models are increasingly misaligned with the dynamic and iterative nature of AI systems, which rely on secondary and often unforeseen data uses for model training, fine-tuning and continuous improvement. These dynamics challenge the GDPR's principle of purpose limitation and its requirements of purpose specification and compatibility. The literature highlights that organizations face significant uncertainty regarding the lawful reuse of data, particularly in cross-border and multi-source contexts, resulting in overly cautious behavior and underutilization of existing data assets (Maryala, 2025; Rana et al., 2024; Şahin & Karayel, 2024). This uncertainty discourages large-scale data sharing initiatives and increases administrative complexity, especially in sensitive domains such as healthcare, where unclear consent requirements and non-standardized practices inhibit interoperability and trust (Kaushik et al., 2024; Bouderhem, 2024). As noted by the European Parliament (2021), "Barriers include uncertainty about rights and obligations in relation to data." This fragmentation ultimately constrains the creation of robust, cross-border datasets essential for scalable AI and GenAI development.

Lack of Interoperability and Common Data Standards

A structural lack of interoperability and harmonized data standards significantly limits data portability and reuse, thereby reducing the volume and quality of data available for AI systems. Studies consistently point to **incompatible data formats, divergent metadata schemas, inconsistent APIs and fragmented documentation practices as key technical barriers** that prevent seamless data exchange across organizations and jurisdictions (Micheli et al., 2023; Schwabe et al., 2024; Borgogno & Colangelo, 2019). This problem is further exacerbated by the overlapping regulatory requirements of the GDPR, AI Act and Data Act, which generate complex compliance environments where technical interoperability does not align with legal interoperability, particularly affecting SMEs and emerging AI startups (Jørgensen et al., 2025). The OECD (2023) explicitly identifies the **"lack of interoperable standards" as a core obstacle to effective data sharing**. The result is reduced dataset scalability, higher integration costs and limited capacity to develop generalizable and high-performance AI models.

Data Quality and Fragmented Governance

Insufficient data quality and weak governance frameworks constitute a critical bottleneck for trustworthy AI deployment. The literature highlights that many organisations lack structured approaches to data sources, lineage tracking, version control and quality assurance, increasing both technical unreliability and regulatory risk (Wach et al., 2023; Li, 2025). Poor data governance is closely linked to biased outputs, reduced model accuracy and erosion of trust in AI systems, particularly in high-stakes domains such as medical AI (Hasan et al., 2024). As a McKinsey's report (2022) notes, "Few companies have the foundational data practices needed to unlock AI value." The absence of clear and operationalised criteria for what constitutes 'high-quality' or 'representative' data further complicates compliance with emerging AI governance frameworks, such as the AI Act, which explicitly require high-quality training datasets. These deficiencies thus directly undermine both the technical performance and regulatory compliance of AI systems.

Commercial and Contractual Barriers

Commercial and contractual frictions, including vendor lock-in, proprietary ecosystems, unclear intellectual property ownership and restrictive licensing practices, significantly hinder data sharing and ecosystem development. Organisations are often constrained by opaque contractual terms and high switching costs, particularly in cloud-based and platform-dominated environments where data access and reuse are controlled by a small number of dominant actors (Borgogno & Colangelo, 2019; European Commission, 2020). **Often, organizations' legacy IT architectures are embedded in specific cloud-based systems that constrain the use of different AI models and experimentation of alternative systems. This reinforces data silos and stifles competition, limiting the emergence of open and collaborative AI ecosystems. Also, it pushes organizations towards micro, use-case experimentation rather than organizational processes' redesign and transformation needed to fully leverage the potential of AI across the entire organization**. Over-reliance on FRAND-based access models and weak enforcement of interoperability obligations reduce the effectiveness of data portability rights and undermine the pro-competitive intent of EU digital policy (Jørgensen et al., 2025). As the European Commission (2020) notes, "High contracting and interface costs and fragmented markets reduce data sharing incentives." These commercial barriers restrict the circulation of data necessary for scalable AI innovation.

