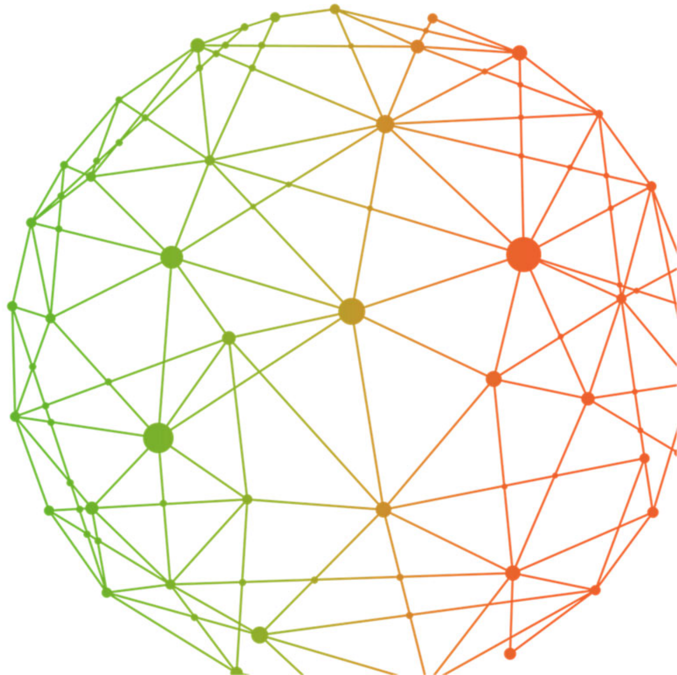
Data Spaces Symposium

9:00

Share data. Unlock value. Boost impact – that’s what data spaces are all about
Welcome to the Data Spaces Symposium 2025!

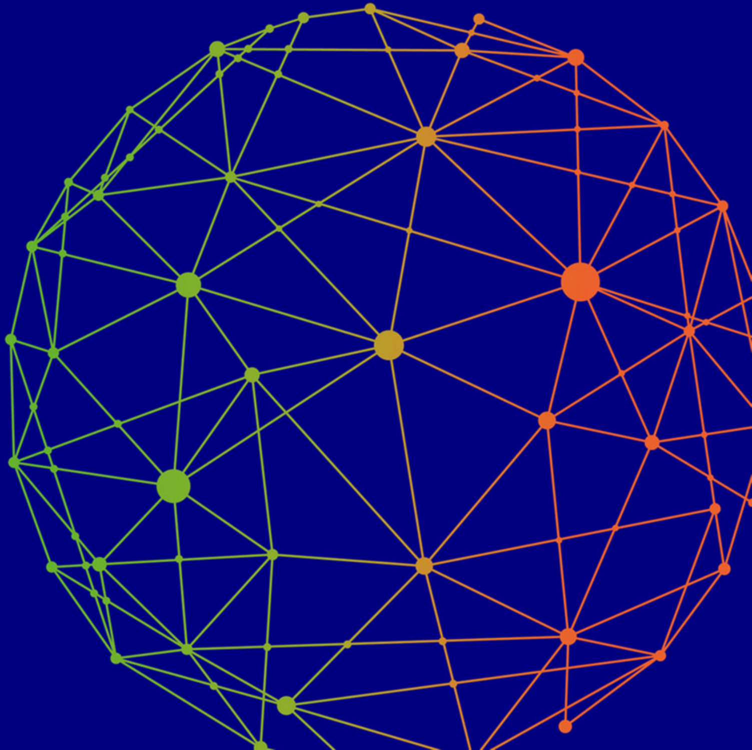
Boris Otto, Hubert Tardieu, Reinhold Achatz,
Thomas Hahn, Yasunori Mochizuki



Share data. Unlock Value. Boost Impact.

Welcome to Data Spaces Symposium 2025

Boris Otto, DSSC
Hubert Tardieu, Gaia-X
Reinhold Achatz, IDSA
Thomas Hahn, BDVA
Yasunori Mochizuki, FIWARE



DSBA ■ BDV BIG DATA VALUE ASSOCIATION FIWARE FOUNDATION gaia-x INTERNATIONAL DATA SPACES ASSOCIATION

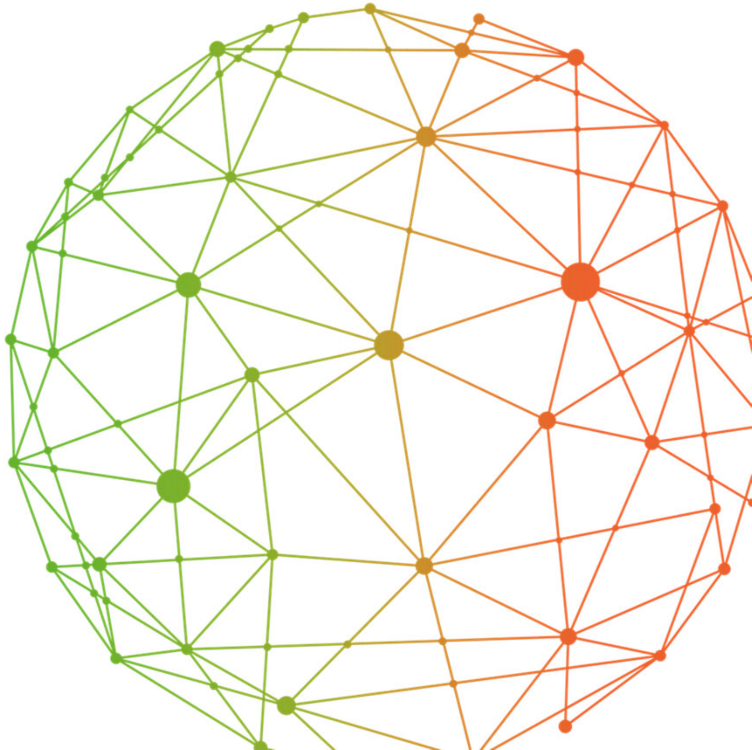
DATA SPACES SUPPORT CENTRE

Funded by the European Union

The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412

2025 is crucial for the deployment of Data Spaces

- Draghi Report: Europe is falling behind in breakthrough digital technologies
- Our business is gravitating more and more around data
- We need to make it happen - globally.
- Cross-border data exchange calls for global standards - fundamental design characteristics and Data Space Protocol

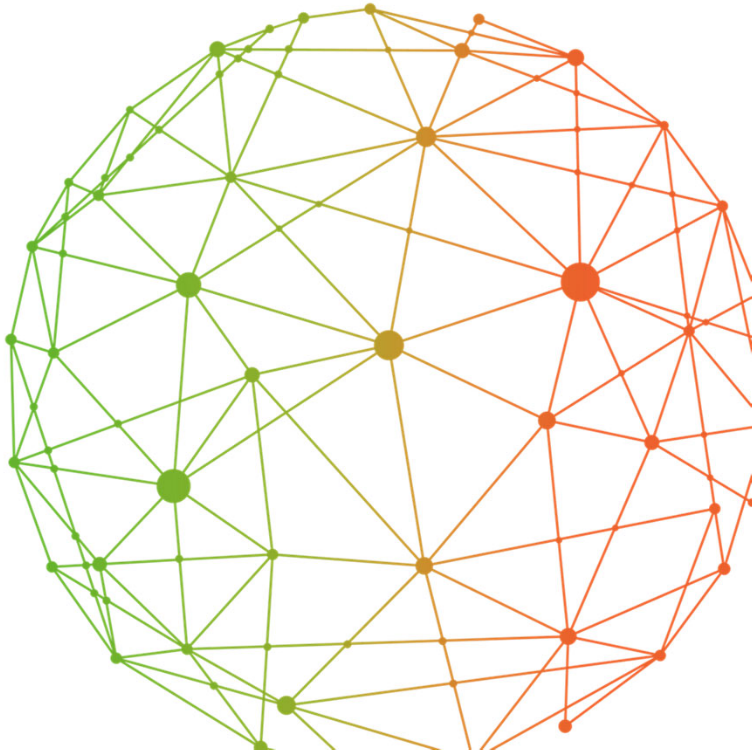


The Demand for Data Sovereignty

- Data travels with the speed of trust

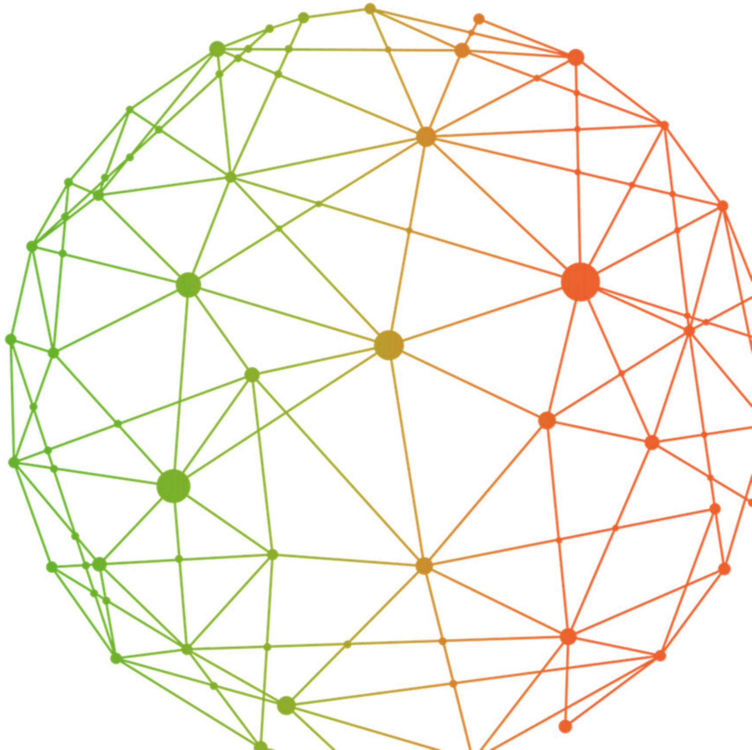
- The Trust Framework is enabling Automated Compliance - based on European values and globally

- Data does not flow on rainbows - creating the Cloud-Edge infrastructure for Data Spaces



The role of communities and the power of Open Source

- The role of communities and the power of Open Source developments
- Being inclusive and interoperable
- Alignment with Global, open & neutral standards
- Global use cases needing Data Spaces

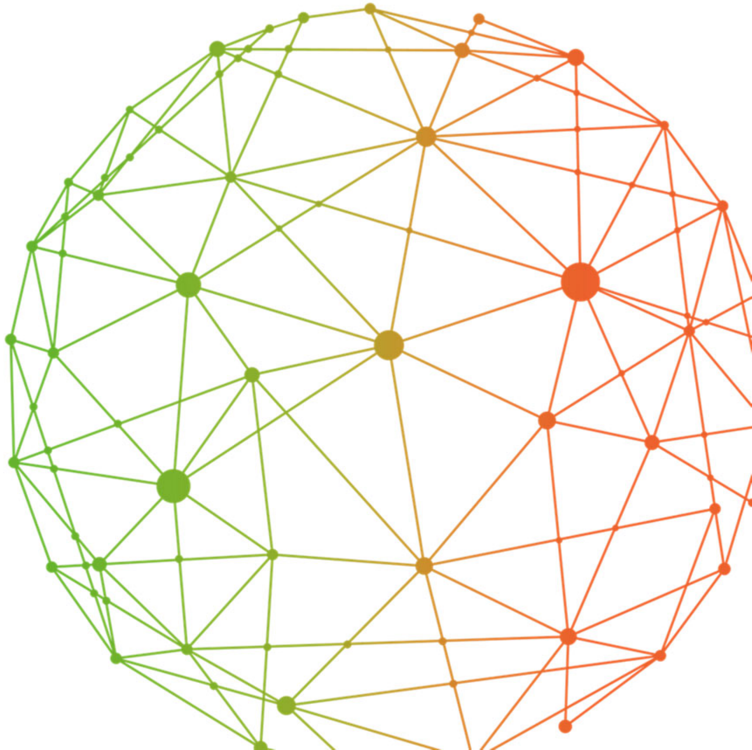


Value creation based on data

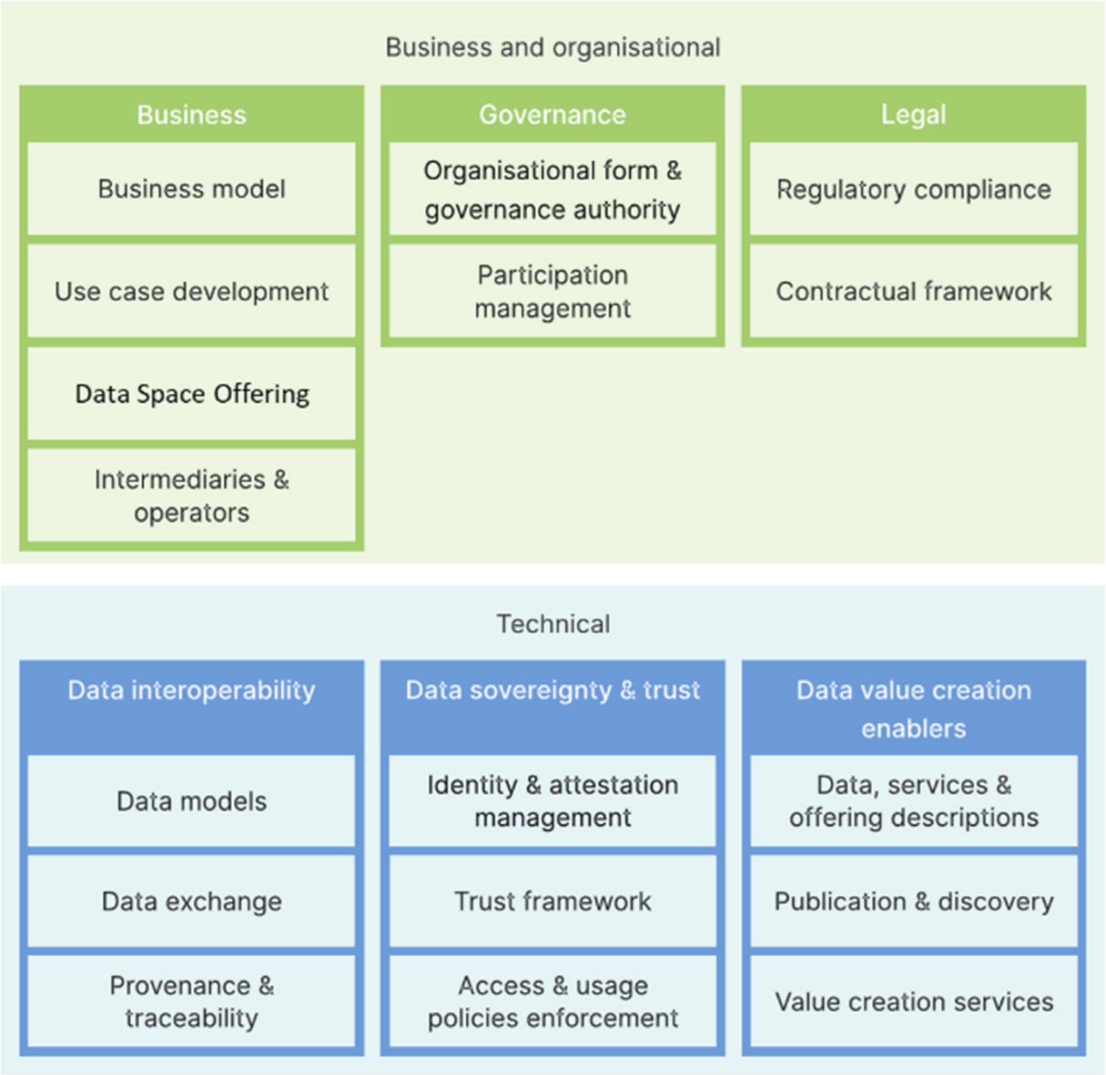
- Value creation based on data is at the heart of data economy - data spaces enabling the future (all kinds of innovation, especially AI)

- Key components to drive AI innovation:
    - AI models
    - Access to data
    - Cloud and high-performance computing
    - Talent

- We are doing this together!
(DSBA, DSSC and extended ecosystem)



The Data Spaces Blueprint 2.0 is now available!



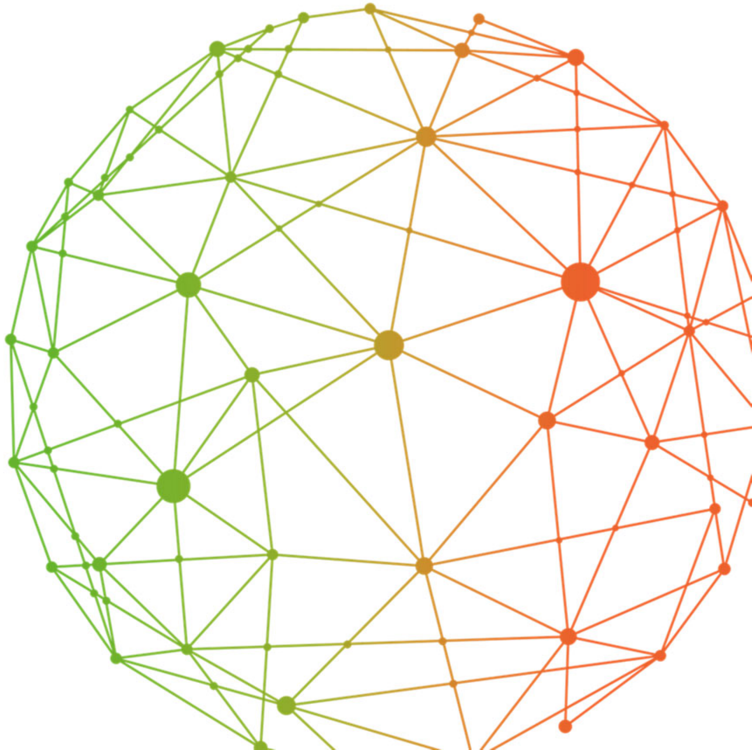
Business and organisational

Business	Governance	Legal
Business model	Organisational form & governance authority	Regulatory compliance
Use case development	Participation management	Contractual framework
Data Space Offering		
Intermediaries & operators		

Technical

Data interoperability	Data sovereignty & trust	Data value creation enablers
Data models	Identity & attestation management	Data, services & offering descriptions
Data exchange	Trust framework	Publication & discovery
Provenance & traceability	Access & usage policies enforcement	Value creation services

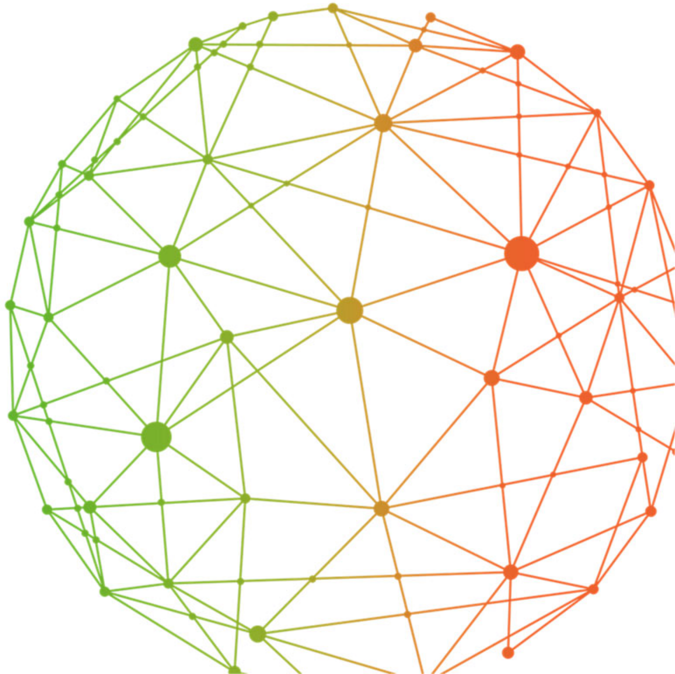
DATA SPACES SUPPORT CENTRE



Data Spaces Symposium

Keynote | The Polish Digital Strategy

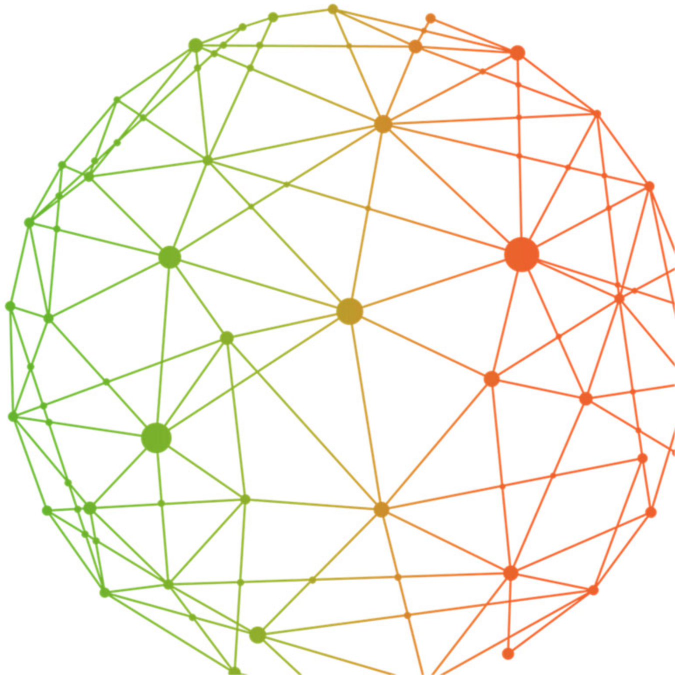
Michał Gramatyka



Data Spaces Symposium

Keynote | Data Economy – The European Way Forward

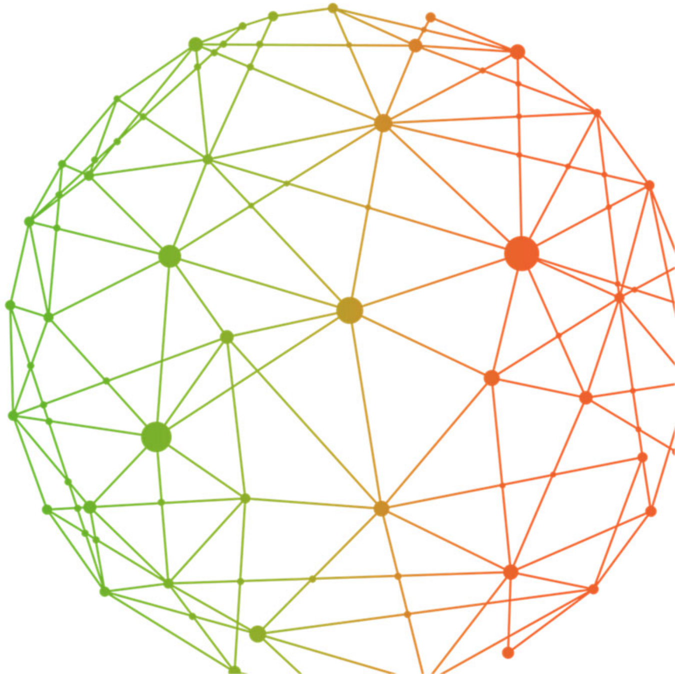
Bjoern Juretzki



Data Spaces Symposium

Keynote | Europe in the Global Data Economy

Hubert Tardieu



Data Sharing

The Key to AI Competitiveness in Europe

Data Spaces Symposium 2025

Hubert Tardieu

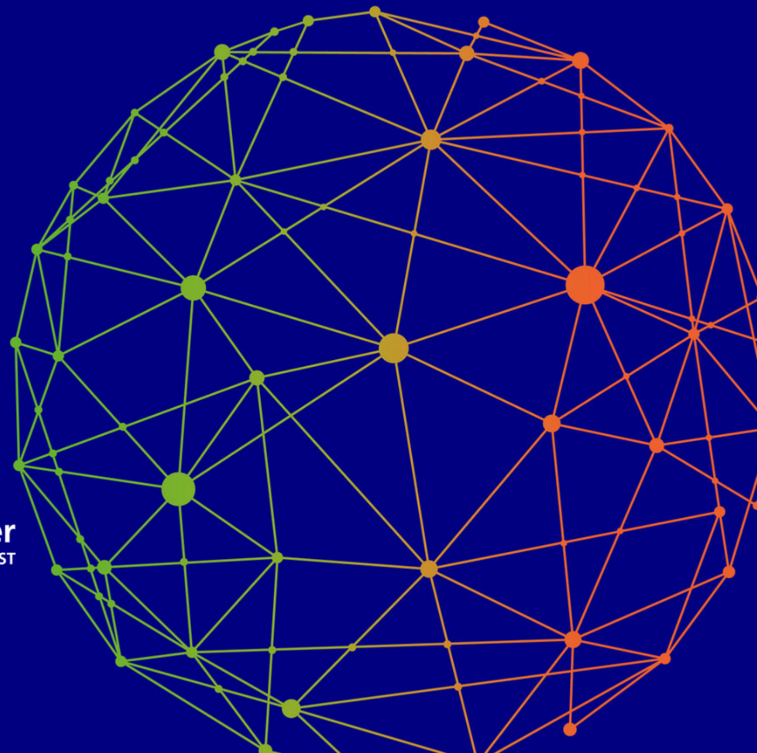
Gaia-X Independant Board Member

DSBA ■ BDV BIG DATA VALUE ASSOCIATION FIWARE FOUNDATION gaia-x INTERNATIONAL DATA SPACES ASSOCIATION

DATA SPACES SUPPORT CENTRE gaia-x INSTITUTE Fraunhofer ISST

Funded by the European Union

The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412



Europe’s Challenge

Draghi Report: The Growing Productivity Gap

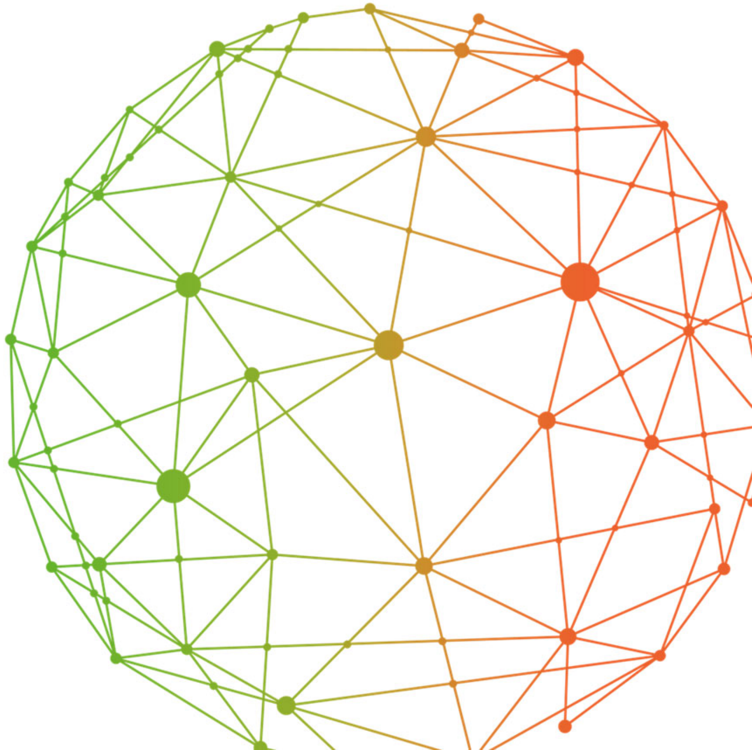
> "The key driver of the rising productivity gap
> between the EU and the US has been digital technology."

