

AI Spaces

From Silos to Sovereign, Cross-Organizational Intelligence

IN COOPERATION WITH



UTokyo

FUJITSU

Management Summary

This paper proposes AI Spaces as a framework for enabling AI systems to collaborate across organizational boundaries under real-world constraints of data sovereignty, confidentiality, regulation, and misaligned incentives.

Many of today's industrial and societal challenges cannot be resolved within the information held by any single organization. Responding to demand fluctuations in supply chains, accelerating product development in manufacturing, enhancing diagnostic support in healthcare, and optimizing operations in next-generation smart buildings — all of these require the secure, sovereignty-preserving utilization of data and knowledge distributed across multiple organizations. AI Spaces address these challenges by integrating Data Spaces infrastructure with AI, enabling cross-organizational intelligence without the need for centralized data aggregation.

Three fundamental patterns of cross-organizational AI collaboration underpin AI Spaces. The first is collaborative model development, where organizations jointly improve models through federated learning while retaining full control over their data. The second is inter-organizational model inference, enabling AI systems to perform reasoning by selectively accessing distributed, sovereign data at inference time. The third is autonomous agent collaboration, where multi-agent systems coordinate decisions and actions across organizational boundaries. These patterns are applicable across domains, and real-world systems frequently combine more than one.

The sustainable operation of AI Spaces cannot be guaranteed by technical feasibility alone. This paper identifies three mutually reinforcing institutional conditions. The first is incentive design: organizations sustain participation not only through financial reward, but through non-monetary value such as access to shared data, participation in governance, and reputation derived from visible contributions. The second is technology standardization: without interoperability at the interaction layer — spanning agent communication protocols, semantic frameworks, policy languages, and trust and identity infrastructure — the network effects essential to AI Spaces cannot be realized. The third is quality management, centered on the AI Bill of Materials (AI BOM), which ensures the transparency, traceability, and auditability of composite AI systems.

At the technology layer, concrete approaches including Federated RAG (F-RAG), the Secure Inter-Agent Gateway, Digital Rehearsal, and Optimization under Incomplete Information demonstrate practical feasibility across diverse domains, from pharmaceutical logistics to healthcare and smart buildings.

AI Spaces represent essential infrastructure for overcoming data silos and harnessing collective intelligence across organizations in a secure, trustworthy, and scalable manner. Realizing this vision requires coherent effort across both technical architecture and institutional design.

Contents

1. Use Cases	4
1.1. Automated Delivery Adjustment in Supply Chain Management	4
1.2. New Product Development: AI-Enabled Data Spaces as Catalysts for Collaborative Innovation	5
1.3. Healthcare and Medical Domains	6
1.4. Next-Generation Smart Buildings	7
2. AI Spaces: An Overview	8
2.1. Three Patterns of Collaborative AI	8
2.2. Three Conditions for Sustainability	9
3. Economic Incentives and Compensation Mechanisms	10
3.1. Incentive Challenges in AI Spaces	10
3.2. Objects of Rights in Collective AI Systems	10
3.3. Incentive and Compensation Mechanisms in AI Spaces	11
4. Technology Standardization and Global Interoperability	11
4.1. Key Standardization Domains for AI Spaces	11
4.2. Global Harmonisation and Ecosystem Alignment	12
5. Quality Management: AI Bill of Materials	12
6. Development Approaches for AI & Data Spaces Integration	13
6.1. Next-Generation Connector Technologies	13
6.2. Retrieval-Augmented Generation (RAG) and Federated RAG (F-RAG)	13
7. Sovereign Multi-Agent Coordination	15
7.1. Challenges in Multi-Agent Collaboration	15
7.2. Enabling Technologies	16
8. Conclusion	16

Authors

Fraunhofer ISST

Boris Otto

Tobias Moritz Guggenberger

Julia Pampus

Fujitsu Limited

Takahide Matsutsuka

Janosch Haber

The University of Tokyo

Noboru Koshizuka

1. Use Cases

This section presents four representative use cases that illustrate how AI and Data Spaces jointly enable cross-organizational intelligence under each participant’s full ownership and control over incentives, constraints and contributions. Together, they form the foundation for the subsequent discussion on the institutional and technical conditions that underpin AI Spaces.

1.1. Automated Delivery Adjustment in Supply Chain Management

In the pharmaceutical supply chain, planning is complicated by limited information exchange, reflecting the inherent restriction that confidential data cannot be shared between companies. Pharmaceutical supply chains however require frequent delivery re-adjustments due to demand fluctuations and sudden order surges caused by widespread disease outbreaks or disasters. As a result, fragmented information among shippers and carriers leads to labor-intensive and sub-optimal ad-hoc coordination.

This use case demonstrates a concrete and operational proof-of-concept in which multiple AI agents collaborate across organizational boundaries under strict information-sharing constraints, showing that cross-company optimization is already feasible using AI-enabled Data Spaces.

In this case study, a shipper-side AI agent and multiple carrier-side AI agents were shown to automatically and rapidly optimize updated delivery plans by securely exchanging only minimal information, such as proposal satisfaction and acceptance status. This mechanism

is enabled by two novel technical solutions: Optimization under Incomplete Information, which derives supply-chain-wide solutions through proposal–response exchanges without disclosing sensitive information such as detailed costs or constraints, and a Secure Inter-Agent Gateway, which protects confidential information and privacy through decentralised AI collaboration under dedicated guardrails (see Section 7. Sovereign Multi-Agent Coordination for more detail).

The FY2025 Promotion Theme Project final report under COCN—a public-private partnership framework aimed at strengthening Japan’s industrial competitiveness—positions this use case as a concrete example for realizing resilient supply chains (COCN, 2026, see Figure 1: Supply chain optimization proposed in COCN project), highlighting its ability to enable optimized inter-company coordination without human intervention during sudden demand spikes, reducing delays and excessive costs. Beyond pharmaceuticals, this use case also provides a model applicable to other industries, with ongoing large-scale validation across end-to-end supply chains.