Misaligned Incentives and Liability Uncertainty

A persistent lack of incentives to share data, combined with uncertainty over liability for downstream misuse or harmful AI outcomes, further discourages participation in data-sharing ecosystems. Data holders often perceive disproportionate risk compared to potential benefits, especially in regulated sectors where reputational and financial consequences of data misuse are severe (Panagopoulos et al., 2022; Taeihagh, 2025). The absence of clear risk allocation mechanisms and liability frameworks creates a chilling effect on collaborative data initiatives. As stated by a CERRE (2021) report, "A lack of incentives for data holders to enter into data-sharing agreements remains a major barrier." This misalignment undermines the formation of shared data infrastructures and weakens the collective innovation potential required for GenAI scaling.

3.1

GDPR-AI Tension Matrix

As organizations accelerate their adoption of AI and GenAI technologies, the interaction between data protection rules and machine-learning practices becomes increasingly complex. While the GDPR provides a robust framework for safeguarding individuals' rights, AI systems depend on data-intensive processes that frequently push against these regulatory boundaries. Understanding where these tensions arise is essential for designing compliant, scalable and trustworthy AI solutions.

The following **GDPR-AI Tension Matrix** (see Table 3) provides a structured overview of where and how the core requirements of the GDPR collide with the operational and data needs of modern AI systems.

As AI models increasingly rely on large, diverse and continuously evolving datasets (often repurposed across different tasks and enriched through profiling) the foundational GDPR principles of purpose limitation, data minimization, transparency, lawful basis, and constraints on automated decision-making create points of friction that organizations must actively navigate.

This matrix examines the data policy implications of AI adoption within the GDPR framework, linking key regulatory requirements to the operational characteristics of machine learning systems. It highlights how core GDPR principles interact with common AI practices such as large-scale data training, dataset reuse, automated profiling, and the deployment of complex models, while also identifying typical friction points, mitigation strategies, and compliance actions. The relative level of data protection risk associated with each GDPR requirement is indicated in the column "Risk", considering the scale of processing, the nature of the data involved, and the potential impact on fundamental rights.

Table 3 **GDPR-AI TENSION MATRIX**

GDPR Article	Requirement	Typical AI Data need	Friction / Issue	Mitigations	Compliance Actions	Risk
ART. 5(1)(B) PURPOSE LIMITATION	Data collected for specified, explicit, legitimate purposes, not further processed incompatibly.	Large-scale reuse of datasets for multiple experiments, transfer learning across tasks.	Reusing training data for new models may be incompatible with original purpose. Weak legal basis for new profiling.	Purpose layering: separate datasets; obtain broad but specific consent where feasible; legal-basis mapping per purpose; implement purpose tags & enforcement in data catalog.implement purpose tags & enforcement in data catalog.	Purpose registry; data provenance logs; contract clauses for secondary use; Data Protection Impact Assessment (DPIA) showing compatibility.	High
ART. 5(1(C) DATA MINIMIZATION	Adequate, relevant, limited to what is necessary.	Need for large, diverse datasets and many features to avoid bias and improve generalizability.	Tension: "collect more to be fair" vs minimization. Over-collection risk.	Feature selection, dimensionality reduction, aggregated / summary stats, use of synthetic data, targeted sampling, on-device feature hashing, federated learning to avoid central raw-data pooling.	Data inventory showing necessity, justification per dataset, technical notes on minimization, model design rationale.	High
ART. 5(1)(E) STORAGE LIMITATION	Keep only as long as necessary.	Long retention for model retraining, reproducibility, audit trails.	Retention conflicts increase risk (older personal data used beyond original retention).	Retention policies for training subsets, versioned datasets with Time-To-Live (TTL); keep pseudonymized snapshots only; store raw data only if strictly necessary.	Retention schedule; deletion scripts; logs proving deletion; dataset access controls.	Medium
ART. 6 LAWFUL BASIS	Must have lawful basis (consent, contract, legal obligation, vital interest, public task, legitimate interests).	Diverse data sources (third-party, scraped, partners) for training.	Consent may be impractical for large/legacy datasets; legitimate interest balancing tests may fail for sensitive profiling.	Use contract/legal basis for B2B data; model-specific legitimate-interest assessment; where consent used, design granular consent UX and consent-tracking. Consider anonymisation to exit GDPR scope.	Lawful-basis mapping per dataset; Legitimate Interests Assessment (LIA) documentation; consent records; contracts with processors.	High
ART. 9 SPECIAL CATEGORIES	Stricter rules; explicit consent or special legal basis for health, biometric, genetic data.	Health, biometric, genetic data used for improved models (e.g., diagnostics).	Such data generally prohibited for profiling unless explicit exception applies → severe legal limits.	Avoid using special categories unless necessary; rely on pseudonymised/ aggregated data; obtain explicit, documented consent; rely on public-interest/ legal basis only when available.	Explicit consent forms; Data Protection Officer (DPO) sign-off; clinical trial approvals/ ethics; Data Protection Impact Assessment (DPIA).	Very High
ARTS. 12–14 TRANSPARENCY & INFORMATION	Inform data subjects how their data is used.	Complex ML models & many training sources; explainability limits.	Difficulty meeting transparency obligations (what data used, automated profiling purposes).	Produce clear, layered privacy notices; model cards and data sheets; training-data summaries; simplified explanation of profiling outcomes.	Published model cards; privacy notice updates; records of communications; transparency logs.	Medium–High