Mario Draghi, Draghi Report (2024)

Key insights from the Draghi Report:

- Europe is falling behind in breakthrough digital technologies.
  - 70% of AI foundation models have been developed in the US since 2017.
  - 65% of the global and European cloud market is controlled by three US hyperscalers.

- AI as key driver of economic growth and innovation.
  - Data is crucial for competitive AI, yet Europe struggles with availability, interoperability, and scaling of data.
  - The paradox: Europe produces massive amounts of industrial data, but it remains siloed within companies and industries.



Europe’s Challenge

Draghi Report: Cross-Industry Data Sharing for Accelerating AI

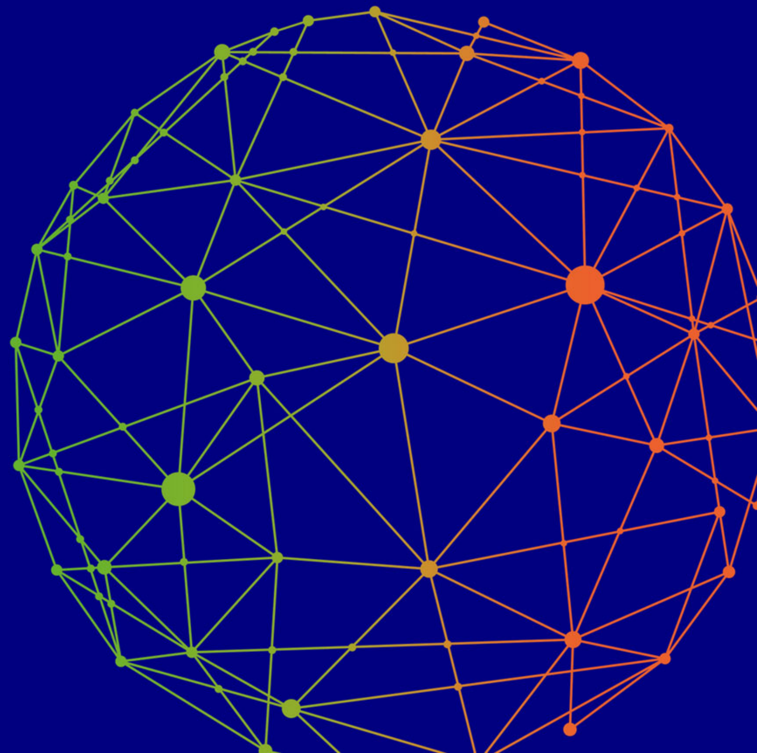
> “The EU should promote cross-industry coordination and data sharing to accelerate the integration of AI into European industry.“

Mario Draghi, Draghi Report (2024)

Draghi Report proposes a sector-specific AI strategy: "EU Vertical AI Priorities Plan":

- Shared AI model development across sectors: Strategic AI integration in 10 key industries (automotive, energy, healthcare, etc.).
- Cross-industry data pooling to overcome Europe's lack of large datasets (“for free”).
- Balance in supporting European cloud industry with securing key technologies amid US dominance.
- Key challenges: Companies hesitate to share data (competition concerns, lack of incentives, regulatory uncertainty).

> The EU must leverage its data-sharing ecosystem to enable the EU Vertical AI Priorities Plan.



AI & Data (Sharing) Value Chain

AI as driver for competitiveness

Five key AI use cases in industry:

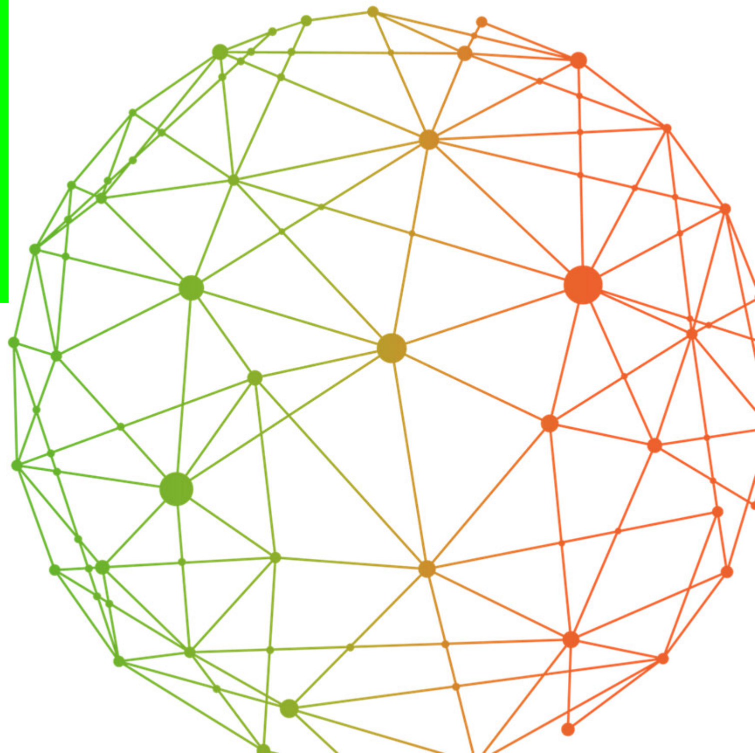
AI-powered digital services (logistics, smart infrastructure)

Predictive AI for industrial operations (maintenance, efficiency improvements)

Generative AI for automation (business processes, marketing, decision support)

AI models enriched with proprietary industry data

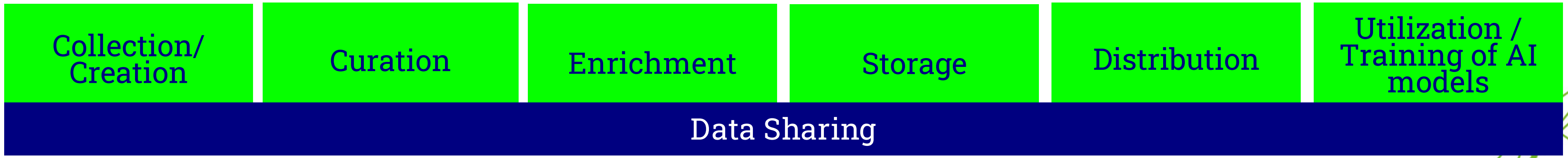
Shared foundation models tailored to specific sectors



AI & Data (Sharing) Value Chain

The Foundation for Accelerating AI

Reducing dependencies and enhancing competitiveness requires mastering the AI and data sharing value chain.



Collection/ Creation

Curation

Enrichment

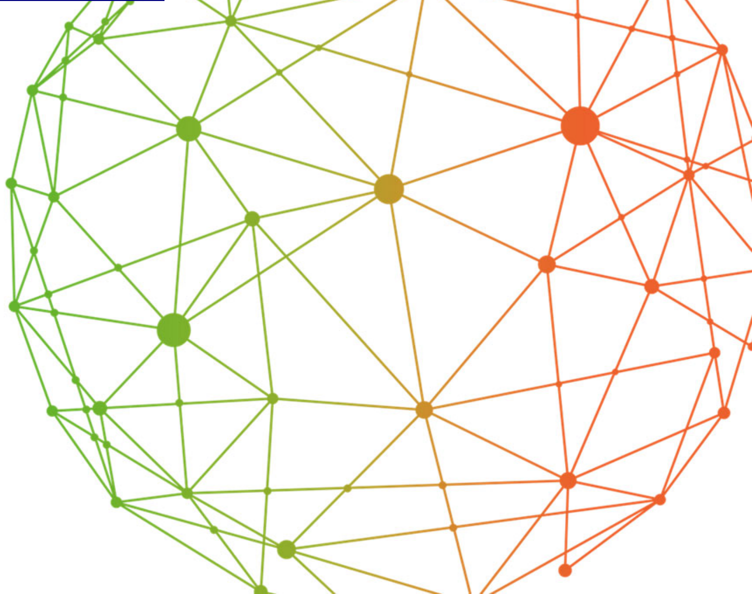
Storage

Distribution

Utilization / Training of AI models

Data Sharing

- Data, large models, and compute infrastructure is distributed across private and public organizations.
- Allocating development resources, e.g. fine-tuning existing open-source models (Mistral AI, Aleph Alpha) for industry-specific use cases.
- Facilitating shared foundation models require EU data standards.
    - Key requirements: trust, data sovereignty, traceability, interoperability, efficiency etc.
    - Alignment with the EU Data Strategy & the "EU Vertical AI Priorities Plan“.



European Data Spaces Ecosystem

Current Status and Progress

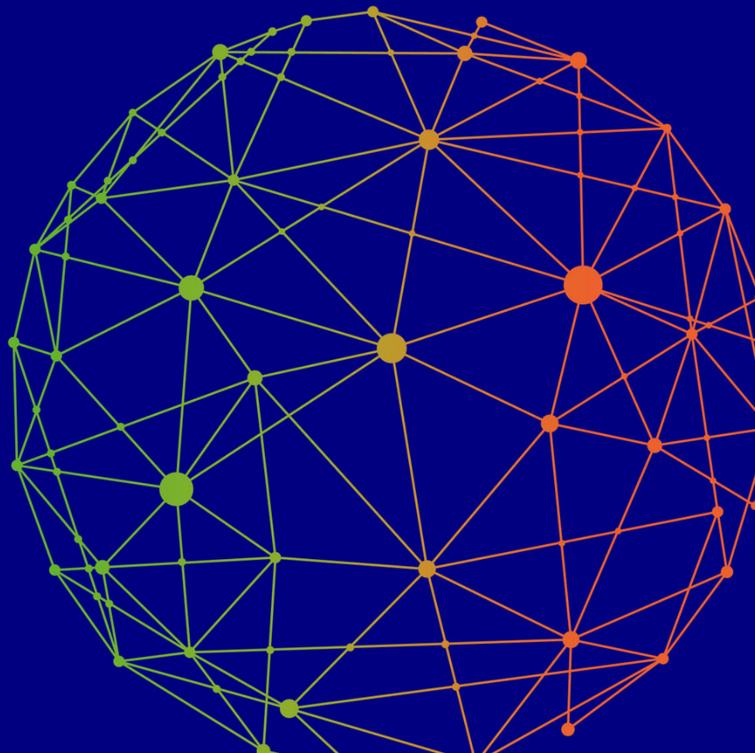
- Significant national and EU funding has supported data spaces since 2019. With technology converging, the focus shifts to adoption, value creation, and data utilization.

- Regulatory framework established: Data Governance Act (DGA), Data Act (DA), AIA (AI Act), and supporting infrastructure like Gaia-X and DSSC.

- While Agdatahub failed due to economic viability, successful projects like Catena-X Aerospace-X/Decade-X, Manufacturing-X, and Energy data spaces optimize supply chains and production.

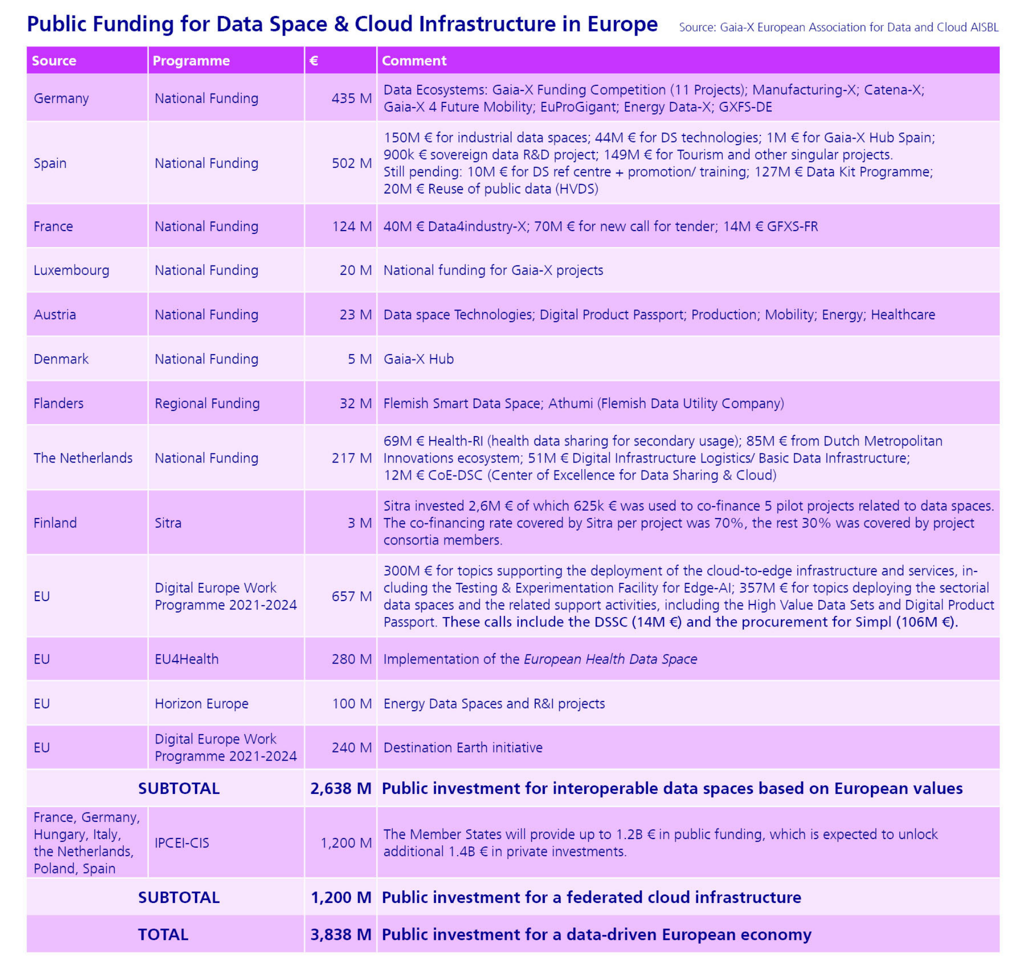
- New European data spaces in key industries (e.g., aerospace, energy, manufacturing) aim for economic viability by 2027.

- From 2028, these data spaces will potentially enable industrial data use for AI training.



European Data Spaces Ecosystem

EU-investments in building data-sharing infrastructure



Public Funding for Data Space & Cloud Infrastructure in Europe

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

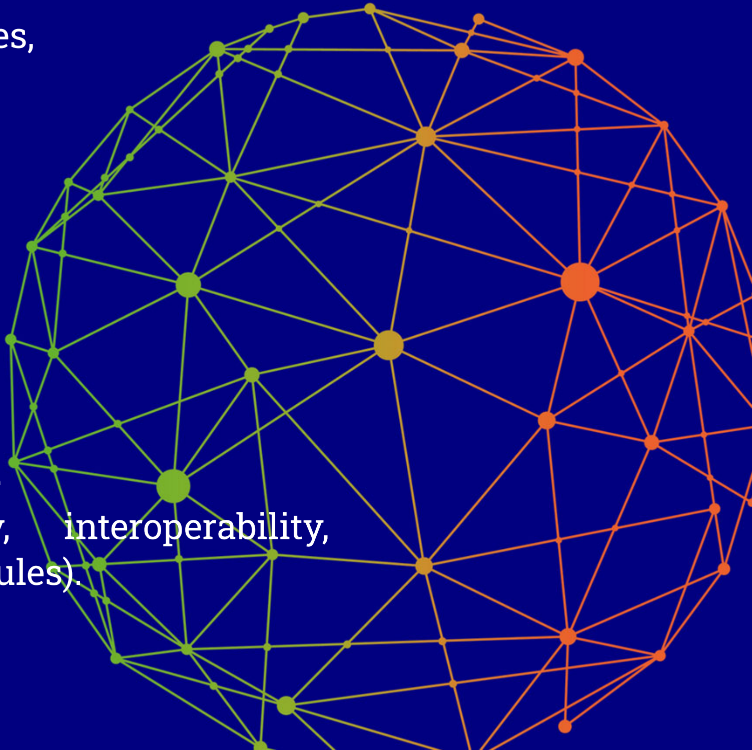
Source	Programme	€	Comment
Germany	National Funding	435 M	Data Ecosystems: Gaia-X Funding Competition (11 Projects); Manufacturing-X; Catena-X; Gaia-X 4 Future Mobility; EuProGigant; Energy Data-X; GXFS-DE
Spain	National Funding	502 M	150M € for industrial data spaces; 44M € for DS technologies; 1M € for Gaia-X Hub Spain; 900k € sovereign data R&D project; 149M € for Tourism and other singular projects. Still pending: 10M € for DS ref centre + promotion/ training; 127M € Data Kit Programme; 20M € Reuse of public data (HVDS)
France	National Funding	124 M	40M € Data4industry-X; 70M € for new call for tender; 14M € GFXS-FR
Luxembourg	National Funding	20 M	National funding for Gaia-X projects
Austria	National Funding	23 M	Data space Technologies; Digital Product Passport; Production; Mobility; Energy; Healthcare
Denmark	National Funding	5 M	Gaia-X Hub
Flanders	Regional Funding	32 M	Flemish Smart Data Space; Athumi (Flemish Data Utility Company)
The Netherlands	National Funding	217 M	69M € Health-RI (health data sharing for secondary usage); 85M € from Dutch Metropolitan Innovations ecosystem; 51M € Digital Infrastructure Logistics/ Basic Data Infrastructure; 12M € CoE-DSC (Center of Excellence for Data Sharing & Cloud)
Finland	Sitra	3 M	Sitra invested 2,6M € of which 625k € 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 2021-2024	657 M	300M € for topics supporting the deployment of the cloud-to-edge infrastructure and services, including the Testing & Experimentation Facility for Edge-AI; 357M € for topics deploying the sectorial data spaces and the related support activities, including the High Value Data Sets and Digital Product Passport. **These calls include the DSSC (14M €) and the procurement for Simpl (106M €).**
EU	EU4Health	280 M	Implementation of the *European Health Data Space*
EU	Horizon Europe	100 M	Energy Data Spaces and R&I projects
EU	Digital Europe Work Programme 2021-2024	240 M	Destination Earth initiative
SUBTOTAL		**2,638 M**	**Public investment for interoperable data spaces based on European values**
France, Germany, Hungary, Italy, the Netherlands, Poland, Spain	IPCEI-CIS	1,200 M	The Member States will provide up to 1.2B € in public funding, which is expected to unlock additional 1.4B € in private investments.
SUBTOTAL		**1,200 M**	**Public investment for a federated cloud infrastructure**
TOTAL		**3,838 M**	**Public investment for a data-driven European economy**

Key initiatives & funding:

- Germany: €435M (Gaia-X, Catena-X, Manufacturing-X).
- Spain: €202M (Industrial & Tourism Data Spaces).
- France: €124M (Data4Industry-X)
- EU: €1,277M (a.o. sectorial data spaces, cloud-to-edge infrastructure)

Challenges:

- Ensuring economic sustainability of data-sharing initiatives.
- Interoperability across industries remains a barrier.
- Companies need incentives to share proprietary data (trust, transparency, interoperability, and clear ownership rules).



European Data Spaces Ecosystem

Exemplary Industry Initiatives

Aerospace

BoostAeroSpace & Aerospace-X/Decade-X

- Joint supply chain optimization among Airbus, Safran, Dassault & Thalès.
- New Aerospace-X/DecadeX project extends to product design & compliance.

Energy

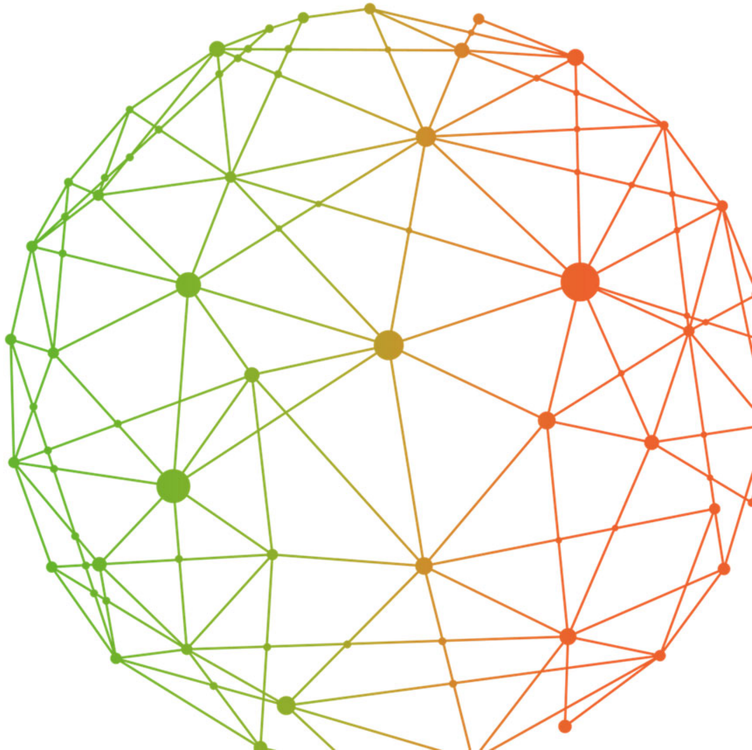
Data-Driven Optimization

- Smart grid data-sharing for real-time energy management.
- France's nuclear data space aims to cut reactor build time to 70 months.

Manufacturing

Factory-X & Catena-X

- Factory-X integrates supply chain AI with shop floor automation.
- Led by Siemens & SAP, ensuring secure industrial data exchange.



Towards a European Data Union

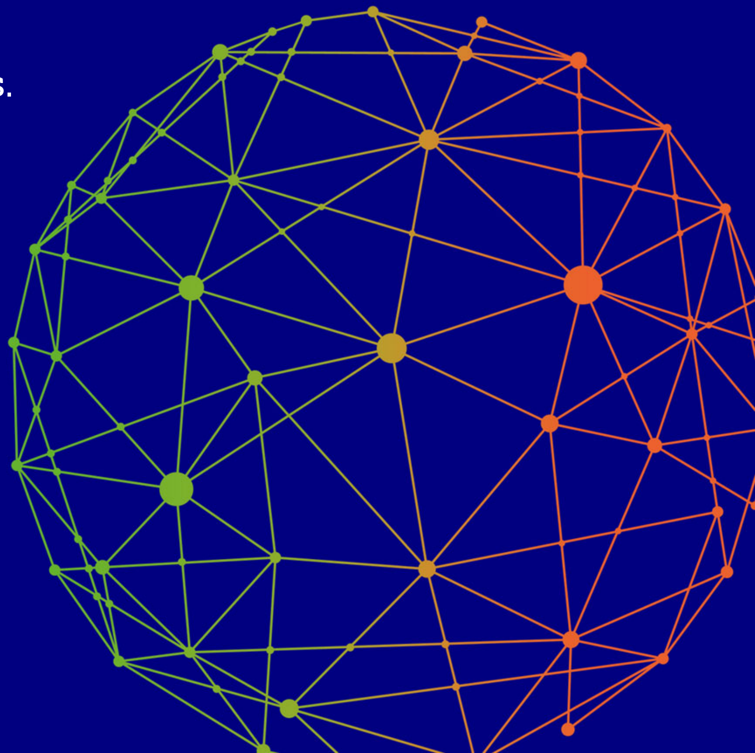
Aligning investment, regulation, and adoption for AI leadership

Draghi Report:

- Trusted Data Intermediaries: The Data Governance Act (DGA) establishes neutral data-sharing platforms.
- Industry-Specific AI Models: Focus on fine-tuning AI models for industrial needs rather than competing with general-purpose AI from the US.
- Standardized EU Data Labels: Introduce certifications for sovereign and interoperable data spaces.
- "AI Sandboxes": Harmonize regulatory test environments to allow GDPR-compliant AI experimentation.