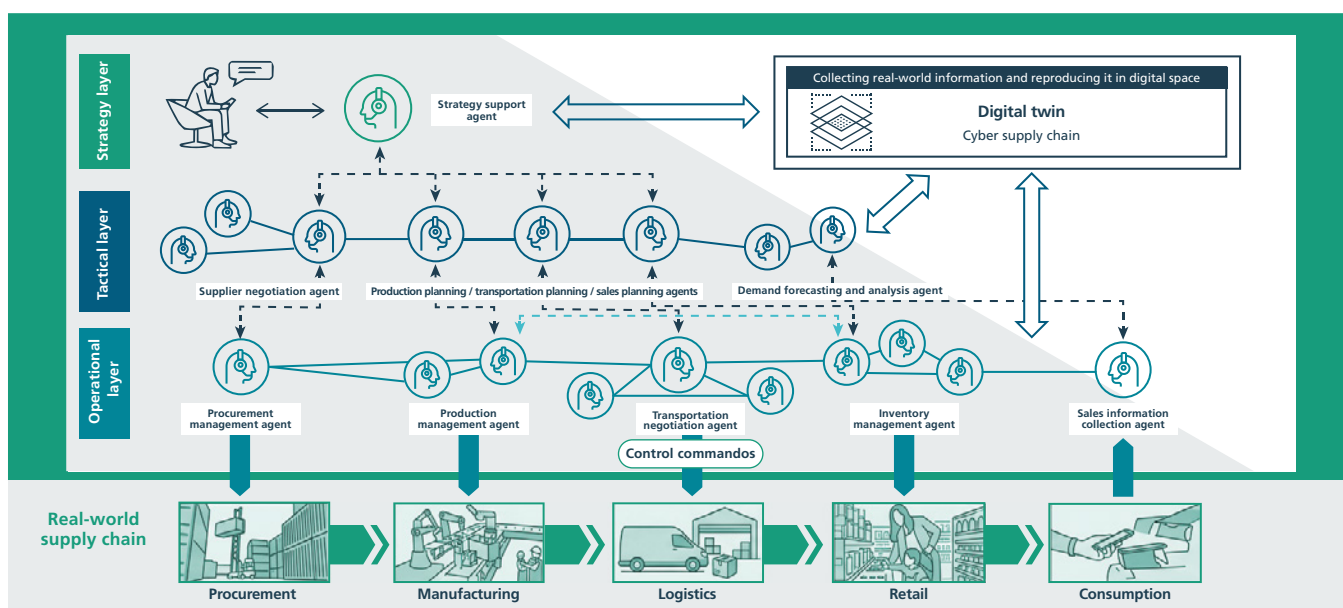


Figure 1: Supply chain optimization proposed in COCN project

1.2. New Product Development: AI-Enabled Data Spaces as Catalysts for Collaborative Innovation

Building on the supply chain example, this use case illustrates how AI-enabled Data Spaces extend from operational coordination to knowledge-intensive, cross-organizational product development. In industries such as automotive, aerospace, and industrial manufacturing, innovation is constrained by fragmented engineering data locked in organizational silos. AI-enabled Data Spaces address this challenge by enabling secure, sovereign data sharing while mobilizing collective intelligence for faster, higher-quality product development. The approach is based on three pillars (Fraunhofer ISST, 2025).

The first pillar is accelerated development through horizontal data integration. Interoperable Data Spaces, based on standardized protocols, enable real-time synchronization of engineering data across organizations. Multi-agent AI systems automatically manage dependencies and propagate design changes, compressing iteration cycles from months to weeks and reducing R&D costs, while enabling broader design exploration and SME participation in innovation networks.

The second pillar is enhanced innovation through collective intelligence. Federated Knowledge Networks connect engineering artifacts, regulations, and design rationales, enabling design reuse, early conflict detection, and large-scale design

space exploration across organizational boundaries. This results in faster development and higher-quality products that incorporate validated solutions and best practices from across the ecosystem.

The third pillar ensures sovereign collaboration through Federated Governance. Machine-readable usage policies, federated learning, and automated compliance monitoring enable cross-organizational collaboration without centralizing sensitive data, while respecting intellectual property, contractual, and regulatory constraints.

Together, accelerated integration, collective intelligence, and sovereign governance position AI-enabled Data Spaces as essential infrastructure for next-generation product development. Early adoption allows industry leaders to capture the full value of collaborative innovation, while policymakers can leverage Data Space infrastructures to strengthen industrial competitiveness and ensure that the benefits of AI-driven innovation are broadly distributed.

1.3. Healthcare and Medical Domains

In healthcare and medical domains, AI has significant potential to enhance diagnostic support, treatment planning, and personalized medicine. By handling vast medical literature, clinical guidelines, and domain expertise, AI can serve as a powerful intellectual partner for medical professionals, helping them make more informed and timely decisions.

To deliver reliable outcomes, AI must access highly specialized and continuously evolving medical knowledge. Rather than relying solely on pre-trained capabilities, this requires the ability to reference the latest expert knowledge and sensitive, institutionally distributed information at the point of decision-making. This includes clinical records, test results, and other medical data that cannot be centralized or freely shared due to privacy, regulatory, and ethical considerations.

In this scenario, an important mechanism is Federated Retrieval Augmented Generation (F-RAG) which combines AI with a Data Space infrastructure. By utilizing F-RAG, each medical institution can retain full control over its own data while enabling secure and selective access under clearly defined conditions. This allows AI to retrieve and integrate insights from multiple institutions and specialized databases without requiring centralized data aggregation, supporting diagnosis and treatment planning without compromising patient privacy or data sovereignty (see [Figure 2: Healthcare and medical AI](#)).

The practical benefits are substantial: clinicians gain access to broader and more up-to-date knowledge for diagnosis and treatment planning; patients receive care informed by a wider evidence base; and institutions can collaborate meaningfully while maintaining trust and regulatory compliance.

