



# Multilingual Knowledge System Use Cases

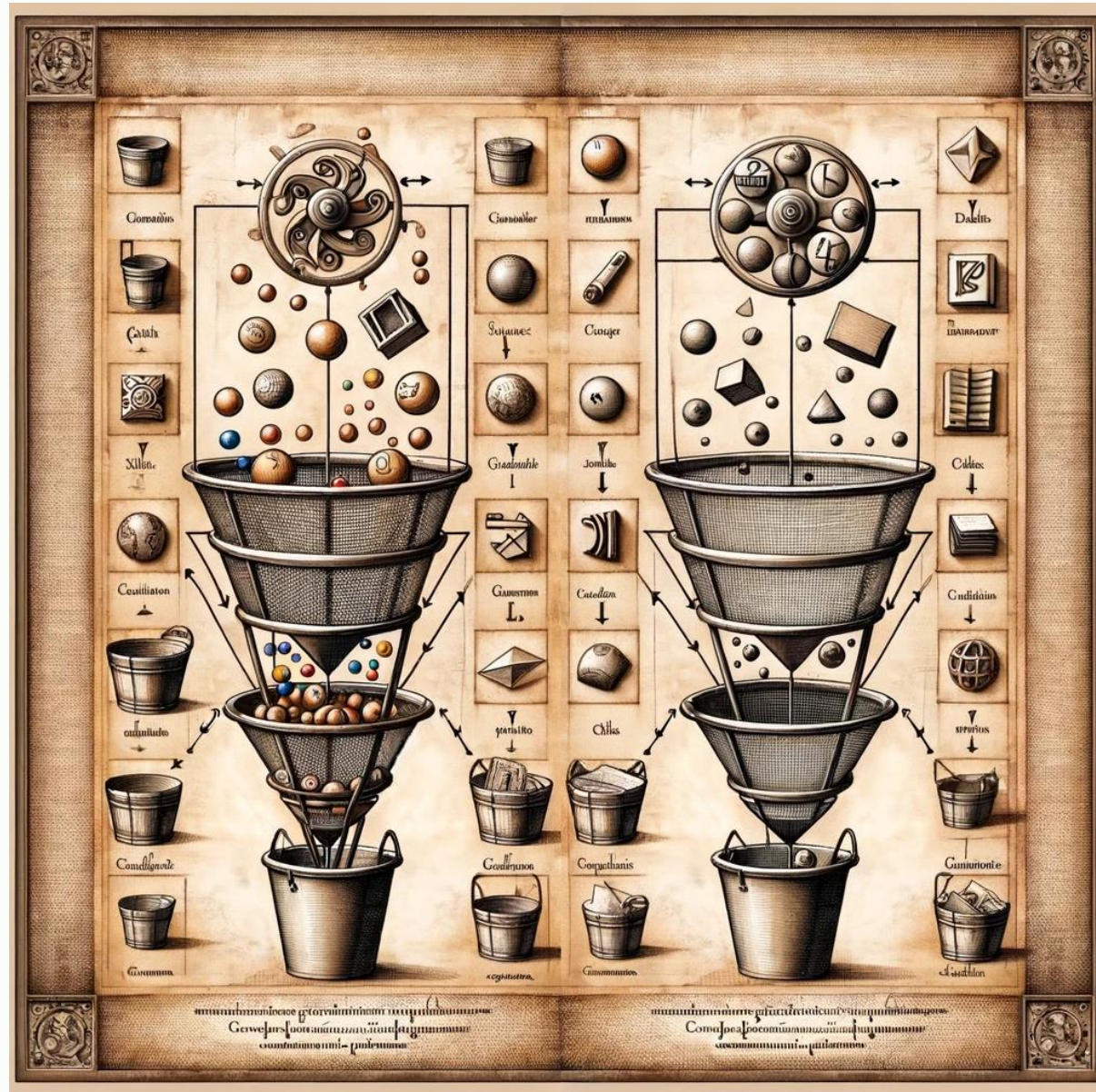
The power of Eurovoc and IATE through semantic integration

An ENDORSE Workshop Hosted by EU Publications Office  
18 April 2024

[michael@coreon.com](mailto:michael@coreon.com) | [jochen@coreon.com](mailto:jochen@coreon.com)