Table 3 **GDPR-AI TENSION MATRIX**

GDPR Article	Requirement	Typical AI Data need	Friction / Issue	Mitigations	Compliance Actions	Risk
ART. 22 AUTOMATED INDIVIDUAL DECISION-MAKING	Right not to be subject to solely automated decisions producing legal/ equivalent effects; right to human involvement & explanation.	High-impact automated profiling (credit, hiring, insurance pricing) or opaque black-box models.	Prohibition/ constraints on fully automated high-impact decisions; need for human review, contestability.	Human-in-the-loop workflows; decision review logs; fallback manual processes; pre-deployment testing to reduce automation; offer meaningful explanations and remediation channels.	Records of human review; logging of automated decisions; contestation procedures; technical documentation.	High
ART. 25 DATA PROTECTION BY DESIGN & DEFAULT	Integrate safeguards from design phase.	Often not followed in agile ML pipelines; data scientists iterate fast without privacy by design.	Late-stage retrofitting of privacy controls; non-compliant model builds.	Integrate privacy gates, model approval workflows, mandatory DPIA signoff, privacy-preserving ML methods.	Design docs, privacy-by-design checklists, approvals in Continuous Integration (CI)/ Continuous Delivery (CD).	Medium
ART. 32 SECURITY	Appropriate technical/ organizational measures (encryption, access controls).	Large sensitive datasets at rest/ processing; risk of exfiltration.	Higher breach impact leading to fines and reputational harm.	Encryption at rest/ in transit, role-based access, secure enclaves, auditing, Single Sign-On (SSO), sequestered environments for training.	Security policies, pentest reports, incident response plan, access logs.	High
ART. 35 DPIA	DPIA required for high-risk processing (systematic profiling, large-scale processing).	Most large-scale profiling/ML projects likely require DPIA.	Failure to carry out DPIA → regulatory action; unassessed high risk.	Conduct DPIAs early; include technical and organisational mitigations; re-assess post-deployment; involve DPO.	DPIA documents, mitigation tracker, monitoring plan.	High
ART. 15–20 DATA SUBJECT RIGHTS: ACCESS, RECTIFICATION, ERASURE, PORTABILITY, OBJECTION	Rights to access, correct, erase, port, object.	Models trained on personal data may embed patterns; subject requests affect model inputs/outputs.	Right to erasure conflicts with model reproducibility; portability hard for model-learned features; objection to profiling impacts model utility.	Minimise identifiable training data; maintain mapping to remove affected records and, if necessary, retrain or use incremental unlearning methods; document processes to handle portability requests.	Procedures for subject requests; logs of actions; tools for data removal; policy on model retraining/ unlearning.	High
RECITALS & **ART. 4** DEFINITIONS	Definitions (e.g., personal data, profiling) that shape scope.	Ambiguity about when model outputs are personal data (re-identification risks).	Edge cases where model outputs or embeddings may re-identify individuals → broader GDPR scope.	Risk assessments re: re-identification; use differential privacy; embedding protections; restrict release of model outputs that leak Personally Identifiable Information (PII).	Re-identification risk assessments; testing reports; DP metrics.	Medium–High