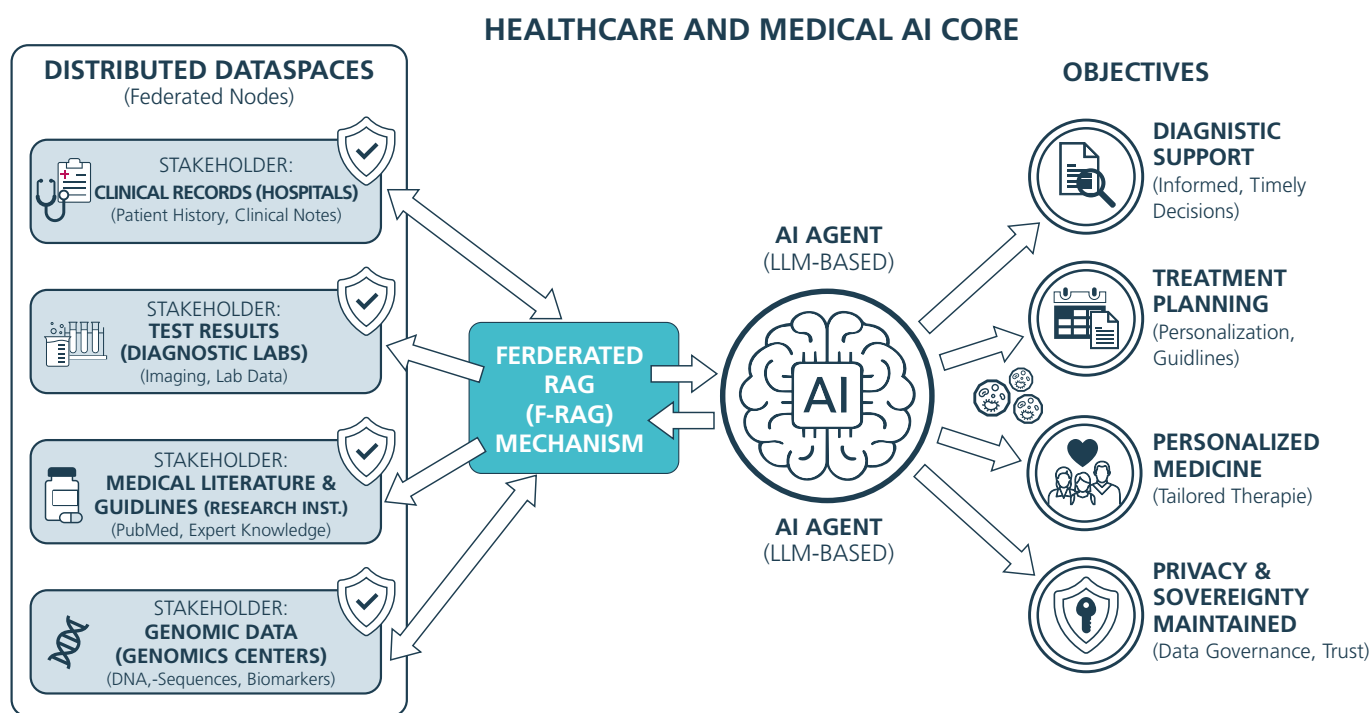


Figure 2: Healthcare and medical AI

1.4. Next-Generation Smart Buildings

In next-generation smart buildings, buildings function as intelligent systems that can perceive, make decisions, and continuously optimize operations. AI serves as the core of this intelligence, enabling flexible decision-making across multiple concurrent objectives — reducing energy consumption, ensuring equipment reliability, enhancing occupant comfort, and maintaining security — even when these goals involve inherent trade-offs.

To deliver real value, AI must access highly diverse and up-to-date information at the point of decision-making: equipment states, real-time sensor data, building utilization patterns, occupant behavior, and security information. Without such access, AI would be limited to abstract recommendations insufficient for real-world building operations.

However, data in smart buildings is inherently distributed across equipment, subsystems, and operational stakeholders. Centralizing all data is often infeasible due to security, governance, and

operational constraints. Data Spaces address this challenge by enabling each stakeholder and system to maintain data sovereignty while sharing data in a secure and interoperable way. By combining this Data Space mechanism with AI and implementing it in an F-RAG system, AI agents can dynamically reference and integrate multiple distributed data sources, enabling flexible operations far beyond static control workflows (see [Figure 3: Next-generation smart buildings](#)).

The value delivered is transformative: building operators gain holistic, real-time optimization across different domains; occupants experience more responsive and comfortable environments; and facility owners achieve greater operational efficiency without compromising data governance.

This integration of Data Spaces and F-RAG therefore represents an essential infrastructure for realizing truly intelligent next-generation smart buildings.

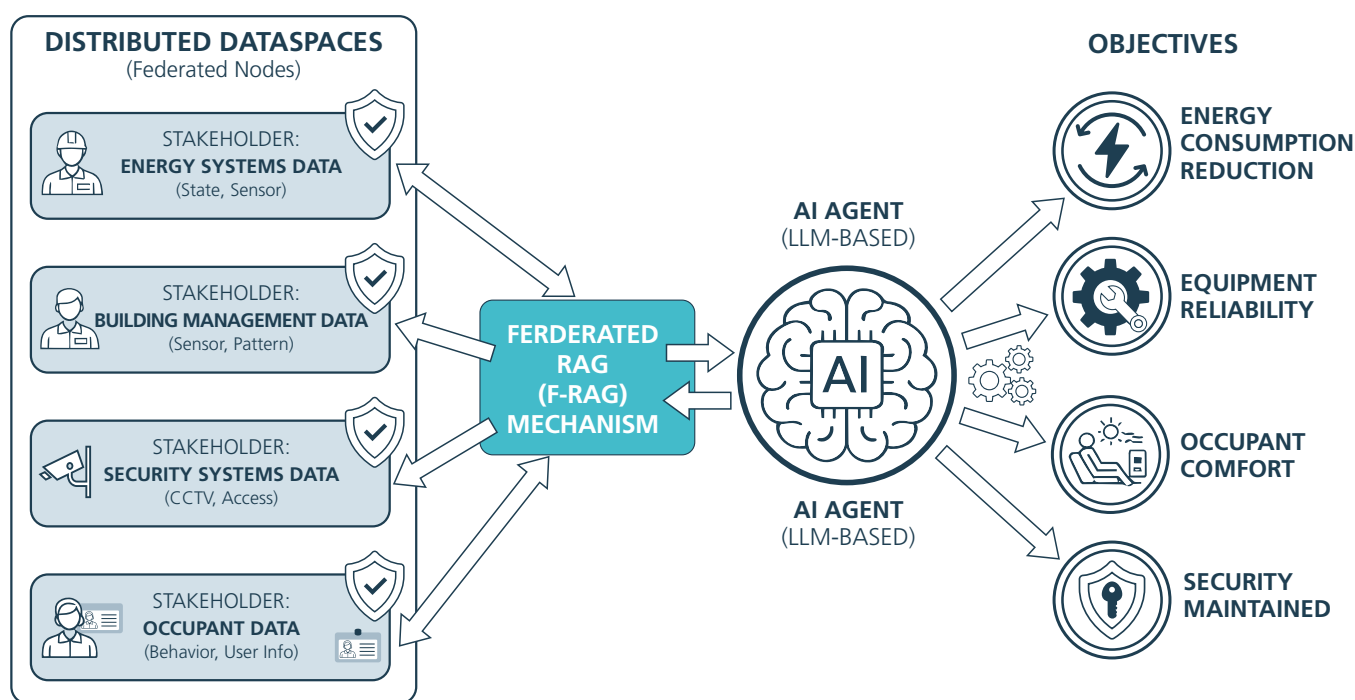


Figure 3: Next-generation smart buildings

2. AI Spaces: An Overview

The four use cases presented in Section 1 —supply chain management, collaborative product development, medical decision support, and next-generation smart buildings — cover four domains with considerably different issues and solution maturity levels. Yet they share a common underlying challenge: how to realize meaningful AI collaboration across organizational boundaries under real-world constraints of data sovereignty, confidentiality, regulation, and misaligned incentives.

Addressing this challenge requires analysis along two dimensions. The first relates to the patterns through which AI systems collaborate across organizations. The second concerns the conditions that enable such collaboration to remain sustainable over time. This section brings both dimensions together to offer a structured account of AI Spaces and to establish the analytical framework guiding the remainder of the paper.

2.1. Three Patterns of Collaborative AI

Across the use cases, three fundamental patterns can be identified through which AI systems collaborate across organisational boundaries (see [Figure 4: Types of collaborative AI technologies](#)). These patterns describe structural modes of collaboration rather than application domains, and real-world systems frequently combine more than one.