Aligning Data Space Investments with AI Priorities:

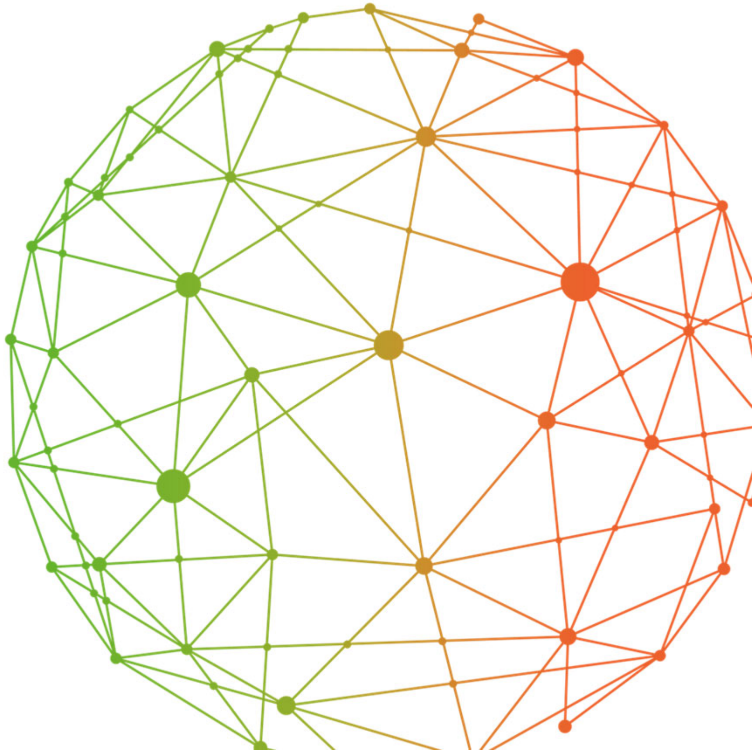
- Targeted Resource Allocation: Focus on high-value areas of the AI and data value chain.
- Shifting existing EU-funded data spaces from pilot projects to scalable applications.
- Focus on ROI: Future AI strategies should ensure data-sharing investments deliver measurable impact (e.g. after 3 years of funding).



Towards a European Data Union

Conclusion

- AI as a Competitive Advantage: Europe must drive AI innovation to reduce economic dependencies and strengthen technological sovereignty.
- European Approach: A shared ecosystem of computing power, large language/foundation models and data for training and fine-tuning AI models.
- Support for open models (e.g., Teuken 7B, Open-R1,...) should be central to the European approach, using internal resources for higher-value segments of the AI and data value chain.
- Need for an Action Plan. that includes technical, governance, and business considerations: Launch of "EU Vertical AI priorities plan".
- European Data Strategy as a Foundation for the transition to a true European Data Union: Season 2 of European data strategy: Maturity Assessment of existing data spaces and launch of new sustainable data spaces



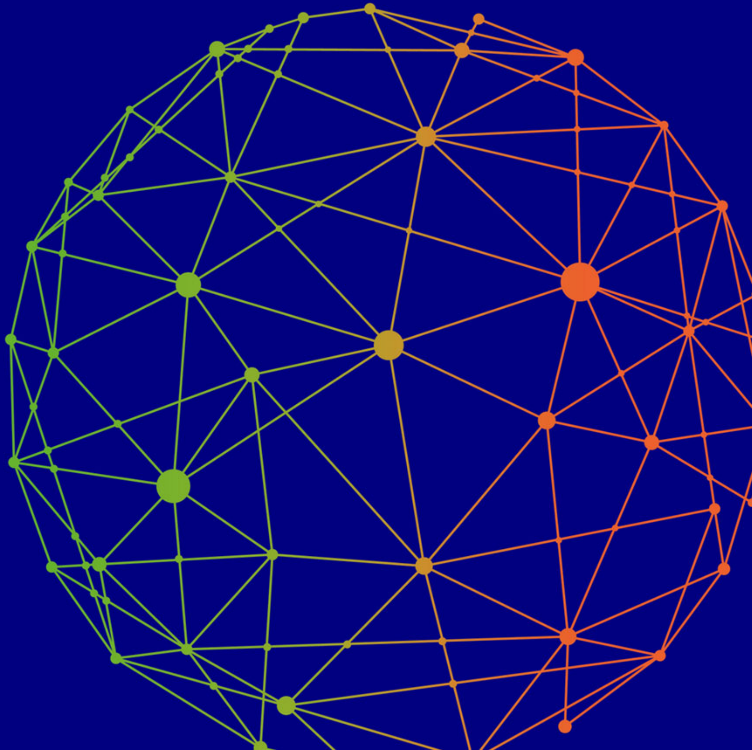
Authors

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.



References

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].

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-morethan-a-market-report-by-enrico-letta.pdf [Accessed: 31 Jan. 2025].

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]

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].

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 [Accessed: 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/thestargate-project-trump-touts-500-billion-bid-for-ai-dominance/ [Accessed: 3 Feb. 2025].

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].

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

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].

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].

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

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].

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].

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

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

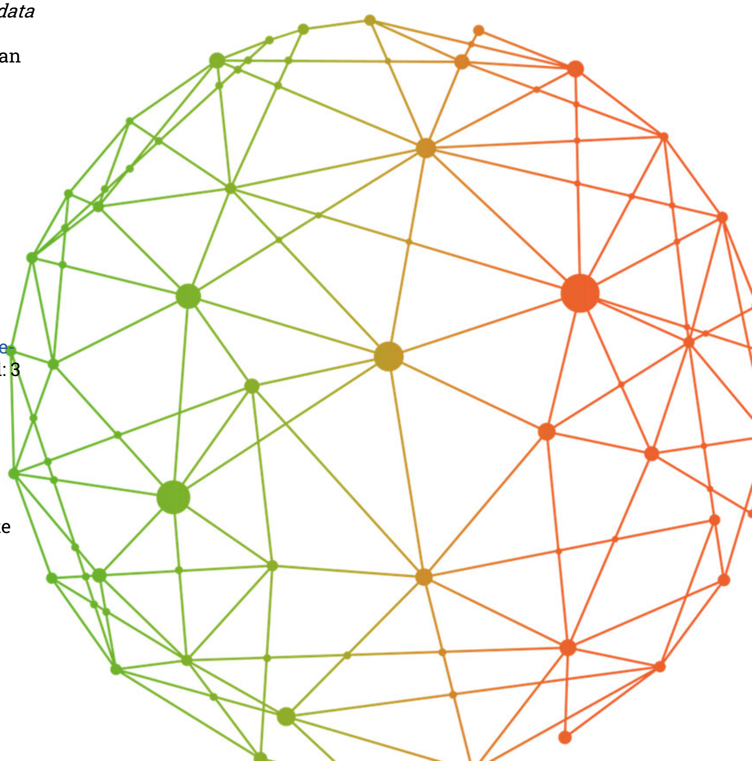
"Given the dominance of US providers, the EU must find a middle way between promoting its domestic cloud industry and ensuring access to the technologies it needs." In: European Commission (2024) *The Draghi report on EU competitiveness* [online]. European Commission. https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en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

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].

HPC: High-Performance Computing.

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-andcreation-of-value-within-data.pdf [Accessed: 3 Feb. 2025].

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.



Thank you

Read the article

English [PDF]:

Spanish [web]:

French [web]:

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".

opean Data Spaces

e Recovery and Resilience Facility Fund (RRF)
per States and the European Commission have
for creating data spaces. The table below
pean and Member State level.

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

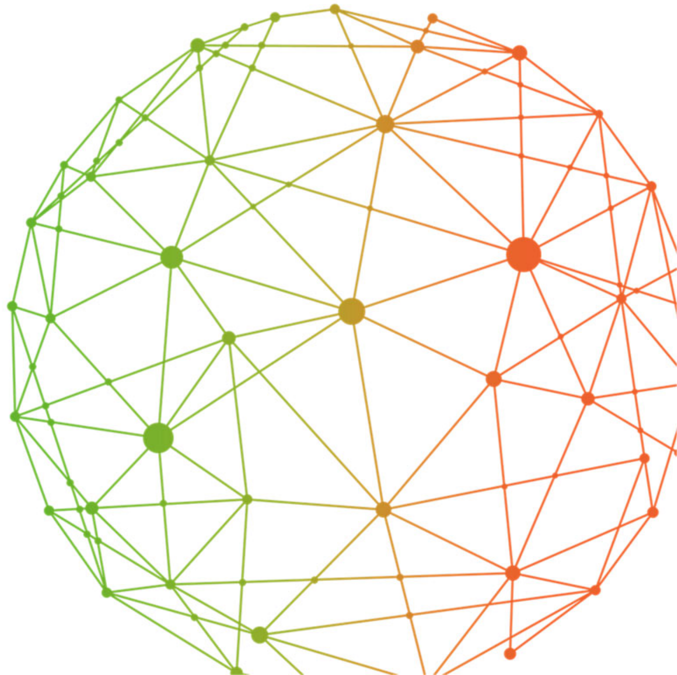
ent
ystems: Gaia-X Funding Competition (11 Projects), Manufacturing-X, Gaia-X 4 Future Mobility, EuProGigant, Energy data-X
ata Spaces Open Call
Spaces Open Call
ustry-X, 70M € for new call for tender
which 625.000 € was used to co-finance 5 pilot projects
e co-financing rate covered by Sitra per project was 70%,
by project consortia members
orial data spaces and the related support activities,
Sets. This set of calls includes the DSSC (14M €)
l data spaces and the related support activities
ct Passport.
Data Space

armer
food trace-
culture in Europe.

Data Spaces Symposium

EDIC: Perfect infrastructure to boost the European Mobility Data Space

Nico Anten and Jon Kuiper,
interviewed by Lars Nagel



Data Spaces Symposium

10:25

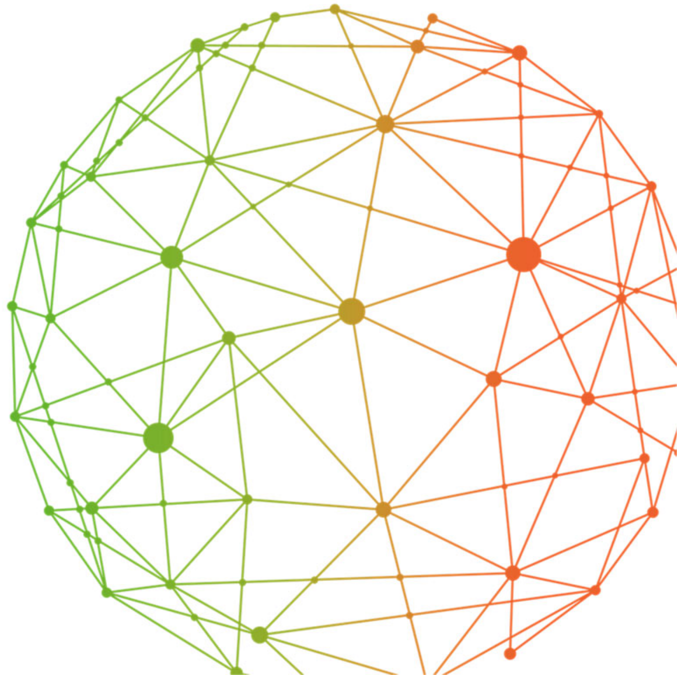
Coffee Break

Take a moment to relax over a cup of coffee

Data Spaces Symposium

The potential and need for data spaces boosting AI

Noburo Koshizuka

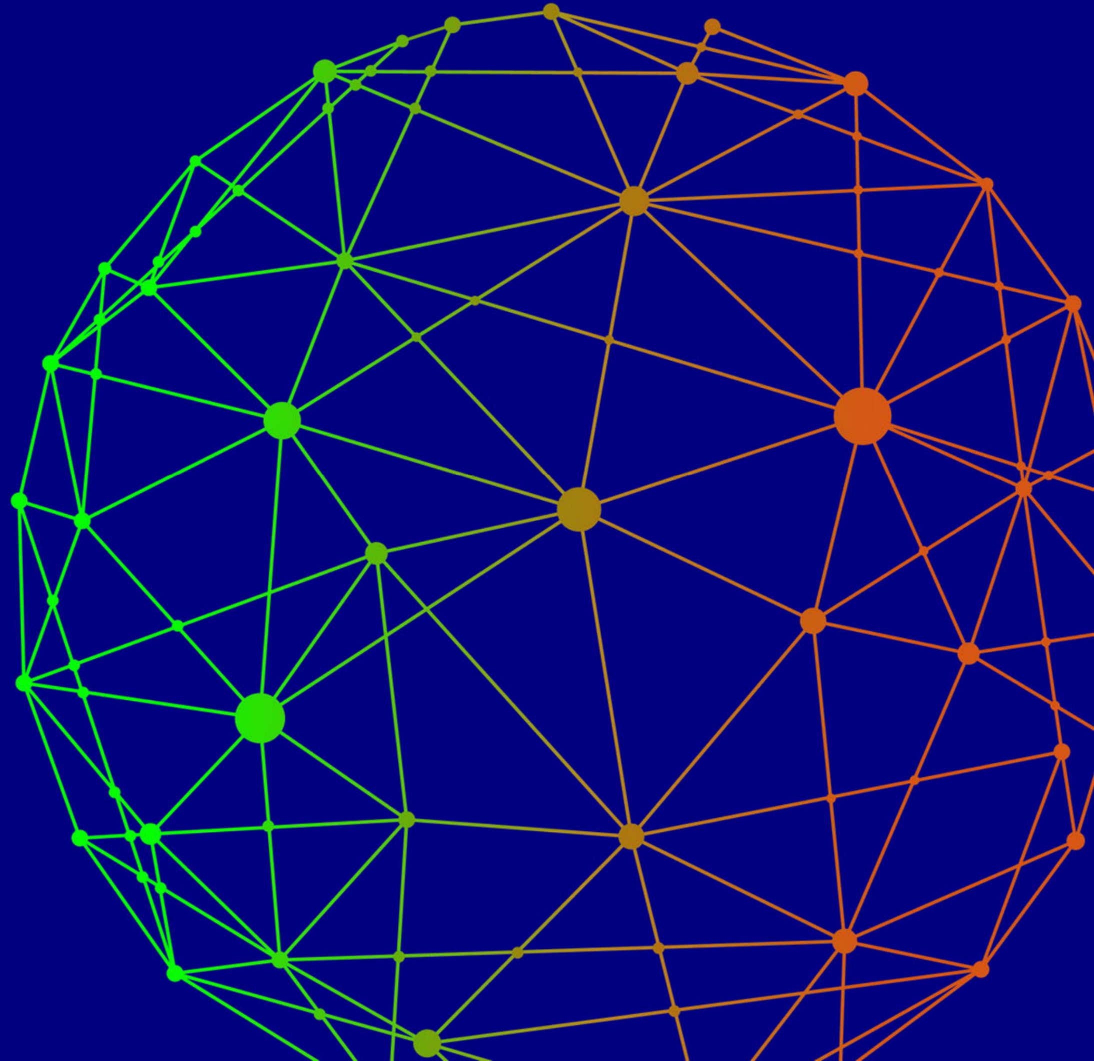


Funded by the European Union

The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412

The Potential and Need for Data Spaces Boosting AI

Noboru Koshizuka 越塚登
Professor, The University of Tokyo
Chair, Data Society Alliance, Japan
Japan Hub Coordinator, IDSA



Introduction

Noboru Koshizuka
越塚　登

Professor
The University of Tokyo

Sub-Project Director
Japan Mobility Data Space

Chair
Data Society Alliance

Chair
Asia Open Data Partnership

Japan Hub Coordinator
Ambassador
IDSA

Director
Green x Digital Consortium

Chair
Weather x Business Consortium

Director
Smart City Social Implementation Consortium

PART 1
Background

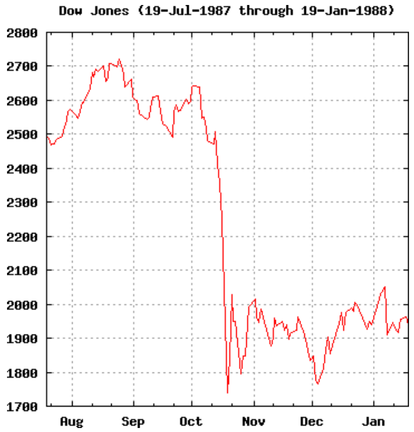
History of AI Technologies

"Artificial Intelligence" proposed by Prof. John McCarthy (MIT)（1957）

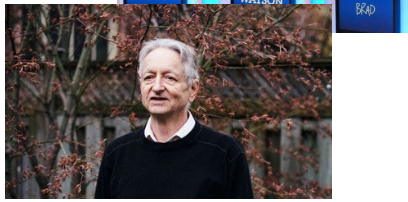
"2001: A Space Odyssey"（1968） HAL 9000

5th Generation Computer Project（1982～1992）

Black Monday（1987）



"Deep Learning" proposed by Prof. Geoffrey Hinton（2006）

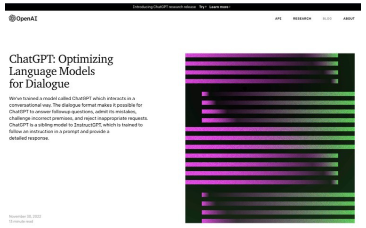


Quiz: IBM Watson won human（2011）

"Shogi" ：AI won professional Shogi Player（2013）

"Go" ：Google Alpha Go won the world champion of "GO"（柯潔）（2017）

ChatGPT (2023)



Introducing ChatGPT research release Try ↗ Learn more >

OpenAI API RESEARCH BLOG ABOUT

ChatGPT: Optimizing Language Models for Dialogue

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests. ChatGPT is a sibling model to InstructGPT, which is trained to follow an instruction in a prompt and provide a detailed response.

November 30, 2022
13 minute read

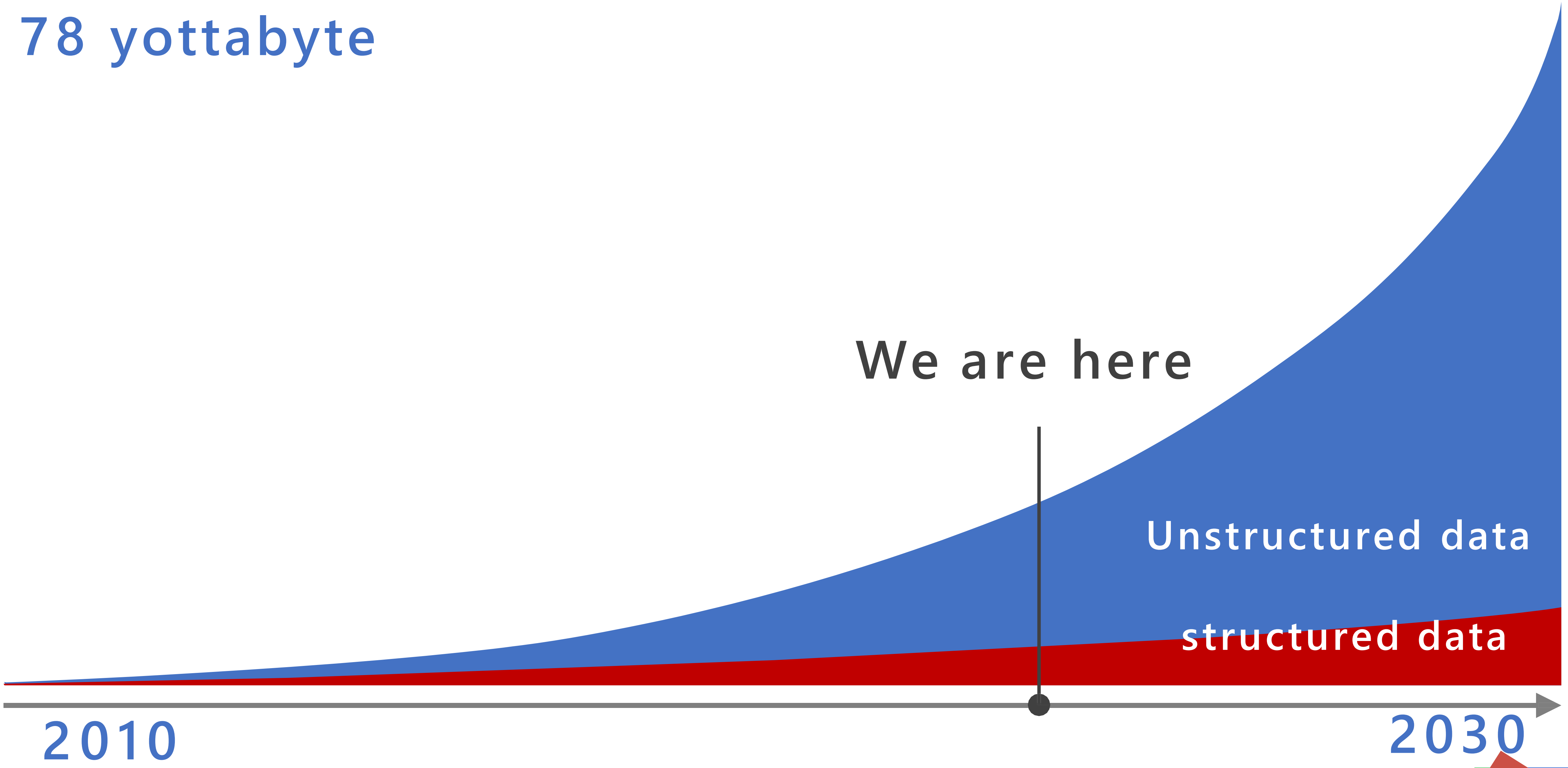
1950's 1960's 1970's 1980's 1990's 2000's 2010's 2020's

AI 1st Wave (1950's~60's)

AI 2nd Wave (1980's)

AI 3rd Wave (2010's~)

Data Growth over the years



Viewpoints from Data Community

Artificial Intelligence

(Machine Learning, Large Language Models, ...)

‖

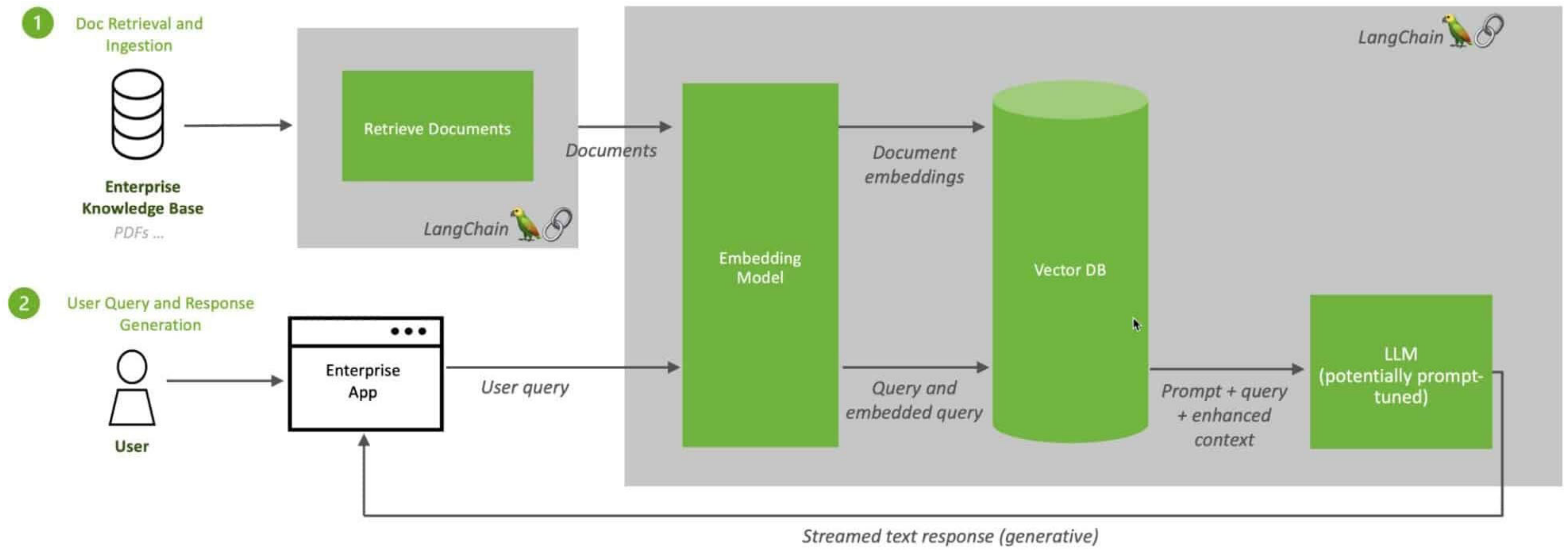
Data Driven Intelligence

“Dataspaces has **big** potential impact for enhancing AI.”