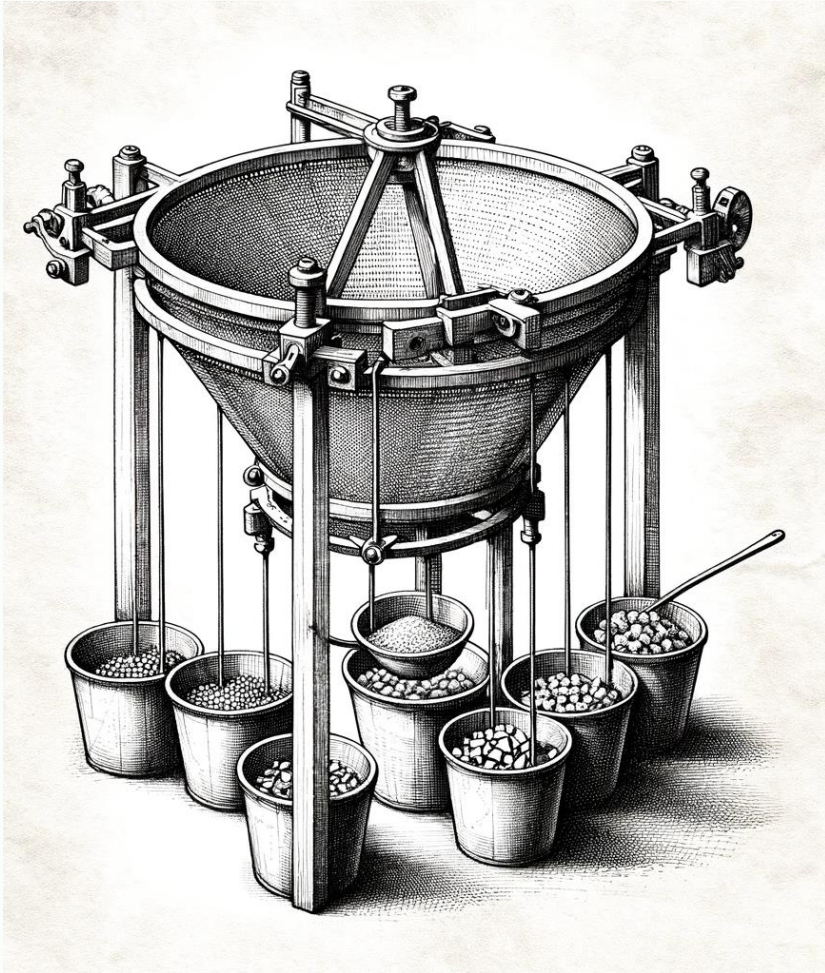
[www.coreon.com](http://www.coreon.com)

# Classification





# Common Classification Methods



## Keywords in Context

Straightforward, but struggles with synonyms, polysemy, cultural context variations, and multilingual.

## Trained Classifiers

Proven Machine Learning approach, but requires extensive, annotated datasets in each language.

## K-Nearest Neighbors

Simple, effective algorithm, but suffers of sparse feature spaces typical of textual data, struggles with cross-language.

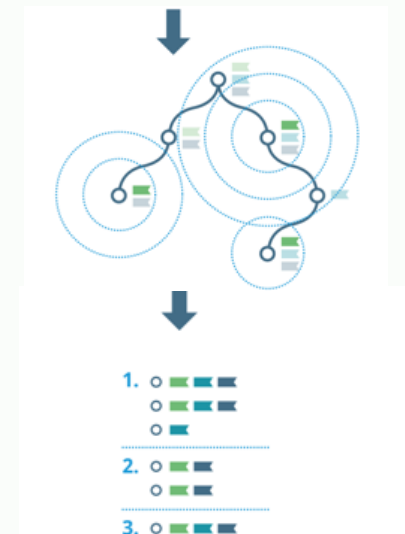
## Word Embeddings

Understands semantic relations, but depends on a balanced training corpus for all languages.

# From Strings to Meaning to Interoperability



Find terms in text using Advanced Linguistic Search.