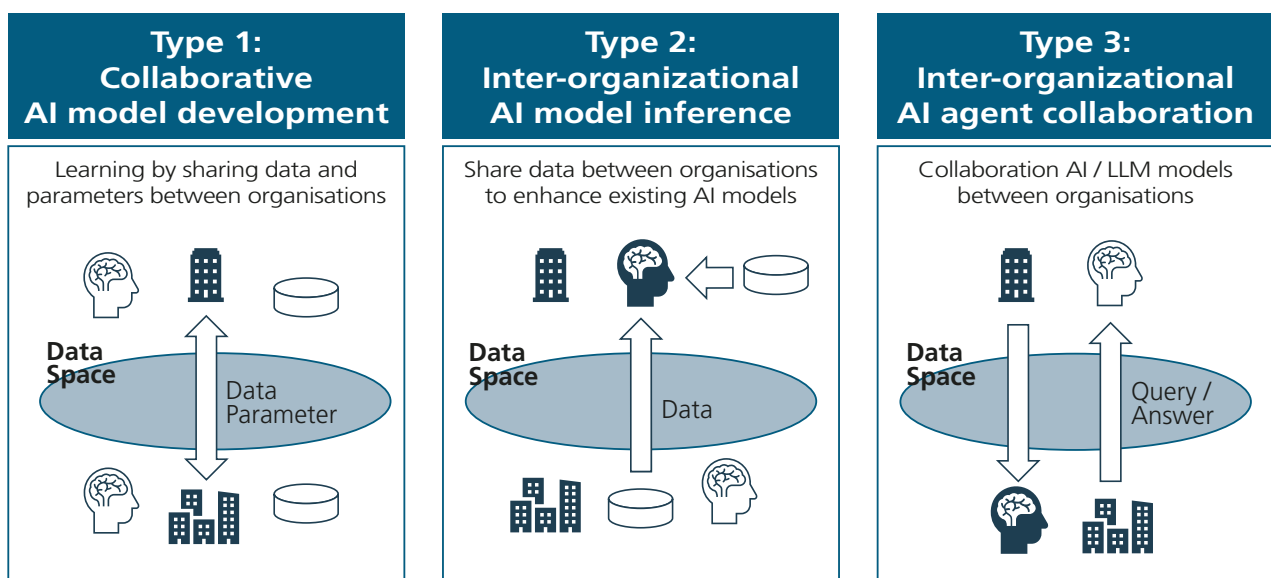


Figure 4: Types of collaborative AI technologies

Type 1 focuses on inter-organizational AI model development, where organizations jointly improve models through federated learning while retaining full control over their data.

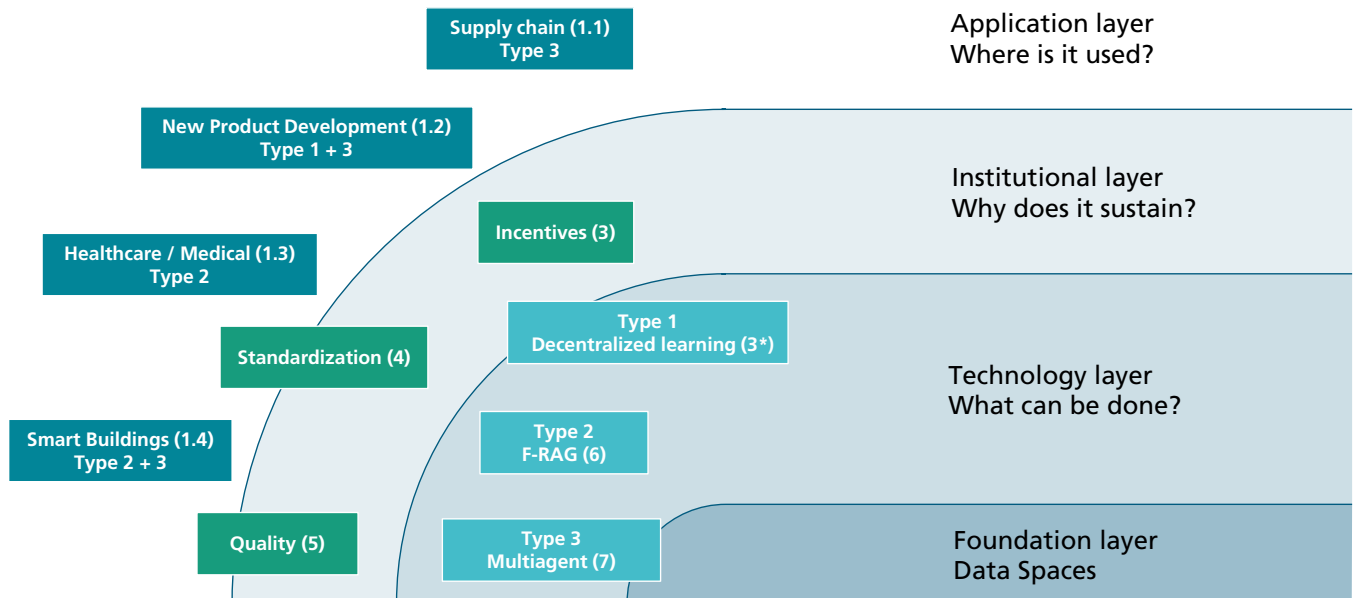
Type 2 addresses inter-organizational AI model inference, enabling AI systems to perform reasoning by selectively accessing distributed, sovereign data at inference time.

Type 3 centers on autonomous inter-organizational collaboration, where multi-agent systems coordinate decisions and actions across organizational boundaries.

For more details on these collaboration patterns, please refer to our previous whitepaper “Decentralized and Collaborative AI for Data Spaces” (Fujitsu, 2025a).

2.2. Three Conditions for Sustainability

The collaboration patterns describe what AI Spaces can do. Implementing them in practice, however, requires more than technical capability alone. [Figure 5: AI Spaces layer structure](#) represents AI Spaces as a layered structure in which each inner layer serves as a prerequisite for the layer that surrounds it.



The numbers in parentheses indicate section numbers
 * Discussed in the context of incentive design

Figure 5: AI Spaces layer structure

At the center lies Data Spaces — the shared infrastructure that connects all layers while preserving each organization’s data sovereignty.

Surrounding it is the technology layer, comprising the three collaboration patterns described above. These patterns define the value that AI Spaces create, yet their operation depends on the institutional conditions that enclose them.

The institutional layer comprises three mutually reinforcing conditions. Incentives (Section 3) ensure that organizations have sufficient motivation to contribute data, models, and operational knowledge on an ongoing basis, rather than free-riding or withdrawing participation. Standardization (Section 4) establishes the common protocols and interoperability rules without which cross-organizational data exchange cannot function at

scale. Quality management (Section 5) provides the transparency, traceability, and accountability necessary for regulatory compliance and for building trust across organizational boundaries.

At the outermost ring are the use cases — the domains in which this layered structure generates tangible value. Their position reflects the principle that application is only possible once the underlying foundation, technology, and institutional conditions are in place.

The layer structure in [Figure 5: AI Spaces layer structure](#) directly informs the organization of the remainder of this paper. Section 3 through 5 address the institutional layer, and Section 6 and 7 the technology layer. Together, they elaborate the conditions under which AI Spaces can deliver reliable, accountable, and scalable cross-organizational intelligence.

3. Economic Incentives and Compensation Mechanisms

3.1. Incentive Challenges in AI Spaces

AI Spaces enable organizations to jointly create AI-driven value by contributing heterogeneous resources — data, models, computational capacity, and operational know-how. Because participation is voluntary and contributions are unevenly distributed, incentive design is a central determinant of sustainability.