PART 2
Recent AI Trends

Recent AI Trends (1): RAG（Retrieval-Augmented Generation）

Retrieval Augmented Generation (RAG) Sequence Diagram

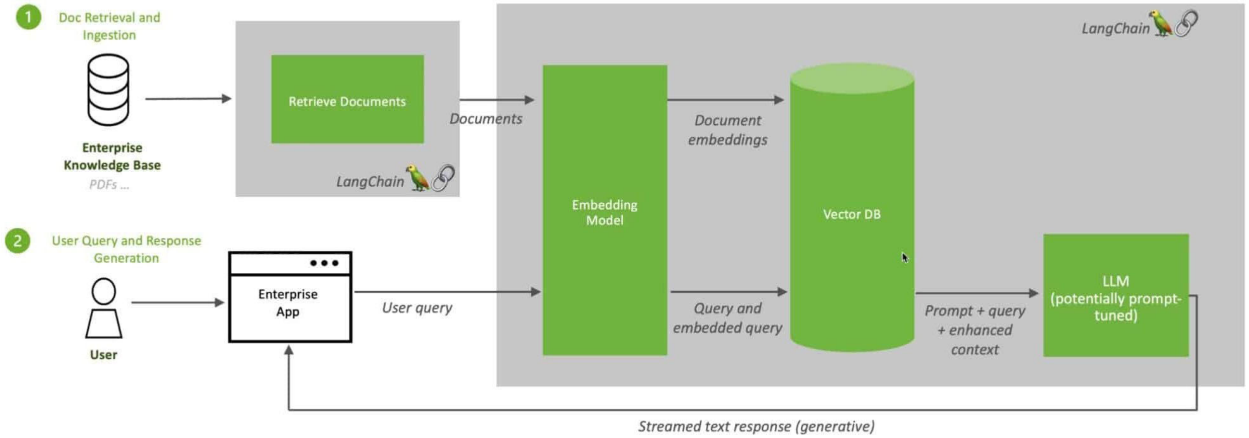


1
Doc Retrieval and Ingestion
Enterprise Knowledge Base
PDFs ...
Retrieve Documents
LangChain
Documents
LangChain
Embedding Model
Document embeddings
Vector DB
2
User Query and Response Generation
User
Enterprise App
User query
Query and embedded query
Prompt + query + enhanced context
LLM (potentially prompt-tuned)
Streamed text response (generative)

https://blogs.nvidia.co.jp/2023/11/17/what-is-retrieval-augmented-generation/

Recent AI Trends (2): AI + Database

AI + Text Database = RAG (Retrieval Augmented Generation)

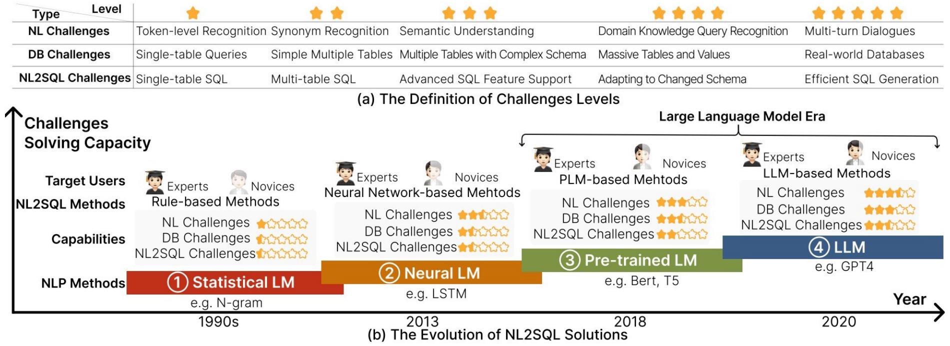

1
Doc Retrieval and Ingestion
Enterprise Knowledge Base
PDFs ...
Retrieve Documents
LangChain
Documents
Embedding Model
Document embeddings
Vector DB
LangChain
2
User Query and Response Generation
User
Enterprise App
User query
Query and embedded query
Prompt + query + enhanced context
LLM (potentially prompt-tuned)
Streamed text response (generative)

RAG + AI Agent = Agentic RAG

- Single Agentic RAG
- Multi-Agent Agentic RAG
- Hierarchical Agentic RAG
- Agentic Corrective RAG
- Adaptive Agentic RAG
- Graph-Based Agentic RAG

AI + RDB = NLIDB (Natural Language Interface to Database)

- Approach: Rule-Based, Neural Network (NN) Based, Pretrained Language Model Based, Large Language Model (LLM) Based, ...



Type \ Level	★	★★	★★★	★★★★	★★★★★
NL Challenges	Token-level Recognition	Synonym Recognition	Semantic Understanding	Domain Knowledge Query Recognition	Multi-turn Dialogues
DB Challenges	Single-table Queries	Simple Multiple Tables	Multiple Tables with Complex Schema	Massive Tables and Values	Real-world Databases
NL2SQL Challenges	Single-table SQL	Multi-table SQL	Advanced SQL Feature Support	Adapting to Changed Schema	Efficient SQL Generation

(a) The Definition of Challenges Levels

(b) The Evolution of NL2SQL Solutions

AI + GraphDB: NL2Cypher

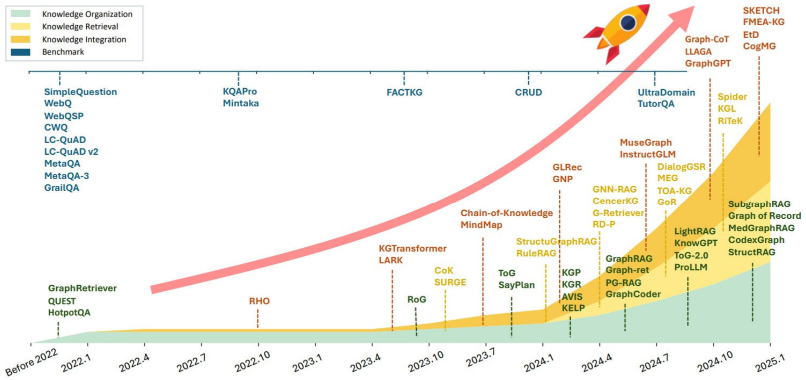
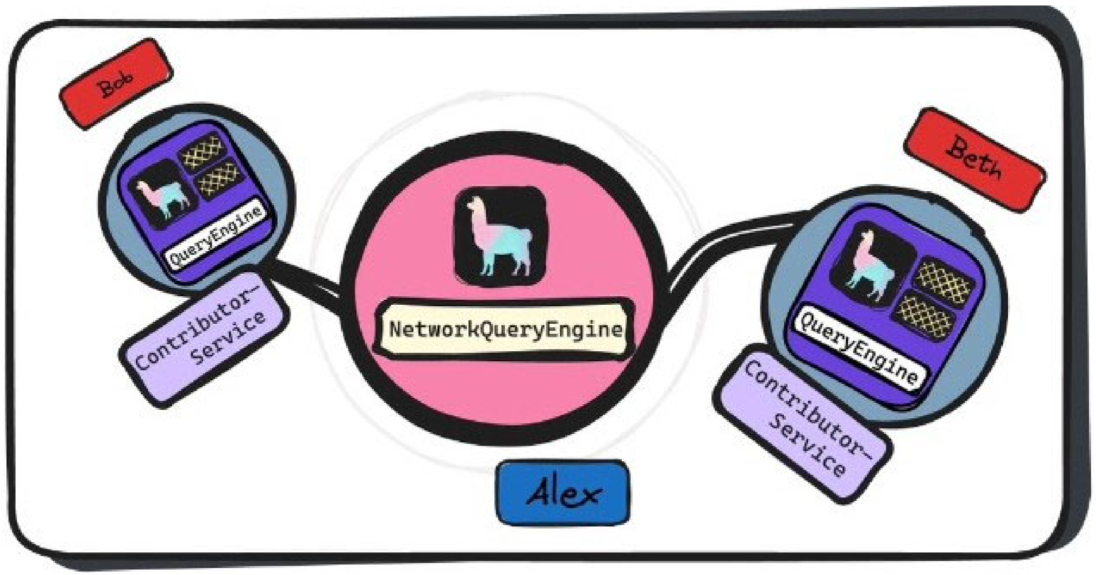


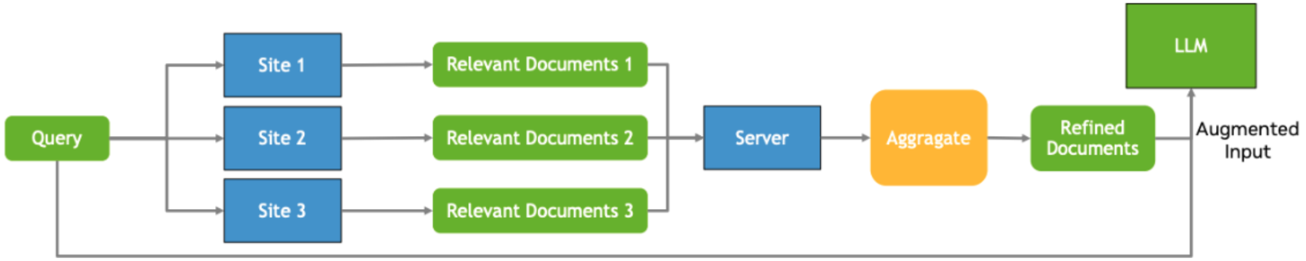
Fig. 1: The development trends in the field of GraphRAG with representative works.

Recent AI Trends (3): AI + Distributed Database

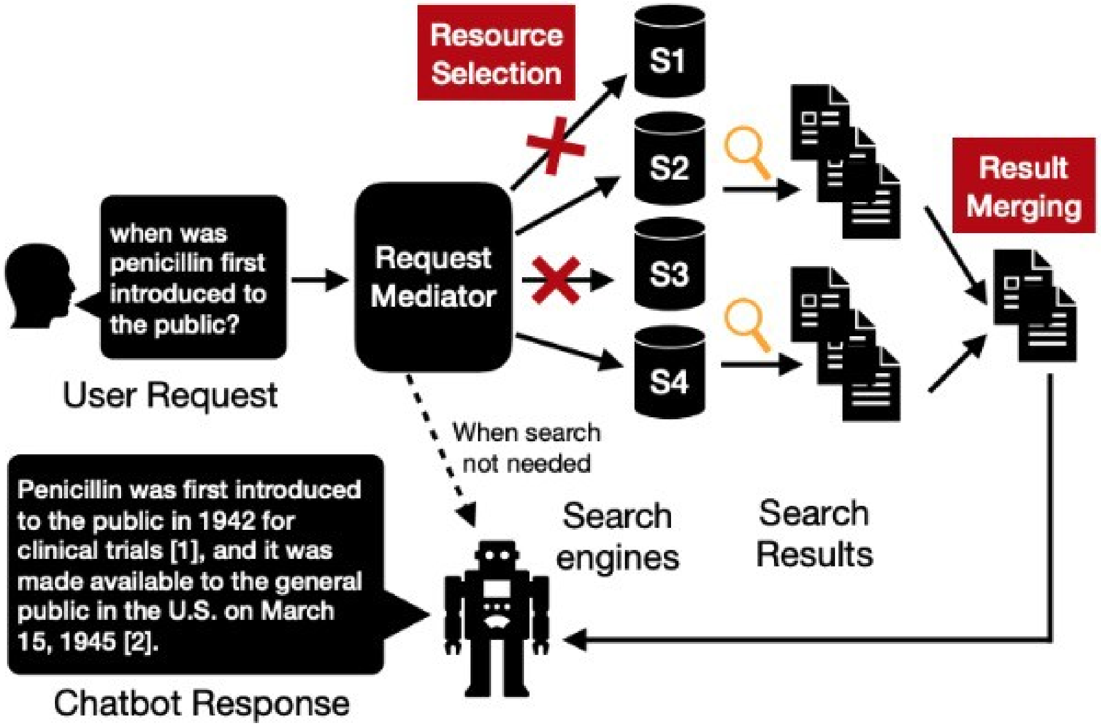


llama-index-networks

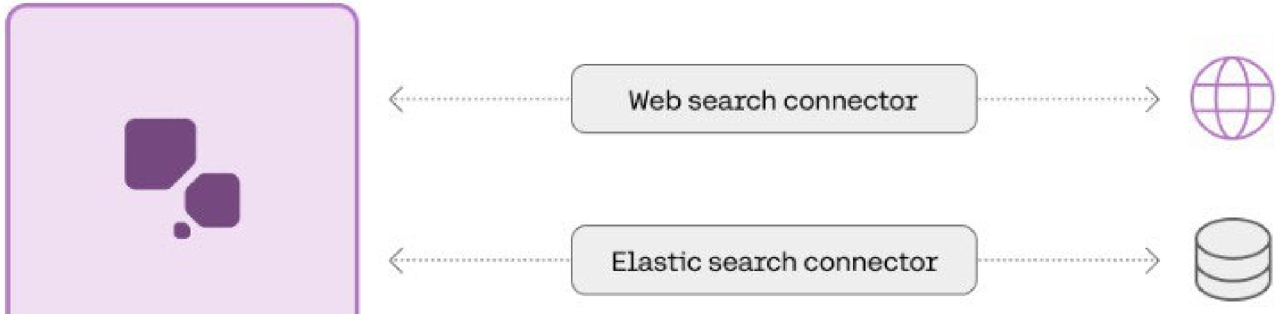
https://github.com/run-llama/llama_index/tree/main/llama-index-networks



C-FedRAG （Confidential Federated RAG）


Resource Selection
S1
S2
S3
S4
Result Merging
when was penicillin first introduced to the public?
User Request
Request Mediator
When search not needed
Search engines
Search Results
Penicillin was first introduced to the public in 1942 for clinical trials [1], and it was made available to the general public in the U.S. on March 15, 1945 [2].
Chatbot Response

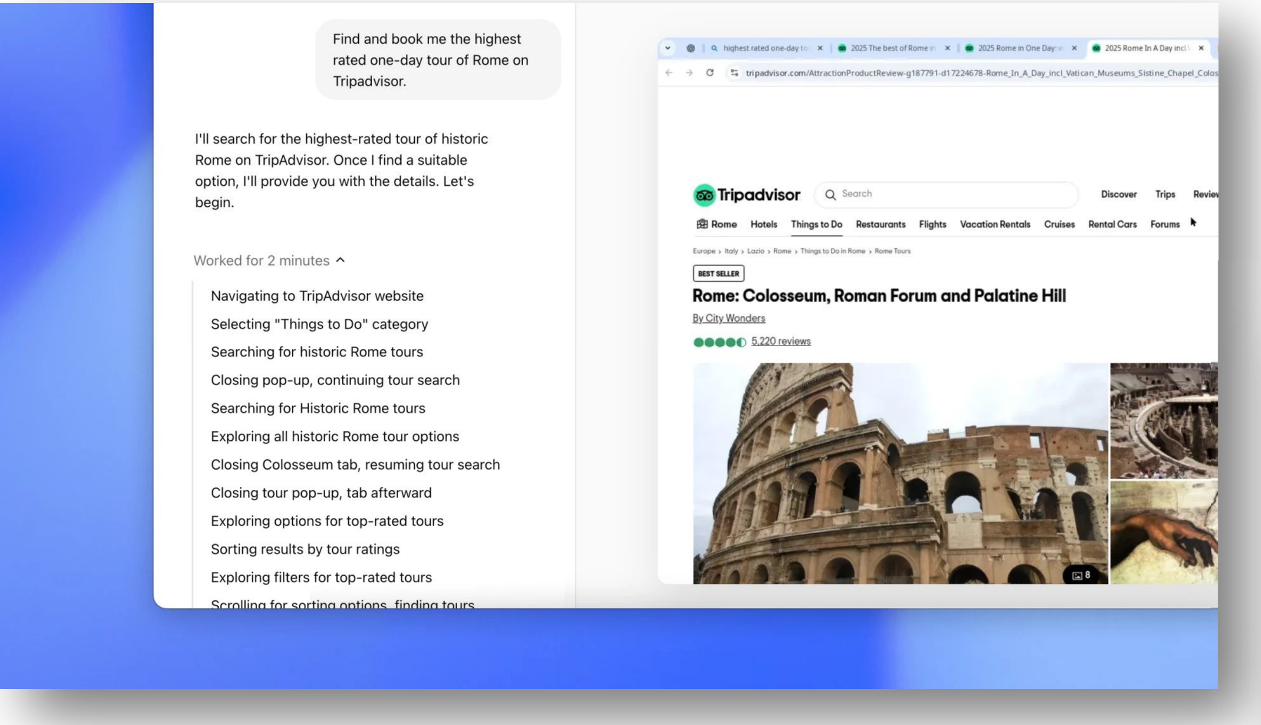
FeB4RAG



RAG Connector

https://cohere.com/llmu/rag-connectors

Recent AI Trends (4): AI Agent



Find and book me the highest rated one-day tour of Rome on Tripadvisor.

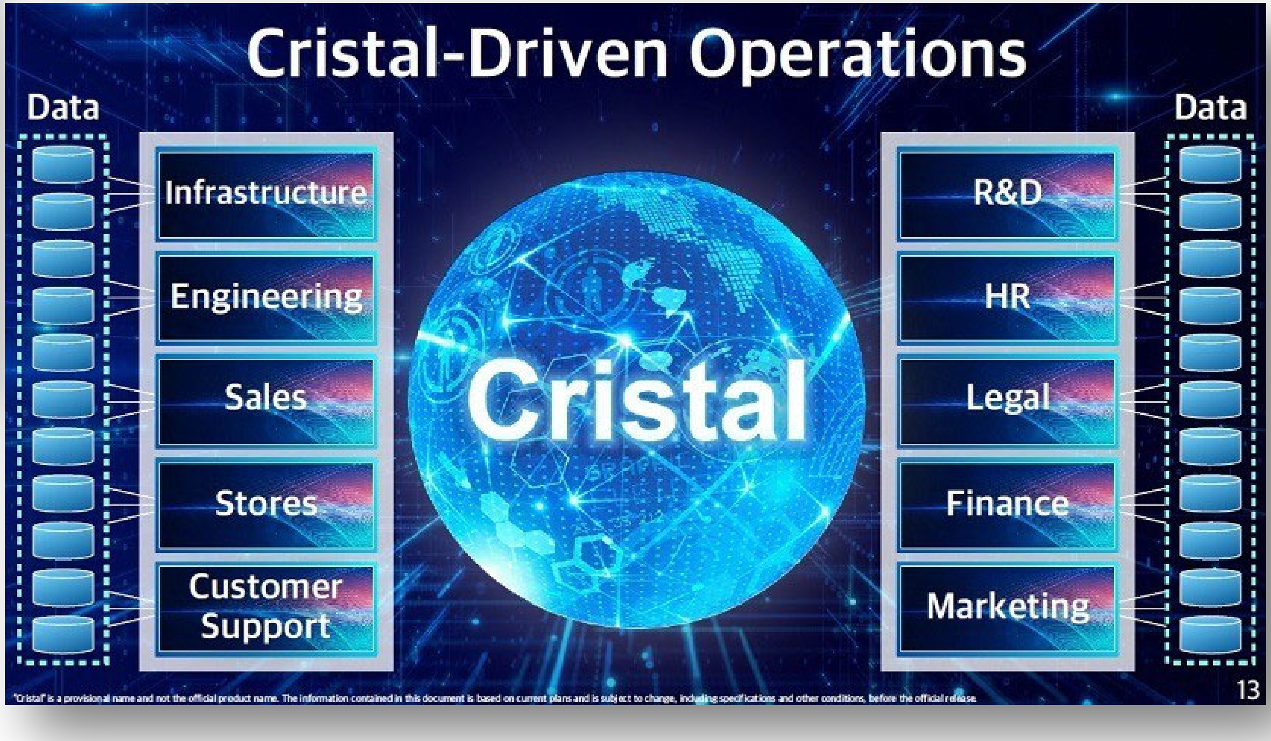
I'll search for the highest-rated tour of historic Rome on TripAdvisor. Once I find a suitable option, I'll provide you with the details. Let's begin.

Worked for 2 minutes ^

- Navigating to TripAdvisor website
- Selecting "Things to Do" category
- Searching for historic Rome tours
- Closing pop-up, continuing tour search
- Searching for Historic Rome tours
- Exploring all historic Rome tour options
- Closing Colosseum tab, resuming tour search
- Closing tour pop-up, tab afterward
- Exploring options for top-rated tours
- Sorting results by tour ratings
- Exploring filters for top-rated tours
- Scrolling for sorting options, finding tours

hiqhest rated one-day to
2025 The best of Rome in
2025 Rome in One Day: w
2025 Rome In A Day incl.V
tripadvisor.com/AttractionProductReview-g187791-d17224678-Rome_In_A_Day_incl_Vatican_Museums_Sistine_Chapel_Colos
Tripadvisor
Search
Discover
Trips
Review
Rome
Hotels
Things to Do
Restaurants
Flights
Vacation Rentals
Cruises
Rental Cars
Forums
Europe › Italy › Lazio › Rome › Things to Do in Rome › Rome Tours
BEST SELLER
Rome: Colosseum, Roman Forum and Palatine Hill
By City Wonders
5,220 reviews
8

"Operator" (OpenAI, 2025)



"Cristal" is a provisional name and not the official product name. The information contained in this document is based on current plans and is subject to change, including specifications and other conditions, before the official release.

"Cristal" (Softbank, 2025)

POINT !

RAG
AI-Agent
etc..

+

Database
ERP
etc...

Integration/Mixture of AI Systems and Legacy Information Systems

Next Step

RAG
AI-Agent
etc..