3.2

Sectoral Constraints Case Notes

Sector-specific regulatory frameworks introduce an additional layer of obligations that frequently exceed the horizontal protections established under the GDPR and materially shape the feasibility of deploying high-risk AI systems.

As highlighted in Section 1.2, sectors such as healthcare, financial services, insurance and pharmaceutical operate under regulatory instruments (including the Medical Device Regulation, Solvency II, critical-infrastructure requirements, aviation and border-control rules, and the Data Act) that impose stringent expectations for data quality, traceability, auditability, and operational safety.

These sectoral regimes restrict not only the types of data that may be collected and repurposed for model training, but also the conditions under which such data may be processed, frequently requiring higher levels of explainability, reproducibility, governance, and liability management than those set out under the GDPR.

As a result, organisations in these domains are often confronted with structural trade-offs between regulatory compliance and technological ambition, with certain high-risk AI applications becoming viable only under tightly controlled conditions or, in some cases, precluded altogether due to prohibitive risk or legal incompatibility.

Case Note 1
Healthcare

In the healthcare sector, the regulatory environment is even more exacting. High-risk AI systems supporting diagnostic or therapeutic decisions must comply with the Medical Device Regulation (MDR), which requires clinical evaluation, post-market surveillance, technical documentation, and full traceability of decision processes. Accuracy alone is insufficient. In fact, interpretability and transparency are critical to enable clinicians and regulators to understand how AI-generated conclusions are reached. The European Health Data Space Regulation (EHDS) further governs the secondary use of health data, imposing strict requirements on interoperability, provenance, completeness, and quality. Large, heterogeneous datasets essential for training advanced AI models are therefore subject to significant constraints, as GDPR and EHDS rules limit cross-use beyond the purpose for which data was collected. Hospitals, for example, cannot automatically incorporate anonymized imaging data collected for routine diagnostics into AI training datasets intended for predictive analytics unless patients have provided explicit consent or the secondary use is legally authorised. Every data point must be fully traceable to its origin, including metadata on collection date, clinical context, and any pre-processing applied. Similarly, laboratory results from multiple institutions must be harmonized to conform to EU-wide standards such as HL7 International FHIR - Fast Healthcare Interoperability Resources, ensuring safe and reliable reuse. Data quality and completeness are also critical. Missing demographic or clinical information must be corrected or excluded to prevent biased predictions and ensure patient safety. **Operationally, healthcare organisations prioritise highly interpretable AI models supported by clinical evidence, impose controlled model updates requiring revalidation and regulatory notification, and implement** strict governance frameworks for data access and auditing. Opaque or continuously adaptive AI architectures may be non-certifiable or economically unfeasible due to these extensive obligations.