This challenge runs through all four use cases in Section 1. In the pharmaceutical supply chain, coordinated delivery optimization is possible only if participants remain motivated to contribute data and models on an ongoing basis. Equally, collaborative product development and federated knowledge sharing in healthcare depend on participants being confident that their contributions will be met with commensurate value.

Each contribution carries direct costs as well as opportunity costs such as information leakage risks, compliance overhead, and operational complexity. When contributions are not clearly linked to the benefits participants receive, rational actors may limit their engagement or free-ride. Empirical studies of

initiatives such as Catena-X confirm that organizations hesitate to participate when this relationship is opaque (Gelhaar et al., 2023), and that participation is driven by a combination of monetary rewards, access rights, ecosystem influence, and future market opportunities (Data Spaces Support Centre, 2024).

Federated Learning (FL) exemplifies these dynamics. While FL enables collaborative model training without sharing raw data, it also allows for — oftentimes undetectable — free-rider behavior. Nakamura (2026) addresses this by detecting non-contributing participants through model weight analysis, while DeFedOblivio framework enables specific contributions to be removed from the global model via decentralized unlearning functionality (Fraunhofer-Gesellschaft, 2026). Though neither is an incentive mechanism per se, both demonstrate that contributions can be assessed on technical grounds while preserving data sovereignty — a prerequisite for credible incentive design.

3.2. Objects of Rights in Collective AI Systems

Incentives in AI Spaces cannot be designed uniformly because collectively developed AI systems are not single, indivisible assets. Collaborative model generation and inference give rise to multiple distinct objects, each subject to different rights and governance arrangements:

- The Training Data Corpus and Synthesized Dataset
- Algorithmic Architectures and Source Code
- Pre-Trained Foundation Weights and Parameter Matrices
- Hyperparameters and Training Methodologies
- Parameter-Efficient Fine-Tuning (PEFT) Adapters
- Prompts, Retrieval Contexts, and Inference Outputs
- Hardware Infrastructure and Digital Twins

These objects differ in how contributions can be attributed and how control can be exercised. Data carries strong sovereignty and usage constraints; weight parameters aggregate contributions in ways that resist individual attribution; model architecture is tied to governance authority. Treating these as distinct objects of rights provides the analytical foundation on which incentive and compensation mechanisms are built.

3.3. Incentive and Compensation Mechanisms in AI Spaces

Incentives in AI Spaces are best understood as structural conditions for sustained participation rather than short-term behavioral triggers. Effective incentive structures combine financial elements with non-monetary value: access to shared data and AI services, participation in governance, and reputation derived from visible contributions (Data Spaces Support Centre, 2024). The healthcare use case illustrates this well — hospitals and research institutions are motivated primarily by access to broader evidence bases and professional trust, not financial reward alone.

Incentive and compensation mechanisms should therefore span monetary, access-, governance-, and reputation-based categories (Gelhaar et al., 2021; Data Spaces Support Centre, 2024), and be positioned as integral components of ecosystem-level governance, coupled with contribution visibility and differentiated rights across the objects that constitute collective AI systems.

4. Technology Standardization and Global Interoperability

Until now, interoperability usually meant that data is collected, structured, and reported in pre-agreed, standardised formats so that it could be utilised across different stakeholders. This approach has driven decades of progress in data exchange standards, from Electronic Data Interchange in supply chains to Health Level 7 in healthcare. With AI Spaces, this paradigm shifts fundamentally: the focus of interoperability moves from the data layer to the interaction layer. This shift does not remove the need for foundational data standards, but adds a new interoperability dimension focusing on the protocols, interfaces, and behavioral contracts that govern how AI systems request information, delegate tasks, negotiate access rights, and coordinate actions across organizational and regulatory boundaries. This transition has significant implications for standardization strategy. Whereas data interoperability could often be achieved through bilateral agreements or industry-specific formats, AI interaction interoperability requires ecosystem-wide coordination to ensure that agents developed by different organizations, using different frameworks, can seamlessly collaborate within the ecosystem.

4.1. Key Standardization Domains for AI Spaces

Achieving robust AI interaction interoperability requires coordinated standardization efforts across multiple domains: agent communication protocols such as Model Context Protocol (MCP) define how agents share context information and coordinate workflows. Semantic frameworks including Asset Administration Shell (AAS) enable cross-domain understanding. Policy languages ensure machine-readable access rights that agents

can interpret and enforce autonomously. Trust and identity frameworks establish authentication and provenance verification for agents. Quality certification standards enable participants to assess partner agent reliability. Each domain addresses a distinct aspect of the challenge, and gaps in any of these areas can undermine the functionality of the overall ecosystem.

4.2. Global Harmonisation and Ecosystem Alignment

The standardization landscape is shaped by regional initiatives, each bringing distinct strengths to the global ecosystem. Rather than viewing these regional differences as barriers, they should be seen as opportunities for global harmonization, since diverse approaches generate practical learnings across ecosystems. However, without deliberate coordination, fragmentation risks would limit the network effects essential to AI Space value creation.

Ensuring interoperability across heterogeneous infrastructures requires meta-level protocols and harmonized policy frameworks that respect regional contexts while enabling

cross-border collaboration. Effective standardization therefore demands multi-stakeholder governance, early alignment between standardization bodies and regulators, and investment in reference implementations. Stakeholders who engage proactively by contributing use cases, technical expertise, and implementation experience will shape the emerging rules of engagement. The decisions made in the next years will determine whether AI Spaces fulfil their potential as infrastructure for collective intelligence, or fragment into isolated islands that reproduce rather than transcend the data silos of the previous technological generation.

5. Quality Management: AI Bill of Materials

The central challenge of quality management in AI Spaces is how to ensure transparency, traceability, and auditability of AI systems that are increasingly composite in nature. With the widespread adoption of generative AI, it has become common for AI systems to be constructed by combining foundation models, fine-tuned models, datasets, external APIs, and RAG pipelines. In such configurations, risks are not confined to individual components; they can materialize through upstream model changes or deficiencies in data provenance. Moreover, in environments where models and external dependencies are frequently updated, manual documentation practices are not sustainable, making integration with CI/CD and MLOps pipelines essential. The AI Bill of Materials (AI BOM) is positioned as the central quality management mechanism to address these challenges.

This positioning is reinforced by converging regulatory and standardization pressures. The EU AI Act mandates risk-based technical documentation (European Union, 2024), while ISO/IEC 42001 requires evidence-based documentation capable of supporting audits and certification (ISO/IEC, 2023). CycloneDX's ML BOM (CycloneDX, n.d.) has emerged as a standard for describing models, data, and configuration information in a machine-readable format, signaling a convergence between AI governance requirements and existing supply-chain management practices.

A practical AI BOM must integrate three layers of information: component-level information on models and data, made explicit through Model Cards that clarify intended use, evaluation conditions, limitations, and provenance (Mitchell et al., 2019); system- or service-level context describing which models and data are used, for which purposes, and under what operational conditions; and operational evidence enabling ex post verification, such as evaluation runs, configuration changes, and approvals.

