Data Sharing

Closing the Gap in the Draghi Report's Call for AI Competitiveness in Europe

Boris Otto

Director of the Fraunhofer Institute for Software and Systems Engineering ISST and member of the boards of directors of the Gaia-X European Association for Data and Cloud AISBL and the International Data Spaces Association

Hubert Tardieu

Independent member and former Chairman of the Board of Directors of Gaia-X European Association for Data and Cloud AISBL

This article was published on February 8, 2025 in 'Le Grand Continent'.

Introduction

The Draghi report[1], solicited by the European Council to the former Italian Prime Minister, proposes a way forward for improving competitiveness in the European Union. It follows the Letta report[2] on the EU internal market—which we have analyzed in a previous article[3]. Letta recommends introducing a fifth freedom on top of the free movement of goods, services, people, and capital because the latter fall short "in addressing the shift from an economy based on ownership to a new one based on access and sharing". Instead, Letta proposes as a fifth freedom the movement and sharing of research, innovation, data and knowledge that have become indispensable drivers of innovation in modern economies.

In this commentary, we shall not address the investment policy and Draghi's recommendations for an integrated capital market but rather focus on the proposed innovation strategy as it is explained in the part B of the report, covering in depth analysis and recommendations both on the sectoral and horizontal policies. **We support the idea of promoting the explicit launching of a second season of the European Data Strategy as it was initially presented in February 2020. A European Data Union will leverage the regulatory framework and the investments in common European Data Spaces and will facilitate Letta's fifth freedom towards AI competitiveness.**

Data sharing in ecosystems in the ten strategic sectors proposed by Draghi is the prerequisite to any useful application of AI in these sectors. The impressive progress of generative AI will not be able to seize its overall potential without access to quality data obtained with the consent of the data holders. Furthermore, open-source AI models such as the DeepSeek R1[4] may be considered a "Sputnik" moment as the example shows that global competition on AI is in full swing. Apparently, a team of 200 engineers is able to develop a more environmentally friendly model that matches the requirements articulated in the Draghi report perhaps better than "brute force" approach announced by the Stargate consortium one week earlier.[5]

In the following, we shall first summarize the Draghi report's position on data sharing, showing that it falls short in acknowledging data sharing as an integral prerequisite for the "EU Vertical AI Priorities Plan". Then, we shall introduce the data sharing value chain as a foundation for competitive advantage through AI, followed by an overview of existing data sharing initiatives in Europe. Afterward, we will explain how to leverage Europe's data sharing investments to support the "EU Vertical AI Priorities Plan".

1. Data Sharing in the Draghi Report

The Draghi report consists of two parts, namely the report itself (Part A) and the recommendations (Part B).

Part A focusses on the analysis of the productivity gap between the EU and the US:

> *"The key driver of the rising productivity gap between the EU and the US has been digital technology and Europe currently looks set to fall further behind".*

The main example of the productivity gap is illustrated by referencing AI:

> *"Europe is lagging in the breakthrough digital technologies that will drive growth in the future. Around 70% of foundation AI models have been developed in the US since 2017 and just three hyperscalers accounts for over 65% of the global as well as the European cloud market... The EU's competitive disadvantage will likely widen in cloud computing as the market is characterized by continuous massive investments, economies of scale and multiple services offered by a single provider. However, there are multiple reasons why Europe should not give up on developing its domestic tech sector. First it is important that EU companies maintain a foothold in areas where technology sovereignty is required such as security and encryption (sovereign cloud solutions). Second a weak tech sector will hinder innovation performance in a wide range of adjacent fields such as pharma, energy, materials, defence. Third AI is an evolving technology in which EU companies still have an opportunity to carve out a leading position in selected segments."*

Later in the report, data sharing is mentioned for what is a key recommendation of the report:

> *"The EU should promote cross-industry coordination and data sharing to accelerate the integration of AI into European industry. [...] To facilitate this cooperation, EU companies should be encouraged to participate in "AI Vertical Priorities Plan". The aim of this plan would be to accelerate AI development across the ten strategic sectors where EU business models will benefit most from rapid AI introduction (automotives, advanced manufacturing and robotics, energy, telecoms, agriculture, aerospace, defence, environmental forecasting, pharma and healthcare). […]* **In particular, to overcome the EU's lack of large data sets, model training should be fed with data freely contributed by multiple EU companies within a certain sector."**

Finally, the importance of cloud computing for data sharing and ways to benefit from the U.S. lead are mentioned:

> *"Given the dominance of US providers, the EU must find a middle way between promoting its domestic cloud industry and ensuring access to the technologies it needs."*

Part B elaborates on and describes the current status of "Digitalization and advanced technologies" in the EU:

> *"The EU's industrial model, so far based on imports of advanced technologies and exports from the automotives, precision mechanics, chemical, materials and fashion industries, does not reflect the current pace of technological change. […]*
>
> *The EU suffers from limited capacity to benefit from 'winners take most' dynamics, network effects and economies of scale in key technologies-except for next generation materials and clean technologies. […]*

Several EU industrial alliances for cloud-based technologies and data exchanges have been created over time with various remits (Andromède, Gaia-X, Catena-X), but results are minimal so far. […]

More recently, several Member States have promoted 'secure' cloud setups where EU-owned Infrastructure-as-a-service providers cooperate with hyperscalers' distribution but retain control over sensitive elements of security and encryption ('sovereign cloud' solutions)."

The second set of recommendations is key to closing the productivity gap and promotes the use of AI in 10 strategic industries:

"Launch an 'EU Vertical AI Priorities Plan'. Within these priorities, the plan would fund key vertical AI models across industrial sectors, built on EU data sharing, safeguarded from anti-trust enforcement for ten strategic industries where European know-how and value capture should be safe-guarded: Automotive Industry and mobility platform, Advanced manufacturing and robotics, Energy for both grid optimization ,as well as the production and integration of sources, Telecom networks ,including edge computing and IoT, Agriculture, including space-generated Earth observation data, Aerospace, Defence, Environmental forecasting, Pharmaceutical, Healthcare"

2. The Data Sharing Value Chain as a Foundation for AI Competitiveness

AI—as digital technology in general—is not an end but is a driver for competitiveness. Thus, it is useful to distinguish typical use cases of Industrial AI:

1. **Use of AI based digital services:** Industrial enterprises in this case "just" consume digital/smart services provided by software vendors etc. that use AI (both predictive and generative AI). Examples can be found in various areas, e.g. AI-enhanced sensor services, estimated time of arrival services in logistics etc.
2. **Use of predictive AI for own industrial digital/smart services:** Industrial enterprises make use of own and/or customer data to provide a digital service to their customers. An example is predictive maintenance or condition monitoring services which make use of data that is created in the use of a product of the respective industrial enterprise. So, private companies (customers) share their data to benefit from better digital services. In case customer data is used for digital/smart services, the European Data Act applies and regulates rights of the data holder.
3. **Use of generative AI by an industrial enterprise:** Activities in business processes such as procurement, accounting, marketing etc. can be accelerated and automated using large language models (LLM). The majority of LLMs is owned by non-European AI companies.
4. **Enrichment of generative AI by an industrial enterprise:** In this case, industrial enterprises do not only simply use LLMs but deploy technologies such as RAG (retrieval augmented generation) to "enrich" LLMs through private data. This approach increases the performance of the LLM. Valuation of shared data is needed because otherwise there would be no motivation for a private company to inject their data into an LLM.
5. **Shared industrial foundation models/LLM:** In this case, several private companies share their data to jointly train and/or fine-tune a foundation model/LLM. We are not aware of one single productive case of this in practice. Intermediaries that facilitate data sharing fall under the European Data Governance Act in order to prevent data pooling at the expense of the data rights holders.

In order to reduce dependencies and increase competitiveness businesses and policy makers must understand the underlying data sharing value chain which extends the traditional data and AI value chain—consisting of collecting/creating, curating, enriching, storing, distributing data

and then using it to train AI models—by the sharing notion. It is a matter of fact, that in Europe no single actor possesses all the resources needed (such as compute infrastructure, data, trust etc.) to operate the entire data sharing value chain.

In the highly dynamic AI ecosystem, it is also of strategic importance to decide where to allocate resources. Should priority be given to developing powerful general-purpose foundation models or instead to fine-tuning and "customizing" existing open models? This question could not be more relevant in the light of the recent Deepseek R1 announcement.

The community is still debating the key factors that make Deepseek R1 superior to alternatives, such as higher data quality, more efficient curation process, the use of reinforcement learning, "distillation" into smaller models etc.