Case Note 2
Banking & Insurance

In financial services, AI applications ranging from credit scoring and fraud detection to underwriting and risk modelling must comply with sectoral regulations that extend beyond the baseline requirements of the GDPR. Frameworks such as Solvency II, PSD2, and the Insurance Distribution Directive (IDD) introduce obligations for secure system architectures, robust audit trails, accurate supervisory reporting, and fully traceable decision-making processes. Solvency II, for instance, mandates full model auditability, extensive documentation, and data integrity sufficient to withstand supervisory scrutiny. As a direct consequence of these regulatory obligations, the use of opaque, black-box AI models is significantly constrained, as their internal logic and decision pathways are often not fully interpretable, reproducible, or auditable. For example, an AI system supporting underwriting or claims management may generate consistent outcomes, yet fail to provide the transparent rationale, documented assumptions, and traceable data lineage required to demonstrate compliance during supervisory reviews under frameworks such as Solvency II or IDD. PSD2 imposes stringent conditions on the use and reuse of payment-account data, requiring explicit consent, strong authentication, and detailed traceability. Even though broad behavioural datasets could theoretically enhance predictive models, regulatory frameworks often prohibit repurposing data for secondary applications such as new profiling or marketing campaigns. IDD further reinforces expectations of fairness, transparency, and consumer information, which limits the deployment of non-interpretable or opaque decision systems in insurance contexts. A particularly salient aspect of regulatory design is the limitation on cross-line data reuse. Banks are generally prohibited from cross-referencing mortgage, investment, and insurance portfolios without explicit consent and supervisory approval. Similarly, insurers must maintain separation between underwriting, claims, and investment advisory datasets to prevent conflicts of interest and protect customer privacy. These data silos shape both technical and operational AI design. **In practice, institutions adopt modular data pipelines to maintain compliance, ensuring that sensitive datasets remain siloed unless lawful access is granted**. Every input, transformation, and prediction must be fully traceable, allowing supervisors and consumers to audit decisions, challenge outcomes, and enforce accountability. Interpretable or hybrid modeling approaches — such as rule-based systems, generalized linear models, or machine-learning models augmented with explainability and governance layers — are therefore preferred, even when this entails trade-offs in predictive performance, flexibility, or scalability compared to fully black-box models. In this context, continuous monitoring, structured reporting, and human oversight remain essential to ensure sustained regulatory compliance as well as operational reliability over time.

Case Note 3
Pharmaceutical

The pharmaceutical sector exhibits similar, if not more stringent, restrictions due to the direct implications of drugs and therapies on patient safety. High-risk AI applications in drug discovery, clinical trial analysis, and pharmacovigilance must comply with frameworks including Good Clinical Practice (GCP), European Medicines Agency (EMA) guidelines, and the EU Clinical Trials Regulation (CTR), alongside GDPR and data-sharing directives. Clinical trial and patient data cannot be shared with insurers, marketing teams, or other commercial divisions without explicit consent or regulatory justification. Pseudonymization or anonymization is often required for AI training, but must preserve clinical relevance to avoid compromising trial outcomes. Pharmacovigilance AI systems must provide explainable outputs for adverse event detection, ensuring that regulators can trace every flagged signal back to its source data, transformations, and model rationale. Data governance requirements under the EU Data Act and GDPR mandate strict segregation of datasets, role-based access

controls, and comprehensive audit logging. This ensures that AI pipelines in drug development remain compliant while preserving the confidentiality and integrity of sensitive patient information. Updates to AI models require careful revalidation to maintain accuracy and regulatory compliance. Limitations on data reuse are particularly significant. In fact, Electronic Health Record (EHR) data cannot be repurposed for commercial purposes, such as marketing or insurance underwriting, without explicit consent. As in healthcare and finance, the operationalisation of high-risk AI relies on modular, auditable architectures that balance innovation with strict adherence to ethical, legal, and regulatory standards.

Across these sectors, cross-selling and secondary data use emerge as highly constrained activities. While AI models could in principle derive additional insights by integrating data across business lines or repurposing existing datasets, legal and ethical constraints impose strong boundaries. Regulatory frameworks prioritise consent, transparency, and purpose limitation over commercial optimisation, requiring organisations to adopt privacy-preserving methods such as federated learning, secure data clean rooms, and synthetic datasets. These approaches allow AI models to benefit from distributed or multi-source data while ensuring that raw or identifiable information is never exposed outside authorised environments. In practice, the combination of modular data pipelines, strict provenance tracking, interpretable models, and privacy-enhancing technologies provides a pathway to leverage high-risk AI in regulated sectors while remaining compliant. Organisations must carefully design AI systems that enable operational efficiency, predictive power, and innovation, yet remain fully aligned with supervisory expectations, ethical obligations, and the constraints imposed on cross-selling and data reuse. This careful balancing act defines the operational and technical architecture of high-risk AI systems across finance, healthcare, and pharmaceuticals, where regulatory accountability, transparency, and trustworthiness are as critical as predictive performance.