In the context of AI Spaces, the importance of the AI BOM is further amplified. In environments where models, data, and agents contributed by multiple organizations are combined in complex configurations, it becomes inherently difficult to trace which components underpin which decisions. This requirement appears consistently across the use cases examined in this paper. In collaborative product development, it means making explicit which models and datasets underpin specific design decisions across organizational boundaries. In supply chain management, it means clearly distinguishing between disclosed and non-disclosed agent components while maintaining regulatory compliance. In healthcare, it means reconciling patient data sovereignty with the traceability of diagnostic reasoning. By providing a structured representation of these composite systems, the AI BOM enables regulatory compliance and cross-organizational accountability to be achieved simultaneously, without undermining data sovereignty.

6. Development Approaches for AI & Data Spaces Integration

6.1. Next-Generation Connector Technologies

There is a strong need for next-generation connectors that enable AI agents and LLMs to directly interact with Data Spaces. Unlike conventional static data provisioning models used during the learning phase, these connectors must allow AI systems to dynamically retrieve data at the reasoning phase according to contextual requirements.

Leveraging MCP is effective in designing such connectors. MCP provides mechanisms that allow models to understand their execution context and communicate securely with external data sources, enabling flexible connectivity between AI systems and Data Spaces.

The target databases extend well beyond text-based embedding search databases used in conventional RAG. They encompass: graph-based knowledge representation databases capable of semantic reasoning; relational databases for numerical data; real-time streaming databases reflecting IoT and operational data; and geospatial databases such as GIS.

To dynamically utilize such heterogeneous sources, connector designs must incorporate AI-interpretable data descriptions, connection control mechanisms, and access policy management. Key research challenges include:

First, techniques to predict which databases likely contain relevant data based on query content, reducing remote queries and improving execution performance. Second, when federated databases are heterogeneous — for example, with differing embedding spaces — techniques for evaluating and re-ranking multiple returned results.

These developments enable dynamic, low-latency data utilization. Continuous tracking of MCP developments and adaptive R&D strategies remain essential for constructing next-generation intelligent digital infrastructure.

6.2. Retrieval-Augmented Generation (RAG) and Federated RAG (F-RAG)

RAG allows LLMs to perform reasoning while accessing external data sources. A vector database containing pre-embedded information is prepared in advance. When responding to a query, the system retrieves data with embeddings close to the query and generates an answer based on the retrieved information.

F-RAG (Amano et al., 2025) extends RAG by enabling secure, federated use of vector databases and other kinds of databases owned by multiple organizations (see [Figure 6: RAG and F-RAG](#)). Data owners dynamically determine whether search access is permitted based on data sovereignty. A user query is issued to each distributed database, and from the aggregated results, the system generates a response using the most relevant data (see [Figure 7: F-RAG Architecture \(Matsunaga, 2025\)](#)).

Several challenges arise in realizing F-RAG in practice. Coordinating retrieval across heterogeneous data sources requires agents capable of dynamically formulating optimal acquisition plans — an approach exemplified by LLMind (Cui et al., 2024). Confidentiality presents a further concern, as retrieved context must remain protected throughout the retrieval and generation process; approaches like C-FedRAG (Addison et al., 2024) address this through confidential computing environments that enable secure collaborative inference without centralizing data.

The use of AI agents further extends this paradigm by dynamically adapting data access workflows rather than following static definitions. When an LLM acting as an agent determines that querying a Data Space would yield a more accurate response, it autonomously performs the search, selectively

determining and executing necessary processing steps. Scalability across distributed and heterogeneous sources remains an active research challenge, with work on federated search methods and approximate vector search contributing to improved performance (Guerraoui et al., 2025).

