European Commission

26 June 2024

AI in support of Interoperability

interoperable europe

innovation ∞ govtech ∞ community

Welcome!

Join us for an engaging and informative workshop session on 'AI for Interoperability'.

Workshop Objectives

- ***Get inspired*** *by the successes, challenges, and opportunities of our speakers in their* ***AI for Interoperability*** *journey and share yours!*

- ***Discover*** *the importance of* ***Interoperability for the development of high-quality AI solutions***

Workshop Practicalities

On-site Participation

- Please **silence your mobile devices**
- The session is divided in **two parts** (9:00-11:00 and 11:30-13:00)
- A **coffee break** is planned at 11:00
- Wait for the **panel discussions** (at 10:15 and 11:15) for asking your questions
- A **microphone** will be passed by the moderators to ensure everyone can hear you
- We will be using **Slido for interactive questions** and polls

Online Participation

- The session is being **live streamed**
- Use **Slido to ask questions and interact** with the speakers during the panel discussions

All sessions are recorded

Wifi connection

Login: SEMIC2024
Password : SEMIC2024!

2: AI in support of Interoperability

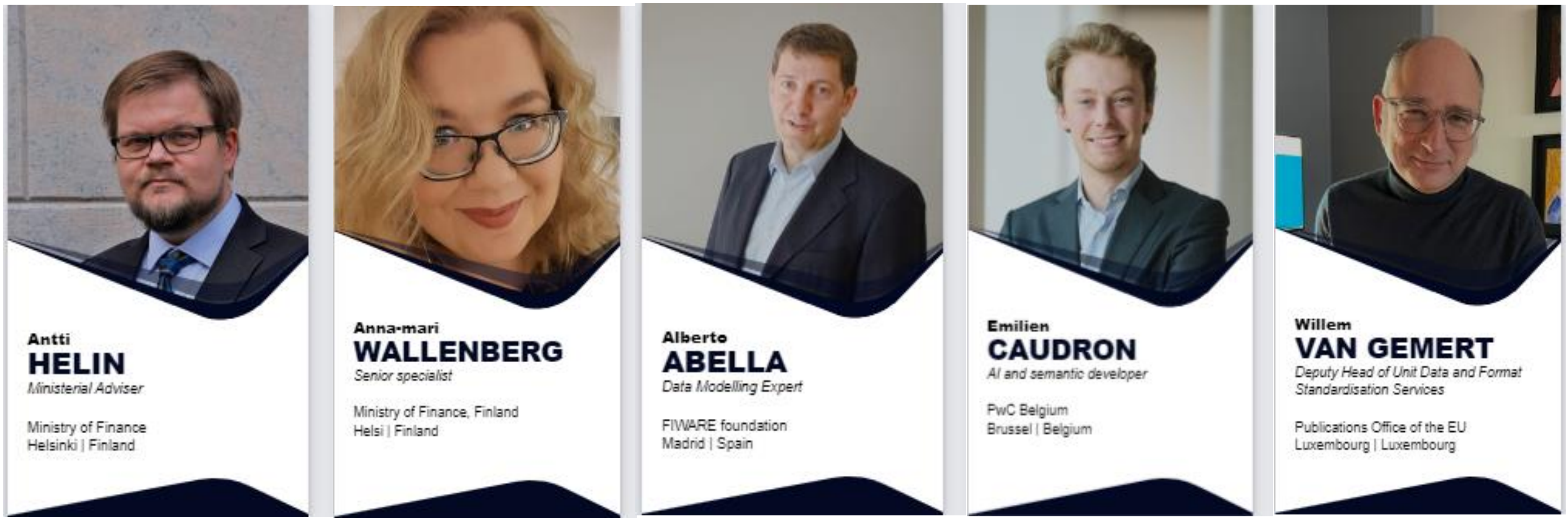
Join us on Slido!

- Use the QR code
- Or go on slido.com
  #SEMIC2024Workshops
- Select the correct workshop

Part 1: AI for Interoperability

SEMIC conference

Workshop’s speakers:



be EU
belgium24.eu

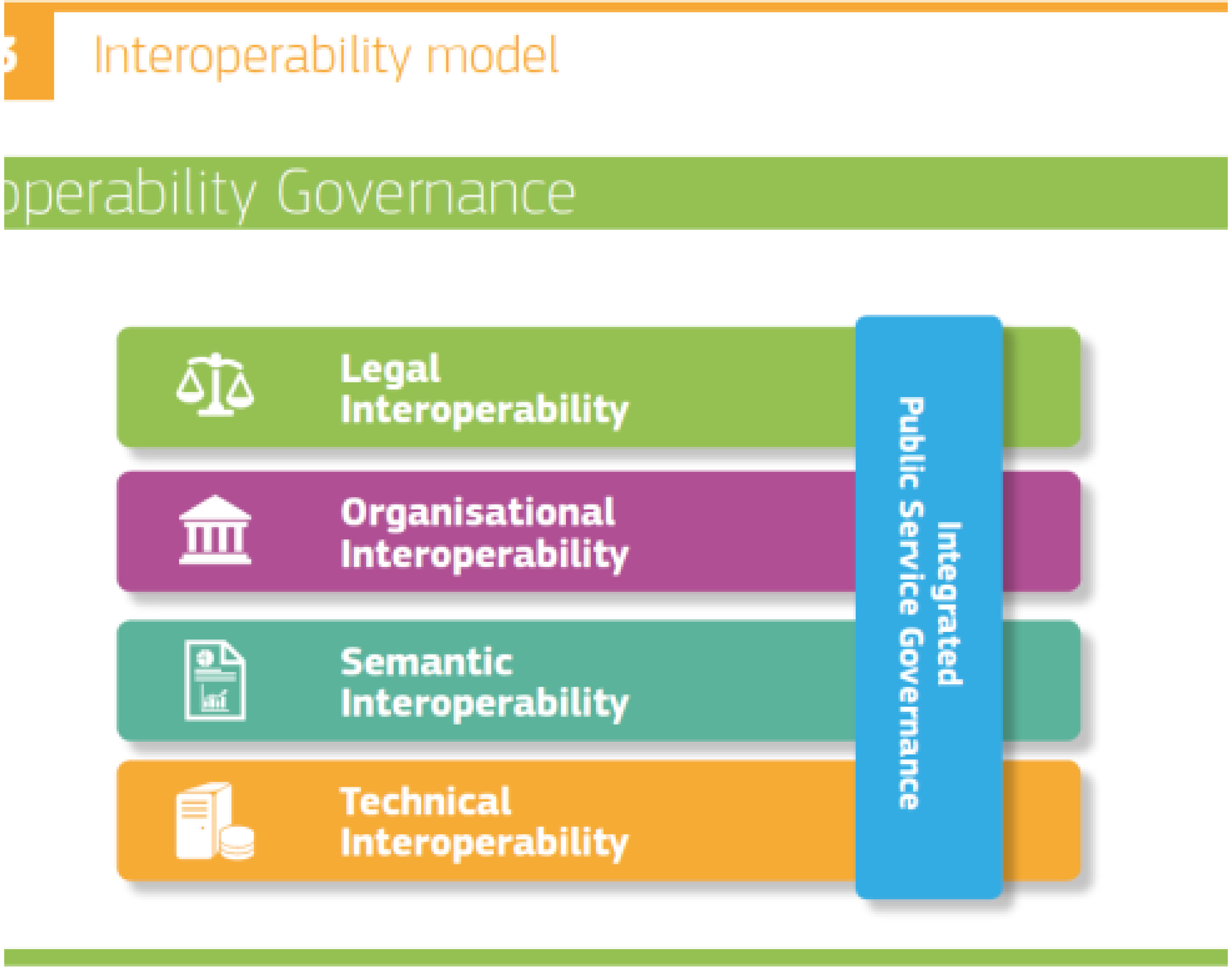
European Commission

AI and Layers of Interoperability:

Use Cases in Finnish Public Administration

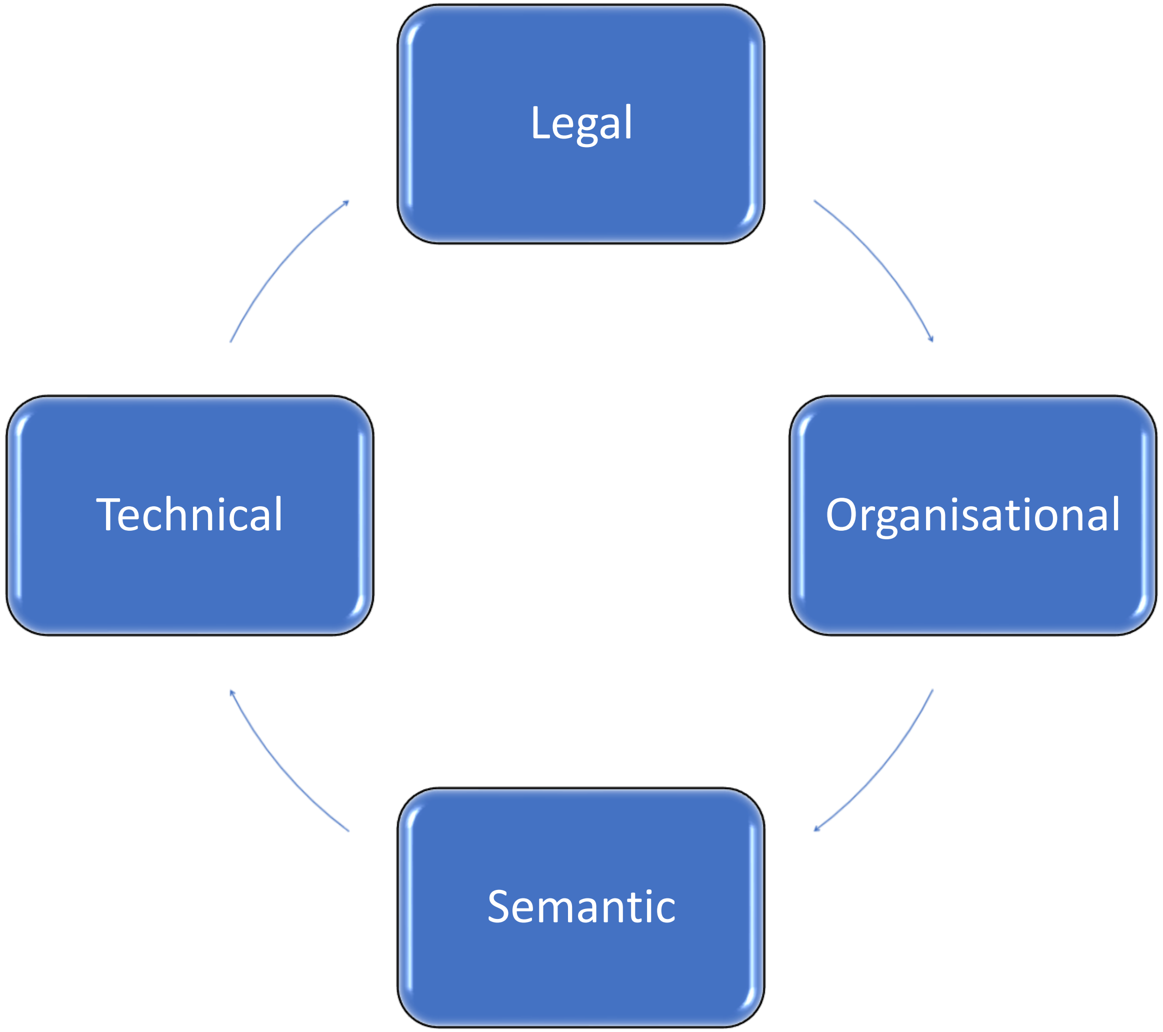
EIF: Layers of Interoperability

- IDEAL: “interoperability model” (IM), which is applicable to all digital public services, an integral element of the interoperability-by-design paradigm.
- TODAY: A concrete example of practical implementation of IM with LLMs in public sector service production
- Highlights challenges in a very legalist country



Interoperability in Design for Public Sector Services (Finland)

- **In public services, legal interoperability comes first**
- Requires sufficient expertise on the legal restrictions, examples:
  - Personal and sensitive information
  - The usage of data and extracted information
  - Security and safety
  - … and the quality of outputs (cf. the problem of hallucination)



Layers of Interoperability: Legal

Legal considerations when using AI

- Rapidly developing technology with rapidly developing applications (lots of skepticism but also hype)
- No information that is confidential or personal data can or should be used as inputs
- Outputs should (must) be reviewed before being used
- Must avoid becoming dependent on a particular system (vendor lock-in)
- Soon: requirements from the AI Act (transparency, HRAI requirements, fundamental rights impact assessments, etc)

Use Case 1: Summarizing the results of a public consultation

- Public consultations are mandatory when negotiating and ratifying binding bilateral agreements
- A consultation can draw hundreds of opinions which must be analyzed and collected before the process can move on
- Generative AI can synthetize and summarize this material, including source tracking to determine which element of the summary originated from which opinion

Use Cases 2 and 3: Establishing legal context and ’regulatory metadata’

- AI can be used to examine a legal draft to determine other relevant legislation (both national and union level)
  - This helps consider implications for legal IO with any new initiative
- All new bills must contain a header section that details the sections and paragraphs that are to be modified (this can be simply or get very complex)
  - Complex, prone to human error, mistakes can have severe consequences
- Government AI pilots are underway to automate these processes using the Semantic Finlex dataset as training data

Eduskunnan päätöksen mukaisesti

muutetaan viranomaisten toiminnan julkisuudesta annetun lain (621/1999)

1 §:n 2 momentti, 4 §:n 1 momentin 4 ja 8 kohta, 11 §:n 4 momentti, 14 §:n 1 momentti, 24 §:n 1 momentin 17 ja 18 kohta, 33 §:n 1 momentti sekä 34 §:n 3 ja 6 momentti, sellaisina kuin niistä ovat 11 §:n 4 momentti laissa 385/2007, 14 §:n 1 momentti sekä 34 §:n 6 momentti laissa 495/2005, 24 §:n 1 momentin 17 kohta laissa 604/2018 sekä 33 §:n 1 momentti ja 34 §:n 3 momentti laissa 853/2020,

seuraavasti:

According to a decision by Parliament

in the Act on Openness of Government Activities (621/1999)

section 1 subsection 2, section 4 subsection 1, paragraphs 4 and 8, section 11 subsection 4, section 14 subsection 1, section 24 subsection 1 paragraphs 17 and 18, section 33 subsection 1 and section 34 subsections 3 and 6, as section 11 subsection 4 is in law 385/2007, section 14 subsection 1 and section 34 subsection 6 are in law 495/2005, section 24 subsection 1 paragraph 17 is in law 604/2018 and section 33 subsection 1 and section 34 subsection 3 are in law 853/2020

shall be changed as follows:

Semantic Finlex

- Semantic Finlex contains current legislation as open data in a machine readable format
  - Approx. 2800 regulations
  - Approx. 60,000 court decisions
- Built on top of Finlex, where all new legal instruments are published in an electronic format, as a research project
- For future applications, Finlex itself will be available as open data in MR format and through an API ’natively’

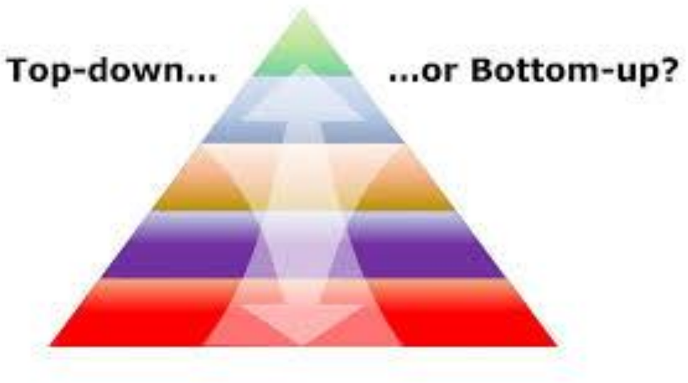
Use Case 4: Machine translation

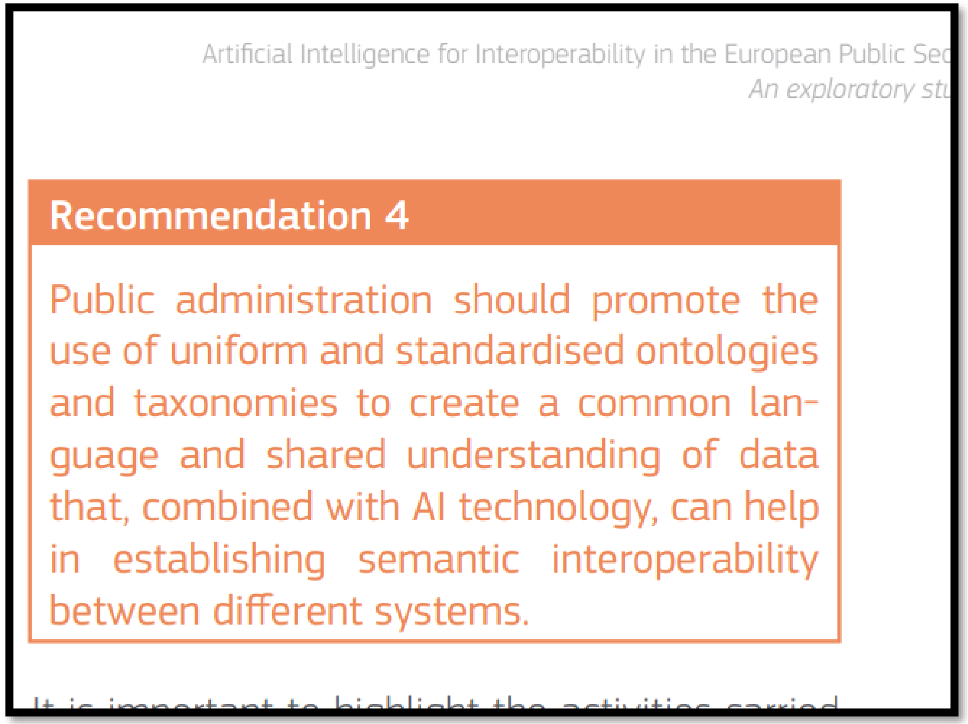
- Finland has two official languages; every major proposal, bill, memorandum or new law must be available in both
- Proposals for new law should include a section on how the issue at hand has been tackled in other countries (usually similar countries like Sweden)
- Government machine translator Aura has been in use since September 2021
  - Trained using specialized datasets and specific vocabularies
  - Available to all central government civil servants
  - Translations from and to Finnish, English and Swedish
  - Up to May 2023, over 68 million words translated (~320,000 A4 sheets)

Layers of Interoperability: Semantic

Layers of Interoperability: Semantic

- **TOP-DOWN:** The development of standardised ontologies (SO) and taxonomies to **create structured data**
- The structured data + AI helps to ”establish semantic interoperability between different systems”
- **A ”chicken-or-egg”- problem**
- The data is often messy, noisy and fragmented, so you can´t use SOs to establish interoperability



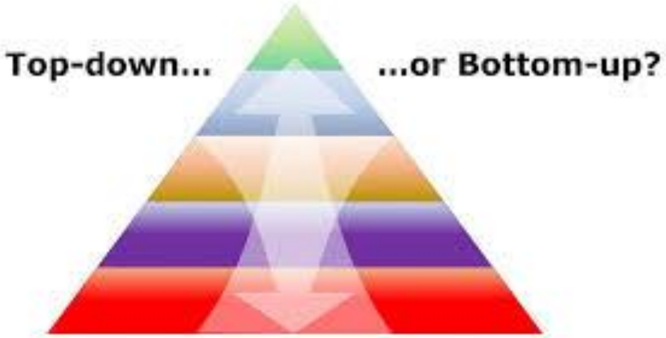

Artificial Intelligence for Interoperability in the European Public Sec
An exploratory stu

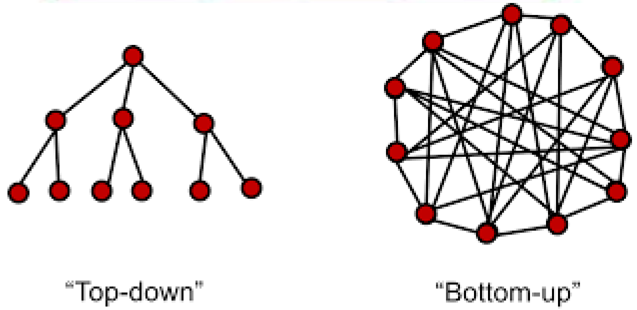
Recommendation 4

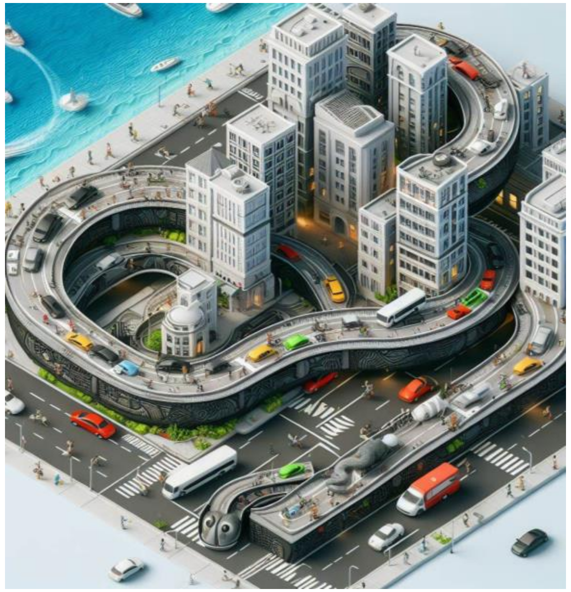
Public administration should promote the use of uniform and standardised ontologies and taxonomies to create a common language and shared understanding of data that, combined with AI technology, can help in establishing semantic interoperability between different systems.

Layers of Interoperability: Semantic

- **Bottom up & data driven approach**: semantic interoperability “extracted” from messy data
- **LLMs**: potential to “extract” quasi-semantic structure from data,
- Words (text) in statistical “semantic” space
  - a representation of a item's statistical “meaning”, while lacks functional meaning (Mahowald & al, 2023)
- **Complementary to the top-down approaches** (standard ontologies and taxonomies)
  - Interoperability by “prompting”







Source: Copilot, 20.2.2024

Layers of Interoperability: Technical

Layers of Interoperability: Technical (example: Pipelines)

- Problem for prompting (or any LLM)-based methods: LLMs: “stochastic parrots”, tendency to extrapolate (“hallucinate”) facts
- **The degree of certainty** for successful execution; an open question
  - A lot of misunderstandings concerning hallucinations
  - Different kinds of hallucinations (not all are legally relevant, cf. Wallenberg under review)
- **Practical turn:** Fine tuning and testing methods may, however, strengthen interoperability (or be disruptive)
  - If pipelines scalable -> increased efficiency
  - Disruptive cases: When pipelines don´t work, case-by-case evaluation?

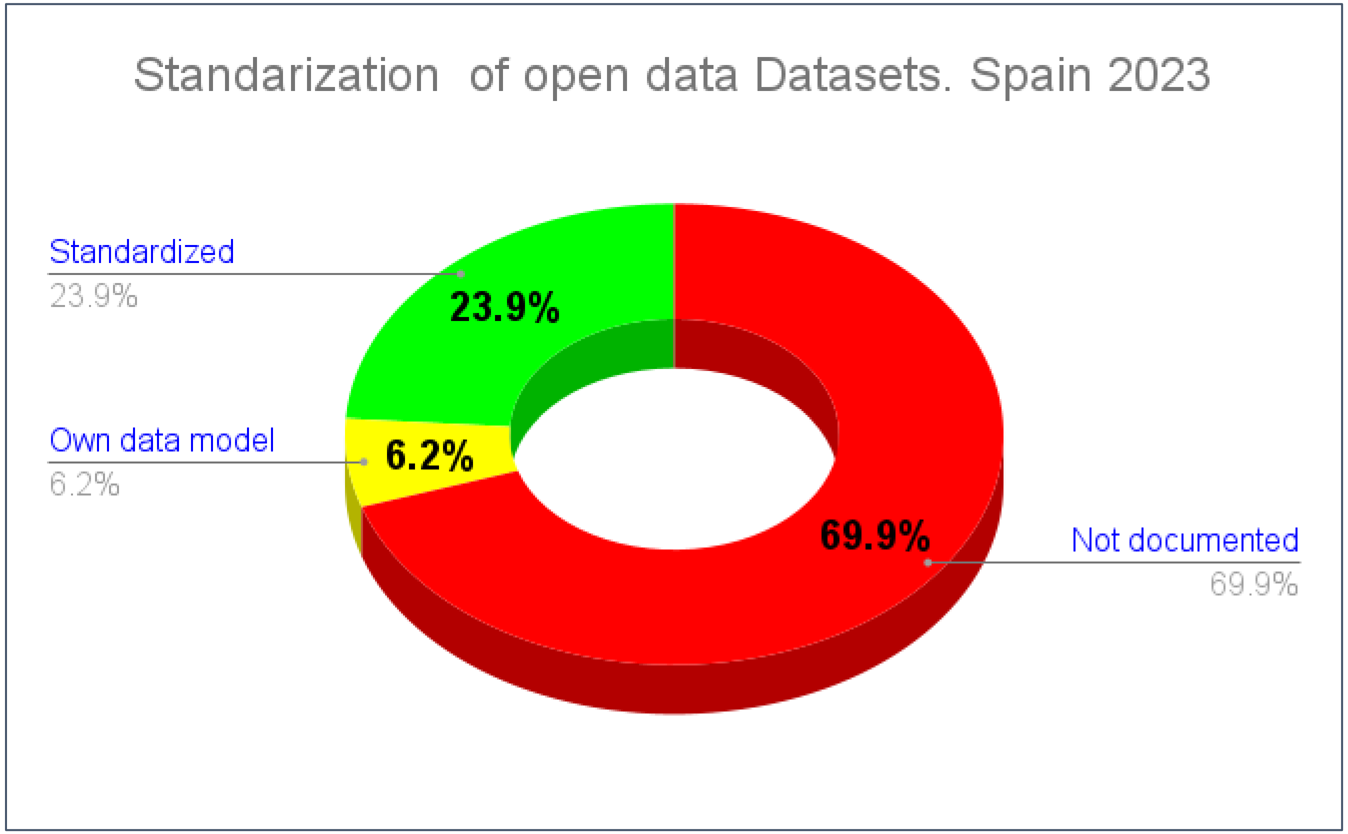
Source: Copilot, 16.6..2024

Thank you

MINISTRY OF FINANCE
FINLAND

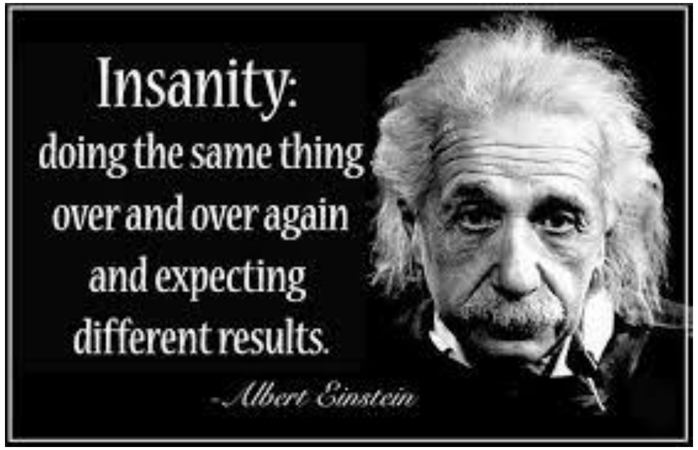
AI for Semantics at FIWARE

Need standardized data models?