In any case, it can be observed that Hugging Face is already offering Open-R1 as a full reproduction of DeepSeek R1 to facilitate the reuse of the foundation model and the development of industry-specific models on top of it.[6]

Against this background, European AI champions such as Mistral AI in France and Aleph Alpha in Germany may now want to consider focusing on use case 4 and 5 described above, and, thus, just using an existing open general-purpose foundation model and focusing on more value-adding industry-specific fine-tuning and augmentation approaches. This strategy is further supported by the fact that mid-size models can also be developed in Europe, as demonstrated by the example of Teuken 7B[7].

Focusing on downstream, higher value-adding parts of the data and AI value chain resembles many data sharing ecosystem examples that run on infrastructure of non-European cloud service providers but ensure data protection and data sovereignty through a regulatory framework and open industry standards (such as the IDSA Dataspace Protocol[8] and the Gaia-X Trust Framework[9]).

Both compute infrastructure and the often cited "data treasure" are distributed among many different organizations, both private and public. Thus, to facilitate shared foundation models, the same requirements as those for the EU data economy in general must be met, i.e. trust, data sovereignty, traceability of data, economics of data etc. These requirements have been at the core of the current European Data Strategy, and therefore must be met in the European Data Union and the AI Vertical Priorities Plan as well.

In addition, the most advanced data spaces, as Catena-X in the automotive industry, are increasingly confronted to the **data interoperability challenge.** Data spaces do not require a common set of data definitions to be adopted in advance by all participants in the data space; to avoid making such an ontology a prerequisite, data spaces have initially concentrated their attention to agree on common data definitions for each of their main use cases; In a second phase, these definitions need to be harmonized using generative AI and existing LLMs fine-tuned with the data catalogs of each partner of the data space. Usable data sets by industry will then be available to feed new AI applications with quality data obtained with the consent of athe data holders.

Thus, AI competitiveness in Europe can only be increased if the investments in common European Data Spaces and the regulatory framework for the data economy are leveraged to the maximum extent.

3. Current Status of the Common European Data Spaces

Over the last four years, taking advantages of the Recovery and Resilience Facility Fund (RRF) after the Covid-19 pandemic, the European Member States and the European Commission have already engaged a significant amount of resources for creating data spaces. The table below provides a snapshot of the funding portfolio on European and Member State level.

Data Space Investments and Subsidies

Source: Gaia-X European Association for Data and Cloud AISBL

Source	Programme	€	Comment
Germany	National Funding	421M	Data Ecosystems: Gaia-X Funding Competition (11 Projects), Manufacturing-X, Catena-X, Gaia-X 4 Future Mobility, EuProGigant, Energy data-X.
Spain	National Funding	150M	Industrial Data Spaces Open Call
	National Funding	50M	Tourism Data Spaces Open Call
France	National Funding	110M	40M € Data4Industry-X, 70M € for new call for tender
Luxembourg	National Funding	20M	
Denmark	National Funding	4,8M	
Finland	Sitra	2,6M	Sitra invested 2,6M € of which 625.000 € was used to co-finance 5 pilot projects related to data spaces. The co-financing rate covered by Sitra per project was 70%, the rest 30% was covered by project consortia members.
EU	Digital Europe Work Programme 2021-2022	206M	For topics deploying the sectorial data spaces and the related support activities, including the High Value Data Sets. This set of calls includes the DSSC (14M €).
EU	Digital Europe Work Programme 2023-2024	151M	For topics deploying the sectorial data spaces and the related support activities including actions on Digital Product Passport.
EU	EU4Health	280M	Implementation of European Health Data Space
EU	Horizon Europe	100M	Energy data spaces and R&I projects
EU	Digital Europe Work Programme 2021-2022	150M	Destination Earth initiative
EU	Digital Europe Work Programme 2023-2024	90M	Destination Earth initiative
	TOTAL	**1,735.4M**	

While the importance of data sharing is now well understood in most of the European countries, the European governments which have been funding the first data spaces, in Germany since 2019 and in France since 2021, are expecting return on their investments. With the technological foundation of data spaces converging, the focus of joint efforts, thus, must now be on adoption, value creation and capture of shared data. Technology is the mandatory condition for success, but economics are the sufficient one.

The community requires support as recent examples show. Founded in 2020, Agdatahub[10], for example has been discontinued in November 2024 because their economic model was not able to deliver a breakeven operation despite 85 000 farmers connected and a strong support of the agriculture unions. The main use cases addressed by Agdatahub were concentrating upon farmer trusted identity to facilitate administrative data exchanges with authorities as well as food traceability and animal genetics, all use cases which are still very strategic for agriculture in Europe.