AI

4 POLICY IMPLICATIONS & RECOMMENDATIONS

High-risk AI systems critically depend on timely access to high-quality data. Yet, across sectors such as healthcare, finance, energy, manufacturing, transportation, and public services, organisations continue to face a complex web of legal, technical, and organisational barriers that inhibit efficient data sharing and cross-entity collaboration. These constraints stem not only from strict regulatory requirements—such as fragmented consent mechanisms, purpose limitation rules, inconsistent data portability, and restrictions on secondary use—but also from structural and operational challenges, including heterogeneous data formats, limited interoperability, and organisational hesitancy to share sensitive or proprietary information.

Addressing these frictions requires policies that go beyond ad hoc or bilateral arrangements, fostering trust-managed, interoperable, and technologically safeguarded ecosystems for data access. Emerging evidence from European Data Spaces, data trusts, privacy-enhancing technologies, and secure computation infrastructures demonstrates that a coordinated mix of governance, technical safeguards, and standards can create legally compliant, operationally feasible, and auditable environments for high-risk AI. Building on these insights, several high-impact policy recommendations emerge:

1 Clarify and harmonize GDPR and related data protection provisions for AI

Ambiguities in the application of GDPR, especially concerning secondary use, purpose limitation, and lawful bases for AI training, create friction for multi-party collaboration. Policymakers should provide sector-specific guidance clarifying how GDPR interacts with privacy-preserving technologies (PETs), federated learning, synthetic data, and secure computation. Clear rules would reduce legal uncertainty, streamline approvals, and facilitate responsible AI innovation while maintaining strong privacy protections.

This recommendation is supported by OECD (2021) analyses and recent academic work on policy-driven AI, which highlight that regulatory ambiguity around GDPR purpose limitation and lawful bases is a major barrier to scalable, multi-party AI development, particularly when using privacy-preserving technologies such as federated learning and secure computation.

2 Standardize consent and authorisation mechanisms

Fragmented consent procedures, coupled with inconsistent data-access authorisation, slow research and increase operational costs. While legitimate interest is emerging as a central legal basis for AI training, including under recent EU initiatives (e.g. the Digital Omnibus), standardized, machine-readable consent frameworks, aligned with purpose-based access models, would allow data subjects and organisations to grant flexible, auditable, and legally compliant permissions, particularly in highly regulated sectors such as health. These mechanisms could also be integrated into governance-driven models, such as data trusts or certified intermediaries, to ensure consistency across sectors and jurisdictions.

Evidence from trusted research environments and initiatives such as Gaia-X shows that standardized, machine-readable consent and authorisation frameworks significantly reduce transaction costs, approval delays, and compliance risks in cross-organisational data sharing.

3 Promote fiduciary stewardship through data trusts and certified intermediaries

Data trusts, cooperatives, and certified intermediaries provide institutional structures that address trust deficits, fragmented governance, and inequitable power dynamics between data holders and users. Policies should incentivize the creation and operationalisation of these entities through legal recognition, funding support, and sector-specific guidance. This would enable multi-party data collaboration under clear fiduciary duties, enforceable usage policies, and transparent accountability mechanisms.

Policy research and pilot implementations of data trusts—particularly in the context of the Data Governance Act and Gaia-X—demonstrate that fiduciary intermediaries can effectively address trust deficits, governance fragmentation, and power asymmetries in data-sharing ecosystems. For example, several Gaia-X-aligned data trust implementations such as EuroDaT, the Mobility Data Space (MDS), and HEALTH-X dataLOFT are already operationalising transparent governance, equitable access, and trusted data exchange across multiple stakeholders in transportation and healthcare domains, providing early evidence of practical success in real-world multi-party data collaboration.

4 Incentivize interoperable, standards-based data infrastructures

European Data Spaces, Gaia-X, and industrial ecosystems demonstrate that interoperability, common ontologies, certification schemes, and standardized APIs are essential for efficient, auditable data sharing. Policy measures should encourage sector-wide adoption of these standards, provide funding and technical support for implementation, and promote transparency artefacts such as model cards, algorithmic impact assessments, and metadata registries. Such measures would strengthen data quality, traceability, and regulatory compliance, enabling reliable cross-sector collaboration.