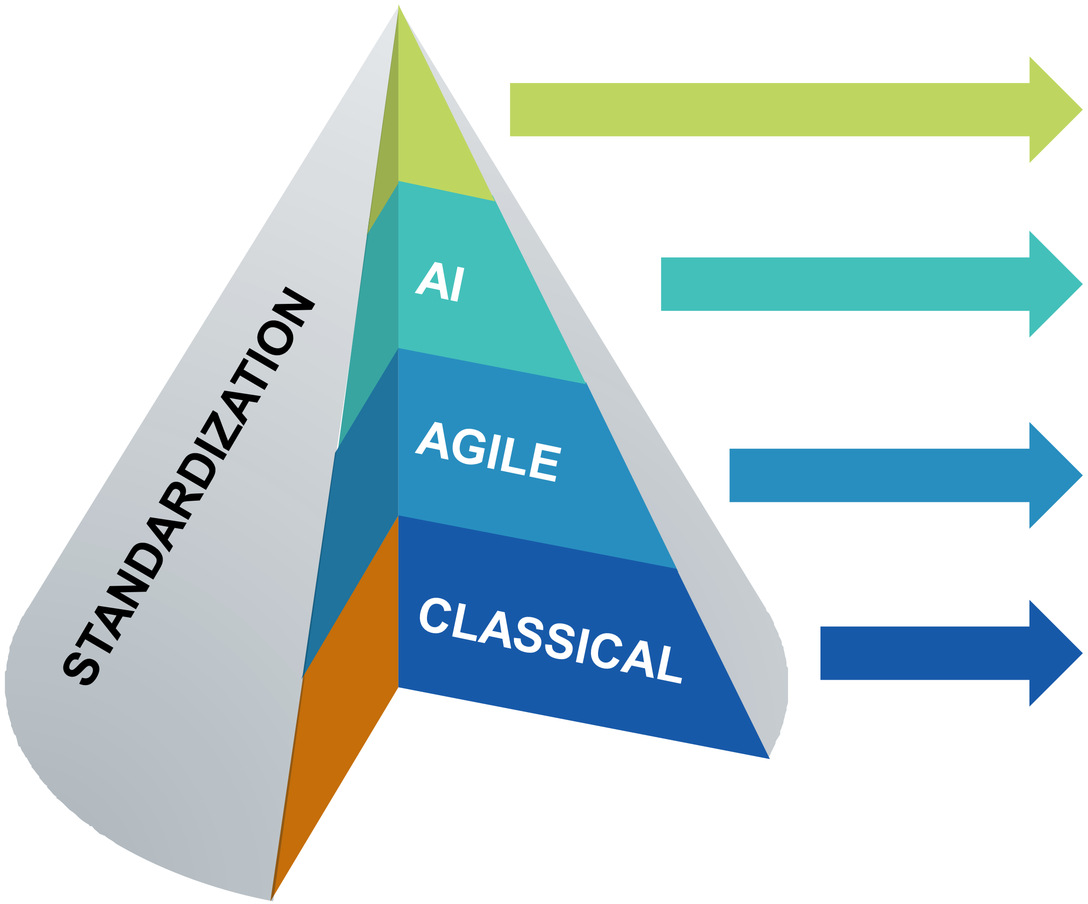
https://bit.ly/opendataspain2023

70% of 71.849 (2023) published datasets are not standardized



Classical standardization initiatives have got 80 datasets' data models in 6 years with low adoption

Standardization on AI-age



FUTURE STANDARDIZATION

¿¿??

AI-SUPPORTED STANDARDIZATION (days)

Connected agents for standards
Supported by AI agents

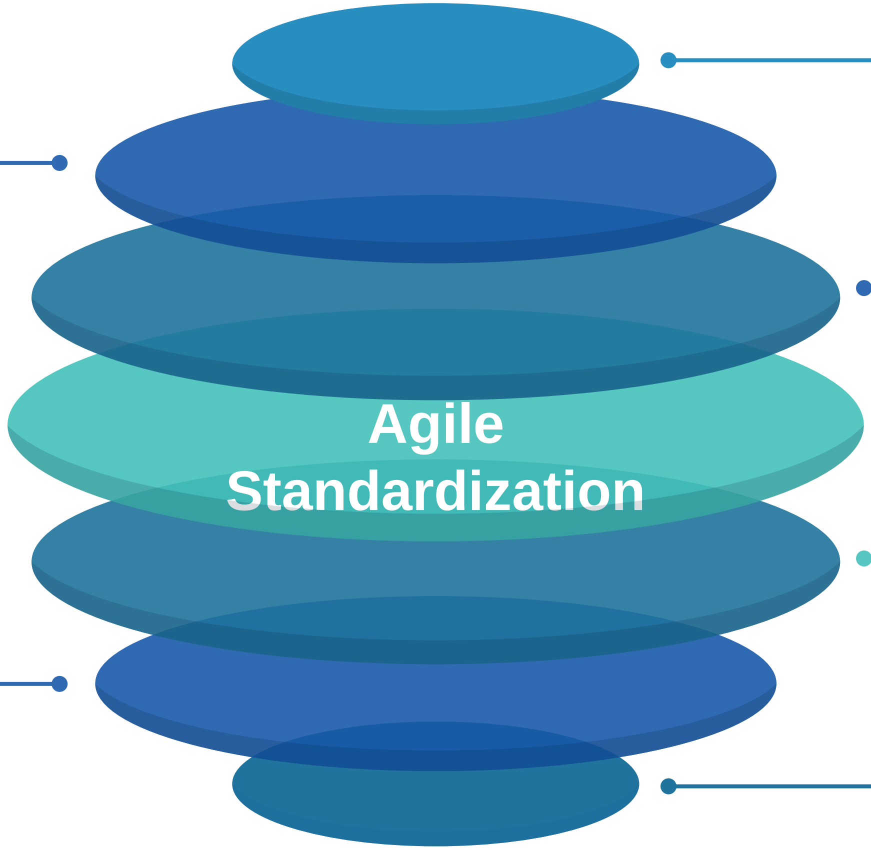
AGILE STANDARDIZATION (weeks)

Based on pioneers and in additive evolution
Simple extension and evolution
Open licensed

CLASSICAL STANDARDIZATION (Tenths months)

Based on consensus and stakeholders involvement
Driven by standardization bodies
Closed and rigid specifications

Seven best practices of agile standardization



Don't just standardize, be agile and standardize

Do it in days/weeks or not do it

Do not reinvent the wheel

if open and adopted standard exists do not reinvent. Adopt.

Normalize real cases

Reality teaches more than a group of experts

Be open

Enable easy adoption and extension

Don't be overly specific

Generalize whenever possible

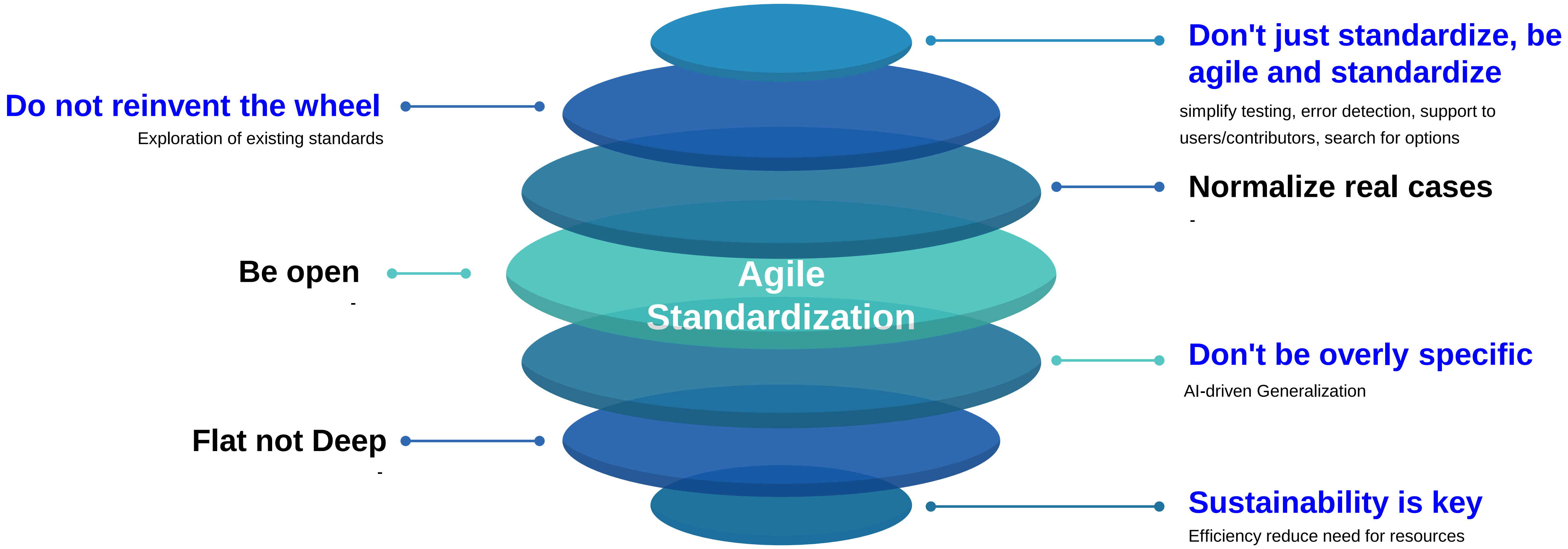
Flat not Deep

Models should be self-contained whenever possible

Sustainability is key

Sustainability mechanisms for the curators of the models is required

With the help of AI



Expected benefits AI for standardization

TESTING

Extended testing of the data models

Speed up testing

Mid-term
6-12 months

SUPPORT

Support to contributors and to users

New exports

Short term
1-4 months

EXPLORATION

New potential data models

Existing regulation to prevent

Long-term
>12 months

Global aim: keep <1 week between first request and data model is published officially

Smart Data Models initiative

Stay in touch https://smartdatamodels.org

@ smartdatamodels

Smart Data Models

alberto.abella@fiware.org

AI 4 Interoperability

Semantic interoperability and artificial intelligence as part of the SEMIC action

AI 4 interoperability

Empower interoperability through AI-driven solutions. Leverage advanced AI models to bridge data silos, foster collaboration, and drive innovation.

Knowledge Extraction

Example of how AI can complement and automate knowledge extraction

Domain Adaptation

Research about the impact of retraining LLM on domain specific data for clustering

Data Modelling Chatbot

Example of how GenAI can support the data modelling process with RAG

Semantic tagging

Exploration of how AI can automate the semantic tagging process

Knowledge Extraction

To harness the potential of AI at supporting interoperability (through automatic semantic alignment, vocabulary extraction, ontology extension, data modelling, …), SEMIC has worked on building a proof-of-concept with DG GROW on the Transition Pathway for Tourism.

The objective was to see how to:

Extract topics from pledges

- **Identify new topics** emerging from the pledges
- **Assign each pledge** to its topic
- **Analyse the coherence** of existing topics

Extract results and implementation dates

- **Identify results** mentioned in the pledges
- **Identify implementation dates**
- **Build timeline** of results

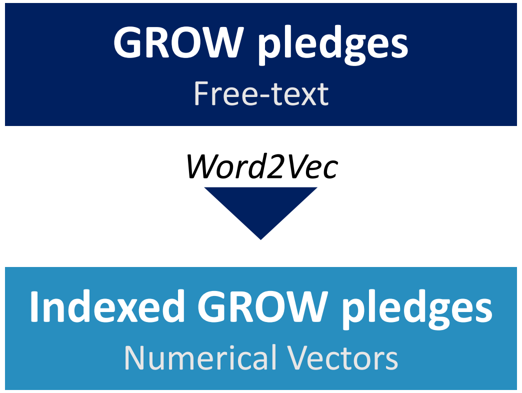
Extracting topics from a corpus of text…

Using **neural language models** and classical **clustering** tools, AI can automatically **identify the main topics** covered by a corpus of texts

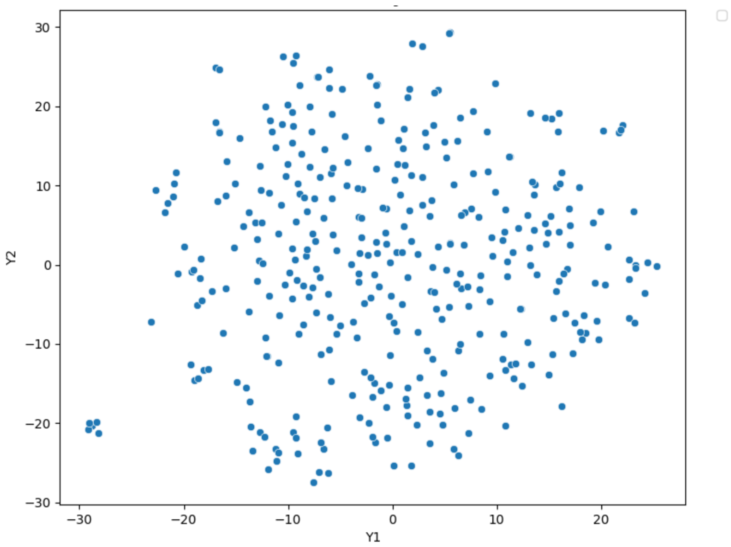
GROW pledges
Free-text

Extracting topics from a corpus of text…

Using **neural language models** and classical **clustering** tools, AI can automatically **identify the main topics** covered by a corpus of texts



1) **Word2Vec**, a neural language model, allows to represent each pledge by a numerical vector while keeping all the semantic and syntactic information.



Extracting topics from a corpus of text…

Using **neural language models** and classical **clustering** tools, AI can automatically **identify the main topics** covered by a corpus of texts

GROW pledges
Free-text

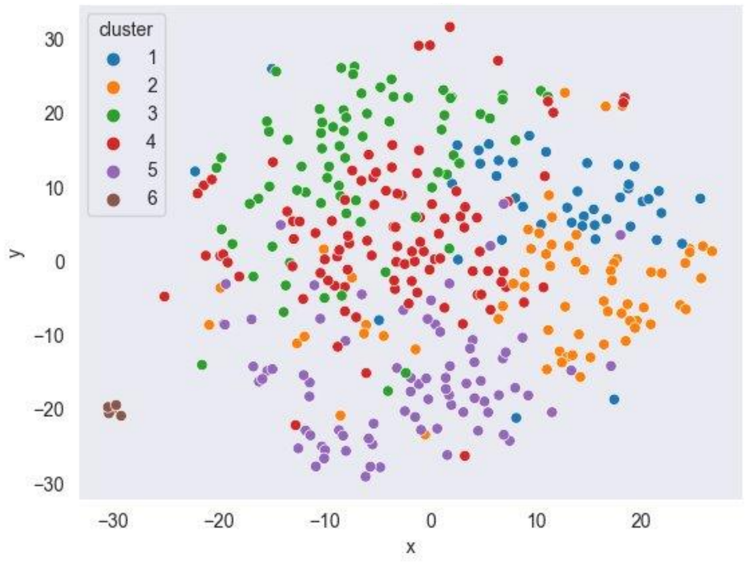
Word2Vec

Indexed GROW pledges
Numerical Vectors

K-Means

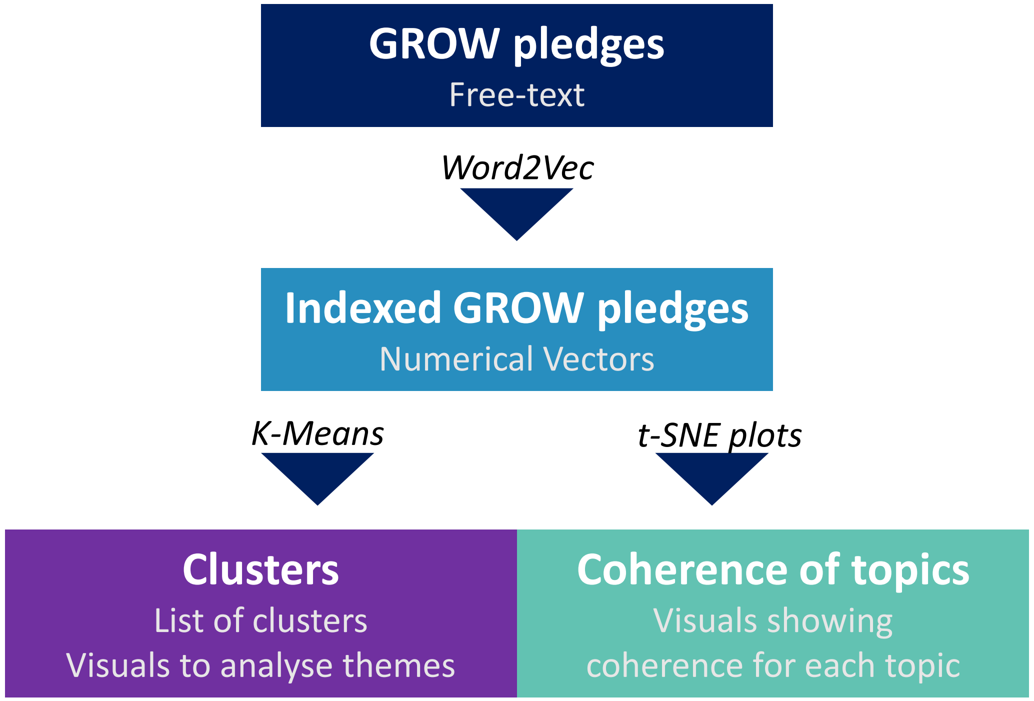
Clusters
List of clusters
Visuals to analyse themes

2) **K-means**, a clustering algorithm, allows to group pledges with a similar semantic meaning into different clusters

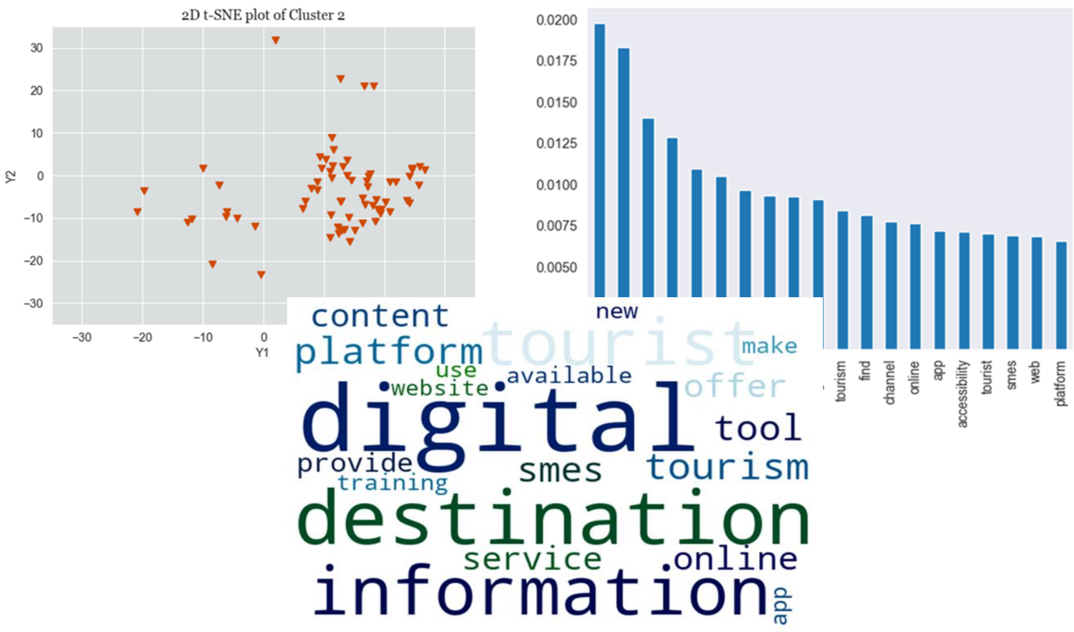


Extracting topics from a corpus of text…

Using **neural language models** and classical **clustering** tools, AI can automatically **identify the main topics** covered by a corpus of texts



3) **T-SNE plots and various word clouds** allow to analyse the theme and coherence of the different topics.



Domain adaptation

To stay at the forefront of AI development, SEMIC has worked on developing analysis and research in the field of Generative AI. A study on the impact of domain adaptation was carried out on two state-of-the-art models:

- How does the retraining of BERT on tourism data impacts the quality of pledge clustering?
- How does the retraining of RoBERTa on tourism data impacts the quality of pledge clustering?

Context and research question

Recent research showed that providing Large Language Models with domain-specific data could significantly improve the performance of these AI models.

For instance, providing legal documents to the LLM could improve its ability to understand legal questions and respond to them more accurately.

Would it be beneficial to fine-tune language models on EU policy documents and legislations?

1. Defining relevant use cases
2. Building a domain corpus
3. Fine-tuning language models
4. Clustering validation

Extrinsic cluster validation: Results

	Extrinsic	
	Accuracy	*F1-score*
Word2Vec	**0.48**	**0.47**
BERT	0.22	0.25
Fine-tuned BERT	**0.29**	**0.34**
Roberta	0.21	0.19
Fine-tuned Roberta	**0.30**	**0.35**

Fine-tuning seems to have a positive impact on the accuracy of the clusters

Overall, the low accuracy and F1-score of the BERT and RoBERTa models tends to indicate that clusters created with these models are harder for "humans" to differentiate. In other words, the increase in complexity also decreased the usability of the results.

These conclusions may provide an interesting perspective on the **trade-off between using more sophisticated models and their usability**. While LLMs showed better quantitative results, they were also found to be less intuitive for human users. Hence, relying on more complex models may not always be the most valuable approach.

Data modelling Chatbot

To support the work of data modellers, SEMIC has developed a prototype of chatbot that can **suggest classes, relations, and properties** to be added to a data model based on **existing concepts**.

Suggest a possible name for a class (CPSV-AP):

" a requirement fulfill a ____" => autocomplete with "Rule"

Suggest a relation (Core Person):

"a person is ______ at this address" = (living, domiciliated, etc.)

Suggest properties (Core Business):

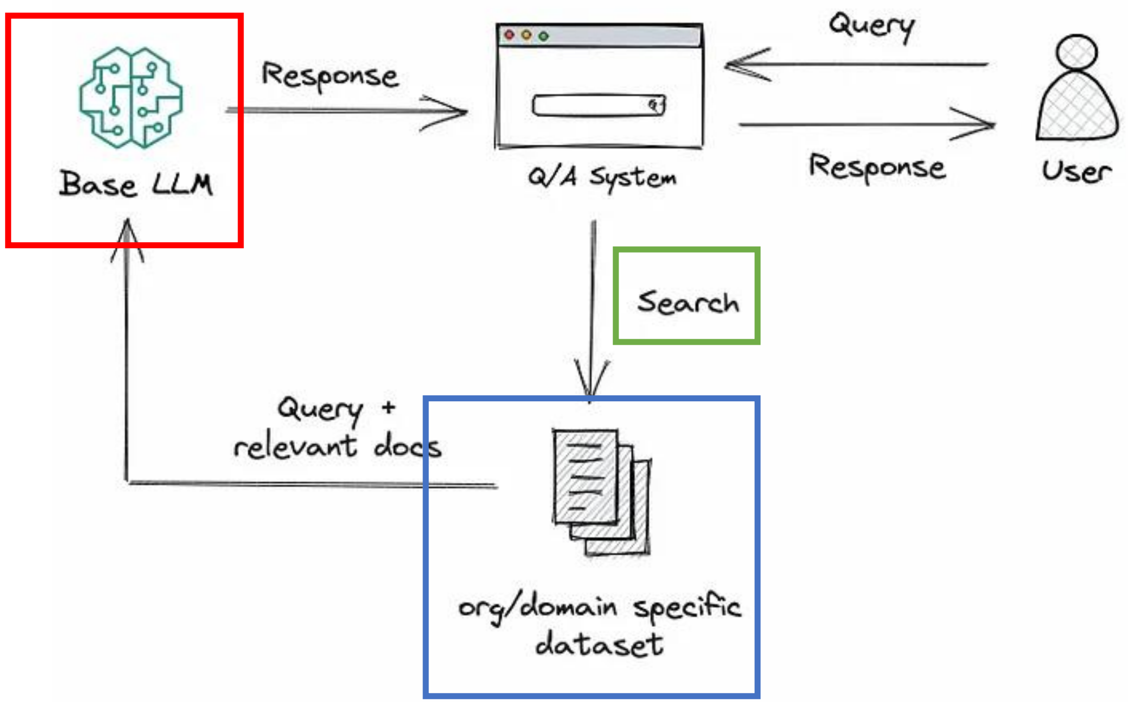
"by what a legal entity can be uniquely identified ?"

Examples of what the prototype could do

Retrieval augmented Generation

A Retrieval Augmented Generation (RAG) chatbot is based on **three main building blocks**:

Architecture of a RAG chatbot



1) **A pre-trained LLM:** A large language model trained on a very large corpus of generic data (e.g. Wikipedia).

2) **A knowledge base:** An index containing relevant documents for helping the data modeller, i.e., specification pages, RDF files, …

3) **A search engine:** A tool able to search the knowledge base for the most appropriate documents to answer a user's query

Building a knowledge base

- **Five relevant sources** for CPSV-AP were identified, scraped, and chunked by the team.
- **Two existing RDF files** were identified as interesting sources.

Status	Data source
☑	CPSV-AP
☑	CPOV
☑	CCCEV
☑	Open Data directive
☑	SDG regulation
☑	LOV
☑	Schema.org

HTML pages containing relevant textual information about CPSV-AP (specification pages and regulations)

RDF files containing existing data models and concepts that could be reused by the chatbot

Building a RAG chatbot – Version 1

Use case 1: **Suggest a relation (Core Person):**

"a person is ______ at this address" = (living, domiciliated, etc.)