The Data Spaces Support Centre offers many tools supporting data spaces on their journey to successful adoption (such as "business and organizational building blocks" and the "co-creation method")[11]. Furthermore, a Gaia-X Institute project on the economics of data sharing has defined four levels of maturity that must be reached before benefiting from a self-sustaining dataspace. In addition, the proposed model[12] establishes that a commonly agreed set of use cases is a pre-requisite for progressing through these maturity levels.

Three examples included in the ten strategic sectors selected by Mario Draghi in his report will illustrate the way forward:

Aerospace sector

The Aerospace sector faces a twofold challenge: (1) a demand of aircrafts which represents a portfolio of 8,000 aircrafts or 10 years of production, (2) a major expected redesign in the next 30 years to deliver low-carbon aircraft. Therefore, the primary use case of the significant European Aerospace project to be launched in 2025, Aeropspace-X, will be around supply chain with the ability of main aircraft manufacturers and their part vendors to share a common dataspace to optimize and derisk part procurement. A first implementation called BoostAeroSpace[13] has been launched 12 years ago, bringing together four major manufacturers in France: Airbus, Safran, Dassault, and Thalès, enabling them to jointly procure the majority of their parts; it took almost 8 years for the system to reach economic sustainability. The new data space will extend the functional reach of BoostAeroSpace to include shared product design and export compliance.

Energy sector

The Energy sector will benefit from data sharing both for optimizing energy distribution as it has started with smart meters and for production and more specifically for nuclear energy production. Only some of the Member States in Europe are involved, with France especially interested in the building of at least 6 new reactors called EPR2. As for Aerospace, supply chain management is the critical use cases with more than 1500 companies involved in the effort to reach the major target of reducing the build time of a new reactor to 70 months, while in China it is already at 60 months.

Manufacturing

In Manufacturing, Germany has recently launched the International Manufacturing-X (IM-X) initiative to increase the competitiveness, resilience and sustainability of the German, European and global manufacturing industry through industrial data ecosystems. As a Manufacturing-X lighthouse project, Factory-X[14] focuses on 11 dedicated use cases that aim to enhance and integrate both horizontal and vertical aspects of the supply chain within industrial operations. These use cases are designed to extend the existing supply chain-oriented solutions developed by Catena-X with a focus on vertical integration, which directly connects and optimizes shop floor operations. The consortium is led by the two main champions of Industry 4.0: Siemens AG and SAP, which have been driving supply chain and factory automation since years.

As rightly mentioned in the Draghi report (Part B): "Several EU industrial alliances for cloud-based technologies and data exchanges have been created over time with various remits (Andromède, Gaia-X, Catena-X) (…)." All the new data spaces are using the foundations laid by Gaia-X and pioneered by Catena which is operational since October 2023. European driven initiatives in Aeronautics, Nuclear, Manufacturing will reuse the distributed technical architecture and the trust framework to coordinate the supply chain and the design of new products in the next 3 to 5 years.

Re-creating a new European cloud industry to compete against hyperscalers is **no longer the preferred option,** and as mentioned in Draghi's report[15], a middle ground has been found: The various industries Aeronautics, Energy, and Automotive have defined their requirements in term of interoperability and sovereignty which have been incorporated into voluntary labels from 1 to 3. These labels are used by Aeronautics, Energy, and Automotive players to issue their requirements to cloud service providers which are offering their services for implementing data sharing. Since the compliance of cloud services to label is difficult to check for users, Gaia-X has set up **clearing houses** (one or two in each country) which are stamping the conformity of the cloud services to label.

The previous European Commission (Dec 2019-Dec 2024) has been wise after defining its data strategy to concentrate on establishing the new regulation leaving industrial associations as Gaia-X the responsibility to build the best options in term of architecture, standards and labels.

Two regulations have been set up covering multiple aspects with two of them especially relevant for helping to create successful data spaces:

- **Data Governance Act** (DGA) has identified the role of **"Data Intermediary"** which is providing the data sharing services to the data space partners; these services are built upon generic cloud services respecting the label required by users of the data spaces and are billed to these users by subscription or by transaction in the way Software as a Service (SAAS) are proposed now for years by the Software Industry. DGA is recommending (or imposing[16]) that none of the data space participants be also the data intermediary (to avoid the bias seen in the B-to-C world where Amazon is benefitting of an undue advantage vis-à-vis the other merchants by operating the platform)
- **Data Act** (DA) is setting up many of the provisions needed for data sharing and which have been incorporated in the labels. Data Act is defining a key principle: **the data generated by an equipment belong to the owner of the equipment and not to the equipment vendor.** This principle is facilitating the data exchange: as soon as the equipment owner is ready to share the data generated by the equipment during its operation (after it has taken ownership of the equipment) the bidirectional exchange of data between equipment vendor and equipment user is prime-pumped, leading to a reduction of operation and maintenance costs.