These large-scale initiatives provide concrete evidence that common standards, shared ontologies, and certification schemes are critical enablers of interoperable, auditable, and regulatorily compliant data exchange across sectors.

5 Support hybrid technological solutions for privacy-preserving analytics

Privacy-enhancing technologies—including federated learning, secure multi-party computation, homomorphic encryption, synthetic data, and differential privacy—enable collaborative AI development without exposing sensitive underlying datasets. Policymakers can accelerate adoption by creating regulatory sandboxes, certification schemes, and financial incentives for organisations deploying these technologies in healthcare, finance, and other sensitive sectors. This approach balances innovation with legal compliance, auditability, and public trust.

Empirical studies and applied research in healthcare, finance, and industrial data spaces indicate that hybrid approaches combining privacy-enhancing technologies—such as federated learning, differential privacy, secure multi-party computation, homomorphic encryption, and synthetic data—with governance and oversight mechanisms offer a practical balance between protecting sensitive data, maintaining analytical usefulness, and ensuring regulatory compliance.

6 Encourage long-term operational sustainability of data-sharing ecosystems

The establishment of secure data environments requires sustained governance resources, technical maintenance, and coordination across multiple stakeholders. Policies should

support operational continuity through multi-year funding programs, institutional support for standardization and certification bodies, and cross-sector partnerships. This ensures that data-access infrastructures remain resilient, legally compliant, and technically robust over time.

Experience from multi-year European data space initiatives and support structures such as the Data Spaces Support Centre shows that sustained funding, institutional governance, and coordinated standardisation are essential to ensure long-term resilience and legal compliance of shared data infrastructures.

7 Foster cross-sectoral and cross-border collaboration

High-risk AI benefits from diverse datasets spanning multiple sectors and jurisdictions. Policymakers should promote harmonized legal frameworks, mutual recognition of certifications, and interoperable infrastructures to facilitate cross-border data sharing. Coordinated European and sectoral initiatives — such as the European Health Data Space[1] and Industrial Data Space[2] — demonstrate the potential of collaborative approaches to accelerate AI innovation while maintaining accountability, privacy, and trust.

Collectively, these recommendations advocate for a multi-layered, hybrid ecosystem in which:

- governance structures define access rights and usage purposes;
- technological solutions enable analytics without exposing raw data;
- synthetic and differentially private datasets provide safe alternatives to sensitive data;
- interoperability standards ensure quality, traceability, and auditability;
- and certified intermediaries maintain accountability and trust.

By combining regulatory clarity, governance innovation, technological safeguards, and standardized infrastructures, policymakers can create an environment where high-risk AI can thrive responsibly and unlock its societal and economic potential while maintaining strong safeguards for individuals and organisations alike.

1 *The European Health Data Space (EHDS) is a cornerstone of the European Health Union and the first common EU data space dedicated to a specific sector as part of the European strategy for data.*
The EHDS Regulation aims to establish a common framework for the use and exchange of electronic health data across the EU. It enhances individuals' access to and control over their personal electronic health data, while also enabling certain data to be reused for public interest, policy support, and scientific research purposes. It fosters a health-specific data environment that supports a single market for digital health services and products. Additionally, the regulation establishes a harmonised legal and technical framework for electronic health record (EHR) systems, fostering interoperability, innovation, and the smooth functioning of the internal market.

2 *Industrial Data Space (IDS) is a digital industrial platform designed to enable secure and sovereign data sharing among different companies and across entire value chains. It connects data from many sectors of the economy into a shared ecosystem in which data flows can be used for smarter services, innovation, and business models, while preserving trust among participants.*

REFERENCES

Borgogno, O., & Colangelo, G. (2019). Data sharing and interoperability in the digital economy: Legal and economic perspectives. European Journal of Law and Economics, 48(3), 451–472.