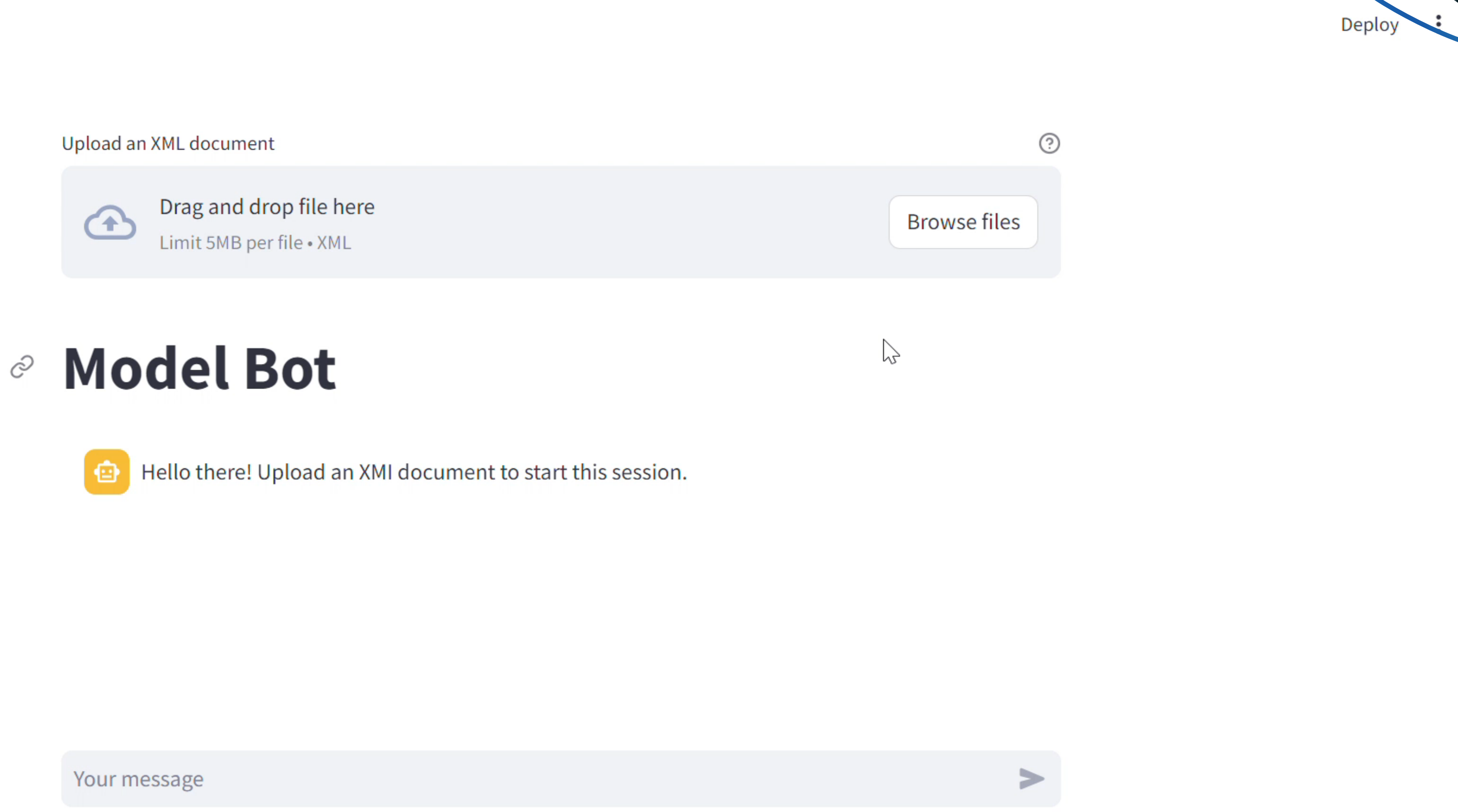
Suggest a possible name for a relation: a person is _____ at this address

Based on the retrieved documents, a possible name for the relation where a person is associated with an address could be "has address of the person" as indicated in [1] . This relation connects a person to their place of residence, business, or employer, and can also include an e-mail or internet address. Therefore, the complete phrase would be "a person is **has address of the person** at this address." However, this is a technical representation from the RDF schema perspective. In a more human-readable format, you might simply say "a person is **located** at this address" or "a person is **resident** at this address," depending on the context of the use case.

1 reference >

AI-generated content may be incorrect

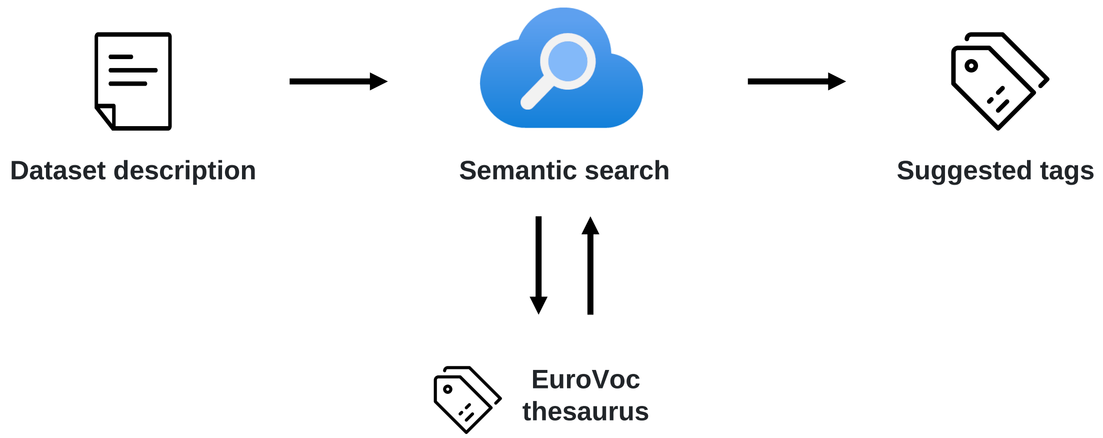
Building a RAG chatbot – Additional features

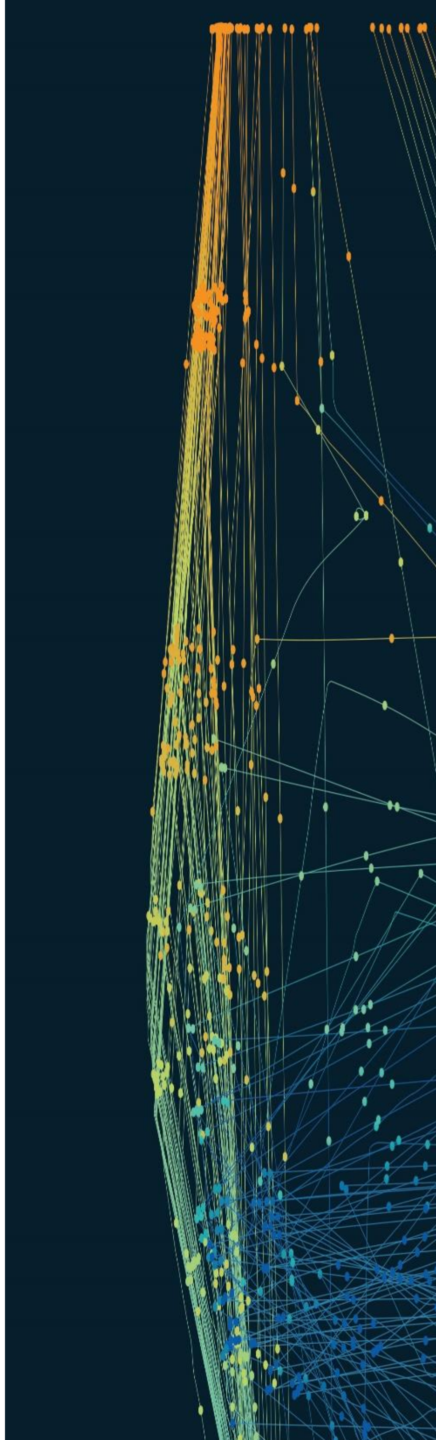


Semantic tagging

To facilitate the research and retrieval of documents on the European open data portal, SEMIC has explored how semantic tagging could be used to **suggest tags from the EuroVoc thesaurus** for a dataset.

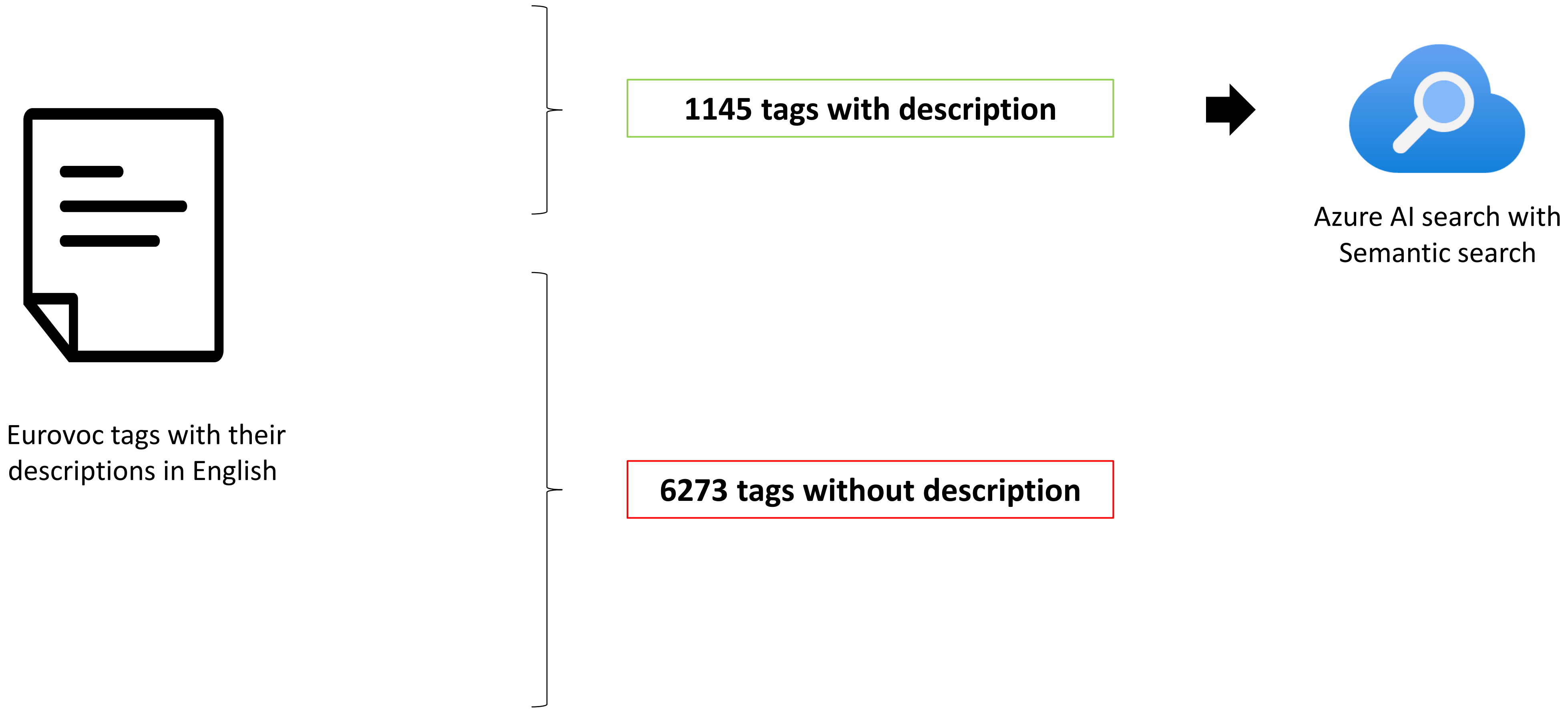
High-level architecture:





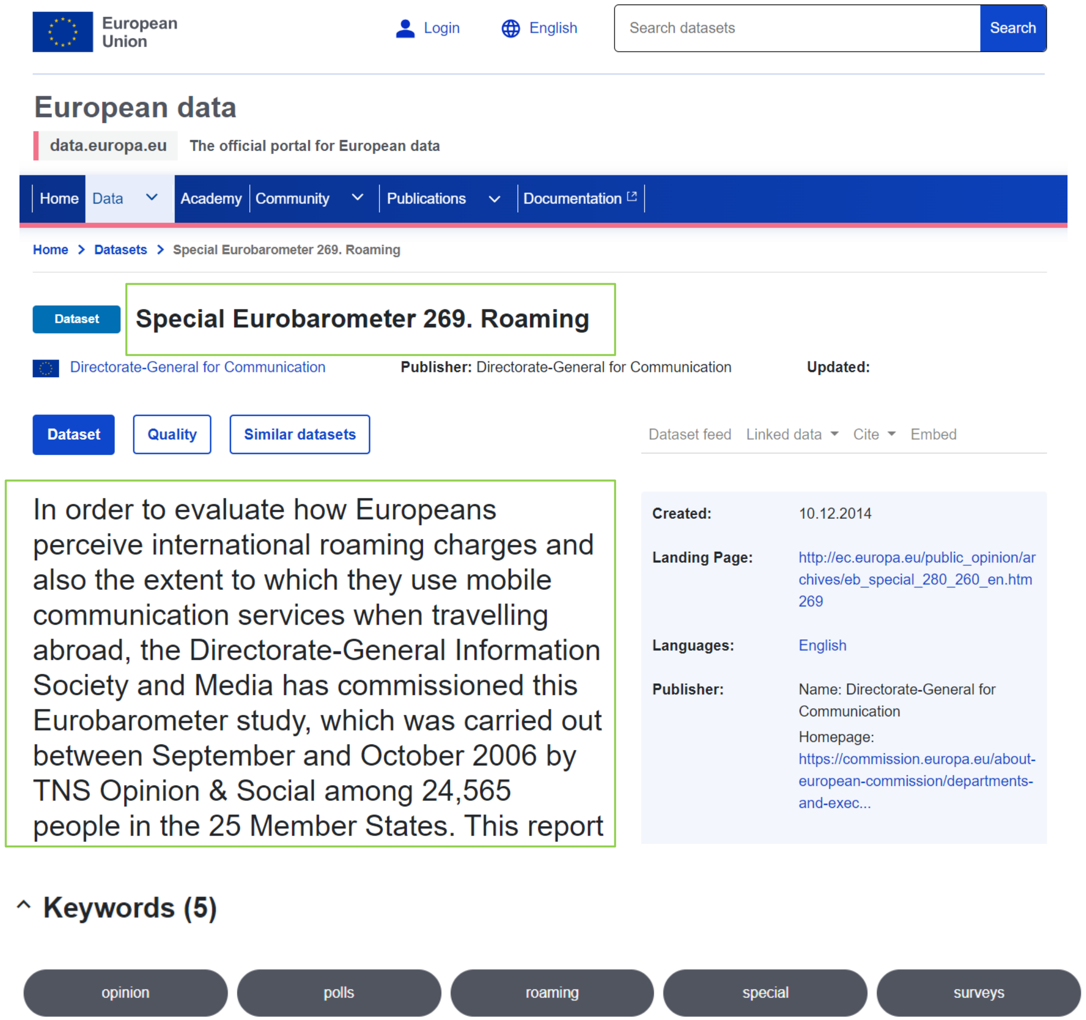
Data source

Starting from the different files available on the OP portal for Eurovoc, we found a zip file containing various XML files. One of those contains the definitions of the different Eurovoc tags in the different EU languages.



First results

Using Azure AI search explorer tool, we tested the efficiency of this approach for two examples. We used datasets found on the data.europa.eu.

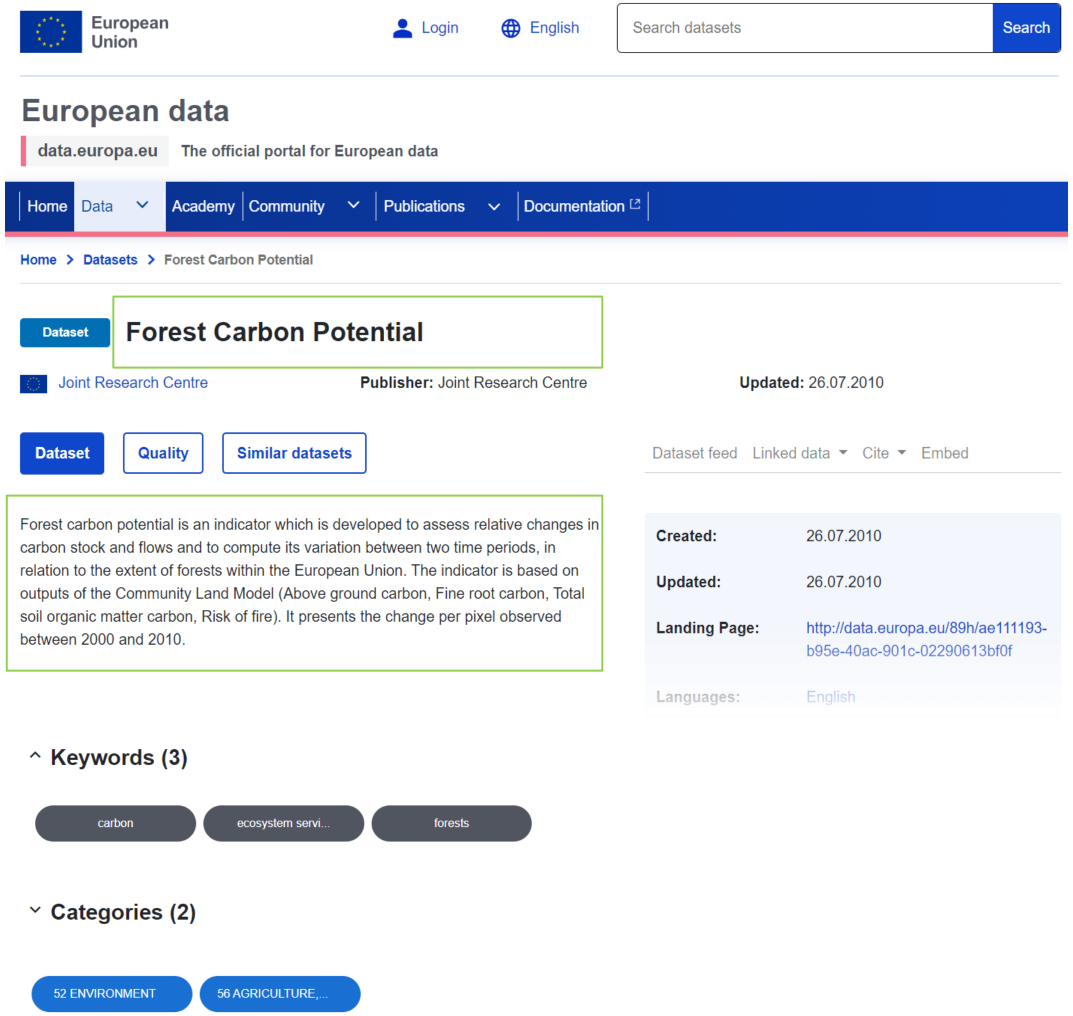


Suggested Tags

Economic governance
EU law – International Law
Roaming
EU market

First results

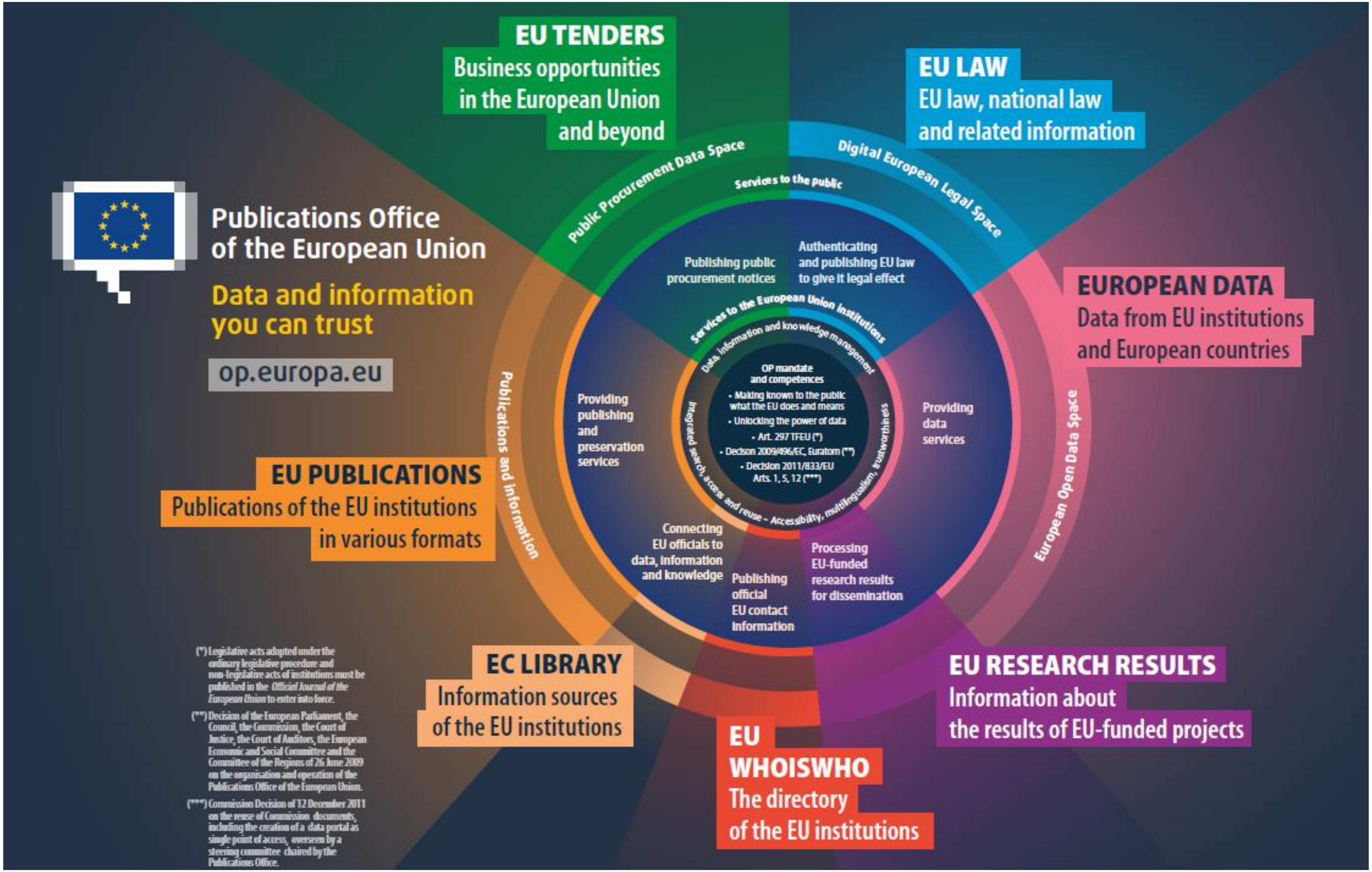
Using Azure AI search explorer tool, we tested the efficiency of this approach for two examples. We used datasets found on the data.europa.eu.



Suggested Tags

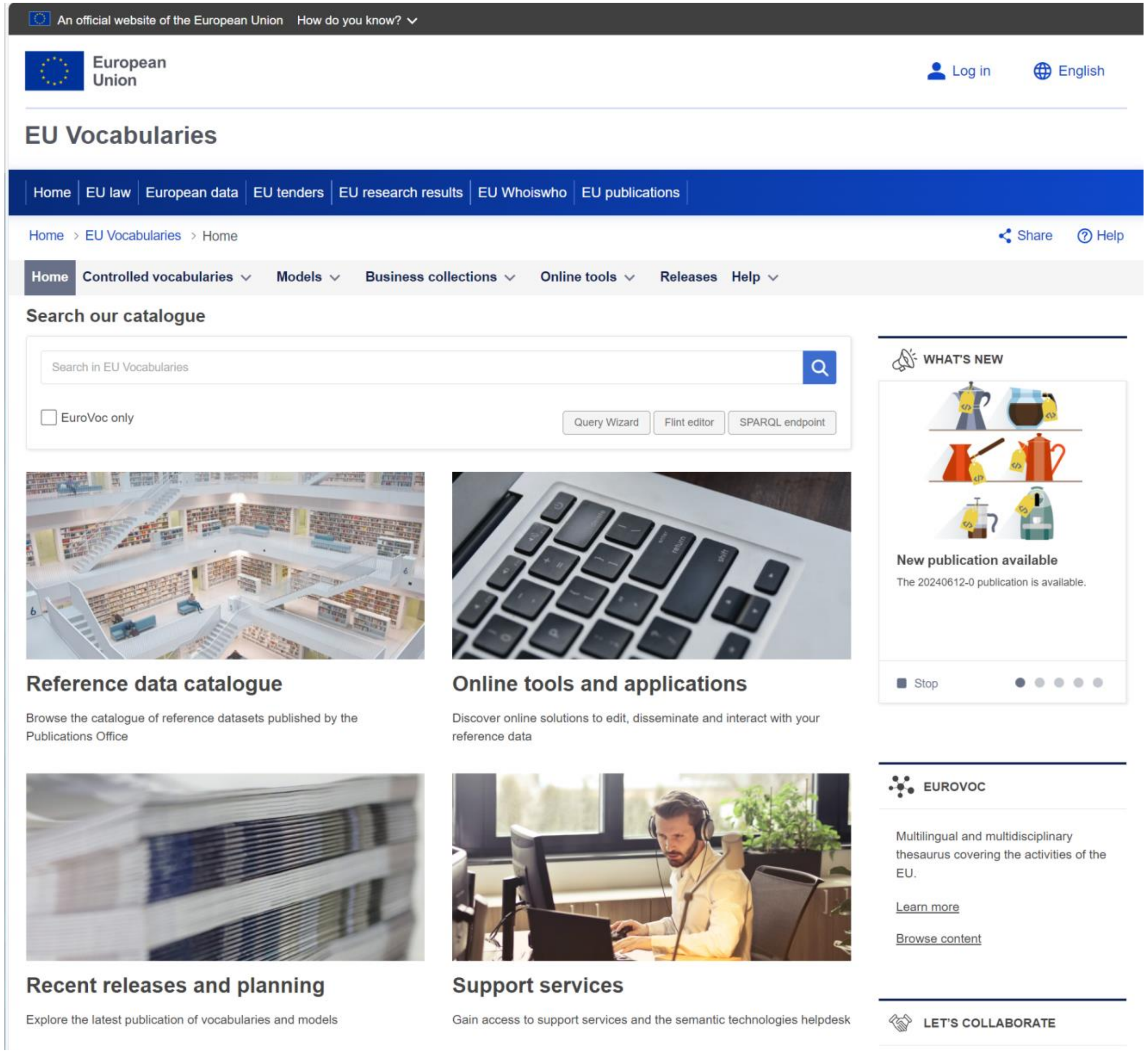
Economic instrument for the environment
European agricultural model
Kyoto Protocol
Carbon neutrality
Carbon capture and storage

EU reference data and AI

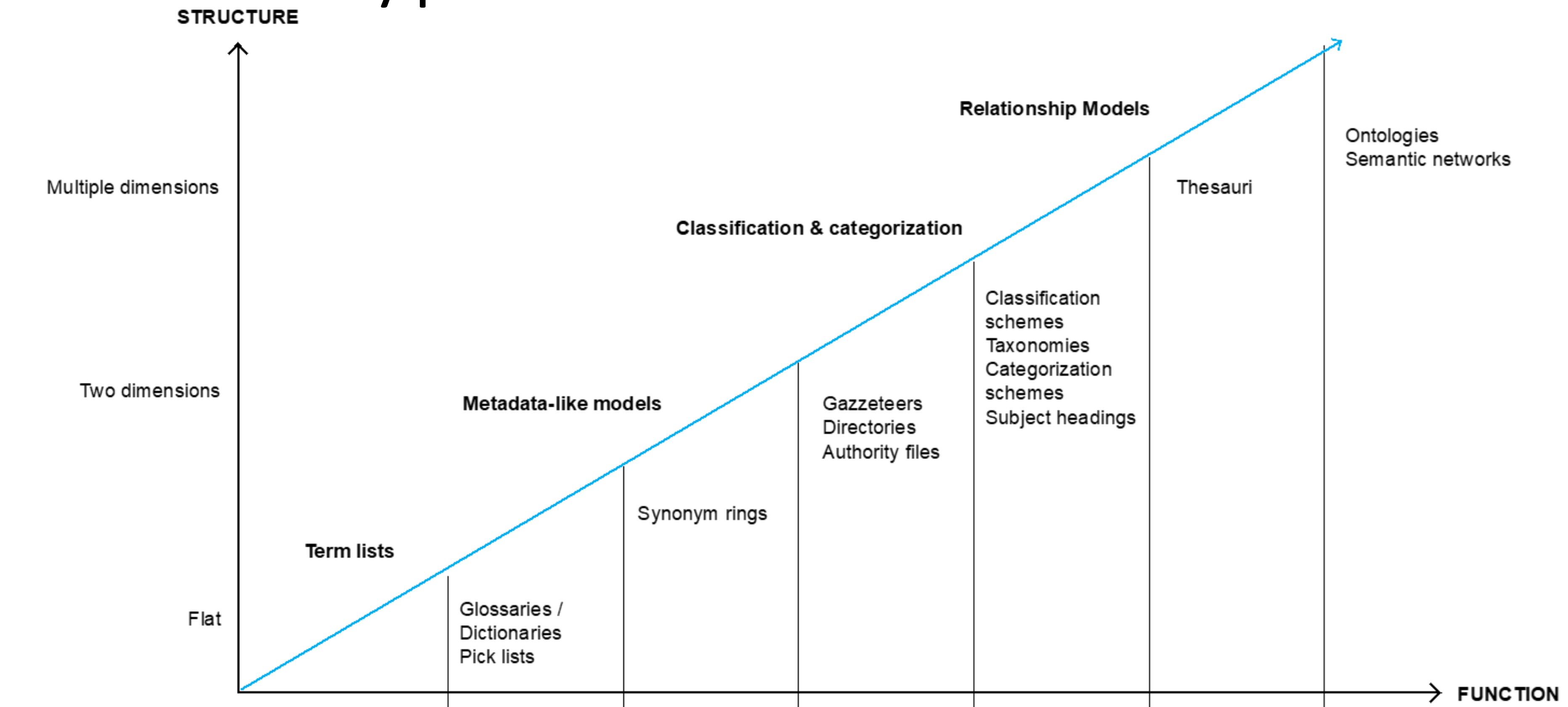


Publications Office of the European Union
Data and information you can trust
op.europa.eu
EU TENDERS
Business opportunities in the European Union and beyond
EU LAW
EU law, national law and related information
EUROPEAN DATA
Data from EU institutions and European countries
EU PUBLICATIONS
Publications of the EU institutions in various formats
EC LIBRARY
Information sources of the EU institutions
EU WHOISWHO
The directory of the EU institutions
EU RESEARCH RESULTS
Information about the results of EU-funded projects
Public Procurement Data Space
Digital European Legal Space
European Open Data Space
Publications and information
Services to the public
Services to the European Union institutions
Data, information and knowledge management
Integrated search, access and reuse – Accessibility, multilingualism, trustworthiness
Publishing public procurement notices
Authenticating and publishing EU law to give it legal effect
Providing publishing and preservation services
Providing data services
Connecting EU officials to data, information and knowledge
Publishing official EU contact information
Processing EU-funded research results for dissemination
OP mandate and competences
• Making known to the public what the EU does and means
• Unlocking the power of data
• Art. 297 TFEU (*)
• Decision 2009/496/EC, Euratom (**)
• Decision 2011/833/EU Arts. 1, 5, 12 (***)
(*) Legislative acts adopted under the ordinary legislative procedure and non-legislative acts of institutions must be published in the Official Journal of the European Union to enter into force.
(**) Decision of the European Parliament, the Council, the Commission, the Court of Justice, the Court of Auditors, the European Economic and Social Committee and the Committee of the Regions of 26 June 2009 on the organisation and operation of the Publications Office of the European Union.
(***) Commission Decision of 12 December 2011 on the reuse of Commission documents, including the creation of a data portal as single point of access, overseen by a steering committee chaired by the Publications Office.