With the experience gained with the launching of the first data spaces—both successful and unsuccessful ones—provide valuable input to identify principles of economics of data sharing.

The data space infrastructure (trust frameworks, data sharing transactions etc.) form must be put in place to now give room to seize the business benefits of data sharing and, thus, reach the break-even points of infrastructure investments and be ready to share data for collective generative AI training and fine-tuning.

Looking at the current portfolio of activities, the new significant data spaces are aiming at this level of maturity and the breakeven position for the major use cases at the end of 2027 enabling an operational reuse of industrial data in 2028.

As we often state, do not expect to harvest trees in the spring if you have planted them the previous autumn.

4. Leveraging Europe Data Space Investments for AI Verticals

The Draghi Report (Part B) is setting up clear objectives and proposals for AI which we heartfully approve:

> *"The EU must have the ambition to be a leader in developing AI for its sectors of strength, regain and retain control over data and sensitive cloud services, and develop a robust financial and talent flywheel to support innovation in computing and AI. To achieve this, the EU should aim to secure a strong position during the next five years in AI embedded in key industrial sectors, such as advanced manufacturing and industrial robotics, chemicals, telecoms and biotech based on a set of EU-developed sectoral Large Language Models and Vertical Models. […]"*

> *"To achieve these objectives, the EU should adopt a new* **"EU Cloud and AI Development Act",** *aimed at enhancing European HPC17,AI and quantum capabilities and infrastructure, harmonizing cloud architecture requirements and procurement processes, as well as coordinating priority initiatives to scale-up private involvement and financing."*

We believe that the proposed timeframe of five years is appropriate and is in line with digital decade targets and objectives: 75% of EU companies are expected to use Cloud, AI, or Big Data by 2030.

However, the way in which the funds are spent during these five years is crucial; we are seriously concerned when we read page 83 of part B about the EU Vertical AI Priorities Plan:

> *"This effort would be fed with* **data freely contributed by EU companies** *and supported with open-source frameworks in data-intensive industries, duly safeguarded from EU anti-trust enforcement, to encourage systematic cooperation between leading EU companies for generative AI and EU-wide industrial champions in key sectors."*

Without discussing the proposed governance for the development of key AI verticals through "CERN-like AI Incubator", we do not want to bet the future of Europe in AI on a mythical assumptions of EU companies- deeply involved in worldwide market- ready to contribute their data for free. The reason why Aeronautic Industry, for example, is willing to share data is not altruistic, it is to increase its delivery volume beyond 800 aircrafts per year and to co-design with the ecosystem the low carbon aircraft, the reason why the nuclear industry has decided to create a dataspace is to meet the challenge to deliver a nuclear plant every 70 months in order to deliver enough clean energy to support European development.

Companies will not contribute their data for free to shared models, but rightfully require trust, transparency, clarity about the ownership of the models etc. Data spaces are the tool at hand to address these fundamental requirements.

In 2025 sharing data is understood by several industries, especially the ones mentioned in the Draghi report, as the only way to address the existential challenges they are facing because of the global warming; in several of the new data spaces just being launched ,reaching the breakeven for the most critical use cases (supply chain, design of low carbon vehicles, respecting environmental regulation) will take 3 years which are a prerequisite for a sustainable industrywide process of collecting industrial data for AI; as mentioned at the end of section 2 data interoperability across the industry shall have been sorted out in between to have industrial data which can be used across industry beyond the perimeter of the use cases for which data have been captured.

We therefore recommend that after 4 years of funding by Europe and Member States, to proceed with a thorough review of data spaces reaching the end of their funding.

Gaia-X Institute and Université Paris Dauphine have proposed a maturity model to assess the economic viability of existing data space[18], following the definition proposed by the EU Data Space Support Center[19]. It is important to anticipate which of the existing data spaces are likely to collapse when funding will stop. Since the strategy for data has mobilized significant energy and funding it is important that this first step be conducted as soon as possible and without complacency, Europe has been pioneering in its strategy of data and mistakes have certainly been done in the past but the next generation of data space being launched in 2025 shall be guaranteed breakeven in the next 3 to 4 years for each of the use cases which have been proposed; a path to data interoperability using generative AI shall be identified, in the meantime, to secure the reuse of sectorial industry data for AI.