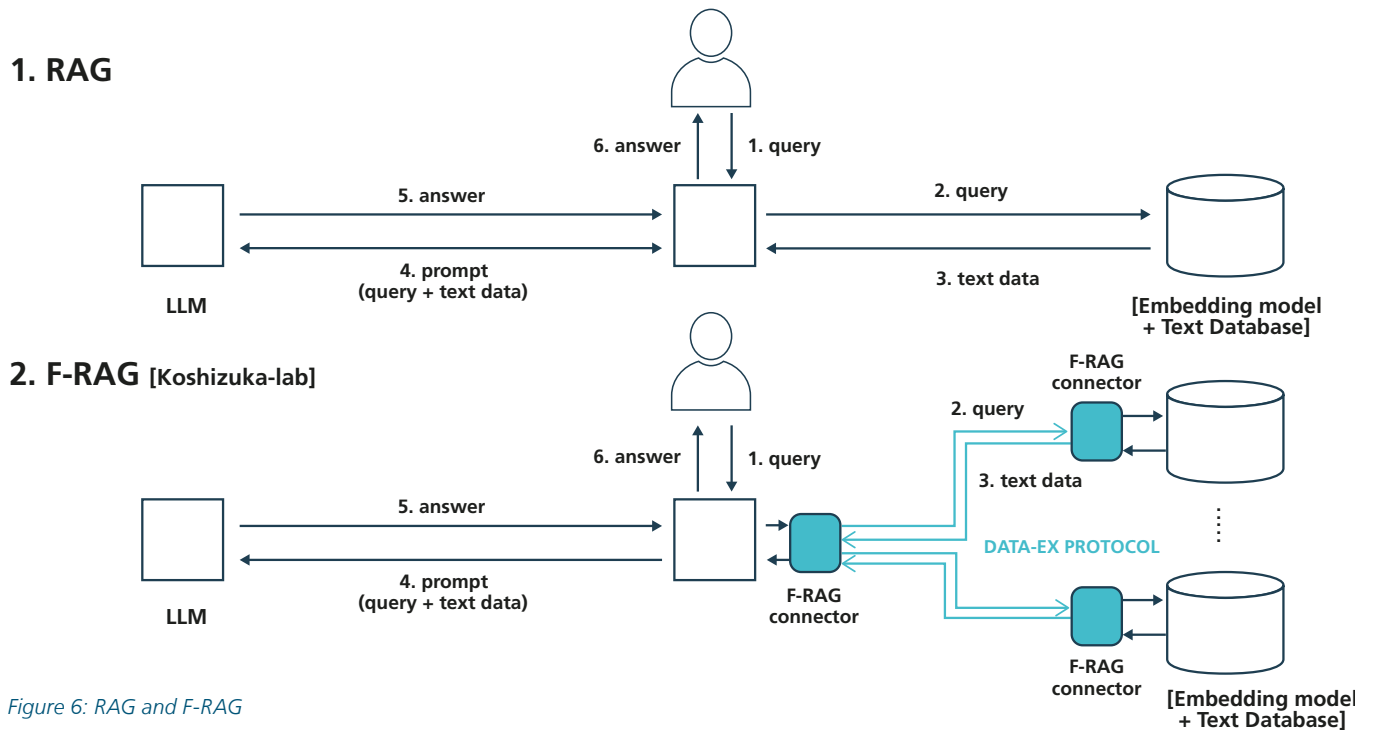


Figure 6: RAG and F-RAG

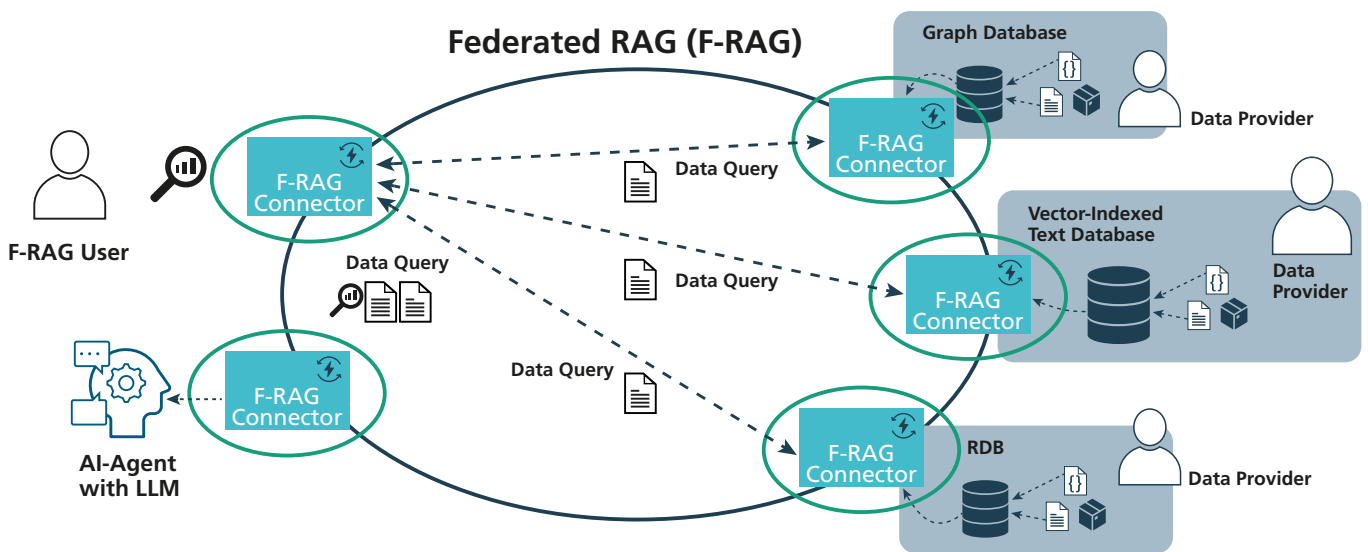


Figure 7: F-RAG Architecture (Matsunaga, 2025)

7. Sovereign Multi-Agent Coordination

7.1. Challenges in Multi-Agent Collaboration

In industrial domains such as supply chains, AI-driven decision-making necessarily spans organizational boundaries. Responding to demand fluctuations, supply constraints, and transportation disruptions requires coordinated judgment across multiple actors — yet real-world environments impose three fundamental constraints that make such coordination non-trivial.

First, information sharing cannot be assumed. Organizations retain data on inventory levels, cost structures, and contractual conditions as sources of competitive advantage, and full disclosure is rarely realistic even where a shared optimization objective exists. AI Spaces must therefore operate under the assumption that only partial information is available at any point.

Second, confidentiality and data sovereignty are hard requirements. Cross-organizational agent collaboration carries the risk that sensitive information may be inadvertently inferred or leaked through learning, inference, or dialogue processes. Effective mechanisms must enable cooperation while ensuring that confidential information remains protected throughout.

Third, decisions must be validated before deployment. In domains where the consequences of decisions propagate across multiple actors, trial-and-error in live operations is not acceptable. Virtual simulation mechanisms are therefore necessary to evaluate outcomes in advance under realistic scenarios.

These challenges cannot be resolved by improvements to data quality alone. They require new technological frameworks capable of supporting sound decision-making under constrained information environments.

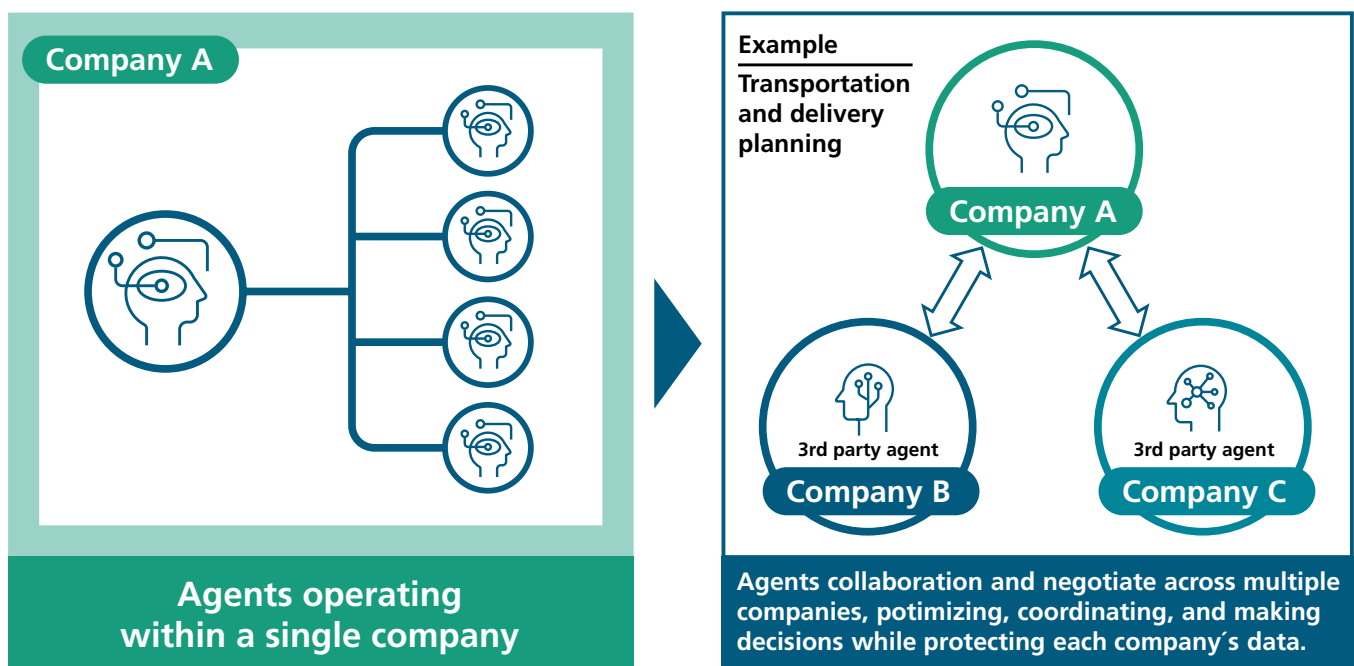


Figure 8: Multi-AI agent collaboration
Source: Fujitsu (2025b)

7.2. Enabling Technologies

Figure 8: Multi-AI agent collaboration contrasts multi-agent systems operating within a single organization with those that collaborate across organizational boundaries. Realizing the latter — enabling agents from different organizations to coordinate under incomplete information while preserving data sovereignty — requires a dedicated technological foundation (Fujitsu, 2025b). Three core technologies underpin this approach.