Central access to EU reference data

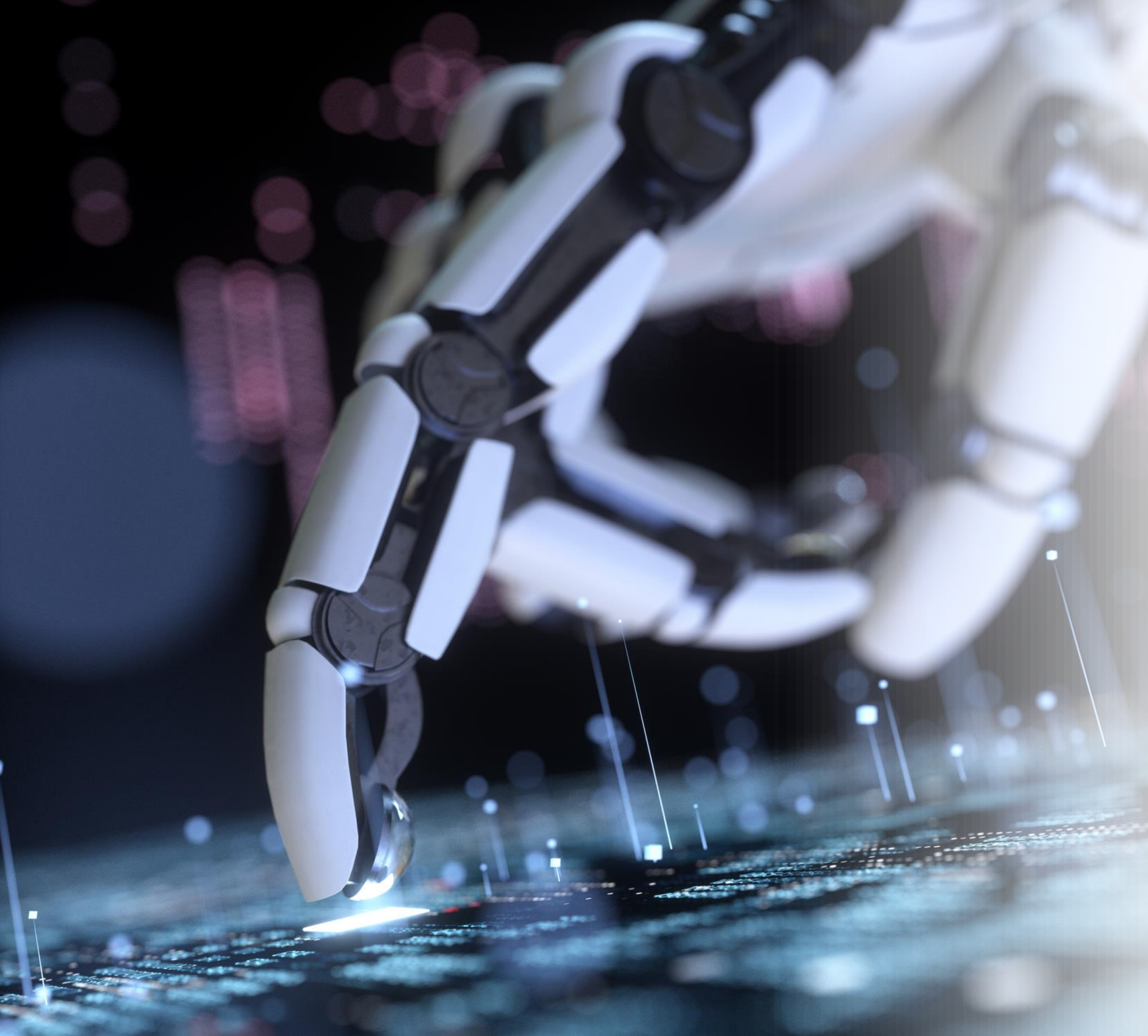


Different types of reference data



Mayor functions

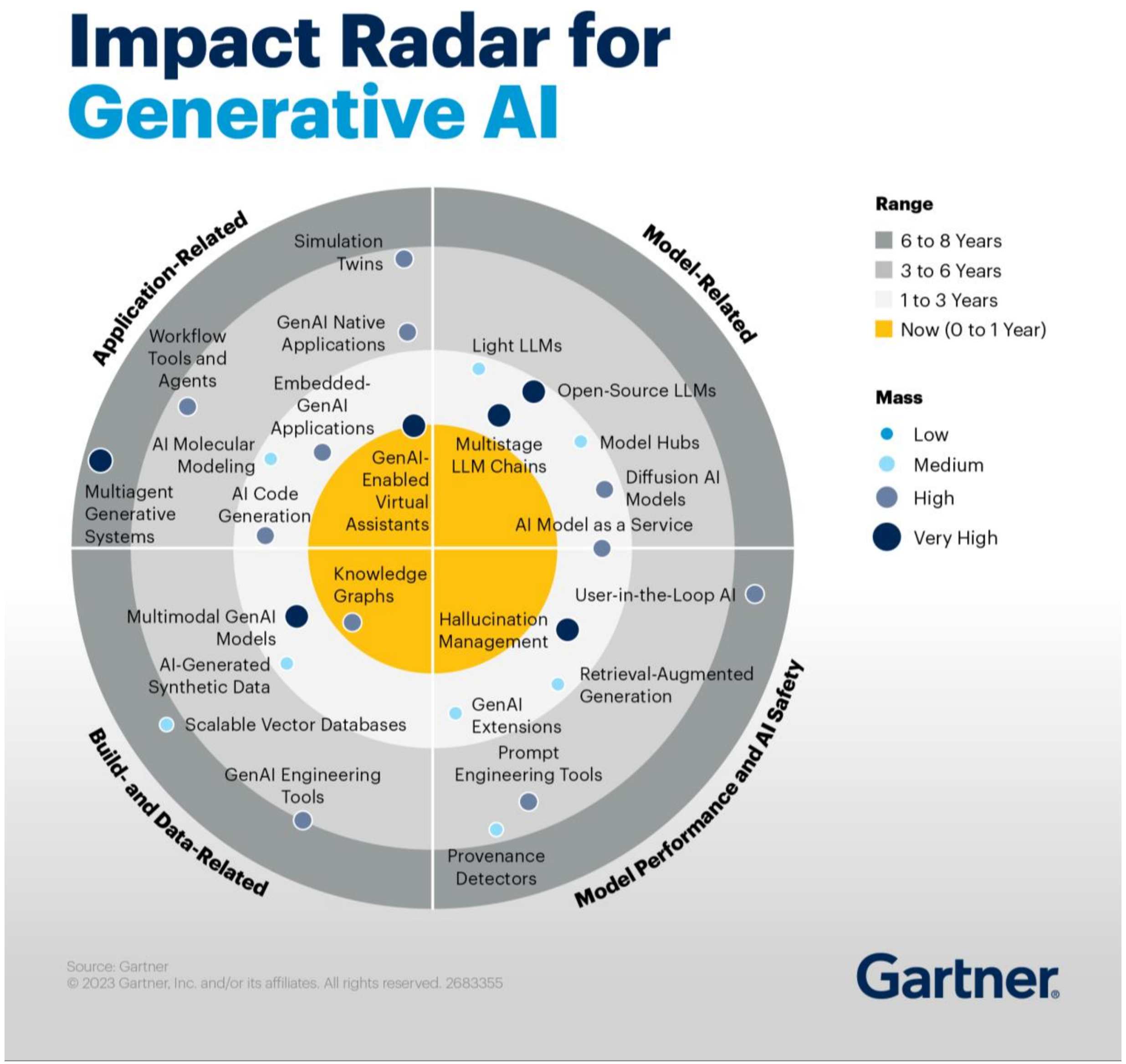
Eliminating ambiguity	●●●○	○○○○	●●●○	●●○○	●●●●	●●○○
Controlling synonyms	○○○○	●●●●	●●●○	●●○○	●●●●	●●○○
Establishing relationships: hierarch ical	○○○○	○○○○	●○○○	●●●●	●●●○	●●●○
Establishing relationships: associat ive	○○○○	○○○○	○○○○	○○○○	●●●●	●●●●
Presenting properties	○○○○	○○○○	○○○○	○○○○	○○○○	●●●●

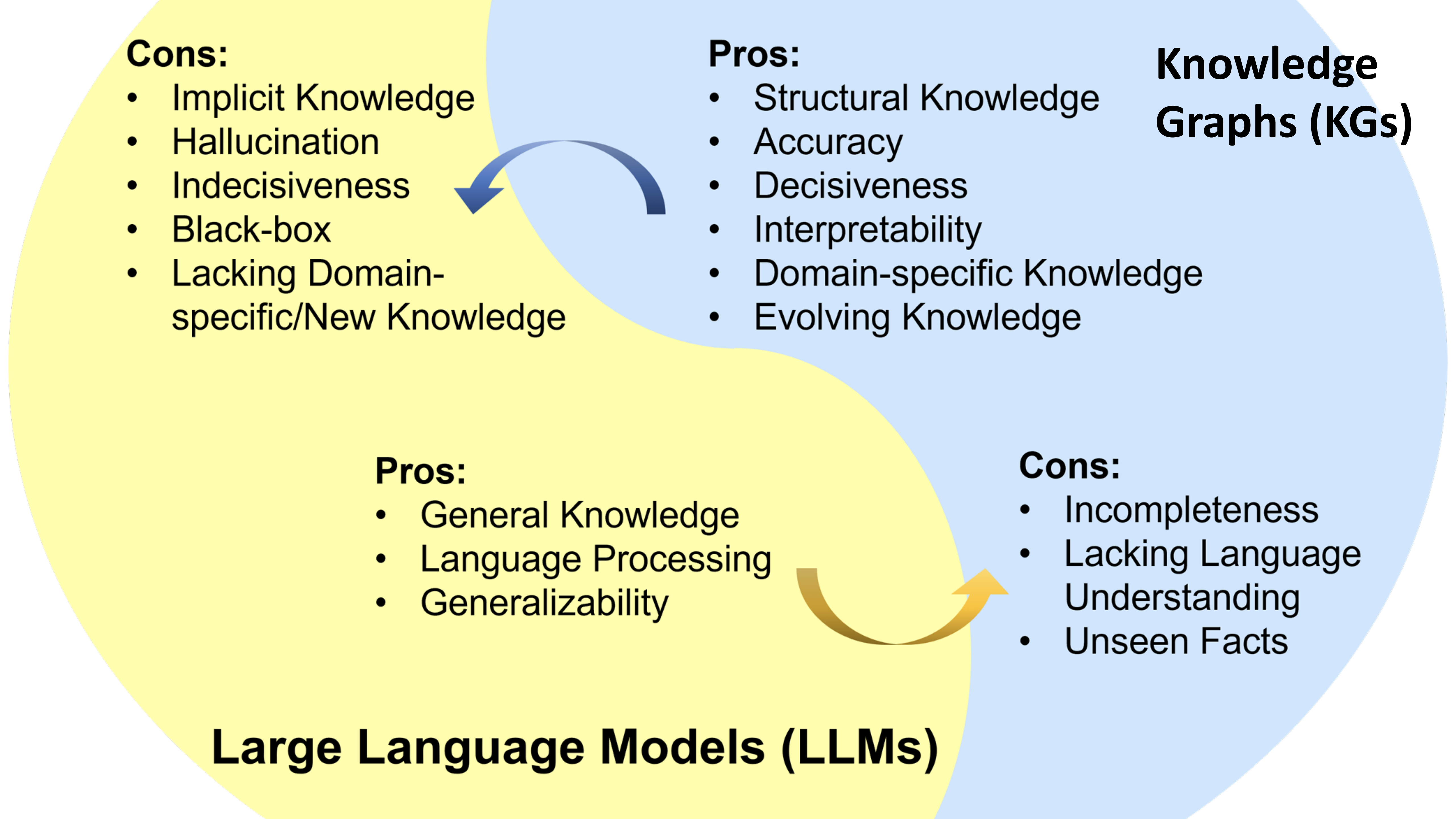


Do we need reference data in the era of AI?

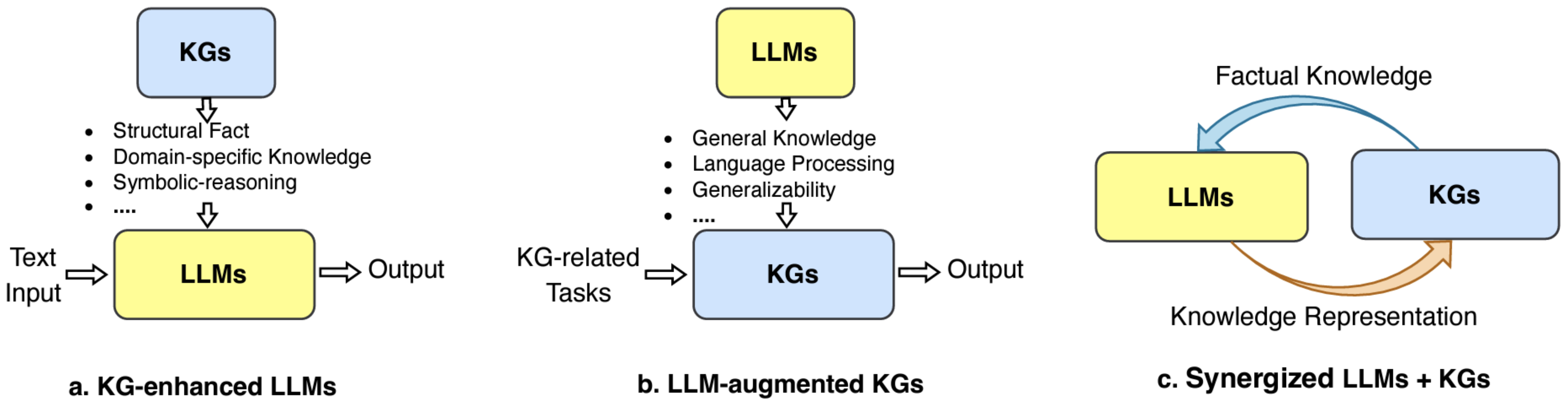
- Trustworthiness of data
- Reference data allow to provide business context
- High-quality data needed to train, validate and test algorithms
- Data quality directly impacts performance and reliability of AI systems

2024: Knowledge graph integration




Cons:
Implicit Knowledge
Hallucination
Indecisiveness
Black-box
Lacking Domain-specific/New Knowledge
Pros:
Structural Knowledge
Accuracy
Decisiveness
Interpretability
Domain-specific Knowledge
Evolving Knowledge
Knowledge Graphs (KGs)
Pros:
General Knowledge
Language Processing
Generalizability
Cons:
Incompleteness
Lacking Language Understanding
Unseen Facts
Large Language Models (LLMs)

Knowledge Graph and LLM unification



Leveraging Generative AI for knowledge graphs (1)

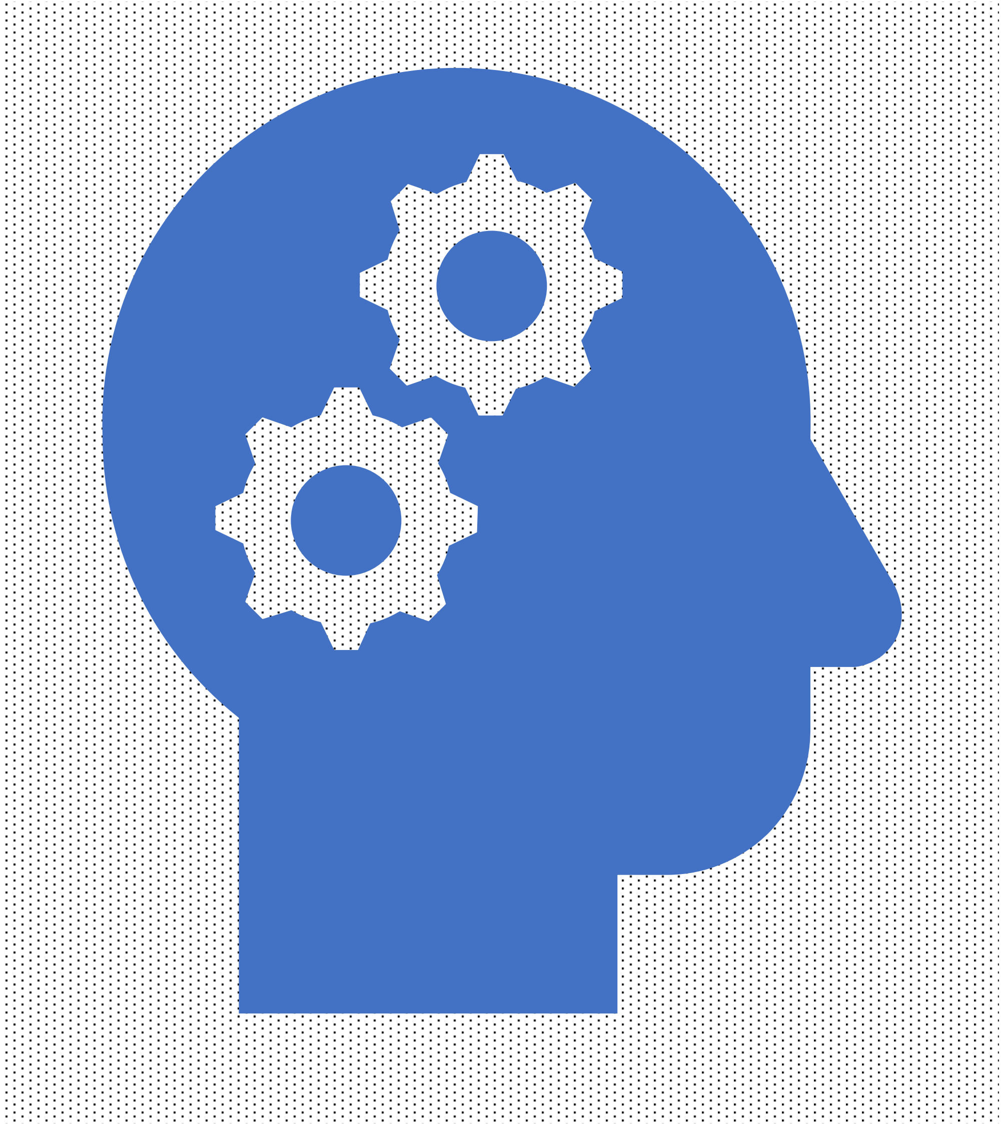
- Quality control and consistency checks
- Automated translation and multilingual alignment
- Automated vocabulary expansion
- Semantic analysis for vocabulary integration
- Custom vocabulary generation for specific needs

Leveraging Generative AI for knowledge graphs (2)

- Natural Language Processing for document tagging
- Chatbot for user interaction with reference data including query creation
- Enhanced search functionality including predictive text and autocomplete features
- Accessibility enhancements

Challenges and points for discussion

- Channel multitude of AI initiatives and share knowledge
- Automatic semantic tagging: From proof of concept to production
- What do you expect from EU reference data?



Thank you

interoperable europe

Panel Discussion

2: AI in support of Interoperability

Join us on Slido!

- Use the QR code
- Or go on slido.com
  #SEMIC2024Workshops
- Select the correct workshop

We have some questions for you too!

COFFEE BREAK
11:00-11:30
HALL 100

Part 2: Interoperability in support of AI

Workshop’s speakers:

Building Trust In AI

Semantic Interoperability and Knowledge Graphs at the Rescue of Large Language Models

Agenda

- Introduction to the Trust Challenge
- Semantic Interoperability Explained
- Knowledge Graphs: AI's Trust Backbone?
- The ABI Project: What We Learned
- Impact on Trust and Efficiency
- The Future of Semantic AI

JEREMY RAVENEL

Founder & CEO

NaasAI

"I'm a farm boy turned data mercenary and now an entrepreneur, I'm passionate about practical innovation to make our digital lives less chaotic."

INTRODUCTION TO THE TRUST CHALLENGE

Key Issues in AI Development and Deployment

01

HIGH PROFILE ERRORS

Autonomous vehicle incidents, healthcare misdiagnoses, hallucinations harms people

02

BIAS AND DISCRIMINATION

AI replicating societal biases in hiring, lending, and law enforcement

03

PRIVACY CONCERNS

Data breaches and unauthorized surveillance

04

ECONOMIC IMPACT

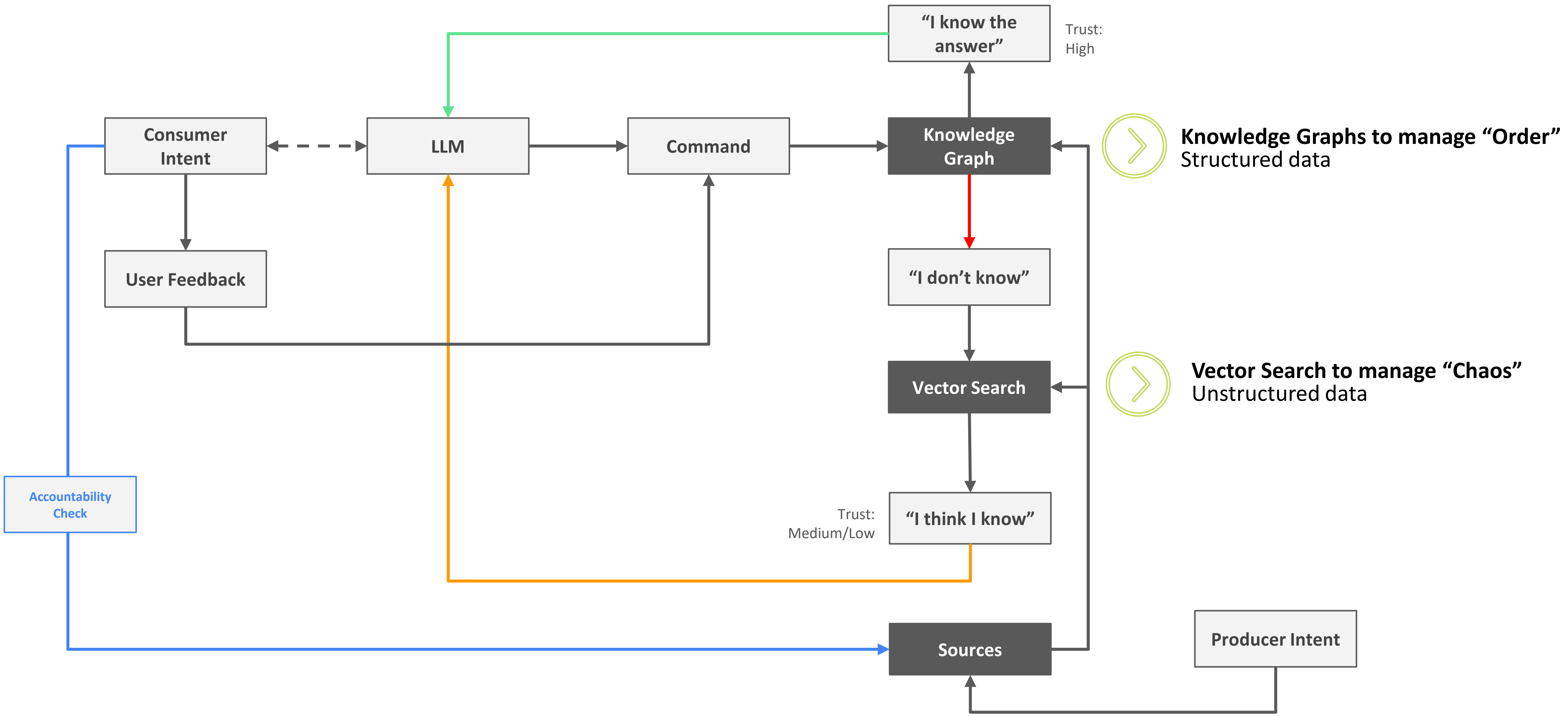
Loss of consumer trust leading to reduced adoption and financial setbacks

05

REGULATORY CHALLENGES

Increasing calls for stringent AI governance and compliance

Large Language Models (LLMs) should not be trusted, they need to be integrated into a broader workflow to maximize their utility and ensure reliability.



Just as musicians and dancers must follow specific rules and cues to perform beautifully, AI systems rely on common language, clear data context, and standardized protocols to communicate effectively

SEMANTIC INTEROPERABILITY EXPLAINED

The Subtle Dance of Grounded AI Communication

Imagine semantic interoperability as the subtle dance of grounded AI communication, essential for ensuring clarity and precision. It's foundational for building a trusted AI ecosystem. It provides:

Common Language: Allows disparate AI systems to speak and understand each other

Enhanced Accuracy: Reduces misconceptions by providing clear data context

Facilitates Collaboration: Connects various sectors and organizations seamlessly

Supports Compliance: Aligns with regulatory standards for data usage and privacy

Knowledge Graphs: AI's Trust Backbone?

Implementing Knowledge Graphs and developing opinionated data models during the pre-processing stage force AI systems builders to create a more structured user experience and journey. This synchronization with the foundational AI model provides several key benefits:

Rich Context: Encapsulates relationships between vast data points, providing deeper insights and more meaningful interactions.