+

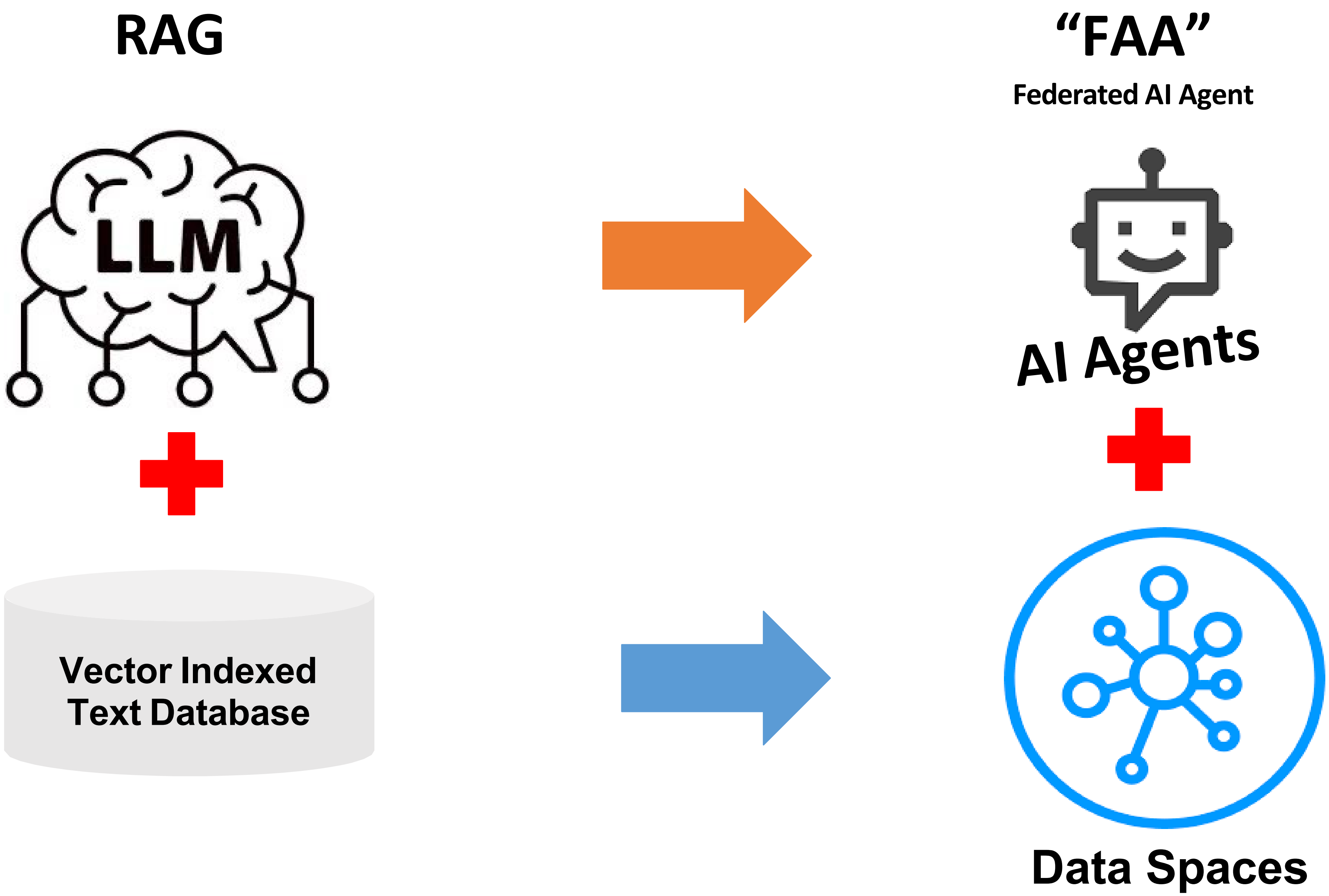
Dataspases

Integration/Mixture of AI Systems and Legacy Information Systems

PART 3
F-RAG (Federated RAG)
AI + Dataspace

Study Trial at Koshizuka-Lab. U-Tokyo

“Federated AI Agent” [Koshizuka-lab, 2025]



RAG

“FAA”

Federated AI Agent

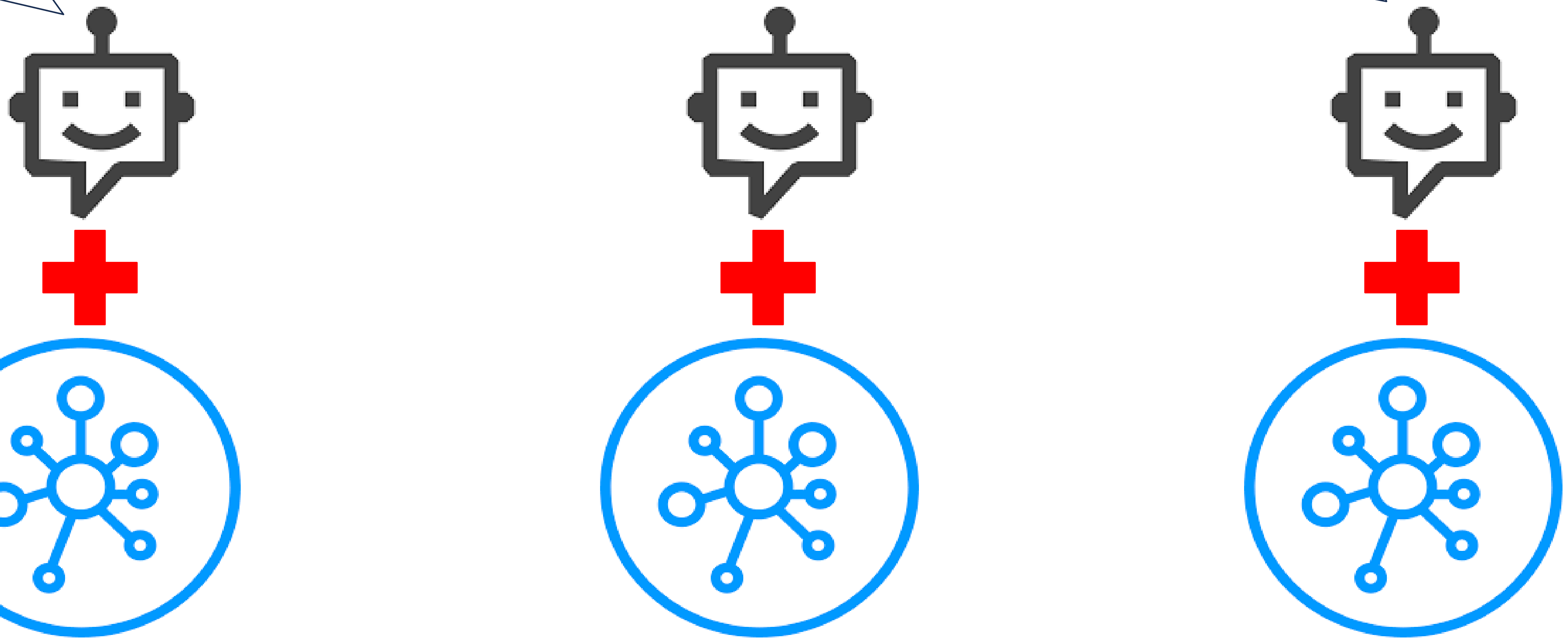
What we can do with F-RAG (1)
Complex queries by Natural Languages…Enhancing findability of data

“Find the data for the calculation of CO2 emission of Product **XDW1029381** ?”

“When does next bus arrive at this bus stop?”

Complex question What are the name and budget of the departments with average instructor salary greater than the overall average?

Searching data by investigating the contents of data

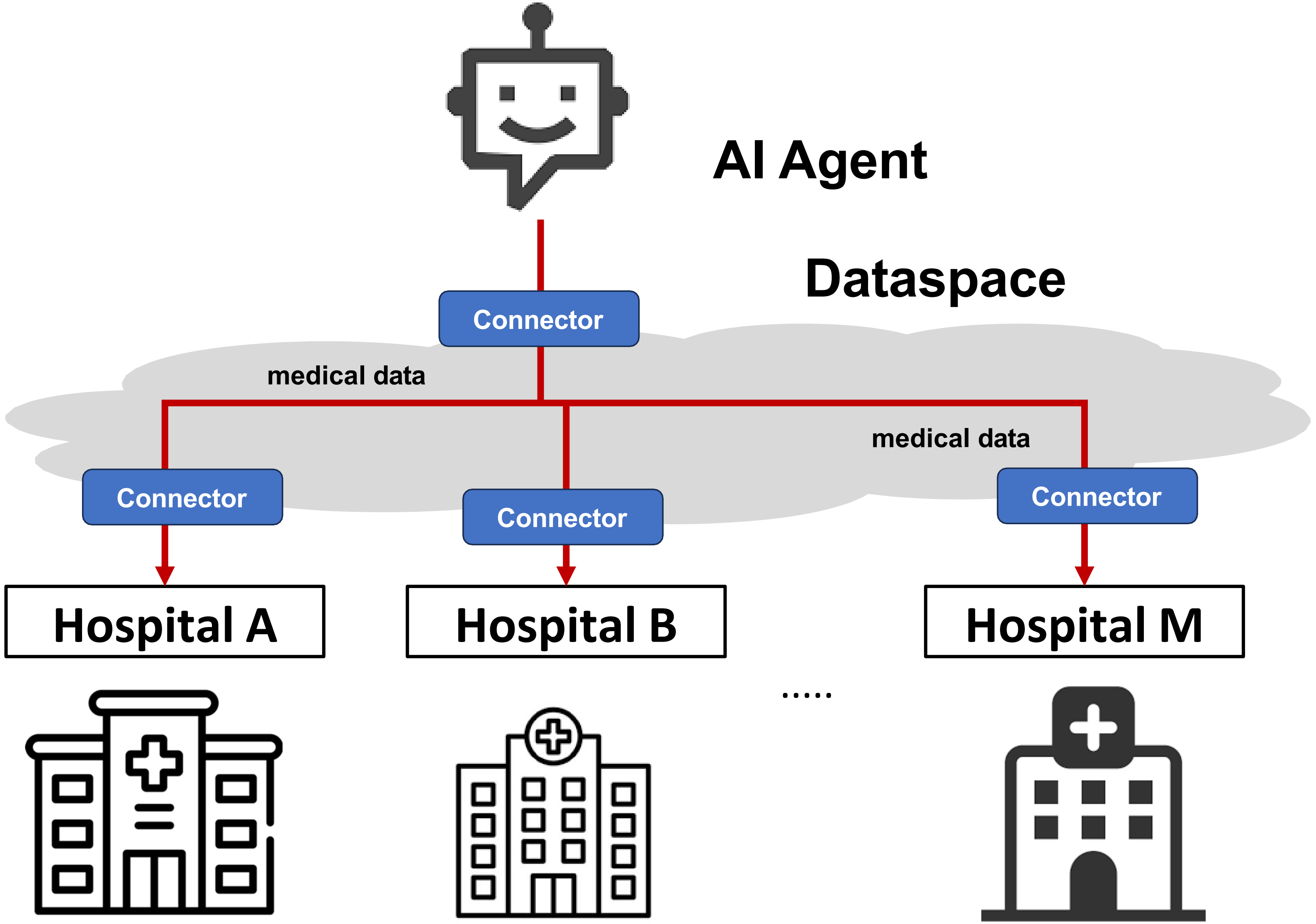


Green Product Dataspace

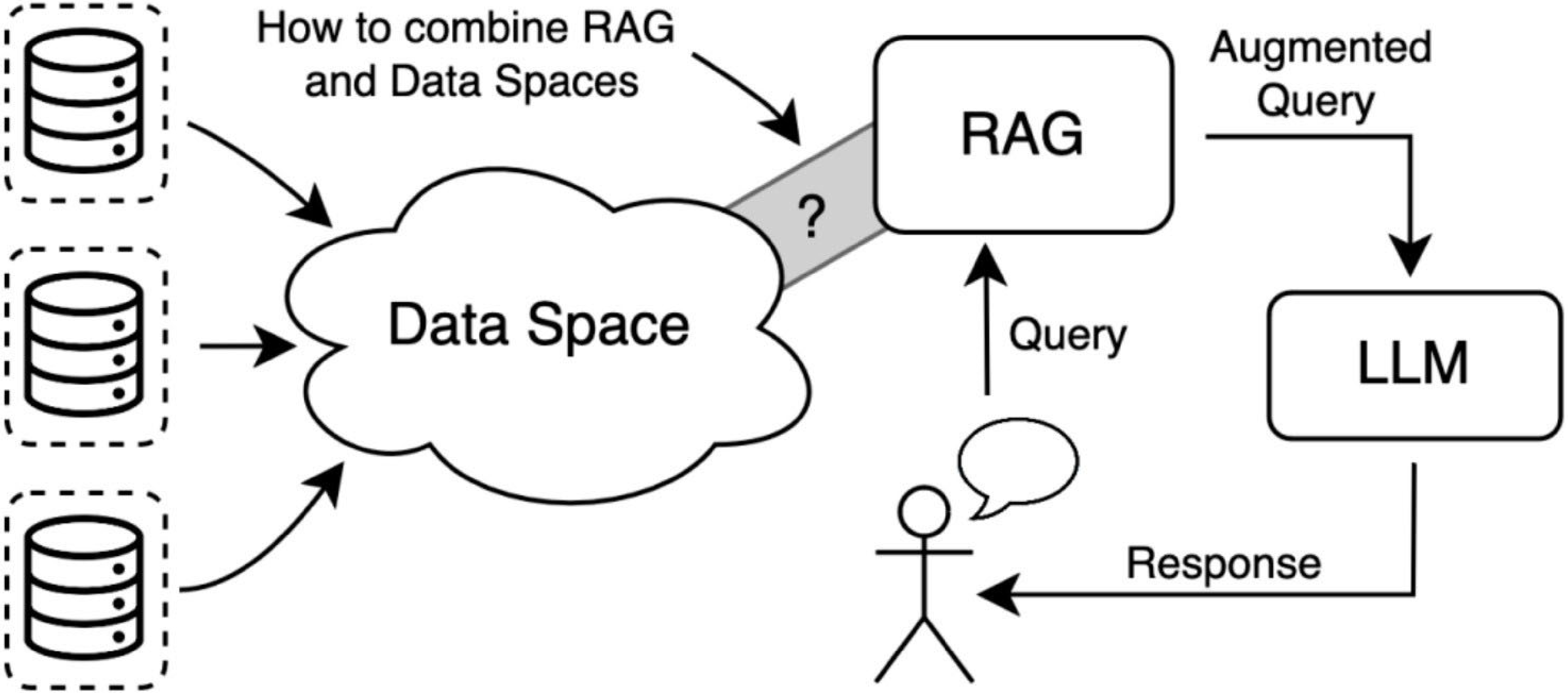
Mobility Dataspace

Private Dataspace of ERP

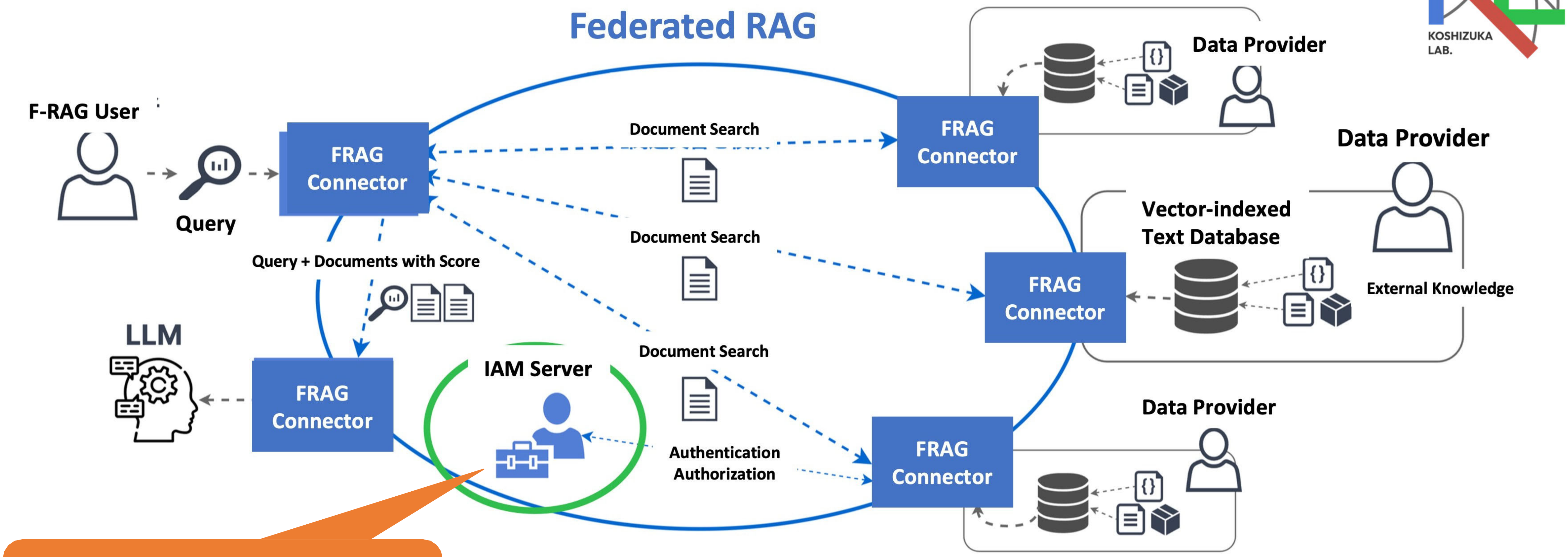
What we can do with F-RAG (2)
RAG/Fine Tuning with Multiple Collaborative Data Sources



DS-RAG [Hermsen et al., 2024]



F-RAG (Federated RAG) [Matsunaga, 2024]

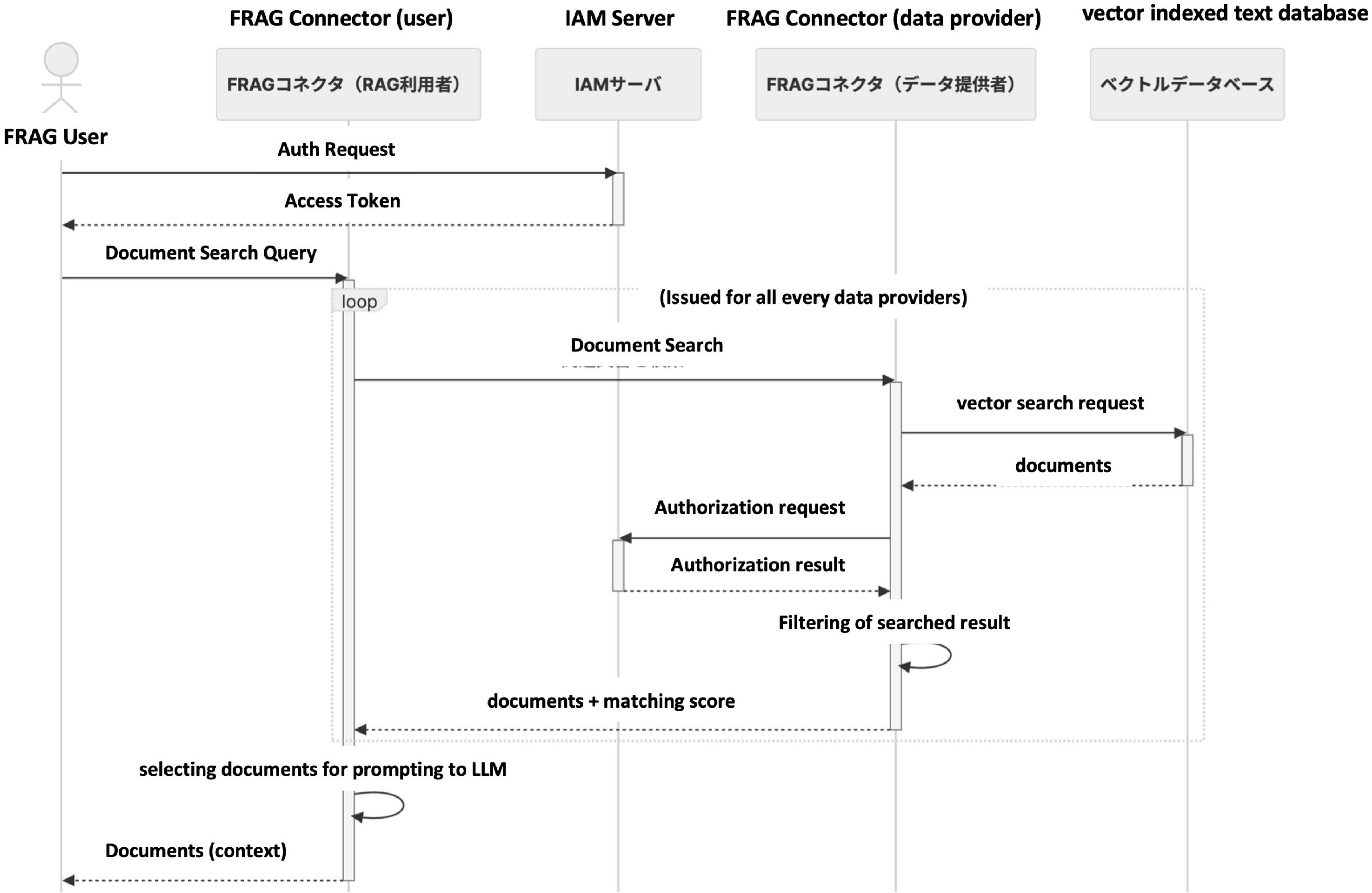


Trust, Usage-control

F-RAG Architecture

important difference between **federation** and **distribution**

New Data Space Protocol for F-RAG



FRAG Connector (user)

IAM Server

FRAG Connector (data provider)

vector indexed text database

FRAG User

FRAGコネクタ（RAG利用者）

IAMサーバ

FRAGコネクタ（データ提供者）

ベクトルデータベース

Auth Request

Access Token

Document Search Query

loop

(Issued for all every data providers)

Document Search

vector search request

documents

Authorization request

Authorization result

Filtering of searched result

documents + matching score

selecting documents for prompting to LLM

Documents (context)

Experiment Result

Data Set	RAG Architecture	Hit Rate	Precision	Recall	MRR
FiQA-2018	**RAG**	0.5633	0.1728	0.373	0.443
	F-RAG	0.566	0.1731	0.3746	0.4427
	Difference (%)	0.4793	0.1736	0.429	-0.0677
NQ	**RAG**	0.5226	0.1127	0.4841	0.369
	F-RAG	0.5383	0.1158	0.4977	0.3798
	Difference (%)	3.004	2.751	2.809	2.927
TREC-COVID	**RAG**	1.0	0.74	0.009541	0.89
	F-RAG	0.996	0.72	0.009167	0.8805
	Difference (%)	-0.4	-2.703	-3.92	-1.067

Precision Test Result

Architecture	Number of Data Providers	Response Time for LLM Query (sec.)
RAG	-	0.0389
F-RAG	1	0.7106
	2	0.9906
	4	1.112
	6	1.179
	8	1.277
	10	1.358
	16	1.679
	24	2.298
	32	2.847

Performance Test Result

! **Inference precision is almost the same.**

! **The is much performance degradation.**

PART 4
Key Insights

Benefits of “Dataspaces plus AI”

Dataspace is useful for AI (LLM)

- In Learning Phase
  - Learning from data sets from multiple collaborative organizations
- In Inferencing Phase
  - Inferencing with Databases (eg. **F-RAG**)
  - Dynamic Learning with Databases (eg. **F-RAG**)
  - Dealing with **Real-time** Information

…

AI (LLM) is useful for Dataspace

- Enhancing findability of data
  - eg. **Data catalog** with LLM provide us natural language interfaces for searching data

…

Next Steps for "Dataspaces plus AI"

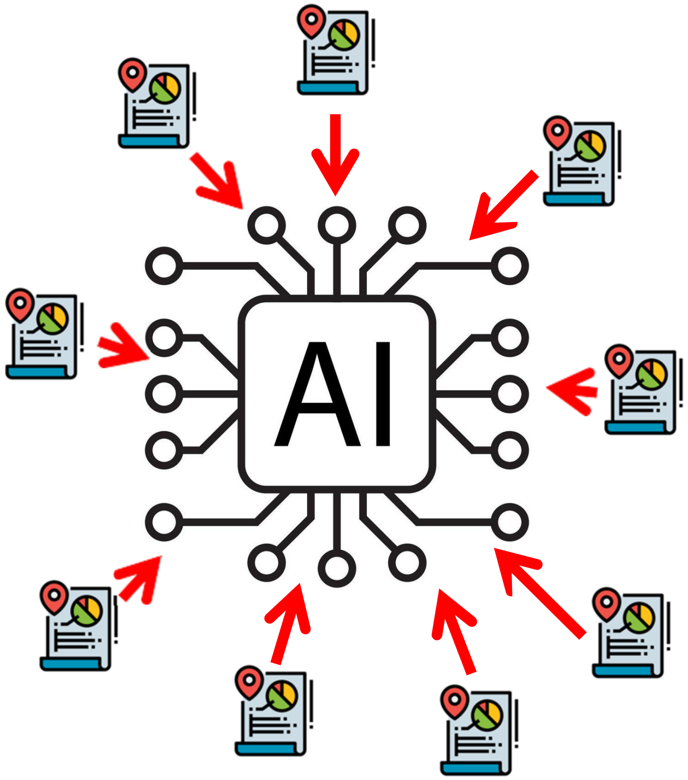
1. New "dataspace protocols" for AI
   - Vector index search protocol for F-RAG

2. New "usage control" for AI
   - eg. data usage for "machine learning" is OK or Not
   - eg. data usage for "F-RAG" is OK or Not

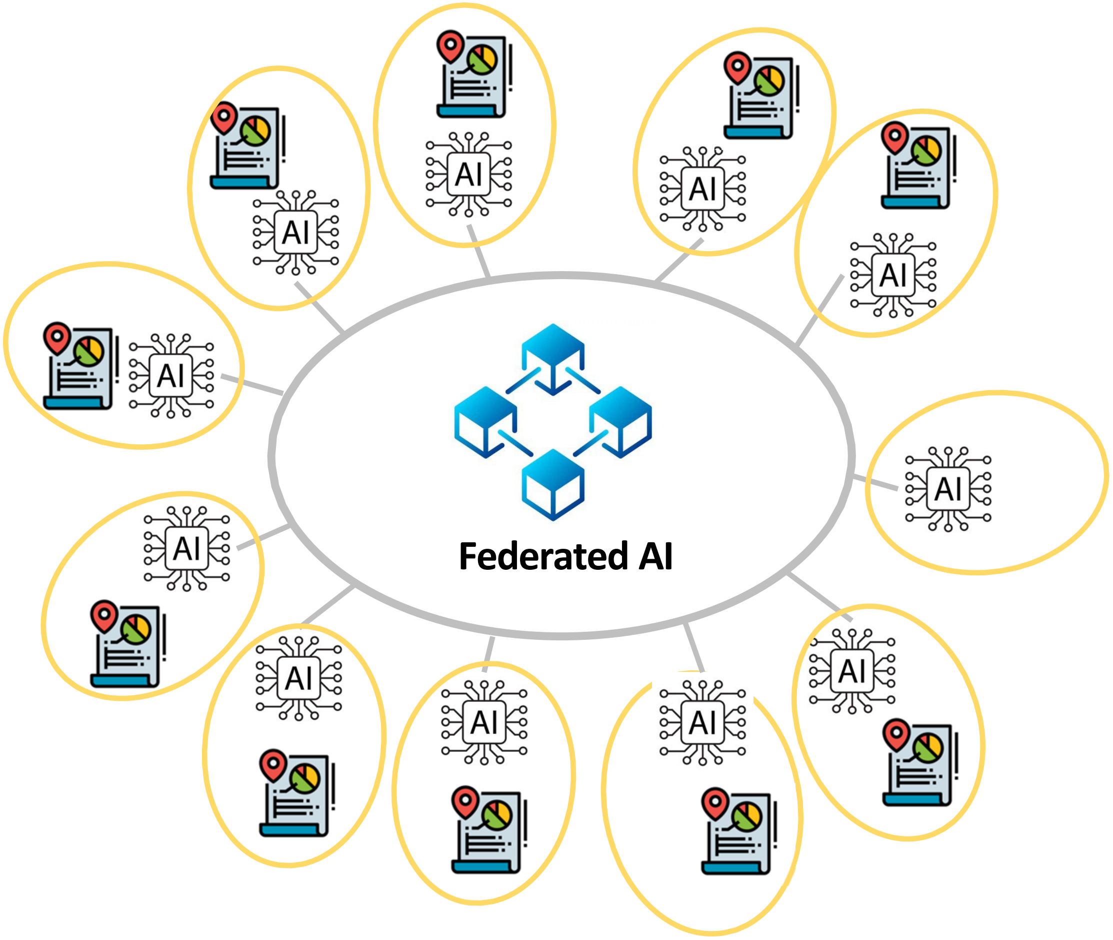
3. Data quality for Readability by Autonomous AI
   - Machine readability ➔ AI readability
   - Open Data "Five Star" is enough?