Bouderhem, A. (2024). Consent and trust dynamics in digital healthcare data ecosystems. Health Policy and Technology, 13(1), 100854.

CERRE. (2021). Data sharing in Europe: Between principles and practice. Centre on Regulation in Europe.

Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market (PSD2), amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010.

Directive (EU) 2016/97 of the European Parliament and of the Council of 20 January 2016 on insurance distribution (Insurance Distribution Directive – IDD).

European Commission (2020). A European strategy for data, COM(2020) 66 final.

European Commission (2021). Fostering a European approach to Artificial Intelligence, COM(2021) 205 final.

European Commission (2022). Proposal for a Regulation on harmonised rules on fair access to and use of data (Data Act), COM(2022) 68 final.

European Commission (2022). Proposal for a Regulation on the European Health Data Space, COM(2022) 197 final.

European Commission. (2020). A European strategy for data. Publications Office of the European Union.

European Data Protection Board (EDPB) (2020). Guidelines 3/2019 on processing of personal data through video devices.

European Data Protection Board (EDPB) (2022). Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk".

European Parliament. (2021). Data sharing and the EU digital governance framework. European Parliamentary Research Service (EPRS).

European Union Agency for Cybersecurity (ENISA) (2021). Cybersecurity considerations for smart grids and critical infrastructure.

Hasan, M., Rahman, S., & Chowdhury, T. (2024). Bias and representativeness in medical AI datasets. Journal of Biomedical Informatics, 150, 104363.

Jørgensen, R. F., Cohen, J., & Mattern, F. (2025). Regulatory complexity and AI innovation in Europe. Policy & Internet, 17(1), 95–112.

Kaushik, A., Verma, P., & Singh, R. (2024). Challenges and opportunities for data sharing in healthcare AI in LMICs. Global Health Informatics Journal, 8(2), 77–91.

Li, Y. (2025). Governance failure in AI data ecosystems. AI & Society, 40(1), 55–69.

Maryala, S. (2025). Rethinking consent models for generative AI systems. Journal of AI Law and Policy, 6(1), 23–41.

McKinsey & Company. (2022). The state of AI 2022. McKinsey Global Institute.

Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2023). The landscape of data and AI documentation approaches in the European policy context. Data & Policy, 5, e24.

OECD. (2021). Interoperability of privacy and data protection frameworks.

OECD. (2023). Data access, sharing and re-use in the digital economy. OECD Digital Economy Papers.

Panagopoulos, G., Kokolakis, S., & Siouna, M. (2022). Legal risk and data-sharing disincentives in regulated sectors. Computer Law & Security Review, 45, 105684.

Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), COM(2021) 206 final, and subsequent amendments.

Rana, S., Gupta, R., & Narayanan, V. (2024). AI governance and consent fragmentation. Technology in Society, 68, 101938.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation – GDPR).

Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices (Medical Device Regulation – MDR).

Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance (Data Governance Act – DGA).

Regulation (EU) 2023/2854 of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act).

Regulation (EU) No 536/2014 of the European Parliament and of the Council on clinical trials on medicinal products for human use.

Regulation (EU) No 575/2013 and Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II).

Şahin, O., & Karayel, D. (2024). Generative artificial intelligence in business: A systematic review. Journal of Business Research, 173, 114612.

Schwabe, D., Meyer, C., & Roth, F. (2024). METRIC-framework for assessing data quality for trustworthy AI. AI Ethics, 4(2), 211–229.

Taeihagh, A. (2025). Governance of generative AI. Policy and Society, 44(1), 1–17.

Van Noordt, C., Medaglia, R., & Tangi, L. (2023). Policy initiatives for artificial intelligence-enabled government: An analysis of national strategies in Europe. Government Information Quarterly, 40(2), 101723.

Wach, K., Weber, C., & Müller, J. (2023). Data governance maturity and AI performance. Information Systems Frontiers, 25(6), 1765–1781.

SDA Bocconi School of Management
Via Sarfatti, 10 - 20136 Milano, Italy
tel +39 02 5836 6605-6606
email: info@sdabocconi.it
www.sdabocconi.it

Follow SDA Bocconi