Dynamic Learning: Adapts and evolves with new information, continuously improving performance and accuracy.

Transparency: Offers clear traceable insights into the AI system response, enhancing trust and accountability.

Interdisciplinary Insights: Bridges gaps between different fields of knowledge, for a more holistic understanding of questions.

THE ABI PROJECT: WHAT WE LEARNED

Building Trust Flywheels Through Semantics and Domain Knowledge

Building ABI, an AI system for businesses, we realized the importance of KG and opinionated data models is not even enough; you need to think in terms of flywheel: interconnected and interoperable elements, a logical flow in the head of the user.

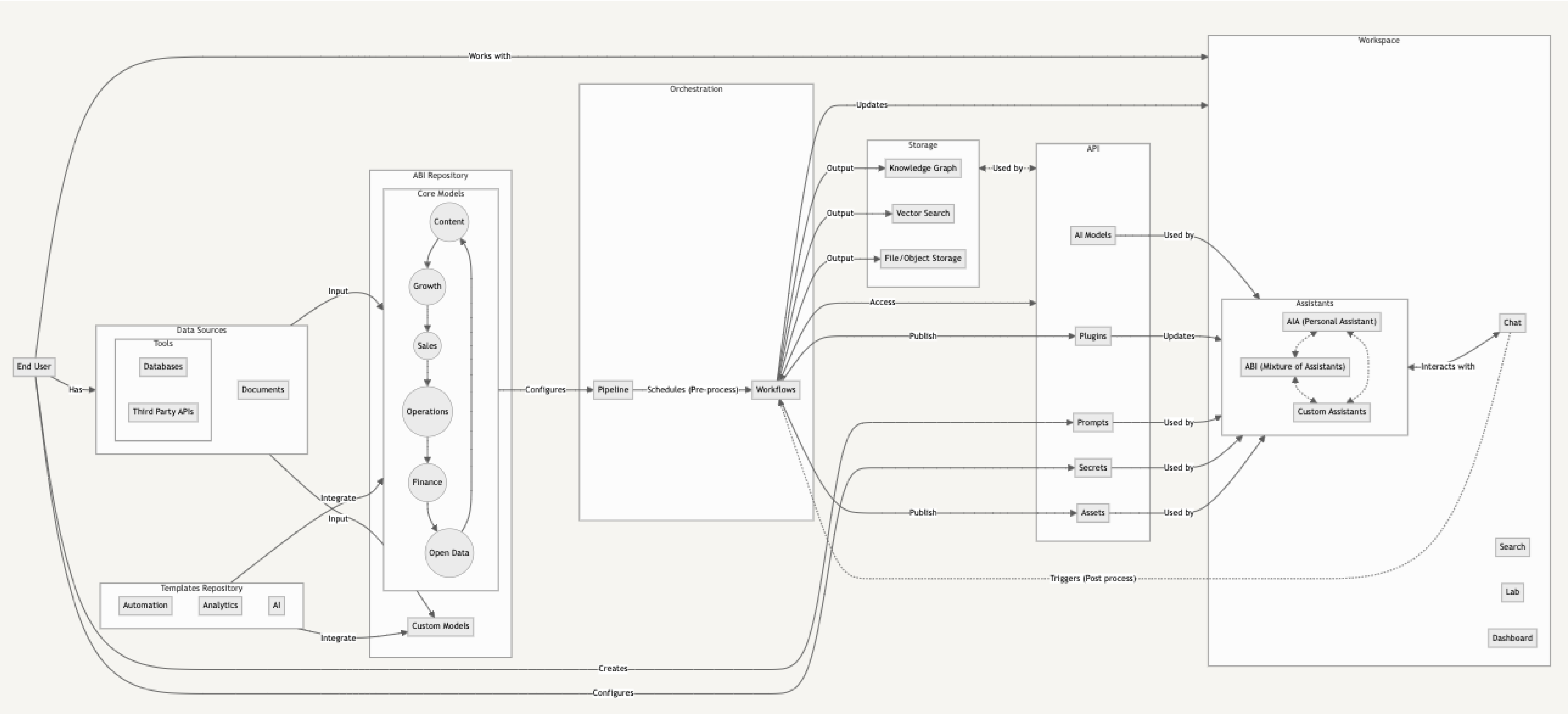
Open Data: External datasets enriches your market understanding, for proactive strategic adaptation.

Content: Standardizing how content is created, tagged, and shared ensures that all communication across channels is consistent and in line with the strategy.

Growth & Sales: Aggregating marketing and sales data create a unified view of qualified contacts and customers interactions and journeys, to build more meaningful relationships.

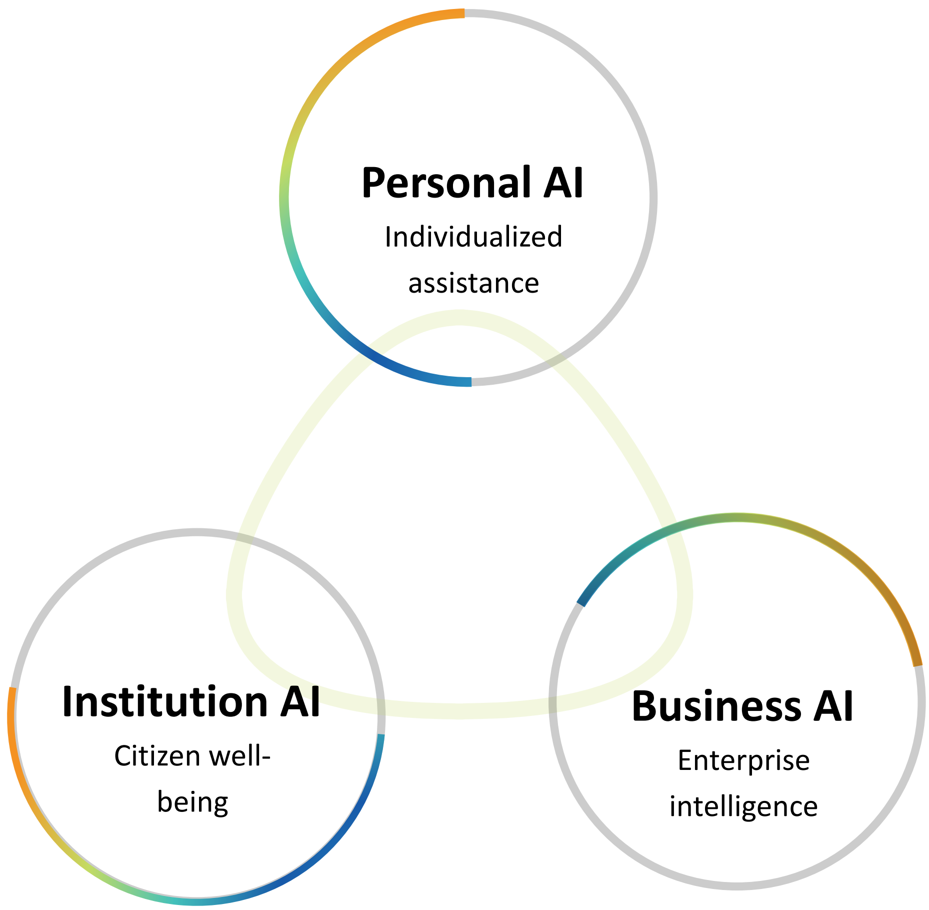
Operations & Finance: Operational tasks to support sales and marketing generates info and assets financial metrics provides a holistic view of business performance.

Our Work Toward a Universal Data & AI Platform Architecture



The Future of Semantic AI

Connecting Personal AI, Business AI, and Institutional AI to enhance personal, business, and societal trusted outcomes.



Personal AI ↔ Business AI:

- Personal preferences and behaviors (from Personal AI) can inform business strategies (in Business AI) for targeted marketing and personalized customer service.
- Business insights and market trends (from Business AI) can help individuals make informed decisions about products, services, and investments.

Personal AI ↔ Institutional AI:

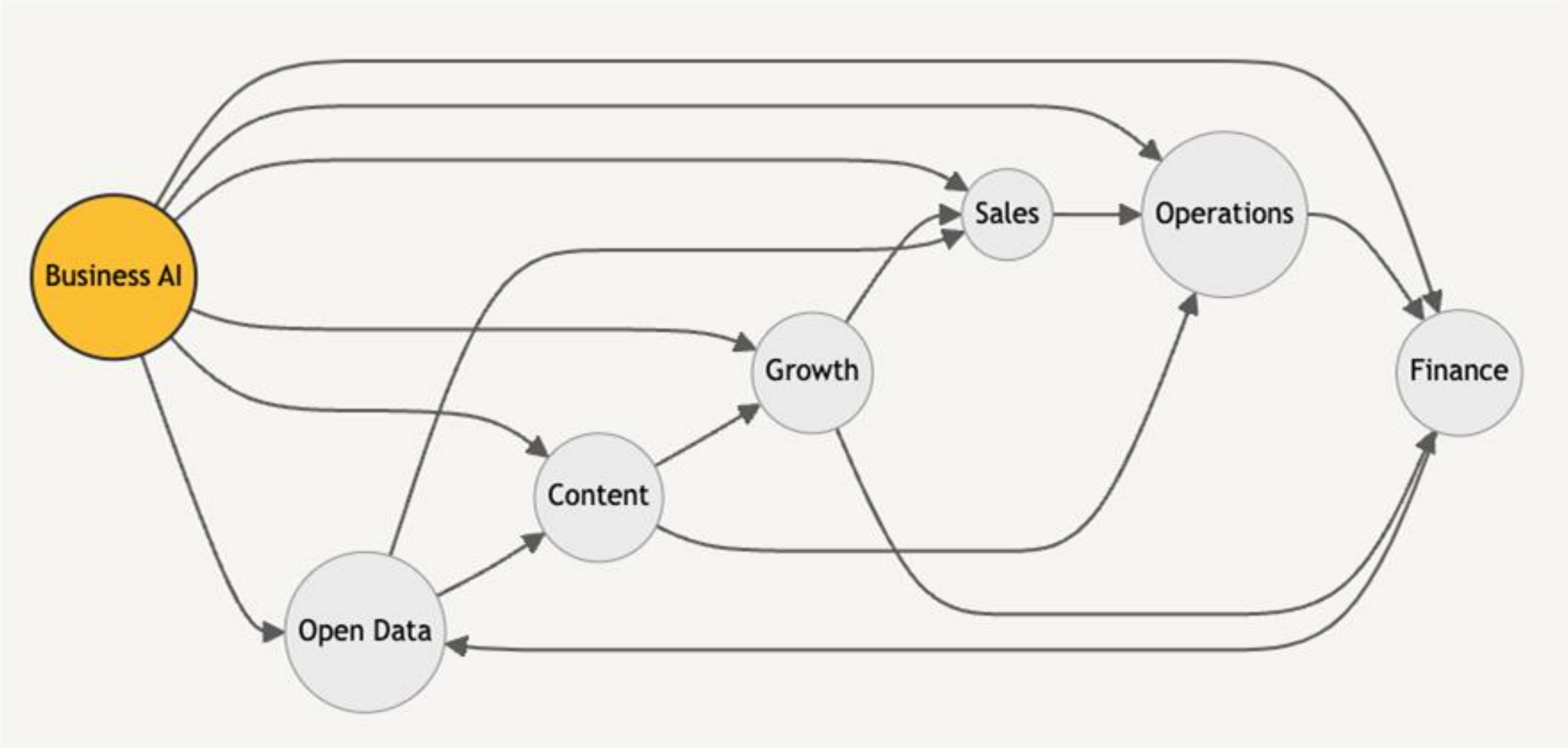
- Educational data and learning progress (from Personal AI) can be shared with institutions to tailor educational programs and resources.
- Institutional research and findings (from Institutional AI) can provide individuals with cutting-edge knowledge and personalized learning opportunities.

Institutional AI ↔ Business AI:

- Academic research and innovation (from Institutional AI) can drive new business opportunities and technological advancements (in Business AI).
- Business trends and operational data (from Business AI) can help institutions align their curricula and research focus with industry needs.

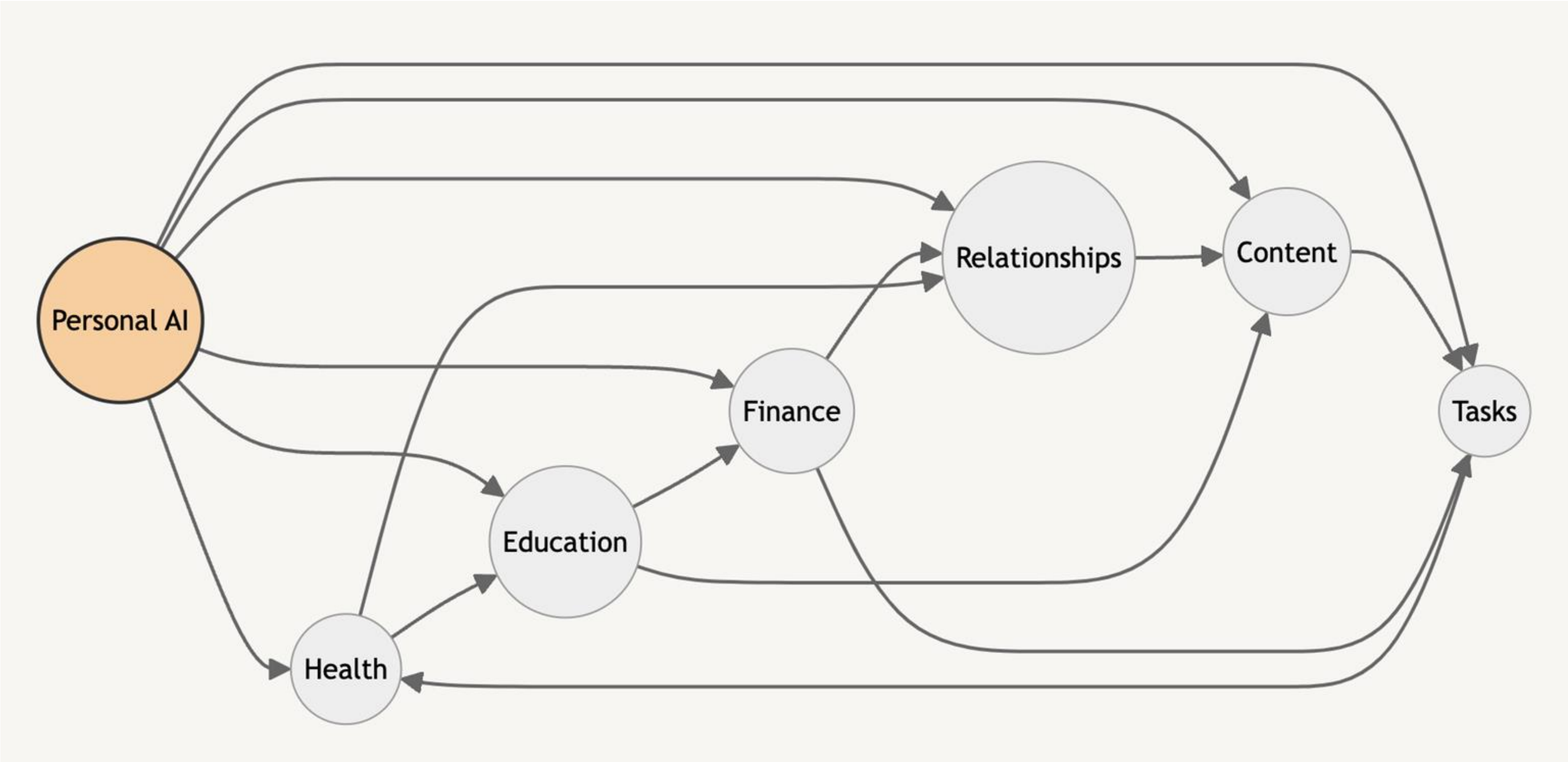
ABI Business AI Flywheel

Open Data enriches market understanding. Standardized Content ensures consistent communication. Growth data unifies audience interactions. Sales data enhances relationship management. Operations support sales and marketing. Finance provides a holistic view for informed decisions. This cycle drives business growth, secures cash flow and drive innovation.



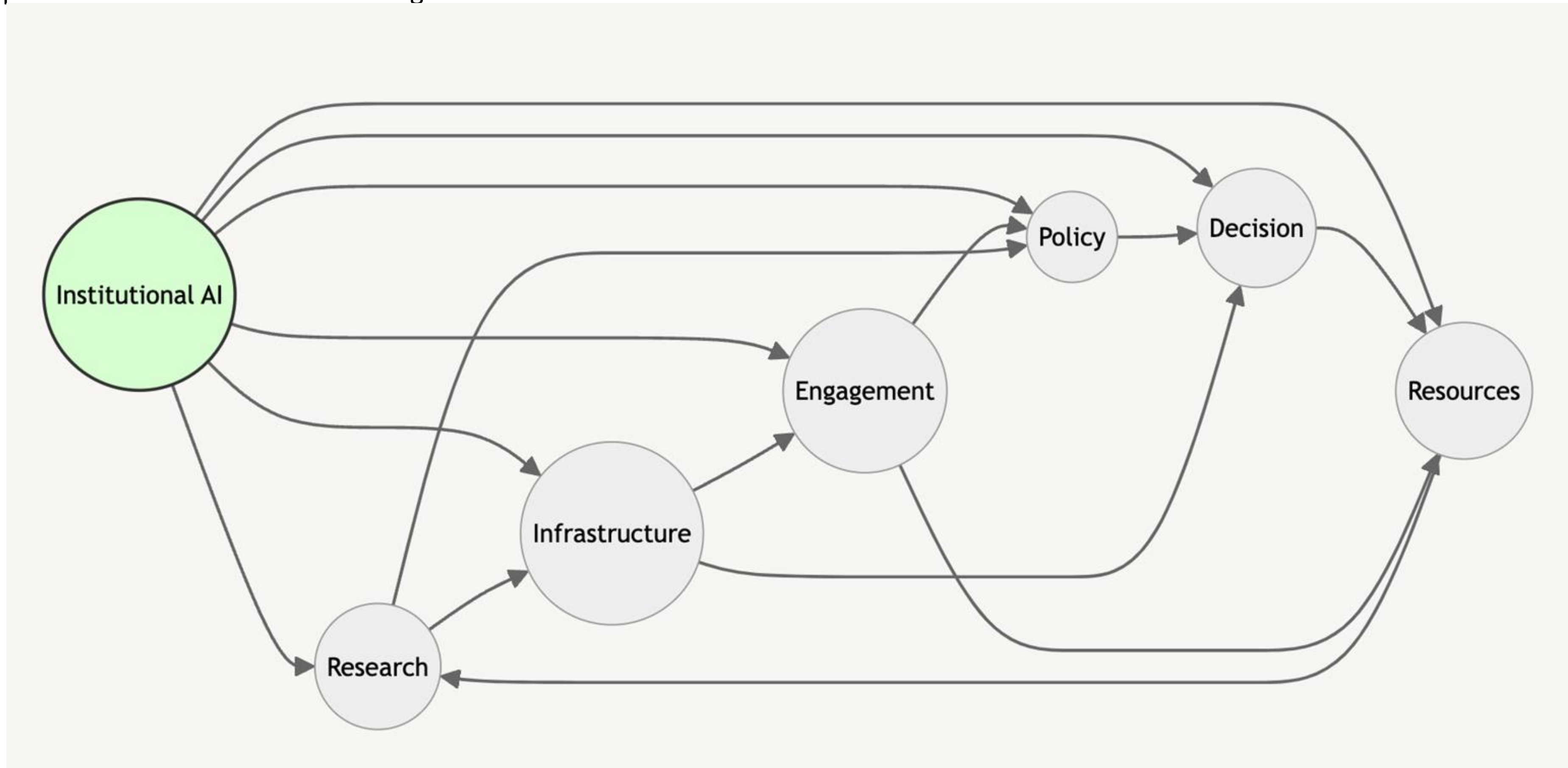
Hypothetical Personal AI Flywheel

In an Personal AI system, Health and Wellness data shapes personalized learning plans. Educational progress influences financial management, leading to financial stability that boosts social interactions. These social interactions refine content recommendations, while productivity tools optimize daily schedules. This interconnected flywheel ensures each aspect of your life enhances the other.



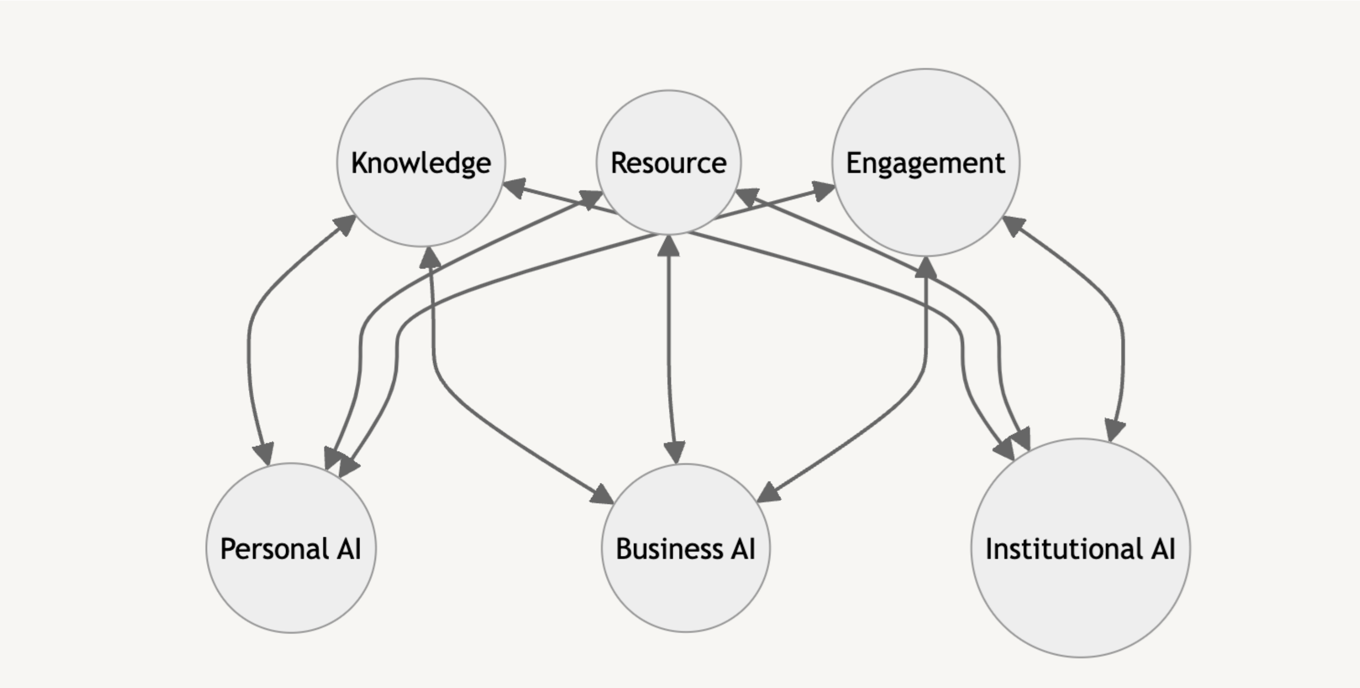
Hypothetical Institutional AI Flywheel

In an Institutional AI system, research and innovation improve the way infrastructure are implemented, boosting engagement by citizens, businesses, and government agencies. Feedback data informs policy updates and ensures compliance with regulations. This compliance data drives decision-making and optimizes resource management, supporting further research and innovation, and enhancing overall institutional performance and citizen well-being.



The Future of Semantic AI: The Case for HyperGraphs

Adopting HyperGraphs, enriched by semantic web technologies (schema.org, LOV, W3C), can create a robust, interconnected ecosystem that drives innovation, efficiency, and personalization, with transparency and accountability .



- **Knowledge and Research Engine**:
  - **Shared across all three domains** to facilitate the exchange of knowledge, research findings, and innovation.
  - **Examples**: Businesses can fund and collaborate with academic institutions; individuals can access the latest research for personal growth.
- **Resource Optimization Engine**:
  - **Manages and optimizes resources** across institutions and businesses, ensuring efficient use of funding, personnel, and facilities.
  - **Examples**: Educational institutions can better allocate resources based on business and personal data insights.
- **User Experience and Engagement Engine**:
  - **Enhances user experience** through personalized recommendations and insights based on aggregated data from businesses and institutions.
  - **Examples**: Businesses enhance customer experience; institutions tailor educational programs.

Key Takeaways

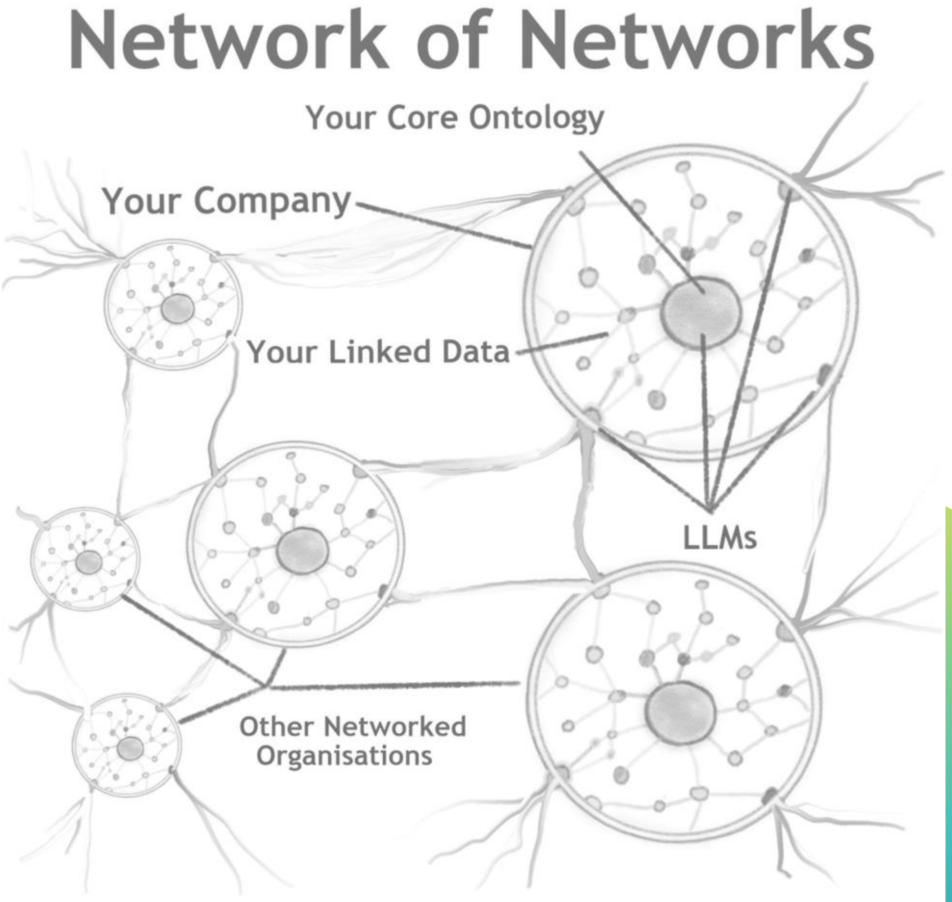
1. Trust in AI is crucial for addressing high-profile errors, bias, privacy concerns, and economic impacts.

2. Semantic interoperability ensures accurate and reliable AI communication by providing a common language and clear data context.

3. Knowledge graphs are essential for providing structure, enhancing transparency, and enabling dynamic learning in AI systems.

4. By structuring our data and building logical bridges between domains, we create unique flywheels that ensure authenticity and independence while creating efficient, interconnected AI ecosystems.