PART 5
Future: From Dataspace to AI Spac

Future: From Data Spaces into "AI Spaces" [Koshizuka-lab, 2024]



VS.



Data Monopolization
Data Hegemony

Big General AI

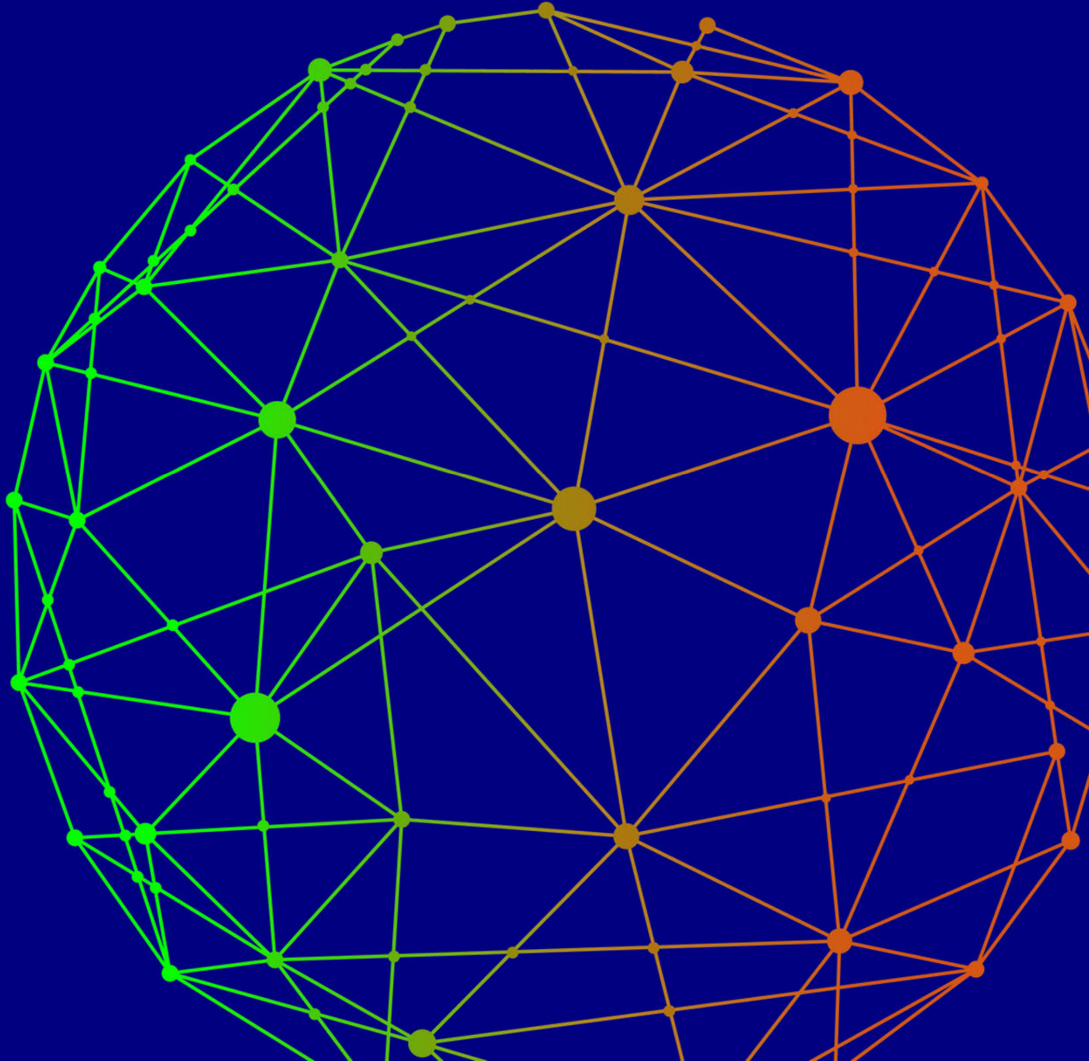
Data Sovereignty

AI Space

Message

“AI is the most important stakeholder of dataspaces.”

Funded by the European Union The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412

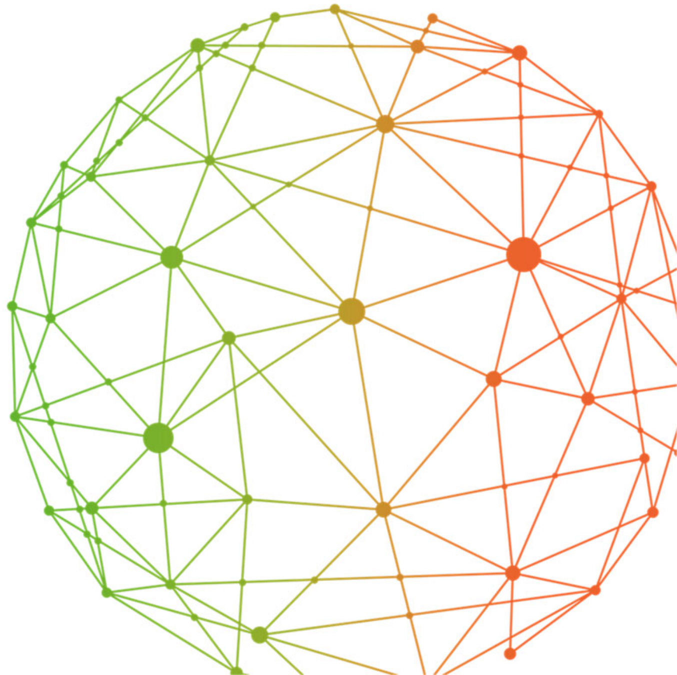


office@koshizuka-lab.org

Data Spaces Symposium

OpenEuroLLM: Building Europe's AI future on open source

Jan Hajic

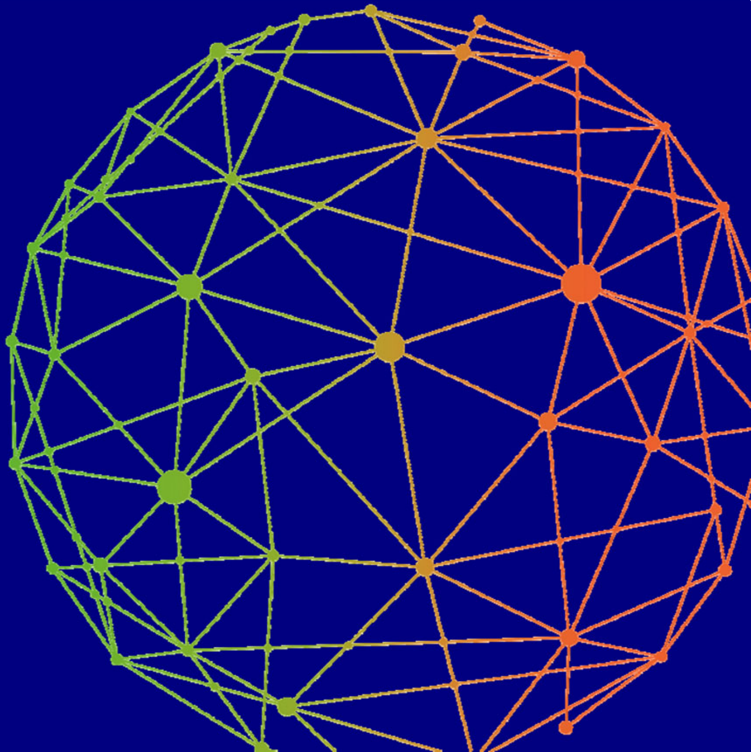


OpenEuroLLM

New generation of open LLMs for Europe

Data Spaces Symposium 2025

Jan Hajič, Charles University, Prague, CZ



DSBA ■ BDV BIG DATA VALUE ASSOCIATION FIWARE FOUNDATION gaia-x INTERNATIONAL DATA SPACES ASSOCIATION

DATA SPACES SUPPORT CENTRE

Funded by the European Union

The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412

OpenEuroLLM

- Our goal:
  - Open
  - Multilingual
  - European
  - Generative
  - Foundational LLM(s)

- Open Source (in full)

  including fully inspectable data

- 32+ languages

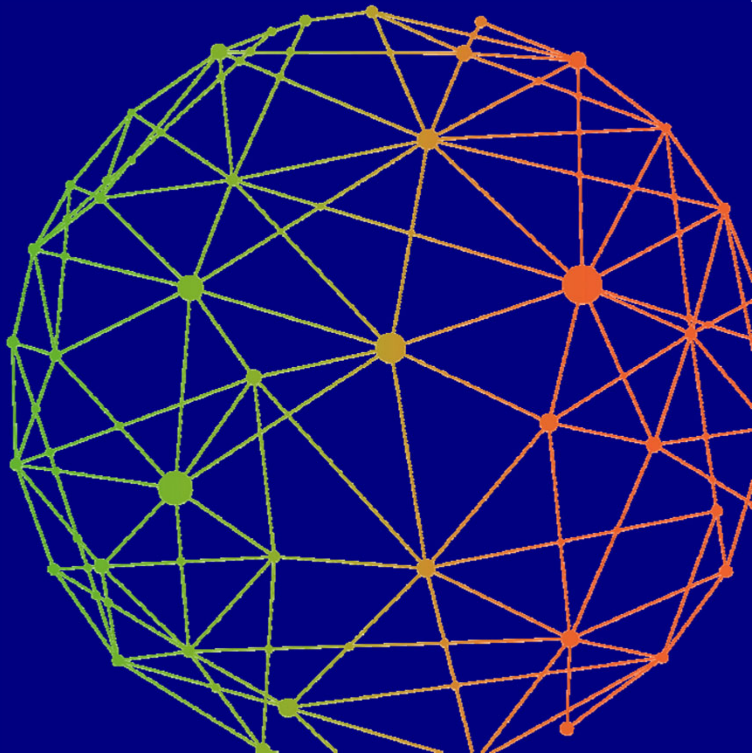
  EU + associated (+ business)

- High-quality

  standard and native benchmarks

- Compliant with EU regulations

OPEN
EURO
LLM



Partners

Charles University

AMD
SILO AI

alt-edic
eL
Da

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

ellis
INSTITUTE
TÜBINGEN

Fraunhofer
IAIS

JÜLICH
Forschungszentrum

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF HELSINKI

UNIVERSITAS OSLOENSIS MDCCCXI

UNIVERSITY OF TURKU

ALEPH ALPHA

AI
SWEDEN

ellamind

LightOn

prompsit

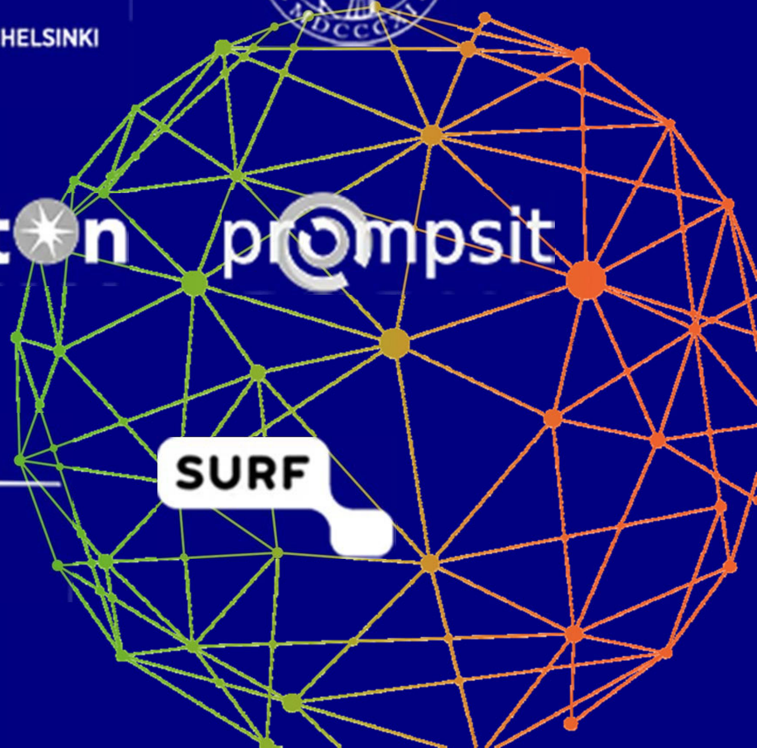
OPEN
EURO
LLM

BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

CINECA

CSC

SURF



Partners

Charles University

AMD
SILO AI

alt-edic
eL
Da

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

ellis
INSTITUTE
TÜBINGEN

Fraunhofer
IAIS

JÜLICH
Forschungszentrum

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF HELSINKI

UNIVERSITAS OSLOENSIS MDCCCXI

UNIVERSITY OF TURKU

ALEPH ALPHA

AI SWEDEN

ellamind

LightOn

prompsit

OPEN
EURO
LLM

BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

CINECA

CSC

SURF

Partners

Charles University

AMD SILO AI

alt-edic elDa

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

ellis INSTITUTE TÜBINGEN

Fraunhofer IAIS

JÜLICH Forschungszentrum

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF HELSINKI

UNIVERSITAS OSLOENSIS MDCCCXI

UNIVERSITY OF TURKU

ALEPH ALPHA

AI SWEDEN

ellamind

LightOn

prompsit

OPEN EURO LLM

BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación

CINECA

CSC

SURF

Partners

Charles University

AMD
SILO AI

alt-edic
eL
Da

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

ellis
INSTITUTE
TÜBINGEN

Fraunhofer
IAIS

JÜLICH
Forschungszentrum

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF HELSINKI

UNIVERSITAS OSLOENSIS MDCCCXI

UNIVERSITY OF TURKU

ALEPH ALPHA

AI SWEDEN

ellamind

LightOn

prompsit

OPEN
EURO
LLM

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

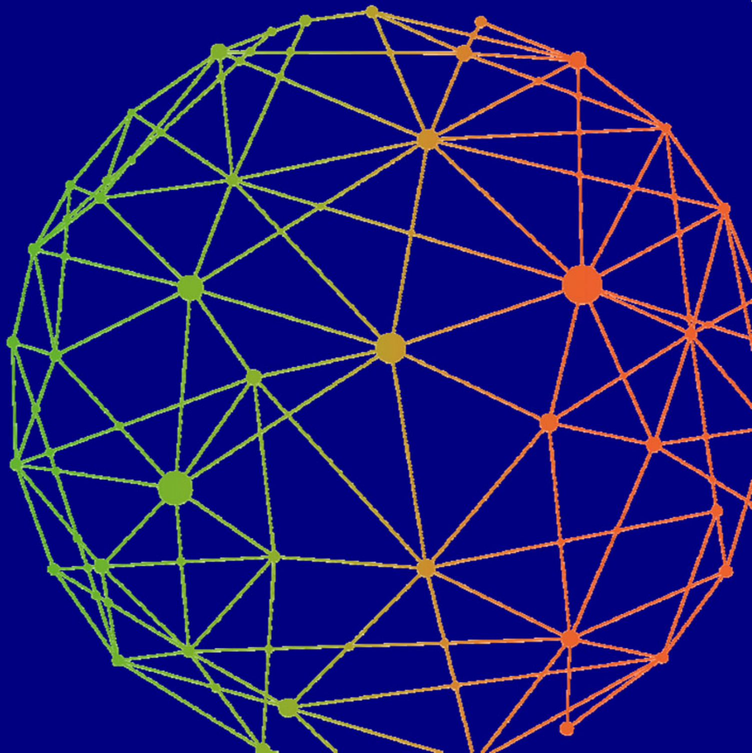
CINECA

CSC

SURF

Partners

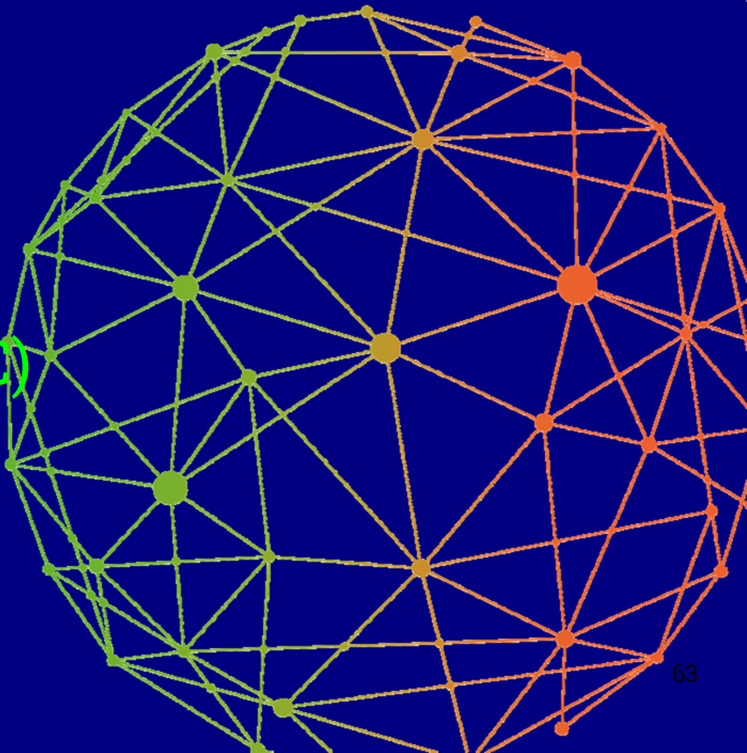
- Programme: Digital Europe (25/50% co-funding)
- Set of AI-06 calls (projects started Jan-Mar 2025):
  - Two large projects: OpenEuroLLM and LLMs4EU
  - Coordination (ALT-EDIC4EU), total ~80 mil. EUR + HPC
  - Part of an ecosystem (Deploy AI, TAILOR, TrustLLM, HPLT, …)
- Together we will
  - Develop open, high quality foundation models
  - Adapt them to applications in all areas, from commerce to egovernment and education
  - Contribute to EU's digital sovereignty



OPEN
EURO
LLM

Open Source / LLM Community

- Open Strategic Partnership Board
  - Open source community members
  - Experts on LLMs (incl. from non-EU ones)
    - Former commercial and/or open source model developers
  - Strategic advisory role
- Experts on legal issues
- Informal cooperations
  - Data side: CommonCrawl, Internet Archive (TBC)
  - Open source models community
    - LAION, open/sci, …

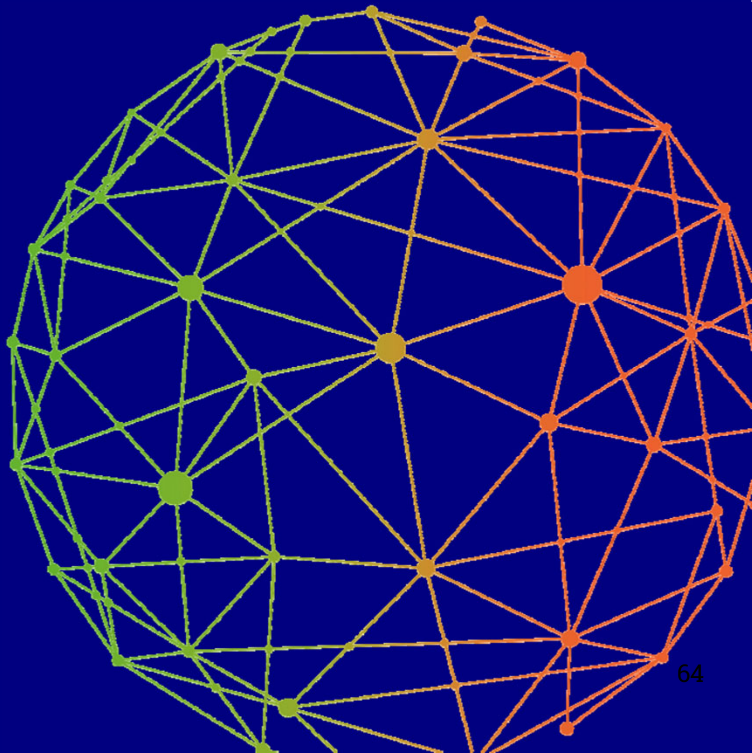


OPEN
EURO
LLM

Computing facilities

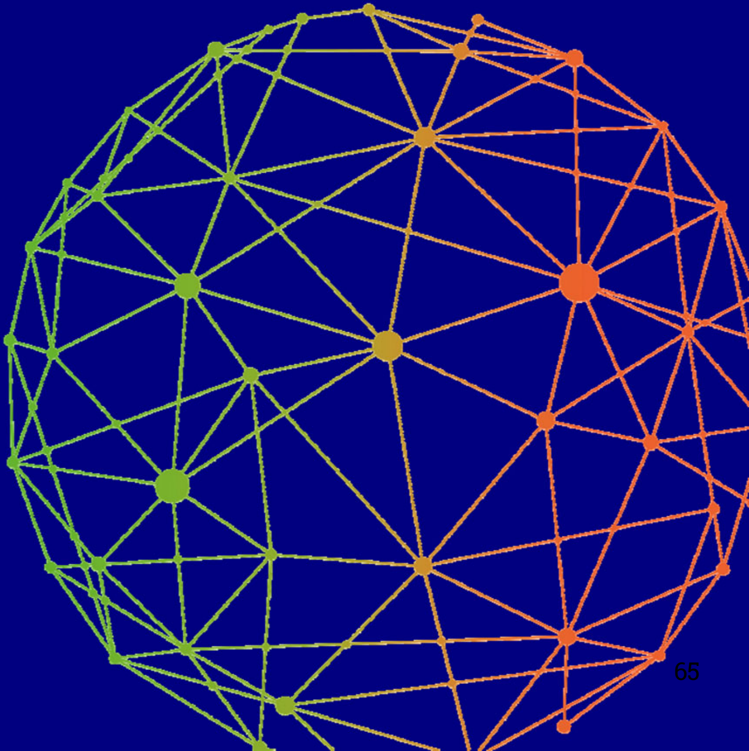
- 5 EuroHPC centers on board (project partners)
  - Technical expertise
    - Jumps start using the respective facilities
- Some compute available from previous projects
- Participation in EuroHPC calls in 2025
  - In line with project plan for the rest of 2025
- Strategic allocations in the future
  - “STEP” seal awarded
  - Using current facilities & new in AI Factories (2026/2027)
  - Estimated capacity needed: 300 mil. GPUh

OPEN
EURO
LLM



Data for 32+ languages

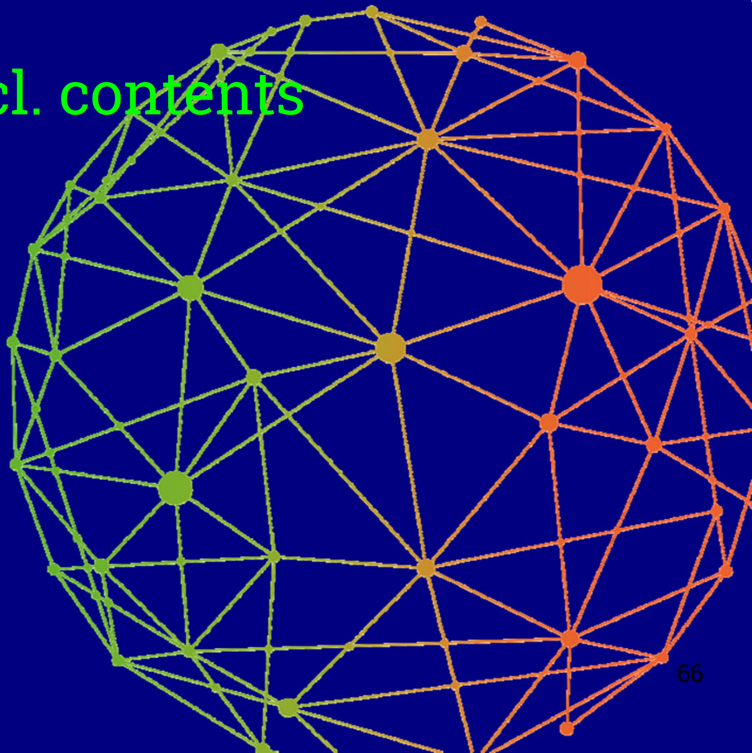
- Using available and Open Source data
    - HPLT 2.0 (HPLT 3.0, July 25), Fineweb2, Cultura-X, …
    - Mixtures to be experimentally determined
        - Ultimate (re)sources: CommonCrawl, Internet Archive
        - OpenWebSearch(?)
- Focus on low-resource languages for additional data
    - Incl. specific cases for very similar languages
- Additional data for
    - Fine-tuning, instruction-tuning, reasoning
        - … if necessary for benchmarking



OPEN
EURO
LLM

Evaluation and Benchmarks

- For initial experiments:
    - Standard benchmarks for base models
- Project longer-term goal
    - Benchmarks for all languages in native form
        - i.e., manually translated or inspected, incl. contents
- Continuous evaluation
- Tests for evaluation data purity
    - I.e., not used in training/SFT/…
- Models released based on evaluation results

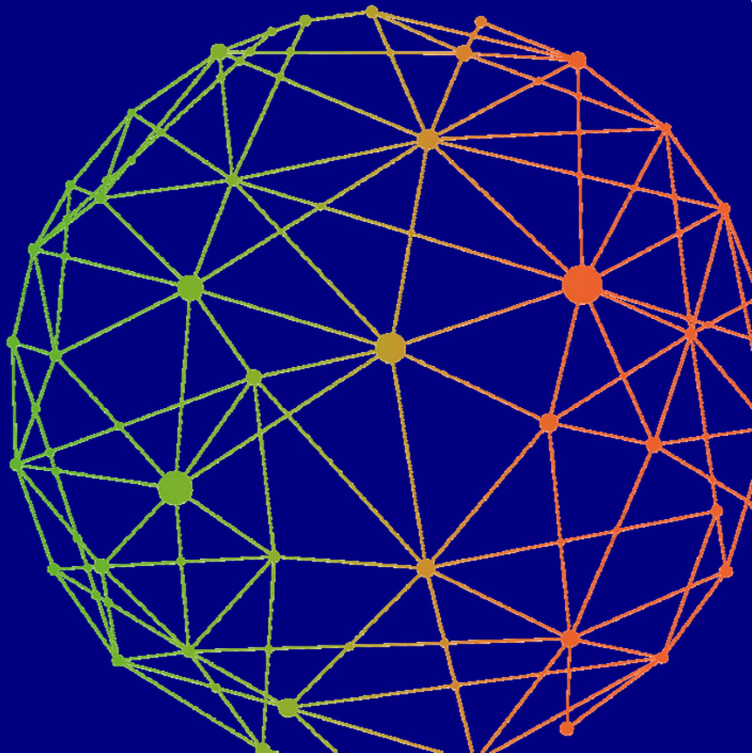


OPEN
EURO
LLM

Thank you!