Optimization under Incomplete Information

Rather than aggregating complete data prior to optimization — an approach that is infeasible in most industrial settings — this technique derives coherent system-level decisions through the exchange of only the minimum information each actor is able to share. Each organization’s agent maintains its own constraints and objectives internally, while estimating the overall situation through interactions with peer agents. In supply chain settings, this enables decisions that avoid inventory imbalances and transportation bottlenecks without requiring full disclosure, producing outcomes that transcend local optimization (Asai, Akima, & Takemori, 2025).

Secure Inter-Agent Gateway

Communication encryption alone is insufficient in cross-organizational settings, as confidential information may still be inferred through agent interactions. The Secure Inter-Agent Gateway addresses this by ensuring that raw data are never shared externally: only abstracted knowledge necessary for peer agents’ decision-making is exchanged. All inter-agent interactions are mediated by the gateway, which enforces controlled collaboration and incorporates guardrail mechanisms to prevent unintended disclosures and manipulation by malicious agents (Fujitsu, 2025b). Organizations can thus participate in collaborative decision-making while preserving their competitive advantages and full data sovereignty.

Digital Rehearsal

To validate decisions before deployment, collaborating AI agents are simulated within a virtual environment, allowing outcomes to be assessed against scenarios such as sudden demand shifts, component shortages, or route disruptions (Fujitsu, 2026). This pre-decision simulation enables organizations to avoid trial-and-error in live operations while building justified confidence in AI-driven collaborative decision-making.

8. Conclusion

As the four use cases examined in this paper demonstrate, the value of AI is no longer determined by analytical capability within a single organization, but by the capacity to collaborate across organizational boundaries. The transition from silos to sovereign, cross-organizational intelligence is not a technical aspiration but an operational necessity across industries.

AI Spaces provide the framework for this transition. Yet their sustainability cannot be guaranteed by technology alone. Incentive design, standardization, and quality management are institutional conditions inseparable from the technology layer — and it is only when all are in place that reliable, scalable cross-organizational AI becomes achievable. The decisions made in the coming years on standards, governance, and incentive frameworks will determine whether AI Spaces fulfil their potential as essential infrastructure for collective intelligence.

References

- Addison, P., et al. (2024). C-FedRAG: A Confidential Federated Retrieval-Augmented Generation System. arXiv:2412.13163, December 17, 2024.
- Amano, K., Seike, H., Hatano, K., & Koshizuka, N. (2025). DP-CR: Differentially Private Centroid Routing for Federated RAG in Data Spaces, The 13th IEEE International Conference on Big Data (IEEE BigData 2025), Macau, December 2025.
- Asai, T., Akima, Y., & Takemori, K. (2025). Inter-organizational multi-agent collaboration technology. fltech – Technology Blog of Fujitsu Research. https://blog-en.fltech.dev/entry/2025/12/02/multi-agent_collaboration
- Council on Competitiveness-Nippon (COCN). (2026). Realization of socially acceptable sustainable engineering through generative AI (FY2025 Promotion Theme Project Final Report). <https://www.cocn.jp/report/d9f052a0da3a6f505cc86bb7fc616a02345b1acd.pdf>
- Cui, H., et al. (2024). LLMind: Orchestrating AI and IoT with LLM for Complex Task Execution. IEEE Communications Magazine, September 27, 2024. DOI: 10.1109/MCOM.002.2400106.
- CycloneDX. (n.d.). Machine Learning Bill of Materials (ML BOM). <https://cyclonedx.org/capabilities/mlbom/>
- Data Spaces Support Centre. (2024). Incentives and synergies between data space participants. In Data Space Design Principles (Version 2.0). <https://dssc.eu/space/DSDPV2/766181556>
- European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
- Fraunhofer ISST. (2025). AI in new product development: Connecting data and unlocking knowledge. https://www.isst.fraunhofer.de/en/publications/press_releases/2025/AI-in-new-product-development.html
- Fraunhofer-Gesellschaft. (2026). Removing corporate data from AI models. <https://www.fraunhofer.de/en/press/research-news/2026/april-2026/removing-corporate-data-from-ai-models.html>
- Fujitsu. (2025a). White paper on concepts and technological directions for realizing decentralised and collaborative AI in dataspaces. <https://global.fujitsu/-/media/Project/Fujitsu/Fujitsu-HQ/technology/research/article/topics/202507-decentralised-ai-for-dataspaces/202507-decentralised-ai-for-dataspaces-en.pdf>
- Fujitsu. (2025b). Fujitsu develops multi-AI agent collaboration technology to optimize supply chains, launches joint trials. <https://global.fujitsu/en-global/pr/news/2025/12/01-02>
- Fujitsu. (2026). Development of supply chain digital rehearsal technology for medium- to long-term supply chain strategy planning to prepare for various uncertainties. <https://global.fujitsu/en-global/technology/research/article/topics/202511-supply-chain-digital-rehearsal>
- Gelhaar, J., Both, J. R., & Otto, B. (2021). Requirements for incentive mechanisms in industrial data ecosystems. Conference on Production Systems and Logistics.
- Gelhaar, J., Bergmann, N., Müller, P., & Dogan, R. (2023). Motives and incentives for data sharing in industrial data ecosystems: An explorative case study. Proceedings of the Hawaii International Conference on System Sciences.
- Guerraoui, R., et al. (2025). Efficient Federated Search for Retrieval-Augmented Generation. arXiv:2502.19280 [cs.LG], 2025. <https://doi.org/10.48550/arXiv.2502.19280>
- ISO/IEC. (2023). ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system. <https://www.iso.org/standard/42001>
- Matsunaga, I., (2025). Federated RAG for Dynamic Discovery of Shared Data Based on Data Sovereignty in Distributed Environments. Master Thesis, Graduate School of Interdisciplinary Information Studies, The University of Tokyo, 2025, in Japanese.
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAcCT '19) (pp. 220–229). Association for Computing Machinery. <https://doi.org/10.1145/3287560.3287596>
- Nakamura, Y. (2026). Dynamic Free-Rider Detection in Federated Learning via Simulated Attack Patterns. arXiv:2604.04611, April 16, 2026. <https://doi.org/10.48550/arXiv.2604.04611>

Image Credits

©Anan – AdobeStock, title

The background of the entire page is a blurred, light blue-toned image of a server room. In the foreground, several server racks are arranged on a grid floor. Lines connect the racks, forming a network topology. The racks are illuminated with a soft blue glow, and the overall scene is out of focus, creating a sense of depth and a high-tech, digital atmosphere.

Contact

Tobias Guggenberger

tobias.moritz.guggenberger@isst.fraunhofer.de
+49 231 97677-439

Fraunhofer-Institute for Software and Systems Engineering ISST
Speicherstraße 6
D-44147 Dortmund
Germany

www.isst.fraunhofer.de