The Draghi report, under the cloud section, offers another recommendation which we support:

> *"Support data brokers (editorial note: such as the Data Governance Act) as 'preapproved' data intermediaries, certifying ex-ante compliance with the EU acquis and guaranteeing regulatory clearance via the 'EU Ombudsman' mechanism. This would help to favour industry-specific solutions promoted by EU companies."*

Beyond the regulatory aspects for the data intermediaries, we have seen in the recent past that the data intermediaries which we also call "orchestrators" are concentrating most of the economic risks which regular members are not willing to take, and which are sometimes hidden through the membership fees which make the data space able to survive as long as it grows creating a de facto "Ponzi" pyramid. Data intermediaries will continue to play a key role on the journey of successful data space and Ai adoption because they address interoperability needs, function as a trust-ensuring entity and coordinate the deployment of funds invested (most often in public-private partnerships).

Finally, the Draghi report address the important point of "AI sandbox regimes":

> *"Harmonize national 'AI Sandbox regimes' across all Member States to enable experimentation and the development of Innovative AI applications in the selected industrial sectors, and ensure harmonized and simplified implementation of the GDPR."*

We believe that this recommendation is key in domain such as Health, Autonomous Driving, but also the whole field of autonomous agents which is likely to develop strongly in the near future.

5. Conclusion

The Draghi Report is right when calling for more economic competitiveness in times of increasing demands for strategic autonomy and technological sovereignty. AI is the technology offering the most powerful source of competitive advantage today, which is why Europe must spur AI innovation and reduce economic dependencies.

Successful AI relies on the availability of key resources such as computing capacity, large language/foundation models, and data. Unlike in other economic areas these resources are not owned by a few market participants or under the control of the state but distributed across various stakeholders.

This is why Europe must find an own way to reach to the urgent goal of competitiveness focused by the Draghi Report. The European approach should be characterized by sharing ecosystems which consists of a distributed landscape of computing capacity, foundation models, and data to train and fine-tune these models. Furthermore, a European way should support and use open models such as Teuken 7B and Open-R1 and leverage own resources mainly for higher value-adding parts of the data and AI value chain. This way will only be successful when an action plan is put into place which considers technical aspects, but more importantly governance and business aspects. The first season of the European Data Strategy implemented the necessary frameworks to move towards a true European Data Union.

Endnotes

1 See, also for quotes of the report further below: European Commission (2024) *The Draghi report on EU competitiveness* [online]. Directorate-General for Communication. https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en [Accessed: 31 Jan. 2025].

2 See Letta, E. (2024) *Much more than a Market – Speed, Security, Solidarity. Empowering the Single Market to deliver a sustainable future and prosperity for all EU Citizens* [online]. European Coucil, Council of the European Union. https://www.consilium.europa.eu/media/ny3j24sm/much-more-than-a-market-report-by-enrico-letta.pdf [Accessed: 31 Jan. 2025].

3 See Otto, B. & Tardieu, H. (2024) *L'Europe dans l'économie mondiale des données: vers un paradigme du partage* [online]. Le Grand Continent. https://legrandcontinent.eu/fr/2024/06/21/leurope-dans-leconomie-mondiale-des-donnees-vers-un-paradigme-du-partage/ [Accessed: 31 Jan. 2025]

4 The announcement of DeepSeek R1 in late January 2025 offers a new opportunity for open-source GenAI, with the cost of model training reported to be 20 times lower compared to ChatGPT-4 while achieving similar performance. Currently, there is a vivid debate ongoing about key factors for the DeepSeek success. See e.g.: Patel, D. et al. (2025) *DeepSeek Debates: Chinese Leadership On Cost, True Training Cost, Closed Model Margin Impacts.* [online]. SemiAnalysis. https://semianalysis.com/2025/01/31/deepseek-debates/ [Accessed: 2 Feb. 2025].