Questions?

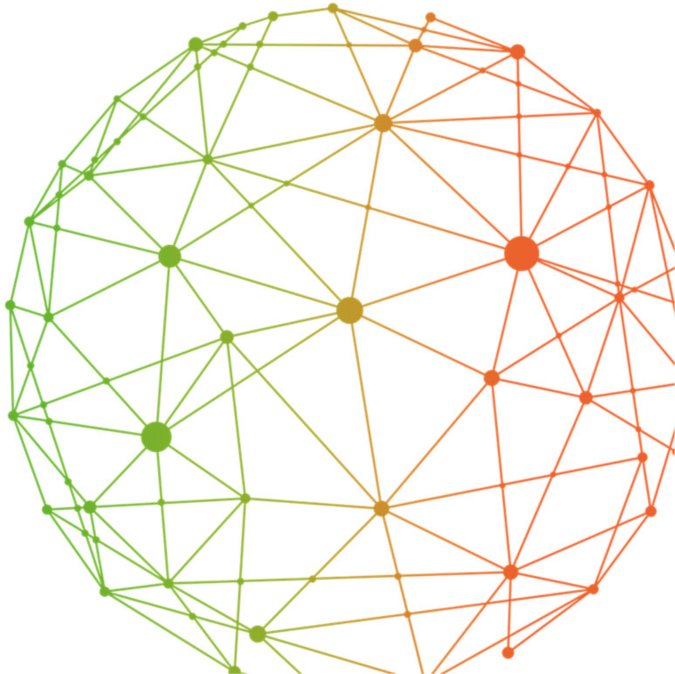
hajic@ufal.mff.cuni.cz



Data Spaces Symposium

Using Generative AI agents in the real world

Roberto Gonzalez



Using Generative AI agents in the real world

Keynote Session

Data Spaces Symposium 2025

Roberto González

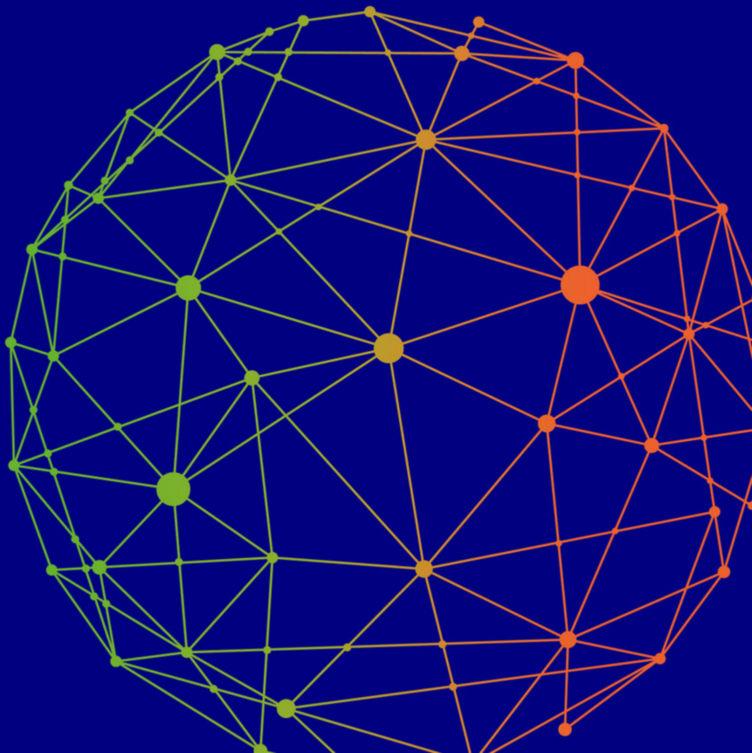
NEC \Orchestrating a brighter world

DSBA ■ BDV BIG DATA VALUE ASSOCIATION FIWARE FOUNDATION gaia-x INTERNATIONAL DATA SPACES ASSOCIATION

DATA SPACES SUPPORT CENTRE

Funded by the European Union

The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement n° 101083412



NEC Laboratories Europe

\Orchestrating a brighter world NEC

Advancing information & communications through **research excellence** and **open innovation**

KEY R&D METRICS

1,500+ Patents

50+ Peer-reviewed publications per year

150+ European projects

40+ University Cooperation’s

OPERATIONAL AREAS

RESEARCH
Leading scientific discovery in Europe

TECHNOLOGY TRANSFER
Commercializing R&D results in existing and new company business segments

STANDARDS
Defining European technology standards and best practices

RESEARCH AREAS

FOUNDATIONAL MODELS

- Human-centric AI
- Generative AI agents

DATA ANALYSIS

- Cyberthreat Intelligence
- Biomedical AI
- Smart districts

DATA DISTRIBUTION

- Decentralized trust

5G & 6G NETWORKS

- Generative AI for RAN
- Integrated sensing and communications
- Technology standards

COMPUTING

- Computational science

CROSS-AREA SYNERGIES

Roberto González – Cybersecurity Program Manager

Experience:

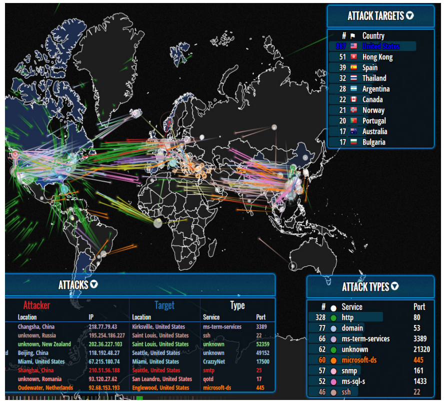
Research Areas: AI, Cybersecurity, Privacy, Big data, networks…
Patents: Holder of more tan 10 patents in Cybersecurity and AI
Publications: Over 50 publications in top-tier journals and conferences
EU Projects: Participated in over 10 EU projects from FP7, including roles as technical coordinator and Work Package Leader

What is Cyber Threat Intelligence (CTI)?

Cyber Threat Intelligence (CTI) refers to the information that organizations use to understand the cyber threats they are currently facing or might face in the future. It's the organized effort to gather, analyze, and disseminate information about these threats, offering a deeper insight into potential attacks, the tactics, techniques, and procedures (TTPs) of adversaries, their motivations, and the vulnerabilities they may exploit.

Problem: collecting and retrieving CTI is difficult

CYBER THREAT ALLIANCE



Shared information about cyber threats
(NEC is a member!)

11`000`000
new reports per month

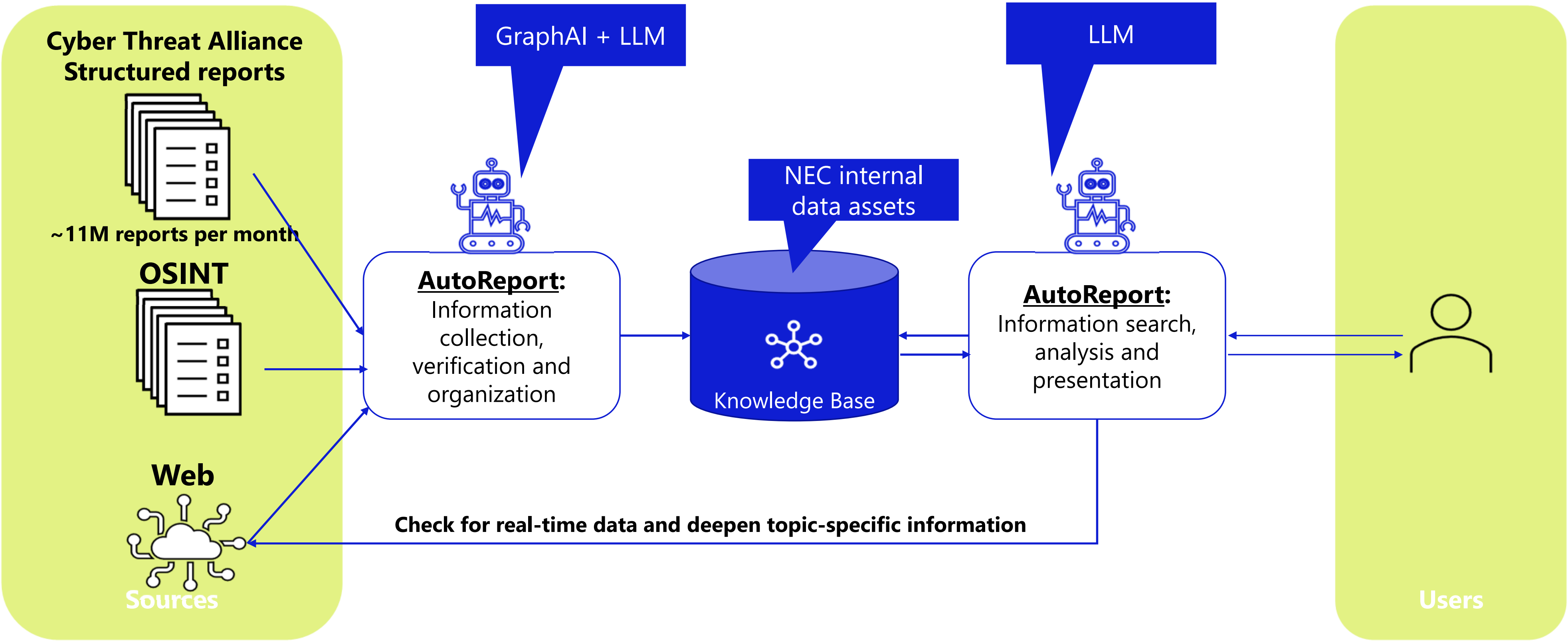
Per company!!

20`000 full-time
Security professionals would be needed to analyze all the reports

Check reports and relate them to the own company (several hours per report)

AutoReport

- An automated Cybersecurity agent that can browse the web, read databases, relate information and generate natural text reports for humans and structured reports for computer systems.
- Started development in 2022; in production since Jan. 2023


Cyber Threat Alliance
Structured reports
~11M reports per month
OSINT
Web
Sources
GraphAI + LLM
AutoReport: Information collection, verification and organization
NEC internal data assets
Knowledge Base
LLM
AutoReport: Information search, analysis and presentation
Users
Check for real-time data and deepen topic-specific information

Example of automated analysis output

SentinelLABS ABOUT CVE DATABASE CONTACT VISIT SENTINELONE.COM

ADVANCED PERSISTENT THREAT

Kimsuky Strikes Again | New Social Engineering Campaign Aims to Steal Credentials and Gather Strategic Intelligence

ALEKSANDAR MILENKOSKI / JUNE 6, 2023

Further, Kimsuky's objective extends to the theft of subscription credentials from NK News. To achieve this, the group distributes emails that lure targeted individuals to log in on the malicious website nknews[.]pro, which masquerades as the authentic NK News site. The login form that is presented to the target is designed to capture entered credentials.

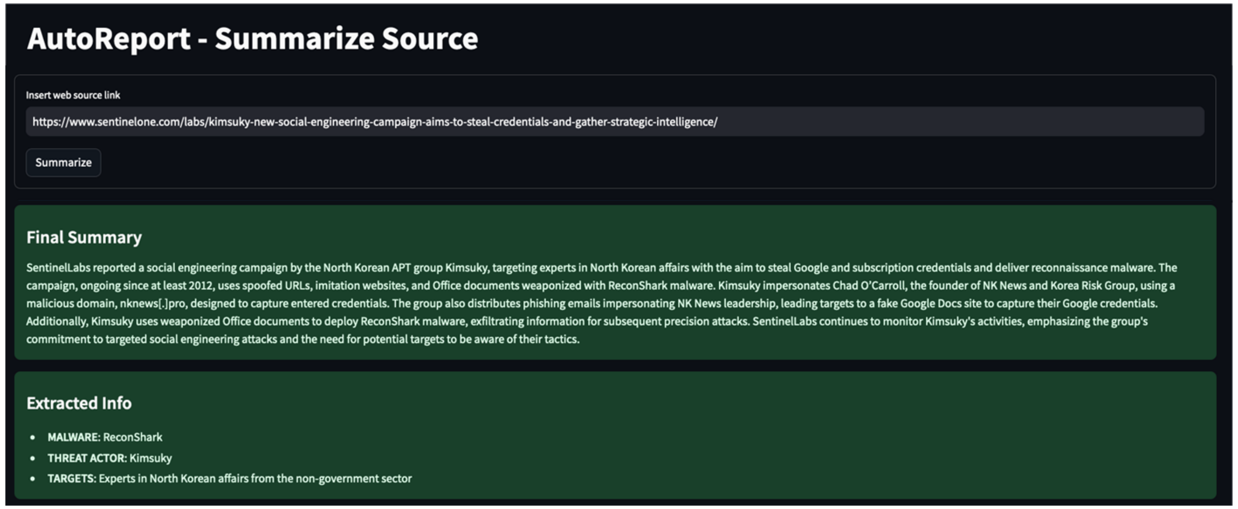
Conclusion

SentinelLabs remains actively engaged in monitoring the activities conducted by Kimsuky. The findings presented in this post highlight the group's persistent commitment to targeted social engineering attacks and underscore the need for increased awareness and understanding of Kimsuky's tactics among potential targets. Maintaining vigilance and implementing effective security measures are imperative to mitigate the risks posed by this persistent threat actor.

Indicators of Compromise

Indicator	Description
nknews[.]pro	Phishing email sender domain
chad.ocarroll@nknews[.]pro	Phishing email sender address
membership@nknews[.]pro	Phishing email sender address
https[://]www.nknews[.]pro	Website impersonating NK News
https[://]www.nknews[.]pro/config[.]php	Website impersonating NK News: b374k login site
https[://]www.nknews[.]pro/ip/register/	Website impersonating NK News: Fake NK News login site
https[://]www.nknews[.]pro/ip/register/login[.]php	Website impersonating NK News: NK News credential theft endpoint
https[://]staradvertiser.store/piece/ca[.]php	ReconShark payload hosting endpoint
https[://]staradvertiser.store/piece/r[.]php	ReconShark C2 server endpoint
162.0.209[.]27	Website impersonating NK News, ReconShark C2 server: IP address
4150B40C00D8AB2E960AA059159149AF3F9ADA09	Malicious document (password-protected): SHA1 hash
7514FD9E5667FC5085373704FE2EA959258C7595	Malicious document: SHA1 hash
41E39162AE3A6370B1100BE2B35BB09E2CBE9782	ReconShark: SHA1 hash

ADVERSARY



AutoReport - Summarize Source

Insert web source link

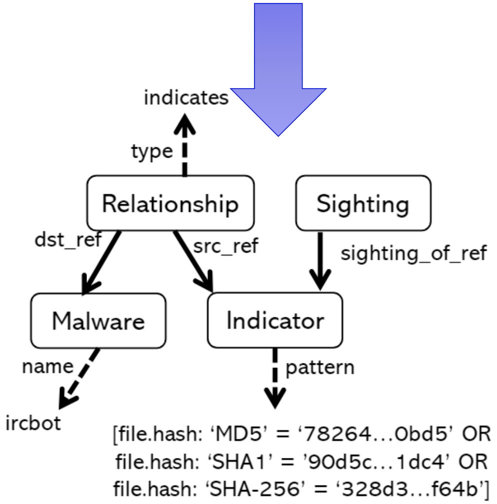
https://www.sentinelone.com/labs/kimsuky-new-social-engineering-campaign-aims-to-steal-credentials-and-gather-strategic-intelligence/

Final Summary

SentinelLabs reported a social engineering campaign by the North Korean APT group Kimsuky, targeting experts in North Korean affairs with the aim to steal Google and subscription credentials and deliver reconnaissance malware. The campaign, ongoing since at least 2012, uses spoofed URLs, imitation websites, and Office documents weaponized with ReconShark malware. Kimsuky impersonates Chad O'Carroll, the founder of NK News and Korea Risk Group, using a malicious domain, nknews[.]pro, designed to capture entered credentials. The group also distributes phishing emails impersonating NK News leadership, leading targets to a fake Google Docs site to capture their Google credentials. Additionally, Kimsuky uses weaponized Office documents to deploy ReconShark malware, exfiltrating information for subsequent precision attacks. SentinelLabs continues to monitor Kimsuky's activities, emphasizing the group's commitment to targeted social engineering attacks and the need for potential targets to be aware of their tactics.

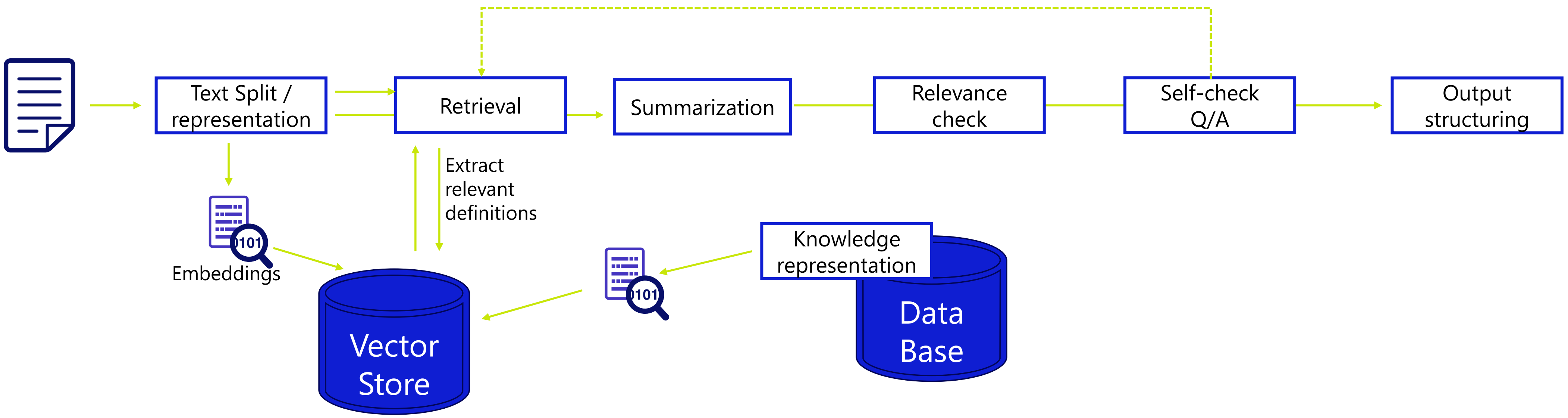
Extracted Info

- **MALWARE:** ReconShark
- **THREAT ACTOR:** Kimsuky
- **TARGETS:** Experts in North Korean affairs from the non-government sector



[file.hash: 'MD5' = '78264...0bd5' OR
file.hash: 'SHA1' = '90d5c...1dc4' OR
file.hash: 'SHA-256' = '328d3...f64b']

How does it work?



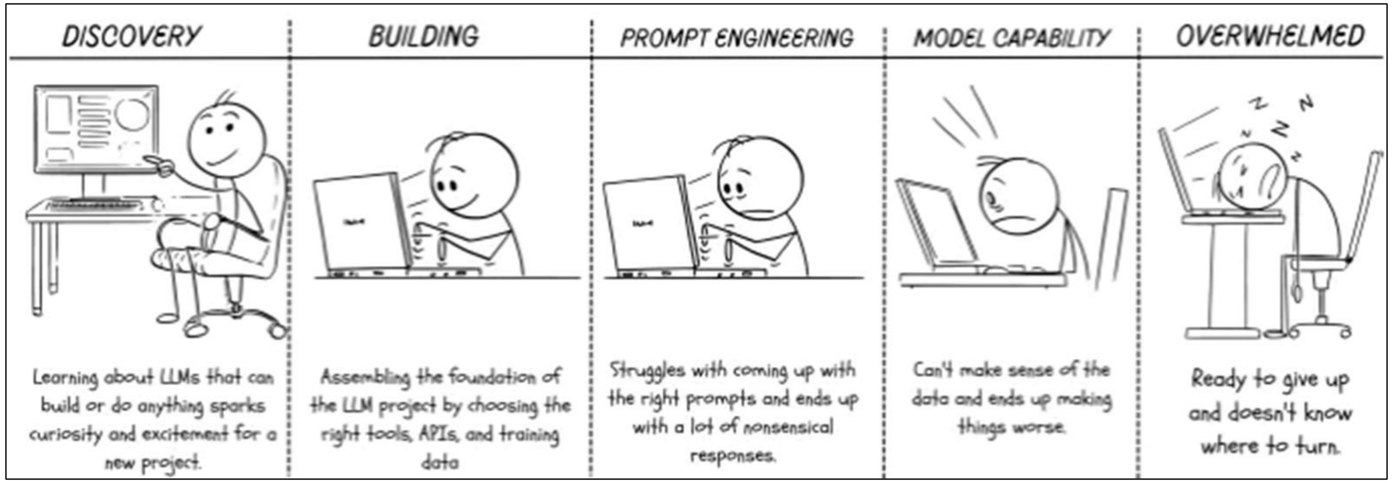
Split a complex task into smaller tasks

- Small task -> Easier task -> constant behaviour
- Integration with external tools and information sources

Pipeline ordering

- LLM chain contain the hardcoded instruction on “how to do” a task

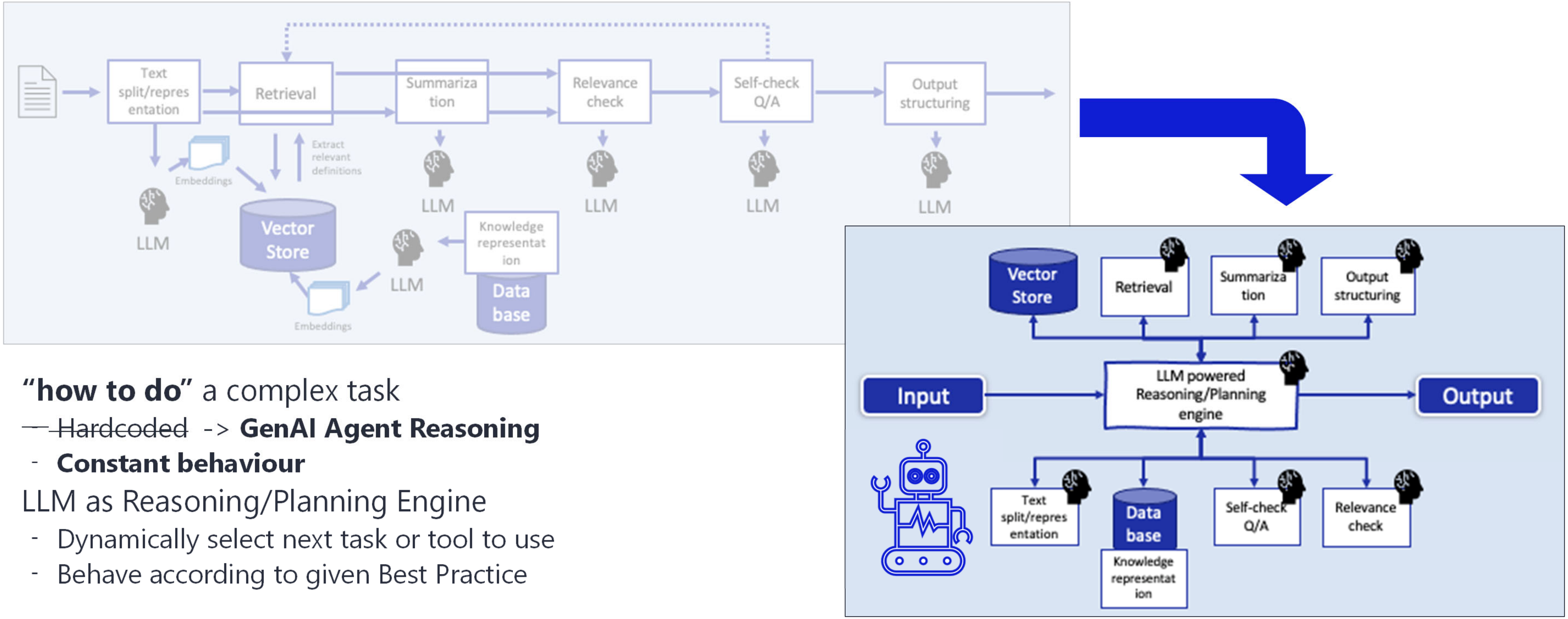
Why can´t we simply use LLMs?