5. The future of AI lies in integrating Personal AI, Business AI, and Institutional AI to drive personalized, business, and societal advancement in a "Network of Networks".



Networks of Networks, by Tony Seal

Thank you

interoperable europe

Discoverability and reusability of AI models

Enhancing the discoverability and reusability of AI models with MLDCAT-AP

Agenda

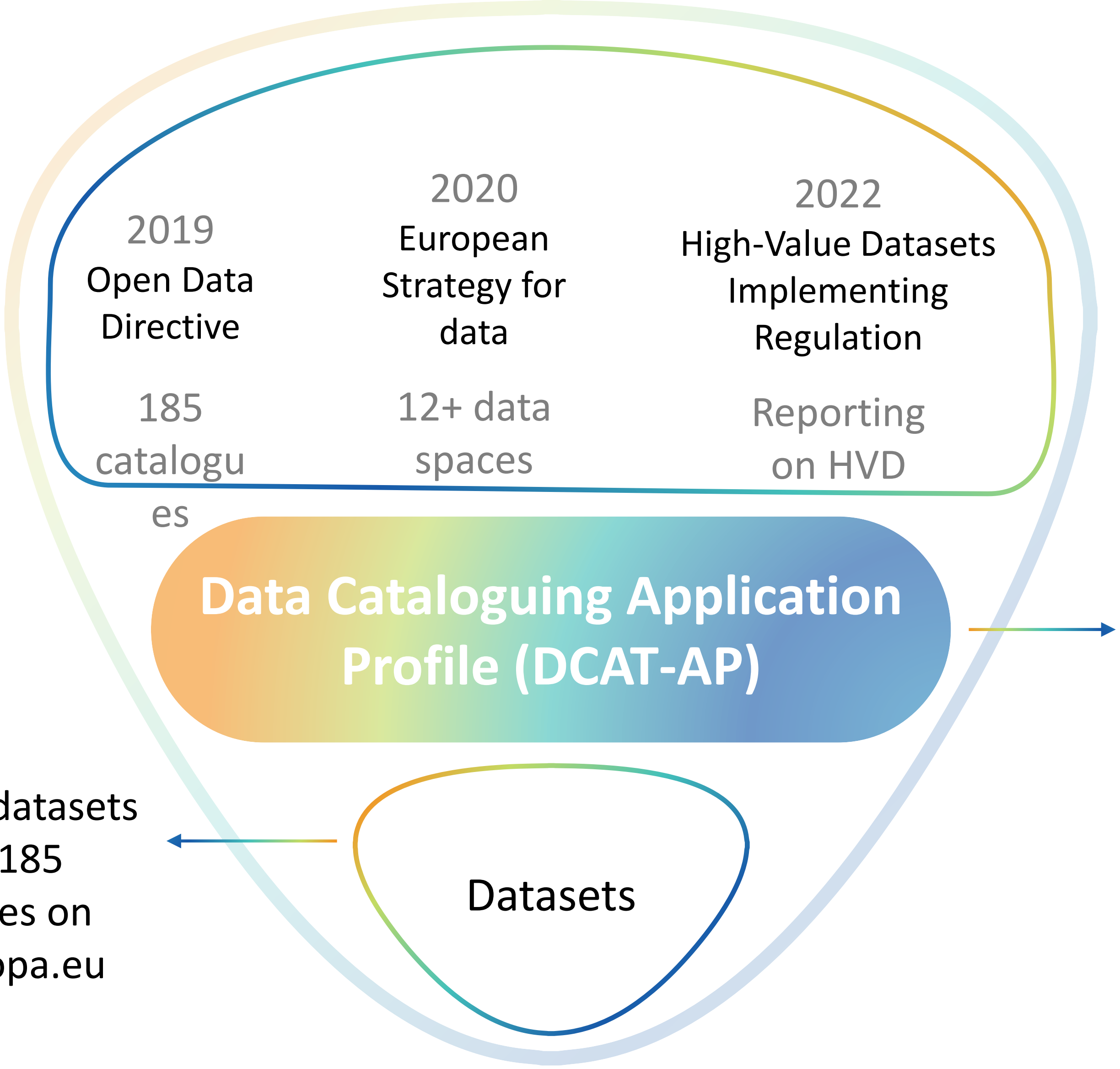
01 The need for MLDCAT-AP

02 MLDCAT-AP as a solution

03 Call for action

The need for MLDCAT-AP

Dataset cataloguing and DCAT-AP



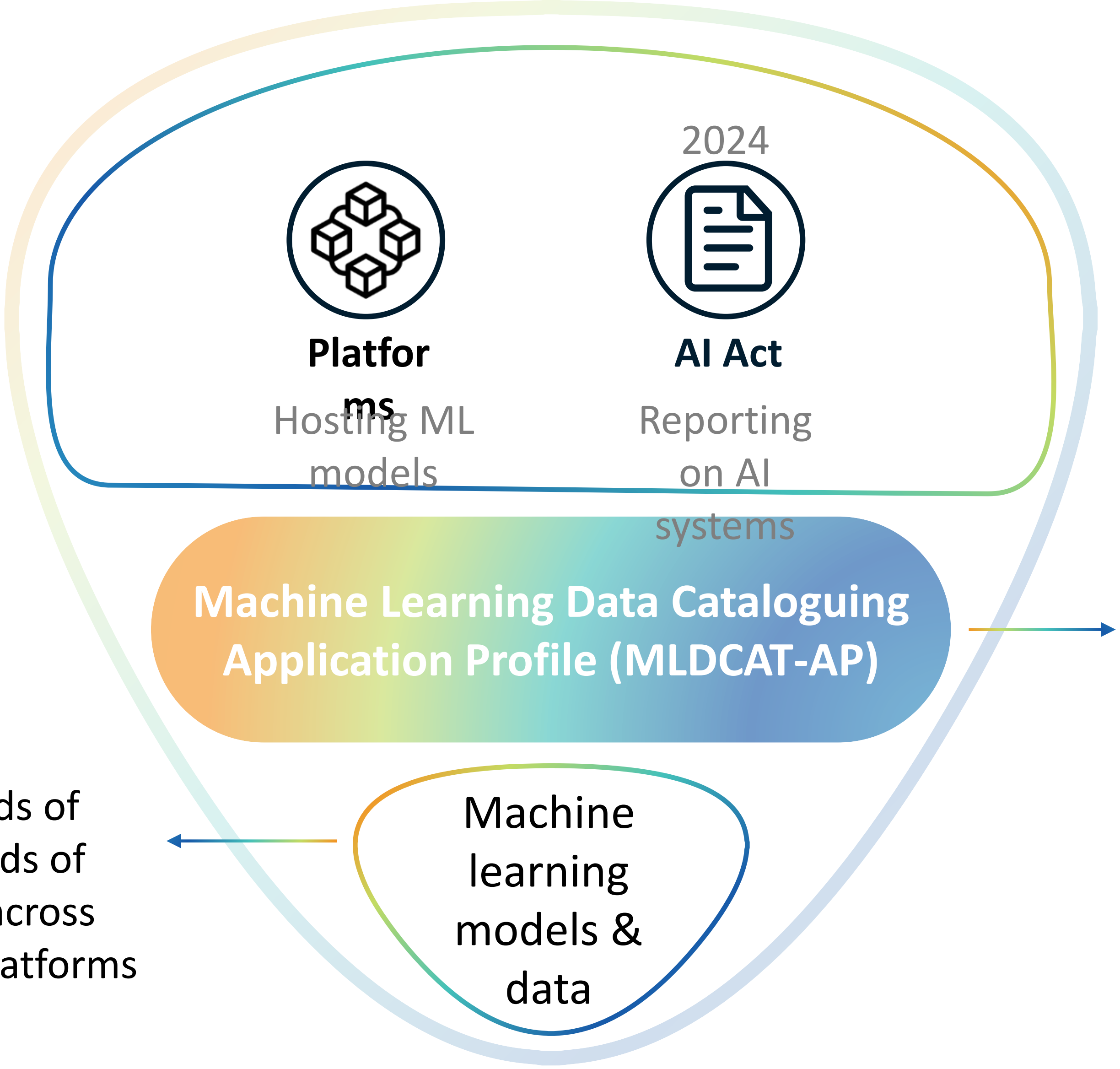
1.745.527 datasets across 185 catalogues on data.europa.eu

Enabler of cataloguing solutions through:

- standardised description of datasets and catalogues
- effective sharing of datasets and catalogues across portals

Resulting in improved findability and reusability of public sector data across Europe.

MLDCAT-AP in the machine learning environment



Hundreds of thousands of models across different platforms

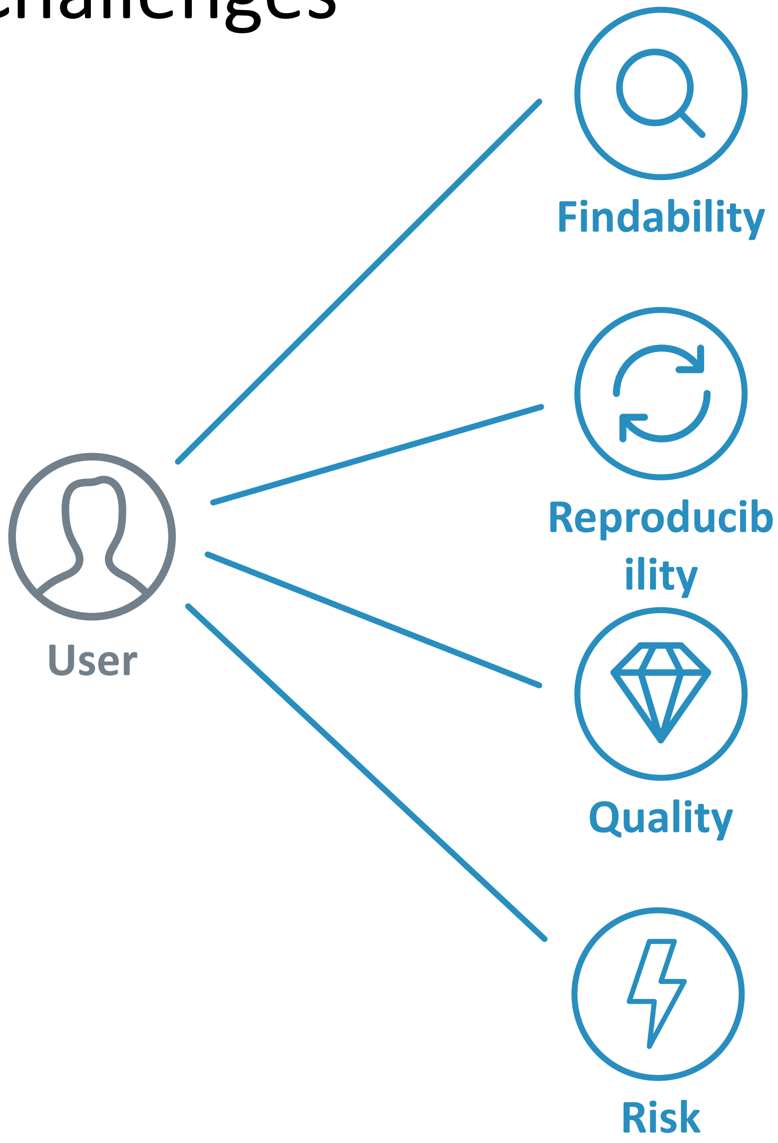
MLDCAT-AP harmonises exchange of data across platforms.

MLDCAT-AP standardises description of ML models and datasets, improving:

- findability
- reusability
- reproducibility
- transparency

while incorporating requirements of the AI act.

Challenges



Model selection difficult because of:

- Poor description of metadata
- Non-standardised values of metadata

Reproduction of experiments proves hard without proper documentation on:

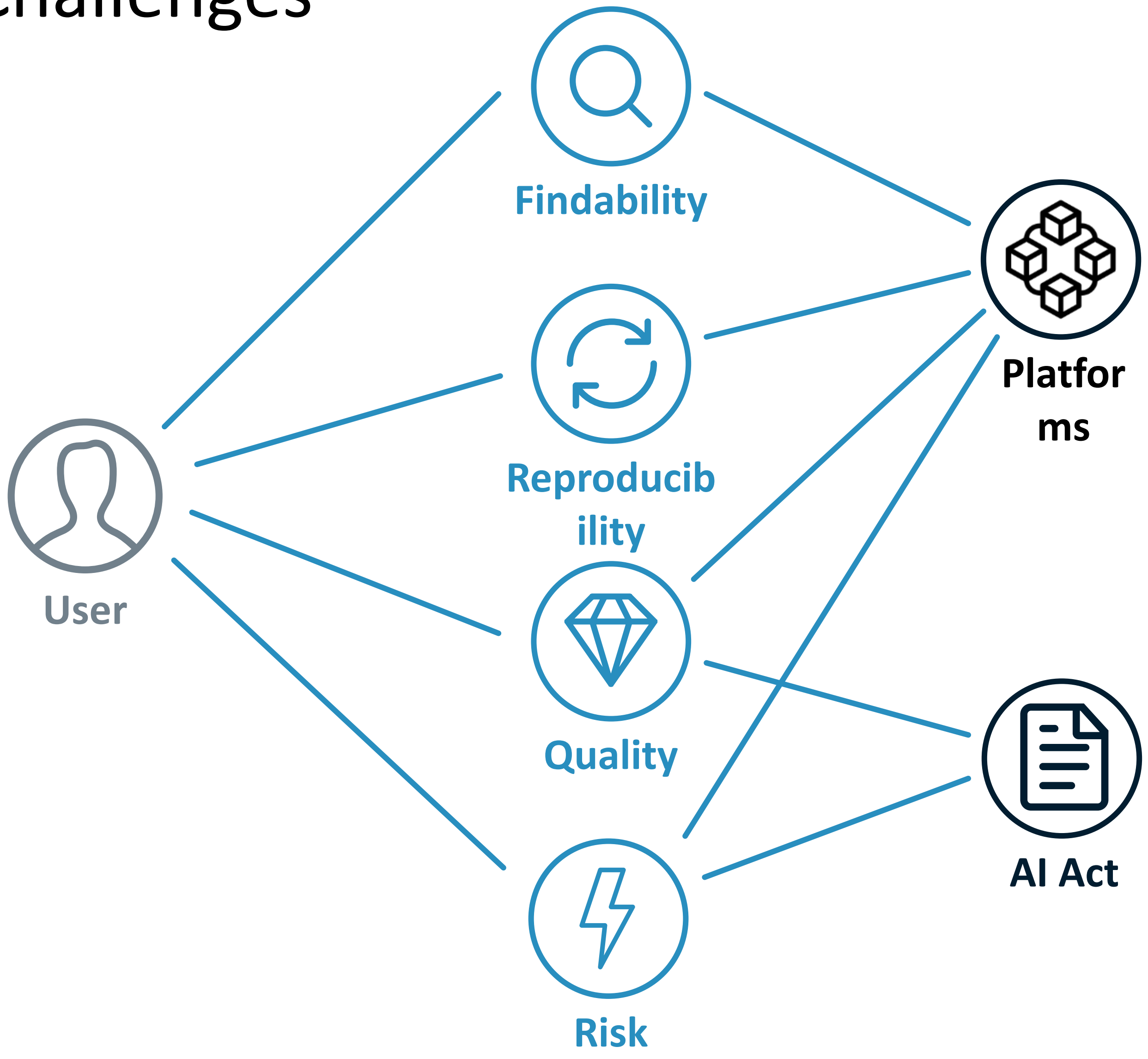
- data splits
- preprocessing
- hyperparameters
- license information

Determining the quality of both data used and results of the model. (Cf. AI act)

Identifying the risks when reusing a machine learning model can be difficult:

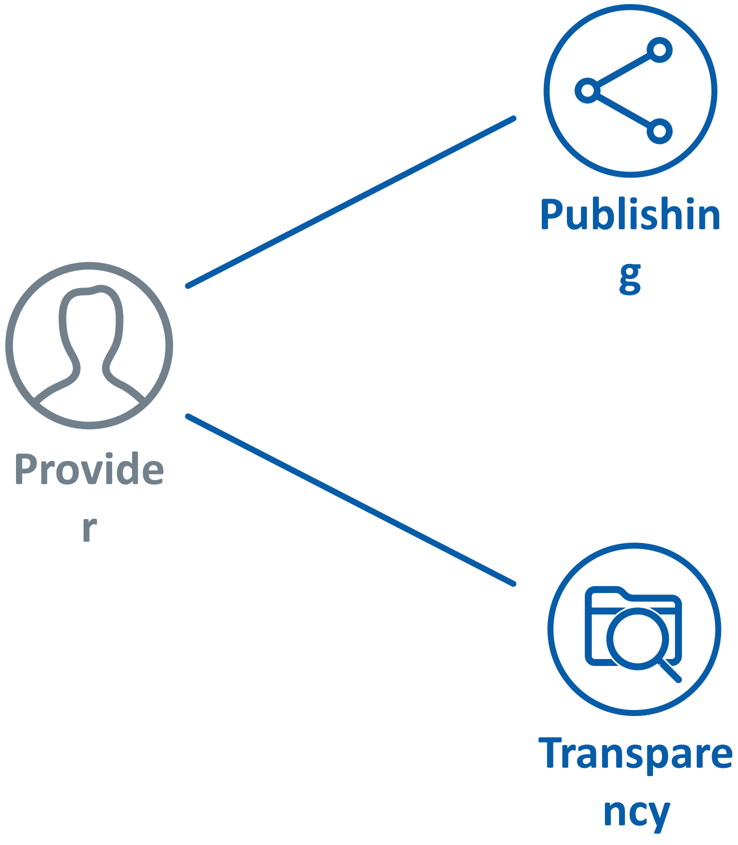
- bias
- limitations

Challenges



European Commission

Challenges

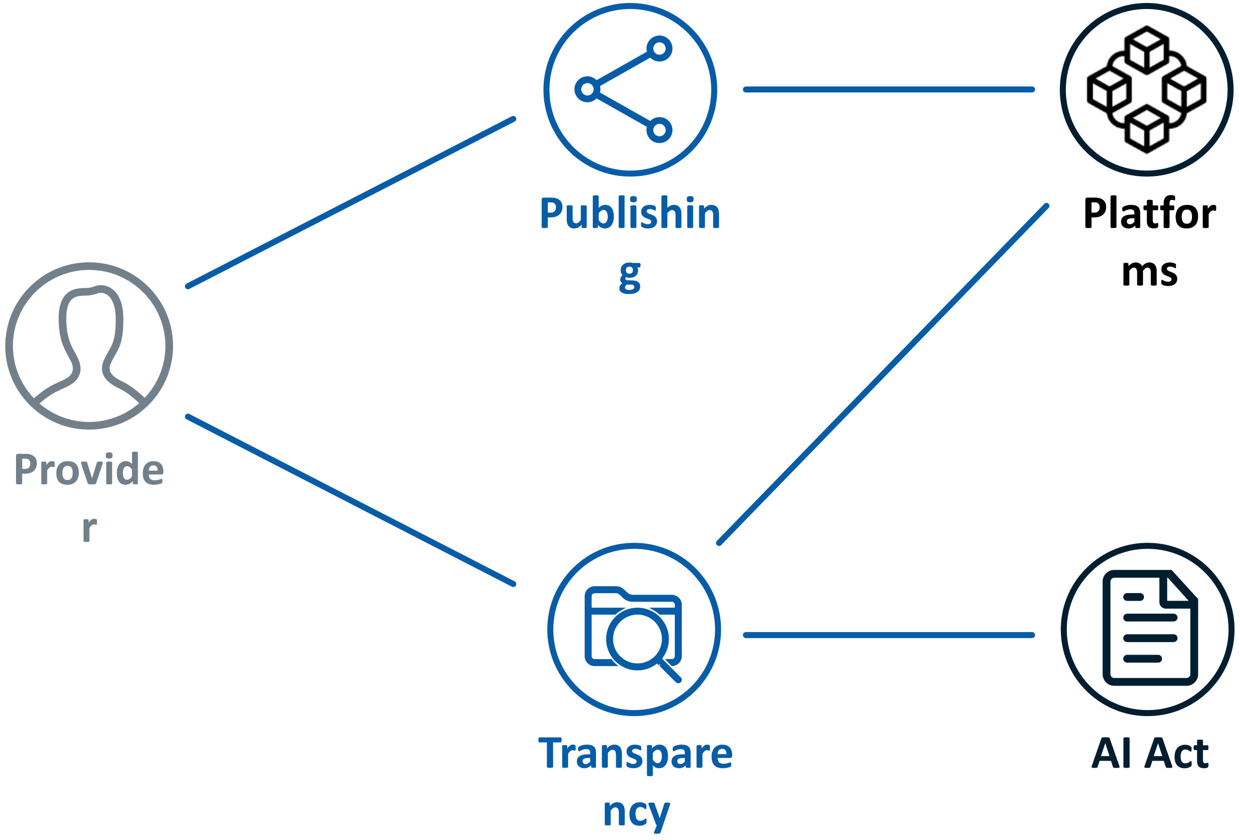


Publication across multiple platforms with different data structures

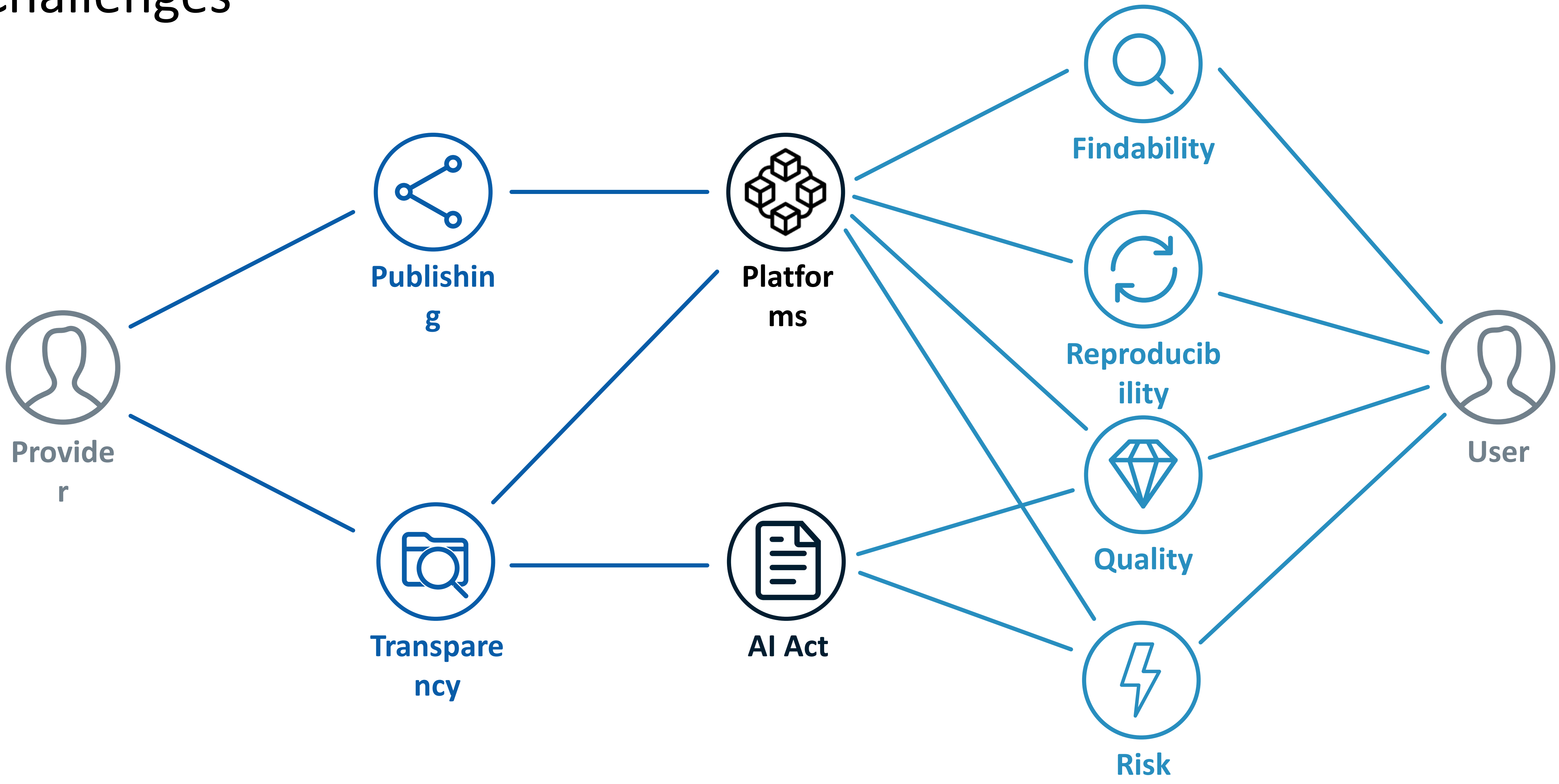
Under the AI act a provider of a (high risk) AI systems will have to disclose information on:

- the algorithm used for training the machine learning model
- the data used to train the model
- instructions for using the model
- etc.