5 See Conroy, G., & Mallapaty, S. (2025) *How China created AI model DeepSeek and shocked the world* [online]. Springer Nature. https://www.nature.com/articles/d41586-025-00259-0 [Acessed: 31 Jan. 2025];
Smith-Goodson, P. & Komball, M. (2025) *The Stargate Project: Trump Touts $500 Billion Bid For AI Dominance* [online]. Forbes. https://www.forbes.com/sites/moorinsights/2025/01/30/the-stargate-project-trump-touts-500-billion-bid-for-ai-dominance/ [Accessed: 3 Feb. 2025].

6 See Bakouch, E., von Werra, L., & Tunstall, L. (2025) *Open-R1: A fully open reproduction of DeepSeek-R1* [online]. Hugging Face Blog. https://huggingface.co/blog/open-r1 [Accessed: 28 Jan. 2025].

7 See Fraunhofer-Verbund IUK-Technologie (Fraunhofer ICT Group) (2024) *Teuken-7B: Multilingual open-source large language model released* [online]. Fraunhofer-Verbund IUK-Technologie. https://www.iuk.fraunhofer.de/en/news-web/2024/teuken-7b--multilinguales-open-source-sprachmodell-veroeffentlic.html [Accessed: 31 Jan. 2025].

8 International Data Spaces Association (2024) *Advancing interoperability: the Dataspace Protocol* [online]. International Data Spaces Asssociation. https://internationaldataspaces.org/offers/dataspace-protocol/ [Accessed: 30 Jan. 2025].

9 Gaia-X European Association for Data and Cloud AISBL (2025) *Deliverables & Gaia-X Standard* [online]. Gaia-X European Association for Data and Cloud AISBL. https://gaia-x.eu/services-deliverables/deliverables/ [Accessed: 30 Jan. 2025].

10 Agdatahub (2024) *Agdatahub applies for bankruptcy proceedings* [online]. Agdatahub. https://agdatahub.eu/en/agdatahub-procedure-collective-liquidation/ [Accessed: 3 Feb. 2025].

11 For details see: Data Spaces Support Centre (2025) *What can DSSC offer to you? Blueprint 1.5 and other resources* [online]. Data Spaces Support Centre. https://dssc.eu/page/knowledge-base [Accessed: 3 Feb. 2025].

12 See Brousseau, E., Eustache, L. & Toledano, J. (2024) *Position Paper: Economics of Data Sharing* [online]. Gaia-X Institute & Chaire Gouvernance et Régulation, Fondation Partenariale Paris-Dauphine. https://gaia-x.eu/wp-content/uploads/2024/03/Study-on-the-emergence-and-creation-of-value-within-data.pdf [Accessed: 3 Feb. 2025].

13 BoostAeroSpace (2025) *Common Solutions for Common Goals* [online]. BoostAeroSpace S.A.S. https://boostaerospace.com/ [Accessed: 3 Feb. 2025].

14 Factory-X (2025) *The digital ecosystem* [online]. Open Industry 4.0 Alliance Implementation GmbH. https://factory-x.org/ [Accessed: 3 Feb. 2025].

15 "Given the dominance of US providers, the EU must find a middle way between promoting its domestic cloud industry and ensuring access to the technologies it needs." In: European Commission (2024) *The Draghi report on EU competitiveness* [online]. European Commission. https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en [Accessed: 31 Jan. 2025].

16 DGA requires data intermediation service providers to be registered and eligible for a EU trust logo to demonstrate that they are meeting all the statutory requirements. See: European Union (2022) *Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA relevance)* [online]. European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868 [Accessed: 2 Feb. 2025].

17 HPC: High-Performance Computing.

18 See Brousseau, E., Eustache, L. & Toledano, J. (2024) *Position Paper: Economics of Data Sharing* [online]. Gaia-X Institute & Chaire Gouvernance et Régulation, Fondation Partenariale Paris-Dauphine. https://gaia-x.eu/wp-content/uploads/2024/03/Study-on-the-emergence-and-creation-of-value-within-data.pdf [Accessed: 3 Feb. 2025].

19 Data Space Maturity Model: A set of indicators and a self-assessment tool that allows data space initiatives to understand their stage in the development cycle, their performance indicators, and their technical, functional, operational, business, and legal capabilities—both in absolute terms and in relation to peers.

Fraunhofer Institute for Software and Systems Engineering ISST

Speicherstraße 6
44147 Dortmund
Germany

info@isst.fraunhofer.de
www.isst.fraunhofer.de/en

Gaia-X European Association for Data and Cloud AISBL

Avenue des Arts 6-9
Bruxelles, Belgium
P.C. 1210

info@gaia-x.eu
www.gaia-x.eu