DISCOVERY
Learning about LLMs that can build or do anything sparks curiosity and excitement for a new project.
BUILDING
Assembling the foundation of the LLM project by choosing the right tools, APIs, and training data
PROMPT ENGINEERING
Struggles with coming up with the right prompts and ends up with a lot of nonsensical responses.
MODEL CAPABILITY
Can't make sense of the data and ends up making things worse.
OVERWHELMED
Ready to give up and doesn't know where to turn.

Challenge

- Easy to get a lucky prompt
  - Extremely complex to generalize it
- Easy to build a first prototype
  - Extremely complex to have constant behaviour
- Engineering effort for each new use case
  - Even when modifying an existing one

Generative AI Agent – From Hardcoded instruction to Reasoning



"how to do" a complex task

~~– Hardcoded~~ -> **GenAI Agent Reasoning**

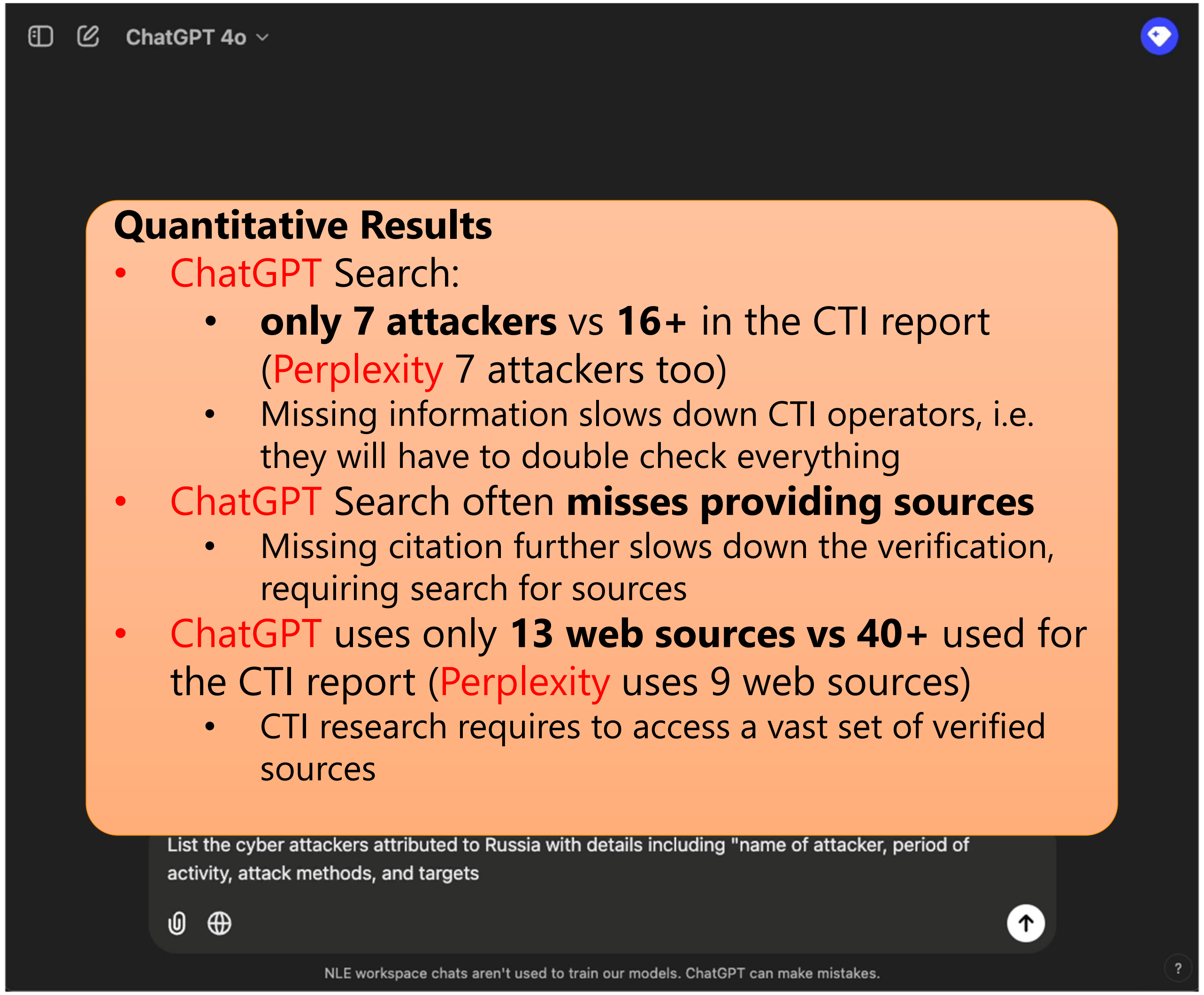
- **Constant behaviour**

LLM as Reasoning/Planning Engine

- Dynamically select next task or tool to use
- Behave according to given Best Practice

Why AutoReport and not ChatGPT search?

Example with ChatGPT Search



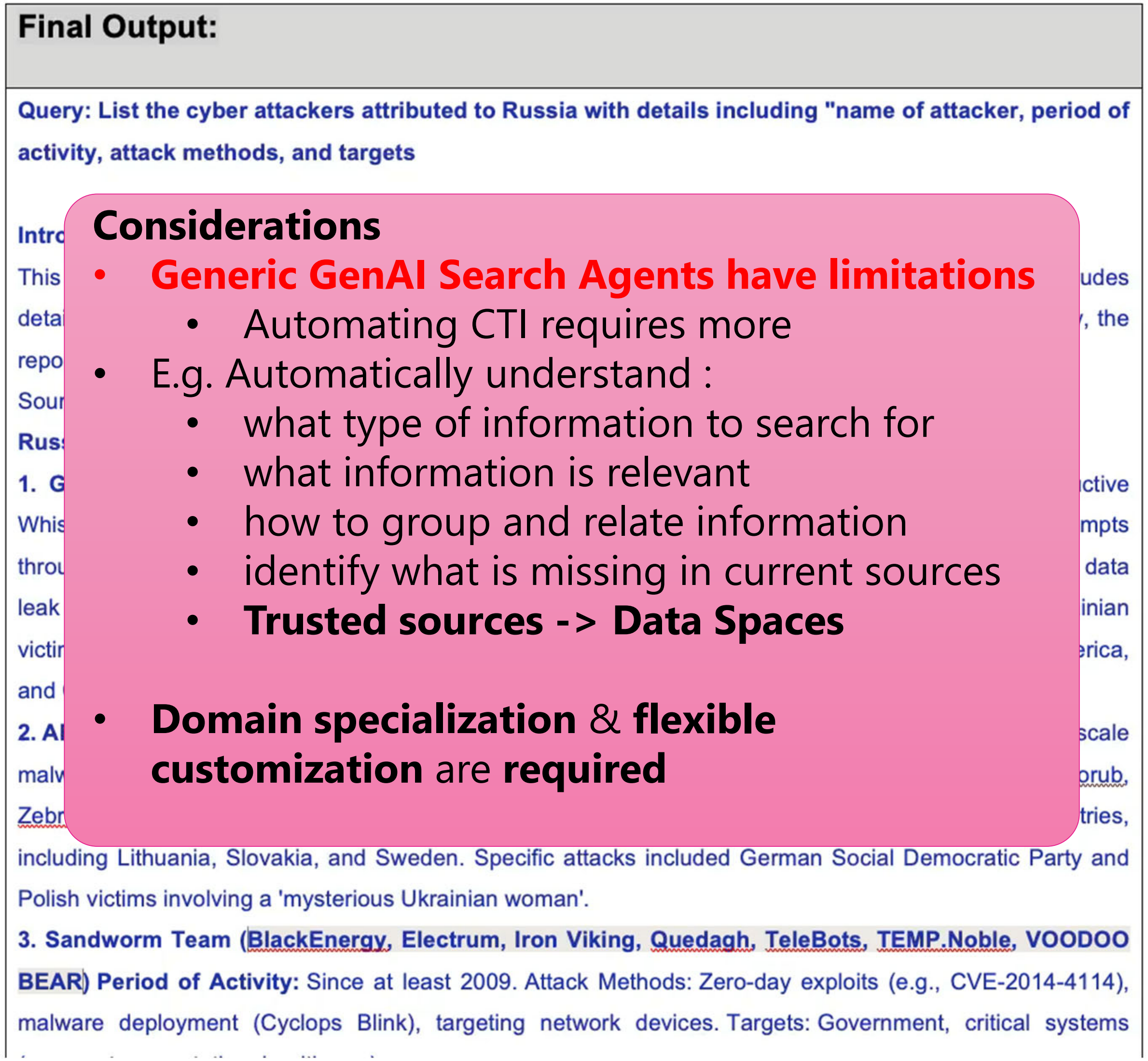
Quantitative Results

- ChatGPT Search:
  - **only 7 attackers** vs **16+** in the CTI report (Perplexity 7 attackers too)
  - Missing information slows down CTI operators, i.e. they will have to double check everything
- ChatGPT Search often **misses providing sources**
  - Missing citation further slows down the verification, requiring search for sources
- ChatGPT uses only **13 web sources vs 40+** used for the CTI report (Perplexity uses 9 web sources)
  - CTI research requires to access a vast set of verified sources

List the cyber attackers attributed to Russia with details including "name of attacker, period of activity, attack methods, and targets

NLE workspace chats aren't used to train our models. ChatGPT can make mistakes.

Actual CTI Report



Final Output:

Query: List the cyber attackers attributed to Russia with details including "name of attacker, period of activity, attack methods, and targets

Intro

This ... udes

detai ... , the

repo

Sour

Russ

1. G ... ctive

Whis ... mpts

throu ... data

leak ... inian

victi ... erica,

and

2. Al ... scale

malw ... orub,

Zebr ... tries,

including Lithuania, Slovakia, and Sweden. Specific attacks included German Social Democratic Party and Polish victims involving a 'mysterious Ukrainian woman'.

3. Sandworm Team (BlackEnergy, Electrum, Iron Viking, Quedagh, TeleBots, TEMP.Noble, VOODOO BEAR) Period of Activity: Since at least 2009. Attack Methods: Zero-day exploits (e.g., CVE-2014-4114), malware deployment (Cyclops Blink), targeting network devices. Targets: Government, critical systems

Considerations

- **Generic GenAI Search Agents have limitations**
  - Automating CTI requires more
- E.g. Automatically understand :
  - what type of information to search for
  - what information is relevant
  - how to group and relate information
  - identify what is missing in current sources
  - **Trusted sources -> Data Spaces**
- **Domain specialization** & **flexible customization** are **required**

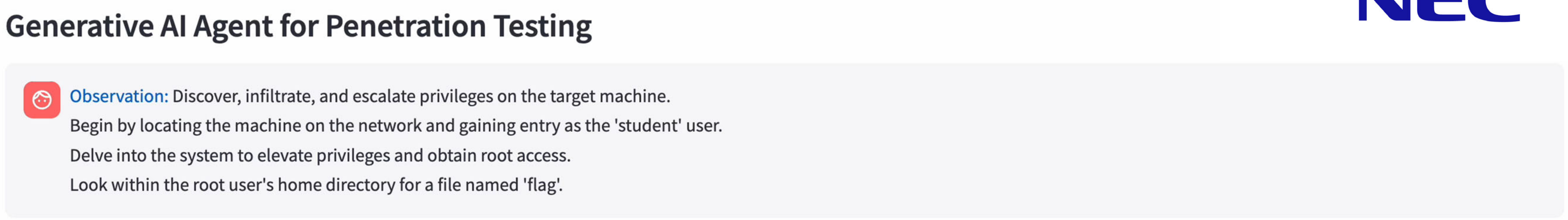
Generative AI agents

Let´s see an example of an agent doing Penetration Testing

Discover, infiltrate, and escalate privileges on the target machine.
Begin by locating the machine on the network and gaining entry as the 'student' user.
Delve into the system to elevate privileges and obtain root access.
Look within the root user's home directory for a file named 'flag'.

Tools:

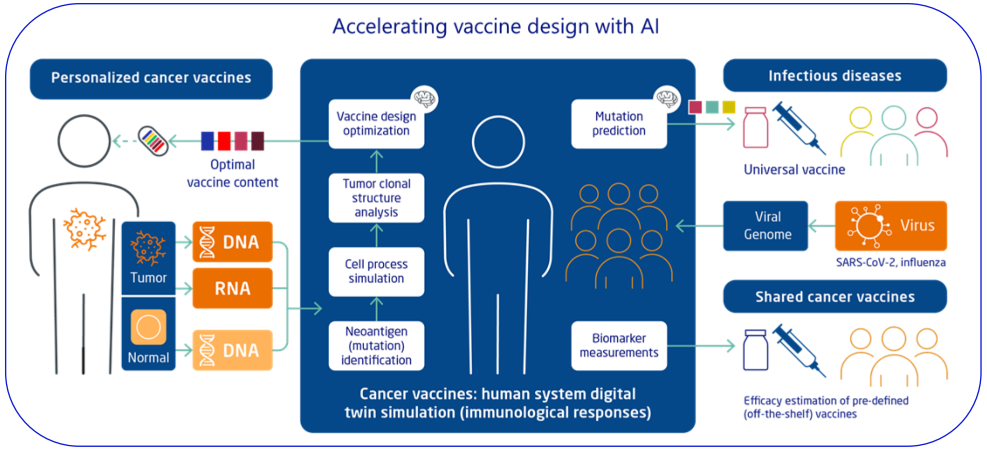
- Access to Data Spaces



Generative AI Agent for Penetration Testing

Observation: Discover, infiltrate, and escalate privileges on the target machine.
Begin by locating the machine on the network and gaining entry as the 'student' user.
Delve into the system to elevate privileges and obtain root access.
Look within the root user's home directory for a file named 'flag'.

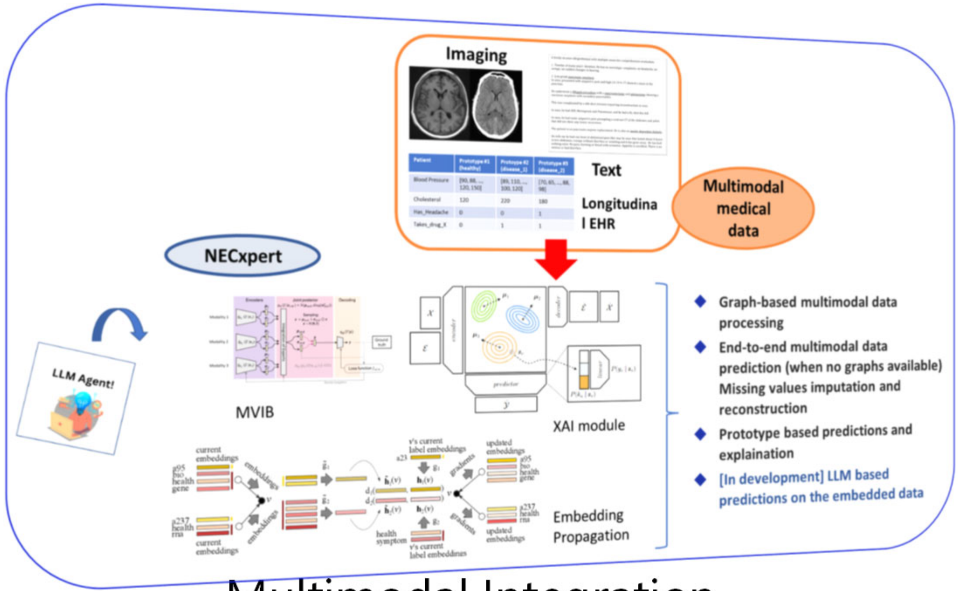
Different verticals


Accelerating vaccine design with AI

Cancer vaccines: human system digital twin simulation (immunological responses)

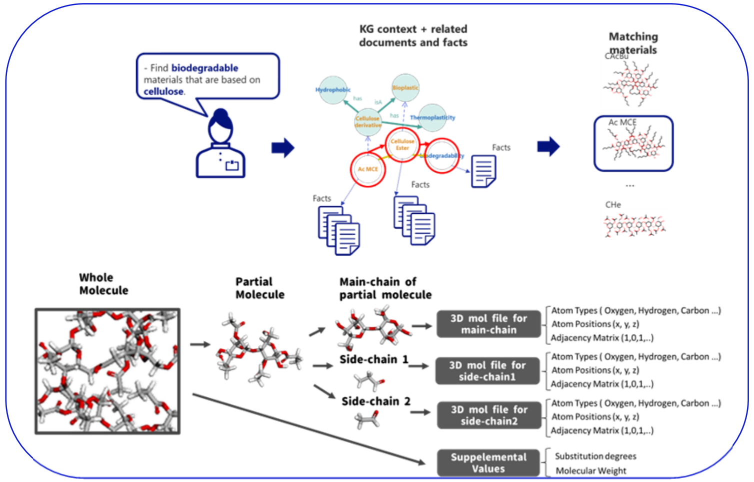
Efficacy estimation of pre-defined (off-the-shelf) vaccines

Drug development



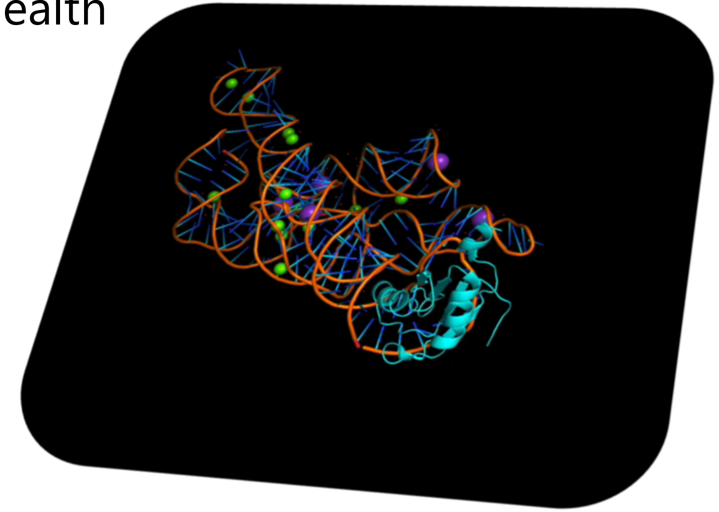
- Graph-based multimodal data processing
- End-to-end multimodal data prediction (when no graphs available) Missing values imputation and reconstruction
- Prototype based predictions and explaination
- [In development] LLM based predictions on the embedded data

Multimodal Integration
For Digital Health



3D mol file for main-chain	Atom Types (Oxygen, Hydrogen, Carbon ...) Atom Positions (x, y, z) Adjacency Matrix (1,0,1,..)
3D mol file for side-chain1	Atom Types (Oxygen, Hydrogen, Carbon ...) Atom Positions (x, y, z) Adjacency Matrix (1,0,1,..)
3D mol file for side-chain2	Atom Types (Oxygen, Hydrogen, Carbon ...) Atom Positions (x, y, z) Adjacency Matrix (1,0,1,..)
Suppelemental Values	Substitution degrees Molecular Weight

Material Informatics



Molecular dynamics
AI based simulations

The Reliable LLM Framework by NEC

The Enabler for High-Risk Use Cases

\Orchestrating a brighter world NEC

Current situation

LLMs hallucinate

- **Performance gap:** need for accurate, dependable AI-generated content.
- **Dangerous:** hallucinations hinder adoption in high-risk domains

Problem

Manual verification

- **Time consuming and inefficient:** humans need to read and check everything.
- **Difficult to find:** hallucinations sound plausible but are incorrect

Solution

Reliable LLMs

Benefits:

- Allow user control and quality assurance
- Reduces error, enhances reliability.
- Increases efficiency, reduces need for manual checks.
- Ensures regulatory compliance, increases trust in AI.

NEC

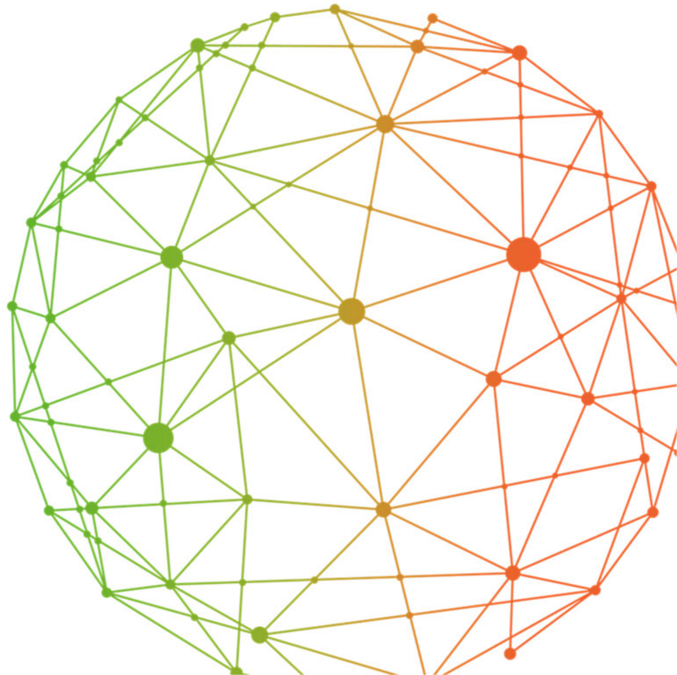
\Orchestrating a brighter world

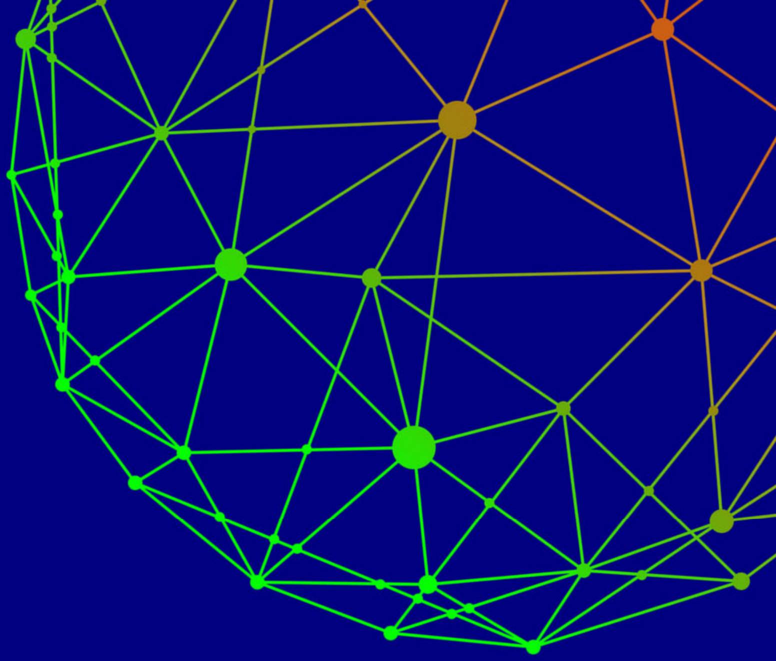
Roberto.Gonzalez@neclab.eu

Data Spaces Symposium

Panel discussion | That's why the economy needs data spaces to evolve

Christoph Mertens, Douglas Ramsey, David Schönwerth, Bettina Tratz-Ryan, Takahide Matsutsuka, Peter Kraemer





Data Spaces Symposium

Share data. Unlock value. Boost impact.

Panel discussion

That’s why the economy needs data spaces to evolve

Christoph Mertens
IDSA

Douglas Ramsey
Axial Global Advisors

Peter Kraemer
Capgemini

Takahide Matsutsuka
Fujitsu

Bettina Tratz-Ryan
Gartner Group

David Schönwerth
Bitkom

Data Spaces Symposium

Enjoy your lunch!

These are the sessions you can choose from at 13:30:

Track 1:

Domain session on energy & green deal data spaces

Accelerating energy transition and realizing green deal – with data spaces as accelerators

Track 2:

Domain session on smart industry data spaces

How data spaces fuel smart industrial solutions

Track 3:

Data space tech session

Designing and delivering the European single market for data

Breakout track:

Business workshop

Assessing the sustainability of business models in data spaces and the role of public policies

Patronat polskiej prezydencji w Radzie UE
Patronage of the Polish presidency of the Council of the EU
Patronage de la présidence polonaise du Conseil de l'UE