Challenges



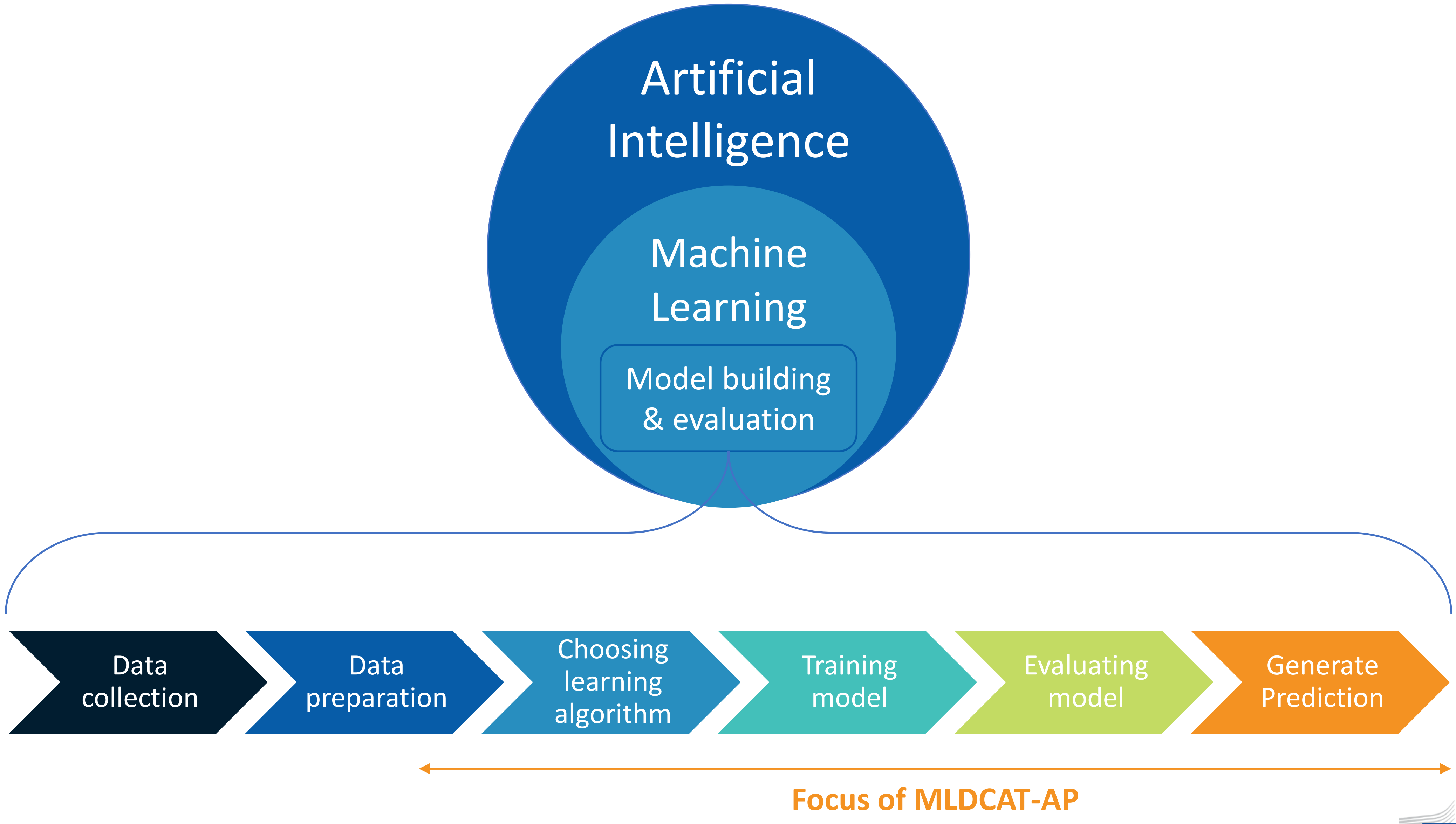
Challenges



Publishin
g
Platfor
ms
Findability
Reproducib
ility
Provide
r
User
Quality
Transpare
ncy
AI Act
Risk

MLDCAT-AP as a solution

Domain of MLDCAT-AP



MLDCAT-AP 2.0.0

Data preparation

Concepts reused from **DCAT-AP**

Classes describing the **quality of the data**

Choosing learning algorithm

Algorithm class (cf. AI Act)

Training model

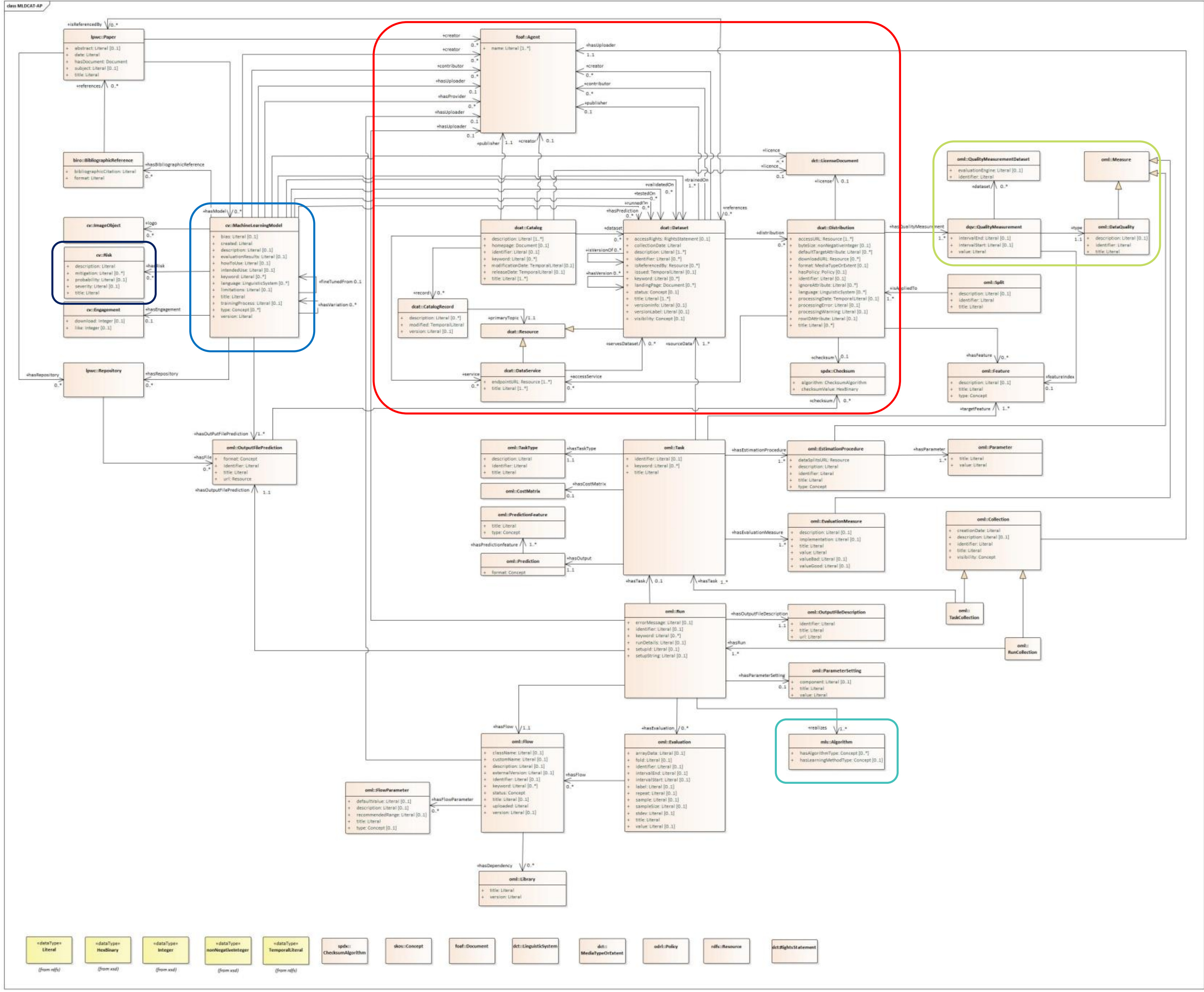
Machine Learning Model class (cf. AI Act)

Evaluating model

Risks of the model (cf. AI Act)

Generate Prediction

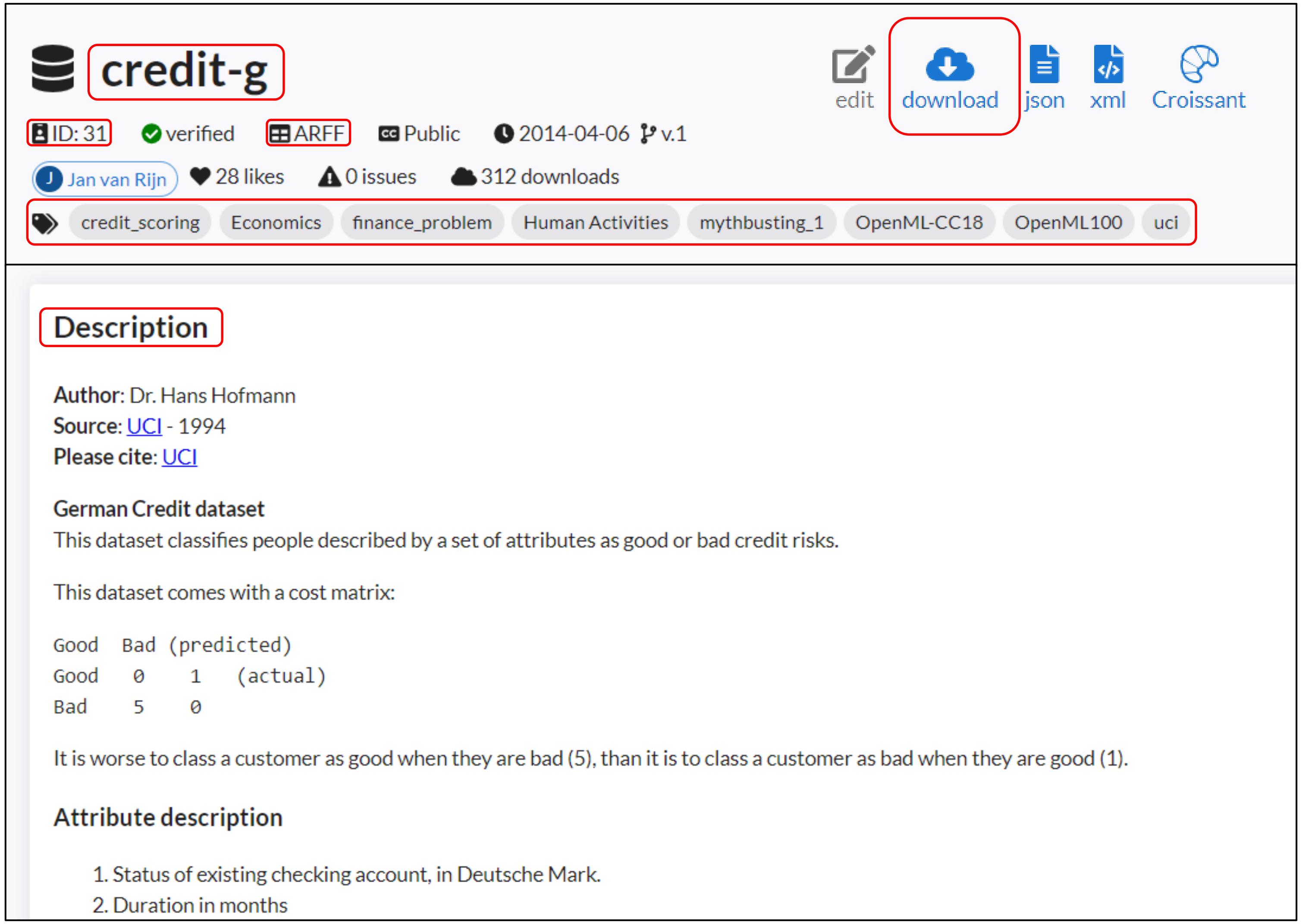
Dataset class as output of a **Machine Learning Model**



MLDCAT-AP – Reusing DCAT-AP concepts

Transparency

Findability



credit-g

edit download json xml Croissant

ID: 31 verified ARFF Public 2014-04-06 v.1

Jan van Rijn 28 likes 0 issues 312 downloads

credit_scoring Economics finance_problem Human Activities mythbusting_1 OpenML-CC18 OpenML100 uci

Description

Author: Dr. Hans Hofmann
Source: UCI - 1994
Please cite: UCI

German Credit dataset
This dataset classifies people described by a set of attributes as good or bad credit risks.

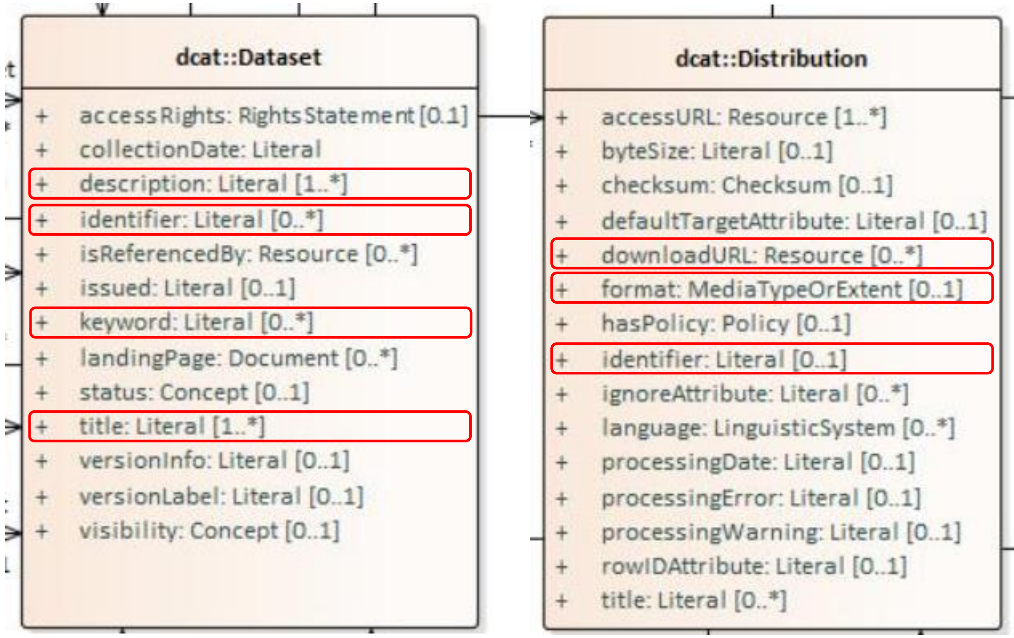
This dataset comes with a cost matrix:

```
Good  Bad (predicted)
Good   0    1   (actual)
Bad    5    0
```

It is worse to class a customer as good when they are bad (5), than it is to class a customer as bad when they are good (1).

Attribute description

1. Status of existing checking account, in Deutsche Mark.
2. Duration in months

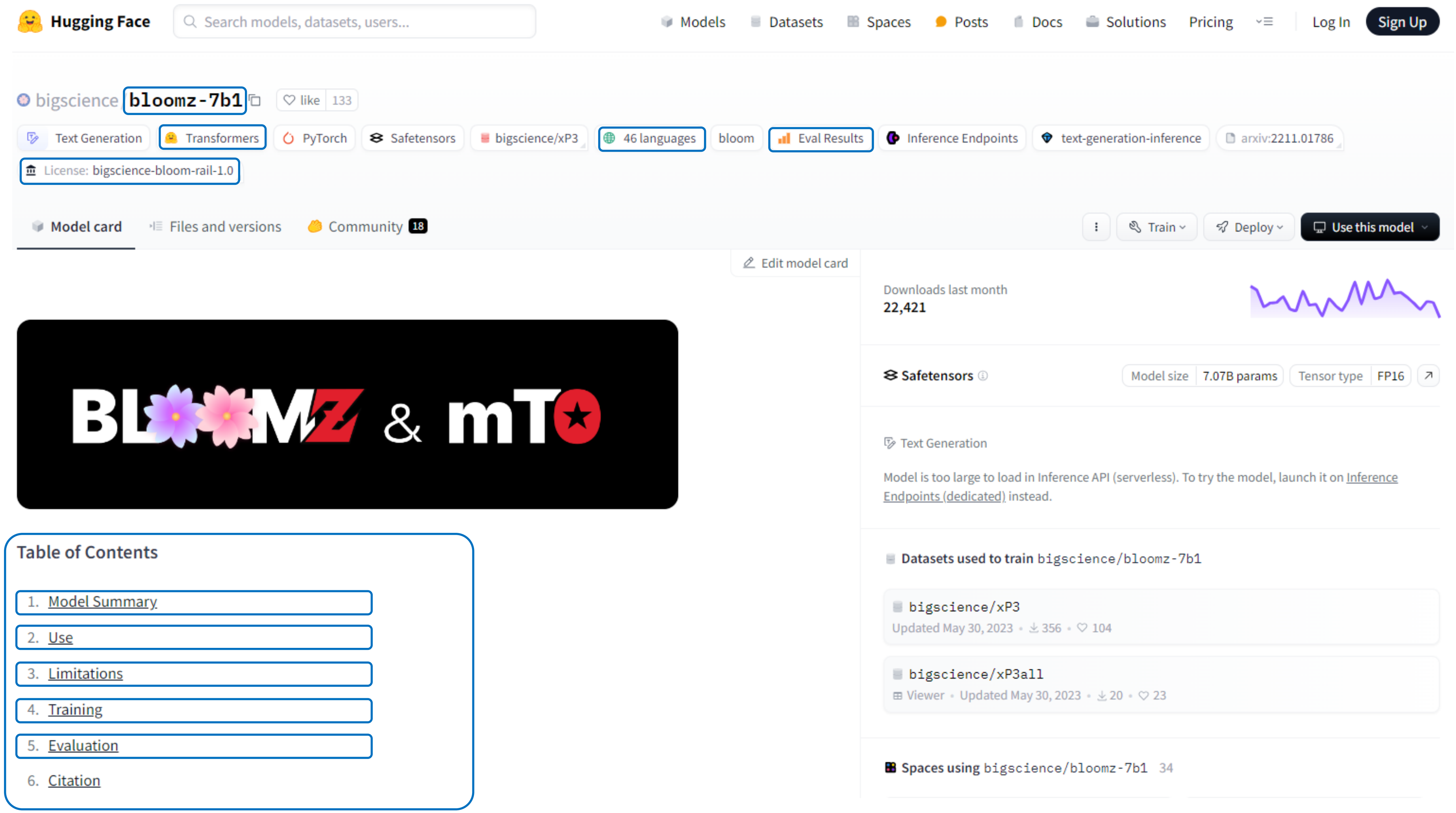


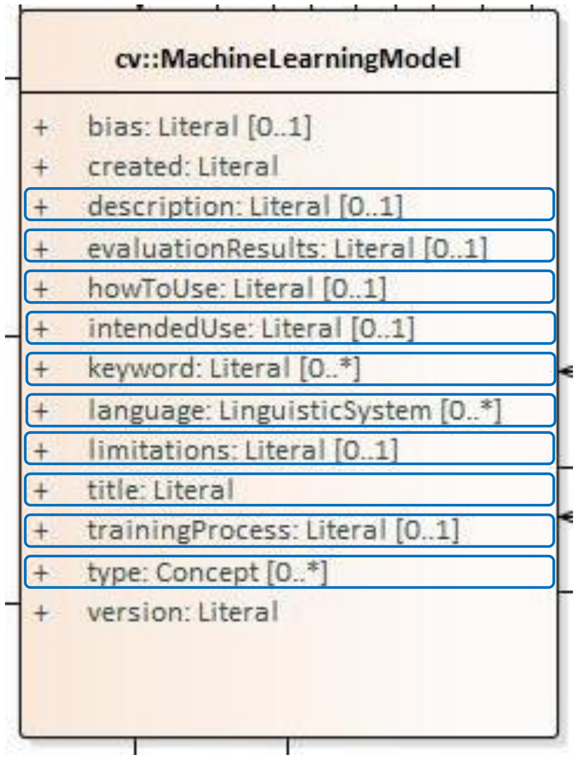
MLDCAT-AP – MachineLearningModel

Quality

Transparency

Findability



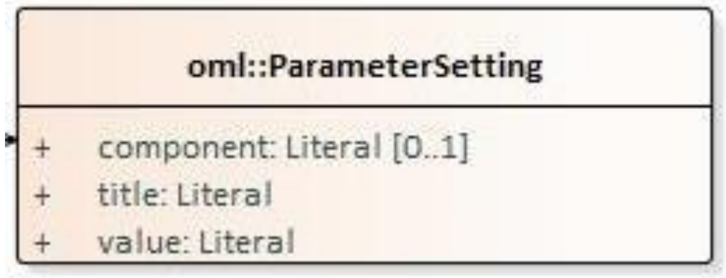


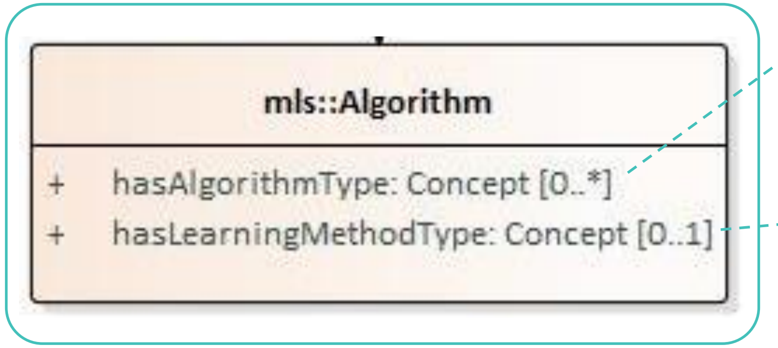
MLDCAT-AP – Algorithm

Reproducibility **Transparency**

Reproducibility is essential for **validation**, **innovation** and **research.**

Transparency, in addition, is required for **regulatory purposes**.





Aligning with the existing class of MLSO.

Both properties are populated with controlled lists, examples of values that are mentioned in the AI act are:

Machine Learning Algorithm
- Bayesian
- DeepLearningAlgorithm
- ReinforcementLearningAlgorithm

Learning method
- BayesianLearning
- Reinforcement_Learning_Algorithm
- Supervised_Learning_Algorithm

MLDCAT-AP – Quality & Risk

Quality

Risk

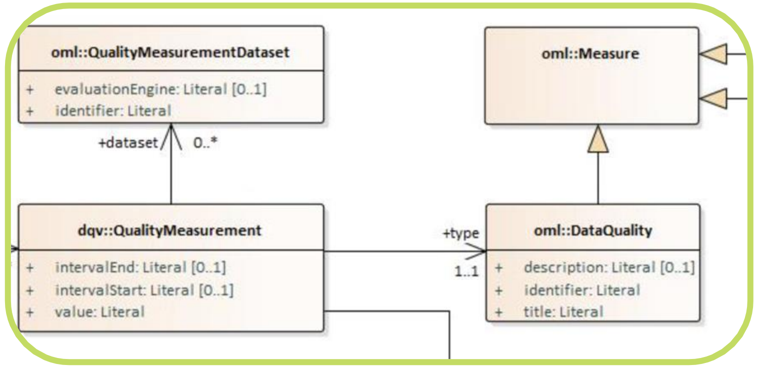
Transpare
ncy

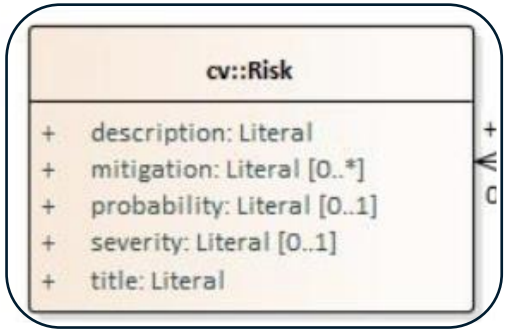
Findability

Quality

AI Act

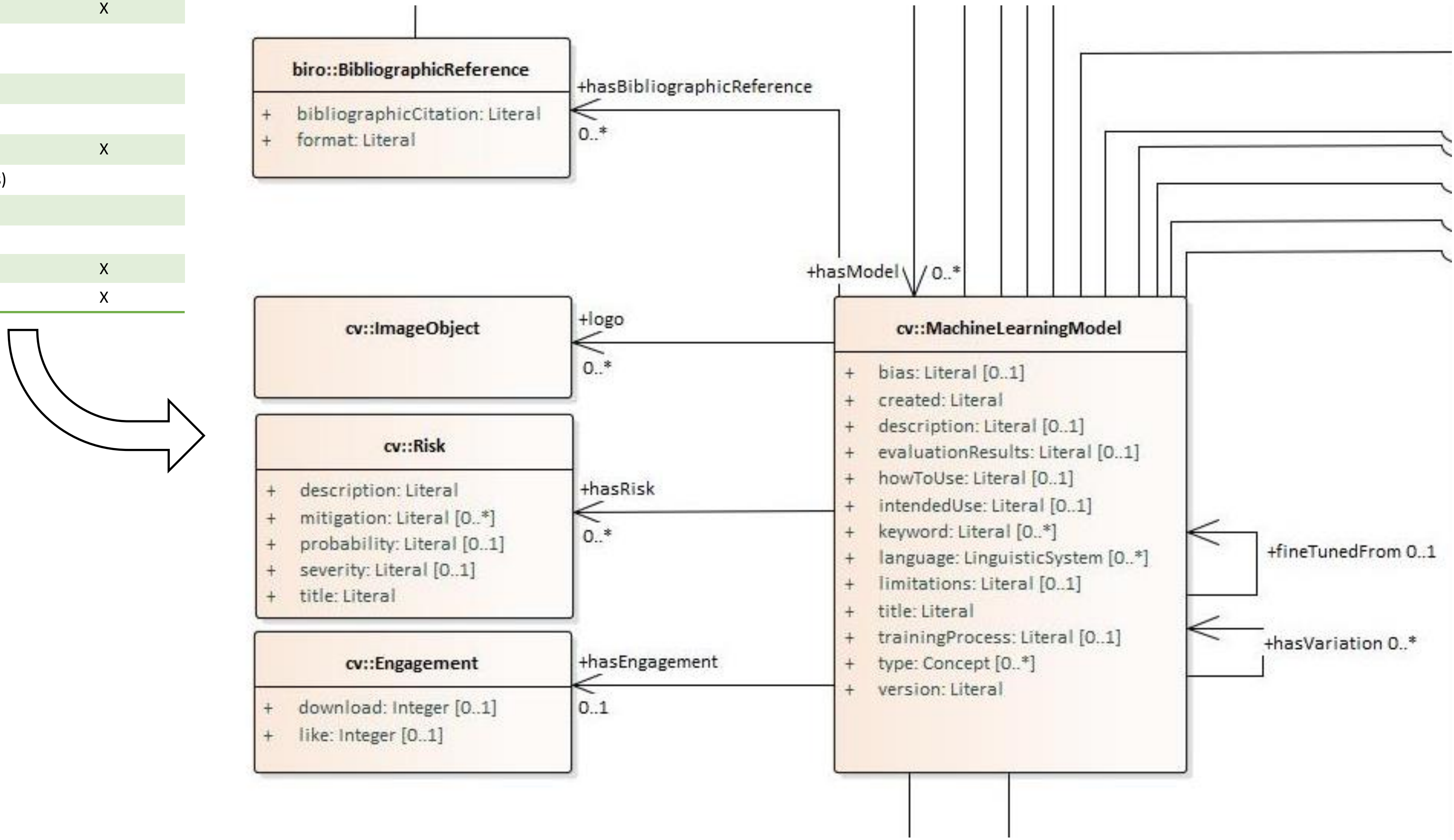
For high-risk AI systems, the requirements of **high quality data**, documentation and traceability, transparency, human oversight, accuracy and robustness, are strictly necessary to mitigate the risks to fundamental rights and safety posed by AI and that are not covered by





MLDCAT-AP – Machine Learning Model

Common Properties	Hugging Face 423K models	Kaggle -TensorFlow 278-2K	Pytorch 55	AzureAI 100+ (in org)
Language	X	X	X	X
References to Paper/Code	X	X	X	
How to use	X	X	X	
Logo	X	X	X	
Files	X	X		X
Likes	X	X	X (GitHub stars)	
Downloads	X	X		
Relation	X (fine tuned from)	X (variation)		
Created date	X			X
Checksum	X			X



Call for action

Call for action

Community building

You are invited to join the MLDCAT-AP community if the topic interest you or can be of use within your organisation by using this link or through helpdesk on the SEMIC Support Centre.

Connecting with platforms

SEMIC is looking for people that can bring us in touch with machine learning platform to facilitate the adoption of MLDCAT-AP, driving interoperability of Machine Learning models and data.


Thank you

interoperable europe

innovation ∞ govtech ∞ community

Stay in touch

(@InteroperableEU) / Twitter

Interoperable Europe - YouTube

Interoperable Europe | LinkedIn

DIGIT-INTEROPERABILITY@ec.europa.eu

https://joinup.ec.europa.eu/collection/interoperable-europe/interoperable-europe

Interoperability of public administrations chatbots

DEP Preliminary study findings – June 2024
Publications Office of the European union

Publications Office of the European Union

Supports EU policies as a **center of excellence for information, data and knowledge management**

Our values

Transparency

We facilitate transparency throughout the policy cycle of the EU institutions to enhance evidence-based decision-making, accountability, civic participation and democracy.

Trustworthiness

We strive to ensure that the content we provide is accurate and reliable so that citizens trust the EU as a provider of information.

Accessibility

We believe access to information is a human right that all citizens should enjoy regardless of language, culture, disability, social status, location, technology or the way they understand information.

Service orientation

We are committed to continuously improving our services to both our institutional stakeholders and EU citizens because we want to contribute to the European project in the best possible way.

Razvan-Petru Radu

Deputy Head of Unit Portal and Web Services

Proposed Agenda

Interoperability study: Current state of development

1. Project background
2. What is interoperability?
3. Current state of interoperability
4. Key considerations for interoperability
5. Viable approaches
6. Regulatory outlook
7. Implementation framework
8. Q&A

AGENDA

1. Project background and importance
2. What is interoperability?
3. Current state of interoperability
4. Key considerations for interoperability
5. Viable approaches
6. Regulatory outlook
7. Implementation framework
8. Q&A

Project background

Who, when, what, why

Why is it important?

Establishing interoperability between organizations focussing on both **EU institutions and member states**

Enhanced **multilingual (integrate language from other chatbot) and transparent access to public data** and services, looking at **latest technologies for virtual assistants**

More **accessible public administrations**

- Seamless, trustable and accurate 24/7 access to public services
- Fully fledged digital services

Readiness to understand and comply with the regulations of the **New Interoperable Europe Act,** which aims to strengthen cross-border interoperability and cooperation in the public sector across the EU.

AGENDA

What is interoperability?

Interoperability is the seamless, effortless ability for users to access, use, and transfer data or functionality across different systems, devices or applications.

Interoperability levels – *European Interoperability Framework (EIF)*

A Interoperability Governance

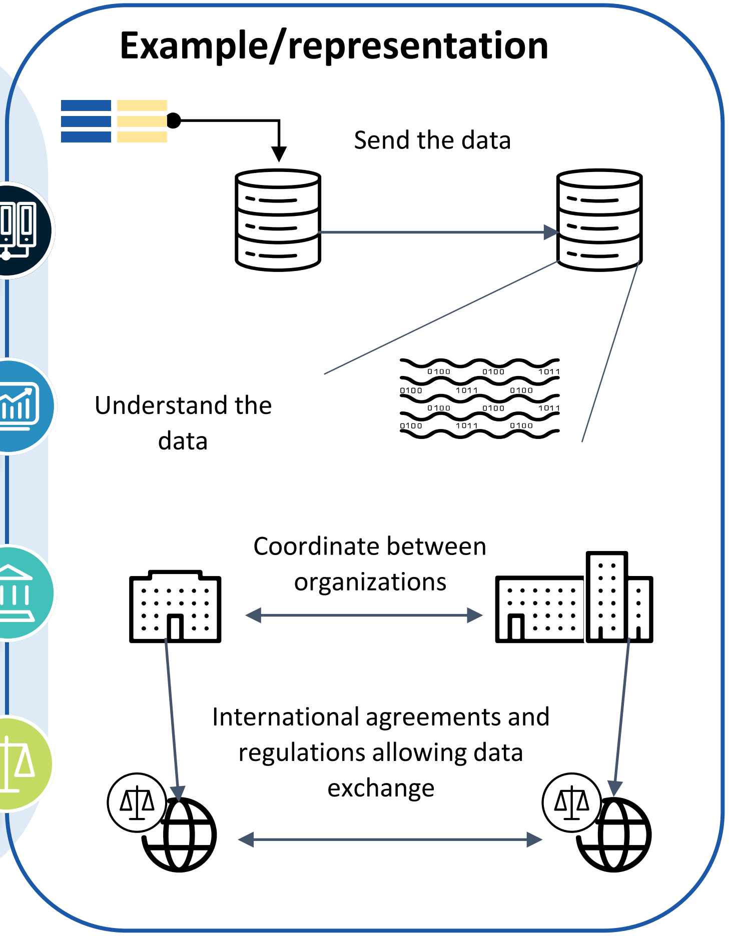
Technical Interoperability – Communication of data between systems, servers, and applications.

Semantic Interoperability – Ability to exchange data but also to interpret and understand the shared data

Organizational interoperability – How well different organizations can work together seamlessly

Legal Interoperability – Ensuring that organizations under different legal frameworks can work together.

B Integrated Public Service Governance

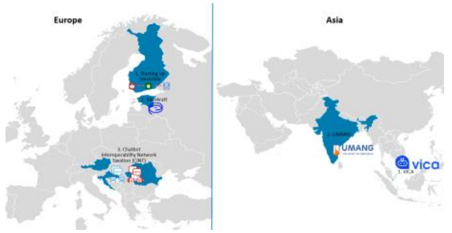


AGENDA

1. Project background and importance
2. What is interoperability?
3. Current state of interoperability
4. Key considerations for interoperability
5. Viable approaches
6. Regulatory outlook
7. Implementation framework
8. Q&A

Current state of interoperability (1/4)

Availability on the market – Existing chatbot interoperability in administrations



Examples of chatbot interoperability implemented across the globe

EU Public Institutions/ Member States with interoperable chatbots

1. **Pilot** – Kamu ⟷ VeroBot ⟷ PatRek – Starting up smoothly: Finland
2. **Not in production** – Burokratt: Estonia
3. **Under development** – CINT-B, CINT-I – Fiscalis: Croatia, Austria, Romania - Chatbot Interoperability Network Taxation (CINT)

Interoperable chatbots or related projects outside of the EU

1. **Production** – vica – Singapore – AlphabotSG: - Virtual Intelligent Chat Assistant (VICA)
2. **Production** – UMANG THE SPIRIT OF NEW INDIA – India - UMANG (Unified Mobile Application for New-age Governance)
3. **Open source** – Open Voice Network – Interoperability of voice-enabled systems. *Findings are globally applicable*

Non-EU chatbots are interoperable in terms of services rather than chatbots. They have incorporated many services into one chatbot/platform by merging the knowledge base into one space.

Reasons for limited interoperability use cases:

- Lack of standardization
- Technology complexity
- Security & Privacy concerns
- Lack of use cases

Current state of interoperability (2/4)

Benefits for institutions implementing interoperable chatbots[1]

Benefits

Improved service delivery
Interoperability facilitates seamless collaboration between chatbots of public services, improving efficiency and accessibility for users

Better Insights and Analytics
Interoperability enables platform providers to gather insights and analytics from multiple chatbot interactions, supporting data-driven decision-making and platform enhancements

Larger knowledge base
Interoperable chatbots broaden organizations' reach by pooling information from various sources, offering an expansive knowledge base for both employees and customers

Collaborative network
Interoperable chatbots promote a collaborative network among organizations with similar challenges, facilitating shared insights and strategies for collective advancement

Cost-effectiveness
Having interoperable chatbots can reduce costs as public institutions can share technologies and leverage shared services

Increased user adoption
Interoperability encourages more businesses and developers to adopt the platform, knowing that their chatbots can work together with other systems, boosting the platform's user base

[1]Non-exhaustive list

Current state of interoperability (3/4)

Challenges to overcome for institutions implementing interoperable chatbots[1]

Challenges

Infrastructural investment
To establish and maintain interoperability technical resources and infrastructural investment required

Governance models
Complexity in coordinating and governing chatbot interoperability initiatives among different organizations or stakeholders

Compatibility and information asymmetry
Incompatibilities and development challenges in diverse systems can hinder interoperability, potentially causing incorrect responses due to varying platforms, protocols, and data formats

Language barriers
Dialogue translation of chatbots with different languages, especially low-resource languages can limit interoperability

Question redirection
A development challenge for interoperable chatbot infrastructure is effectively managing question redirection and corresponding answer display

Conversation flow & handoff
Managing conversation flow, context retentivity, and user experience consistency during cross-bot interactions presents technical and design challenges

Error handling & recovery
Handling errors during cross-bot interactions is challenging due to diverse chatbot error strategies and recovery capabilities

Consent and session management
Obtaining user consent for data sharing across multiple chatbots, while managing session lifecycles for personalized cross-bot interactions, presents a significant challenge

[1]Non-exhaustive list

AGENDA

Current state of interoperability (4/4)

Risks to avoid for institutions implementing interoperable chatbots[1]

Risks

Inadequate interoperability standards
Absence of universally accepted interoperability standards and guidelines creates complexity in uniform data handling and communication among chatbots

Performance and scalability issues
Adding additional bots can result in response delays and potential system crashes owing to poor interoperability

Data breaches
Cross-system data exchanges may compromise sensitive data security due to weak integration, leading to potential unauthorized access or malicious activities

Misalignment of responses
Uncoordinated chatbots can produce contradictory responses. Lack of context may lead to repetitive or incomplete information, affecting the user experience

User experience inconsistencies
Poor interoperability may cause inconsistent branding and user experiences across platforms, leading to trust issues and limited chatbot capabilities

Dependence on vendors/ service providers
Chatbot interoperability could increase vendor dependence, limiting the ability to switch vendors without incurring cost or putting network at risk when parties change vendors.

Dependency on other bots in network
A failure, disruption or discontinuation in one bot can affect the functionality of interdependent systems, hindering the entire chain

[1]Non-exhaustive list

AGENDA

Key considerations for interoperability (1/2)

Interoperability requirements considerations

Chatbot types

- Different types of chatbots: rule-based (based on predefined rules on which they have been initially programmed) & AI/ML powered (use ML and AI to provide more accurate and personalized chat experiences)
- Depending on the chatbot type some of the following considerations need to be applied differently

Dataset

- Integration compatibility
- Data quality & variety

Conversational flow

- Best practices
- Direction of conversation flows between chatbots (one vs. multi-directional)
- Voice considerations

Licensing & contract

- Pre-contractual (NDA - Non-Disclosure Agreements)
- Onboarding (Project contract, SLA - Service Level Agreement)
- Data privacy (Data processing agreement, Data protection & privacy, Privacy notice)
- Other terms (IPRs - Intellectual Property Rights, Copyright)

Language

- Low resource languages
- Approaches and techniques in language processing and translation: Indirect & Direct

Security

- Communication with desired chatbot
- Protecting chatbots from abuse

AGENDA

Key considerations for interoperability (2/2)

Impact of UX/UI principles on interoperability

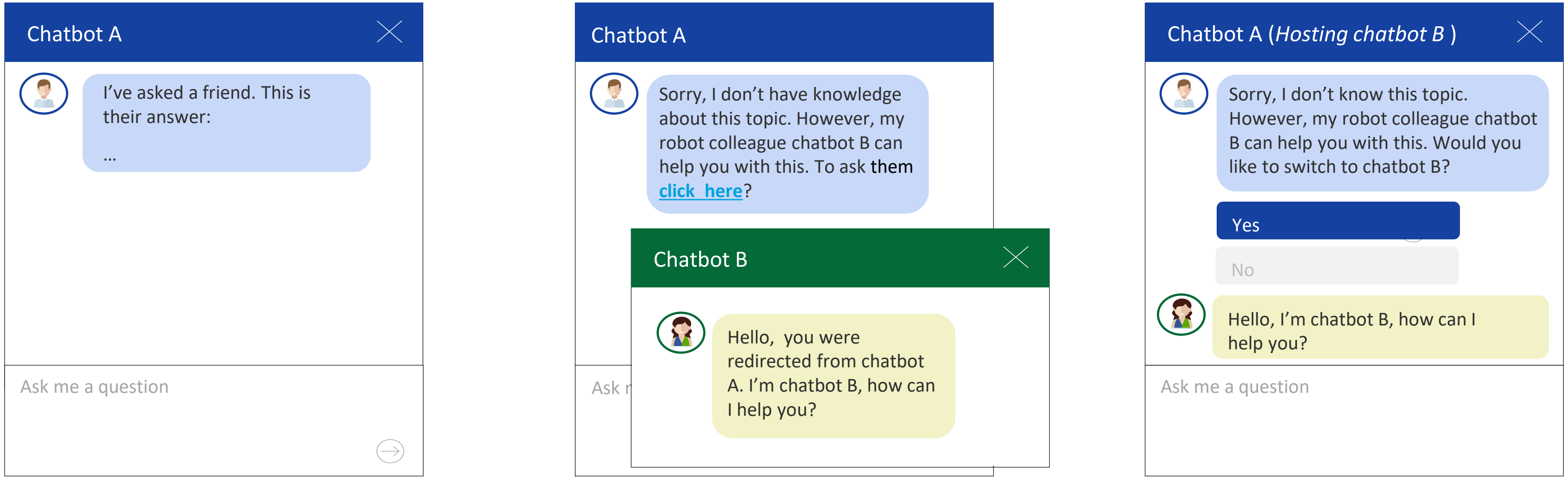
Important terminology
Host bot: Initial bot interface replying to user (Chatbot A)
Contributing bot: Bot transferring content to host bot (Chatbot B)

1. Conversational flow elements - redirections

Option 1: Host bot provides answer from contributing bot

Option 2: Host bot redirects user to contributing bot's interface *(a. sharing content, b. Not sharing content)*

Option 3: Host bot hosts contributing bot in same interface



2. Conversational style (Design & interaction features)

Chatbot design
- Number of characters in text
- Feedback buttons
- Media support

Chatbot personality
- Use of emojis
- Restricted words and topics
- Conversational style (e.g., legal vs. casual)

User interaction
- Transfer transparency
- Expected interaction speed
- Response time & latency

AGENDA

Viable approaches – Interoperability mechanisms (CKD/ IL)

Sharing domains of knowledge: Central Knowledge Directory (CKD) vs. Interoperability layer (IL)

There are two mechanisms to trigger interoperability:

Central Knowledge Directory (CKD)

- CKD is a **comprehensive catalog of available chatbots**, showing their expertise areas and query handling skills. It enables chatbots to understand each other's knowledge domains within the network.
- CKD **stores metadata from different chatbots,** providing information about their **capacity, purpose, and functionality, fostering interaction/transfer** between them.

Interoperability layer (IL)

- The Interoperability layer is **an intermediate layer** between the **host bot and the contributor bots.**
- Unlike the centralized directory, it **does not store the metadata** of all chatbots, however if knows all bots in the network and can refer to them
- It works through a **referral mechanism, taking thresholds of** a question's likelihood to be answered by each bot and choosing who to select from the network

Examples… interoperability transfer mechanisms in action:

*Chatbot A which is a tax chatbot is asked on how to file an online complaint for theft. Based on this user input **the CKD** will see which other bot is best to connect to – chatbot B an immigration chatbot or chatbot C a Police chatbot (all this information is stored in the CKD). The CKD transfers the user to chatbot C, the Police chatbot that can deal with topics on theft.*

*Chatbot A (host) forwards the question from the user to chatbot B, C and D **through the IL**. In return, they send the confidence score for this query (indicating how likely each can answer the question). The IL calculates the highest score (if higher than the minimum threshold set) and makes sure that Chatbot A is connected to that bot for getting the related information.*

AGENDA

1 Project background and importance

2 What is interoperability?

3 Current state of interoperability

4 Key considerations for interoperability

5 Viable approaches

6 Regulatory outlook

7 Implementation framework

8 Q&A

Viable approaches – Interoperability mechanisms (Centralized/ Decentralized)

Sharing domains of knowledge: Centralized vs. Decentralized Interoperability

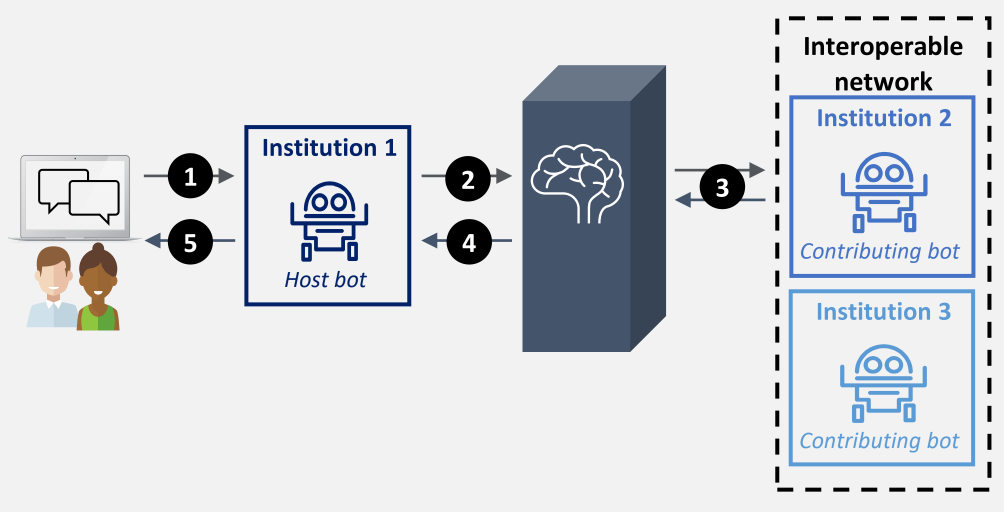
To compare approaches several factors can be considered such as:

- Storage optimization
- Maintenance
- Scalability
- Response time
- Disruption handling
- Setup

Interoperability can be deployed Centralized or De-centralized:

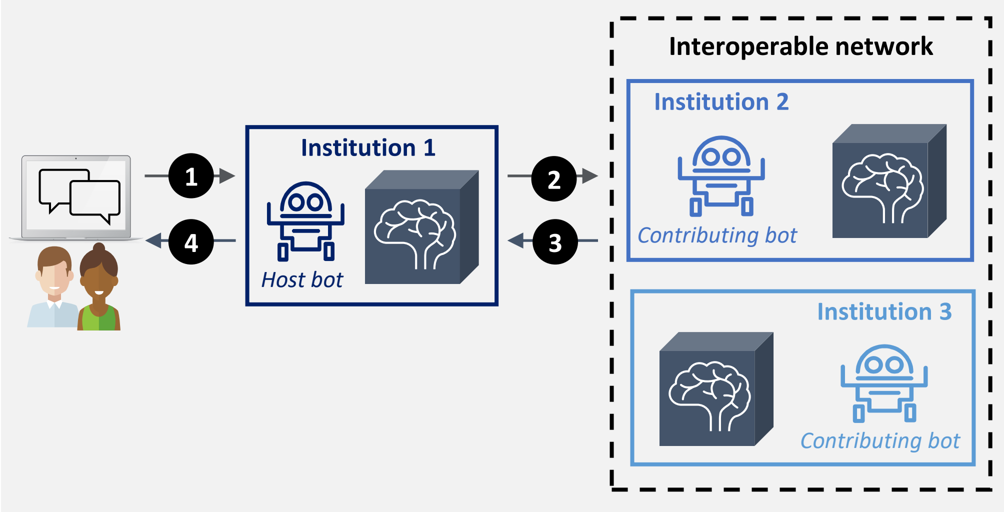
Centralized

This process begins with a user input which is sent directly to the **central interoperability mechanism**. This mechanism processes and controls the task, then sends the appropriate output to the host bot.



Decentralized

The user input is sent to an **individual chatbot** which processes the task independently. Each chatbot **maintains their own interoperability mechanism** to determine if other bots should be queried for information



The interoperability mechanism:

- Central Knowledge Director (CKD)
- Interoperability Layer (IL)

AGENDA

1 Project background and importance

2 What is interoperability?

3 Current state of interoperability

4 Key considerations for interoperability

5 Viable approaches

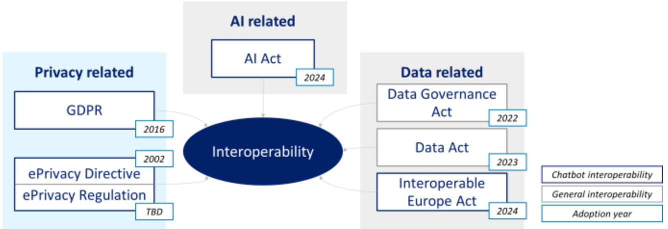
6 Regulatory outlook

7 Implementation framework

8 Q&A

Regulatory outlook (1/3)

A view on relevant regulation and their impact on interoperability: **Privacy**



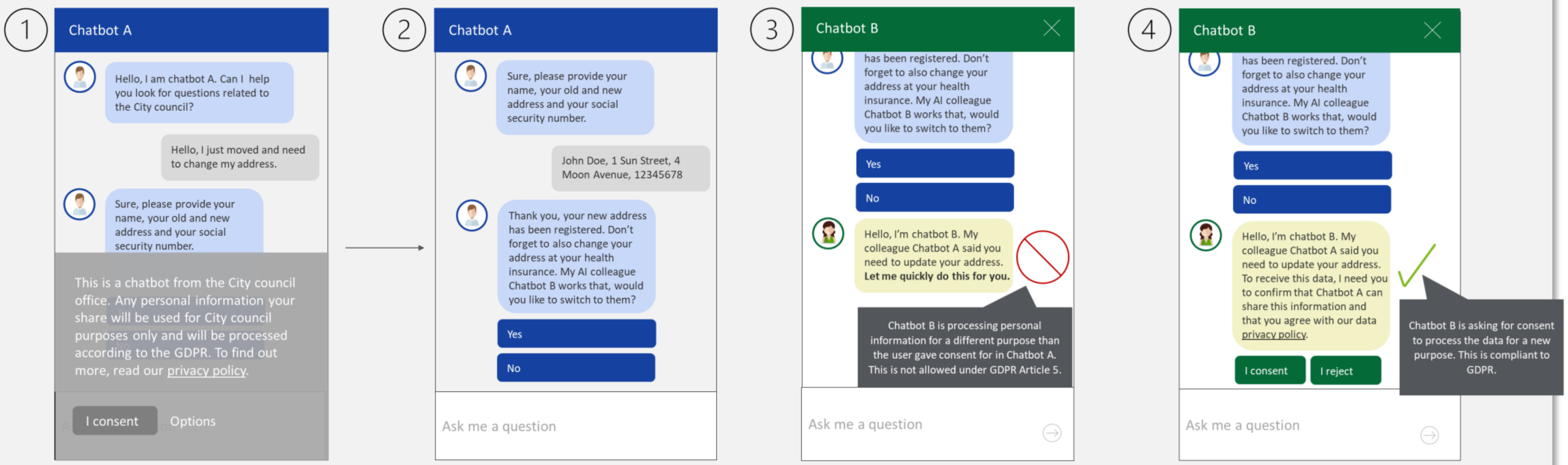
2016

General Data Protection Regulation (GDPR)

EUDPR

Enforces strict privacy rules, protecting individuals' rights to control their personal data.

Principles relating to processing of personal data (Article 5): Application toward reusing data for another purpose



DISCLAIMER: Relevance to interoperability shown through selected examples, these present one specific impact, more are or could be applicable for each Act

AGENDA

1 Project background and importance

2 What is interoperability?

3 Current state of interoperability

4 Key considerations for interoperability

5 Viable approaches

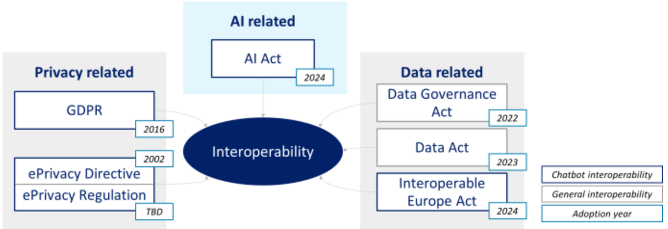
6 Regulatory outlook

7 Implementation framework

8 Q&A

Regulatory outlook (2/3)

A view on relevant regulation and their impact on interoperability: **AI**



2024

Artificial Intelligence (AI) Act

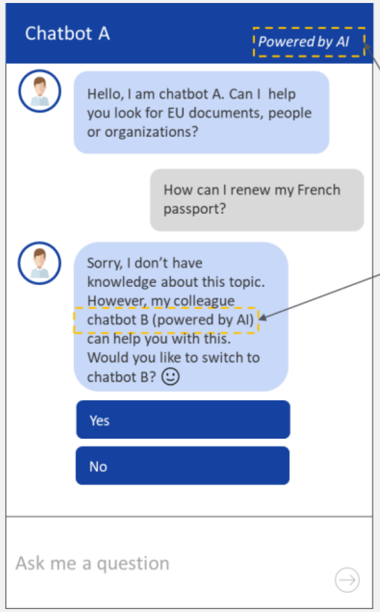
Regulatory framework for AI technologies using risk-based approach.

Risk-based approach

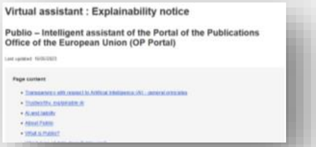
CATEGORIES	Prohibited	Requirements & obligations	Transparency obligations	No restrictions or voluntary	Obligations for providers
LEVELS	Unacceptable	High-risk	Limited	Minimal	GPAI
EXAMPLES	Biometric identification based on race, religion, etc, social scoring, emotion recognition in education/workplace	Biometrics, critical infrastructure, education & vocational training, employment, law enforcement, border control	Emotion recognition, chatbots, generating of manipulating images	Predictive maintenance, meeting schedulers, spell checkers, spam filters	Generative AI including large language models
TIME TO COMPLY	6 months	24 months	24 months	N/A	12 months

Generally, chatbots generally fall under transparency obligations but can also fall under high-risk and GPAI depending on the application and technology of the system used

Transparency obligations



Transparency obligation: Mandatory to display to the user that they are interacting with an AI system



***Transparency use case: Publications Office Explainability Notice for End Users**: https://op.europa.eu/en/web/about-us/explainability-notice*

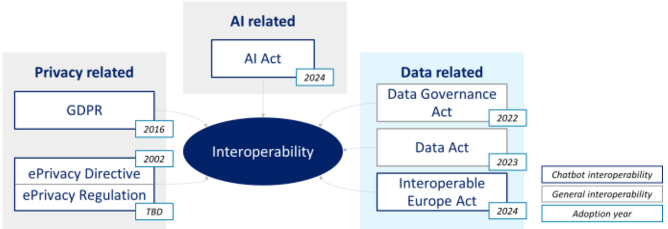
DISCLAIMER: Relevance to interoperability shown through selected examples, these present one specific impact, more are or could be applicable for each Act

AGENDA

1 Project background and importance
2 What is interoperability?
3 Current state of interoperability
4 Key considerations for interoperability
5 Viable approaches
6 Regulatory outlook
7 Implementation framework
8 Q&A

Regulatory outlook (3/3)

A view on relevant regulation and their impact on interoperability: **Data**



2016

Interoperable Europe Act

Aims to strengthen cross-border interoperability and cooperation in the public sector across the EU.

Chapter 4: Improved governance

Existing chatbots with interoperability capabilities do the interoperability assessment and get shared to the portal

Interoperable solutions are shared and can be re-used by other public institutions

Other public institutions can re-use existing interoperable solutions and thus achieve interoperability by design

DISCLAIMER: Relevance to interoperability shown through selected examples, these present one specific impact, more are or could be applicable for each Act

AGENDA

1 Project background and importance
2 What is interoperability?
3 Current state of interoperability
4 Key considerations for interoperability
5 Viable approaches
6 Regulatory outlook
7 Implementation framework
8 Q&A

Implementation framework

End-to-end process in achieving interoperability between chatbots

This broad framework is adaptable and can assist with the implementation of diverse interoperability projects

General framework to accelerate interoperability of bot *(to adjust on project specificities)*

Project phases	INITIATION	POC[1] DEVELOPMENT	TESTING	DEPLOYMENT & MONITORING
Activities	• Define project goals • Describe interoperable network • Define & align on governance • Define functional design of interoperable network • Define technical structure • Prepare testing • Finalize contractual alignments • Consolidate project discovery	• Set-up interoperable architecture • Build interoperability structure • Complete PoC	• Conduct testing • Validate testing	• Define deployment pipeline • Deploy to production • Manage & monitor release
Deliverables	• D1.1 Functional design plan • D1.2 Technical architecture plan • D1.3 Testing plan • D1.4 Contractual documentation • D1.5 Project plan	• D2.1 Working system	• D3.1 Sprint reports & test logs • D3.2 Sign off completion certificate	• D4.1 Deployment pipeline • D4.2 Live interoperable chatbots • D4.3 Monitoring reports

[1]Proof of Concept

Thank you for your attention!

Any questions?

Panel Discussion

2: AI in support of Interoperability

Join us on Slido!

- Use the QR code
- Or go on slido.com
  #SEMIC2024Workshops
- Select the correct workshop

We have some questions for you too!

Thank you

interoperable europe

LUNCH BREAK
13:00-14:00
HALL 100