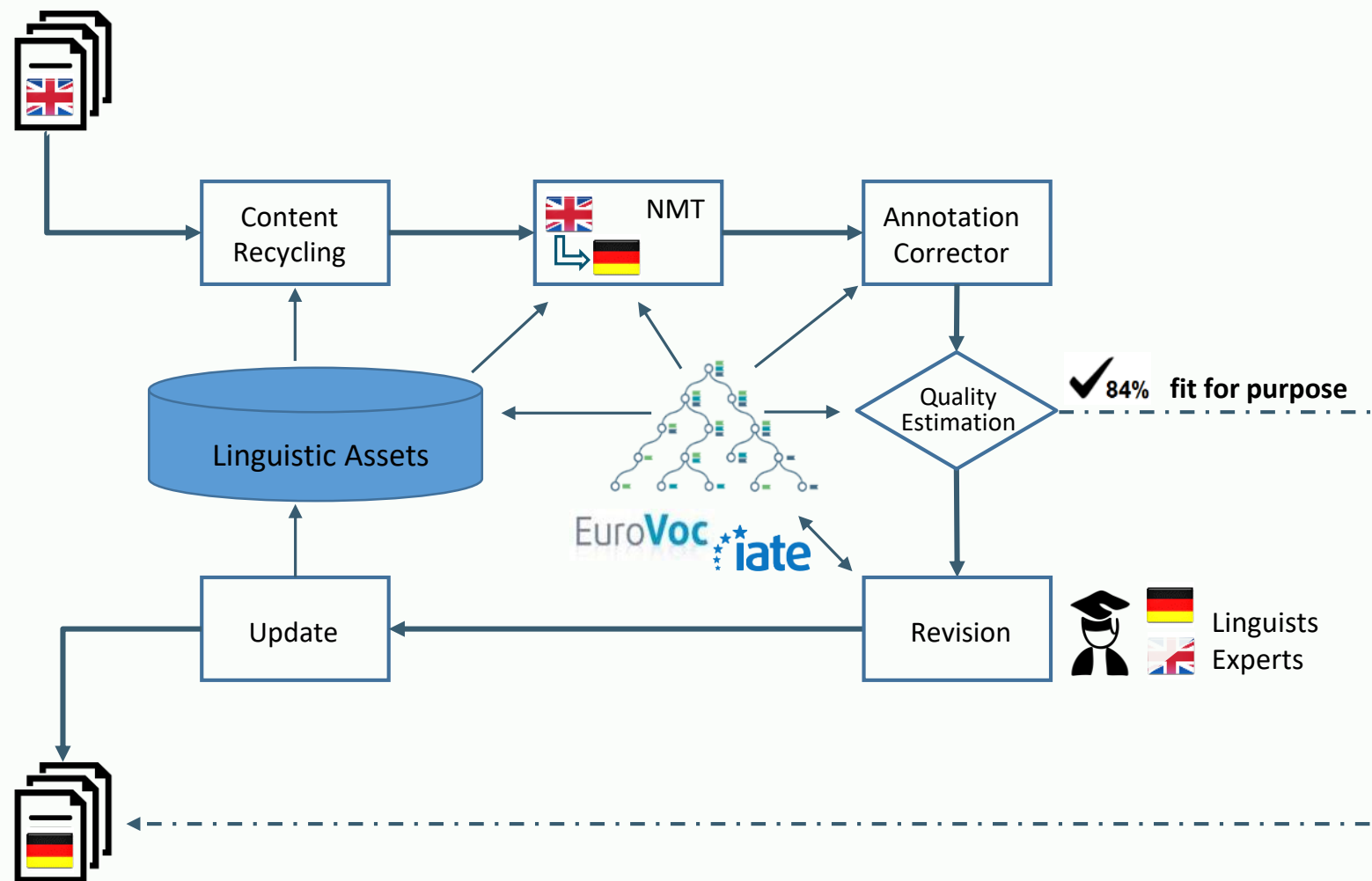


Locate them in the knowledge graph.

Create domain heat maps.

Link to other datasets (HR, DMS, QA rules).

# Translation in the Language Factory



# Key Innovations of the Language Factory

Content  
Recycling

- Replaces segment based, 1:1 source/target Translation Memories
- Allows safe recycling in variable granularity
- Easy maintenance

Annotation  
Corrector

- Identify domain specific terminology
- Find relevant, similar content
- Use Generative AI to correct Machine Translation

Quality  
Estimation

- Use just another AI to judge translation quality
- Constantly trains model with every revision cycle

Revision

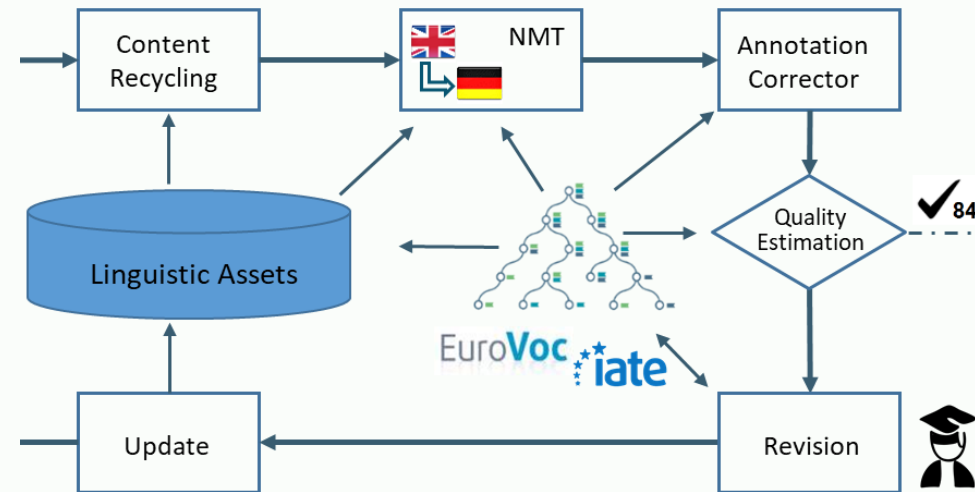
- Revise instead of translate
- Subject matter experts instead of linguists

# Workflow Questions



- ¿ Content to recycle?
- ¿ Translations to train the NMT?
- ¿ NMT engine to pick?
- ¿ Terms correct or wrong?
- ¿ Style rules to apply?
- ¿ Quality fit-for-purpose?
- ¿ Best experts?
- ¿ If busy, second best?

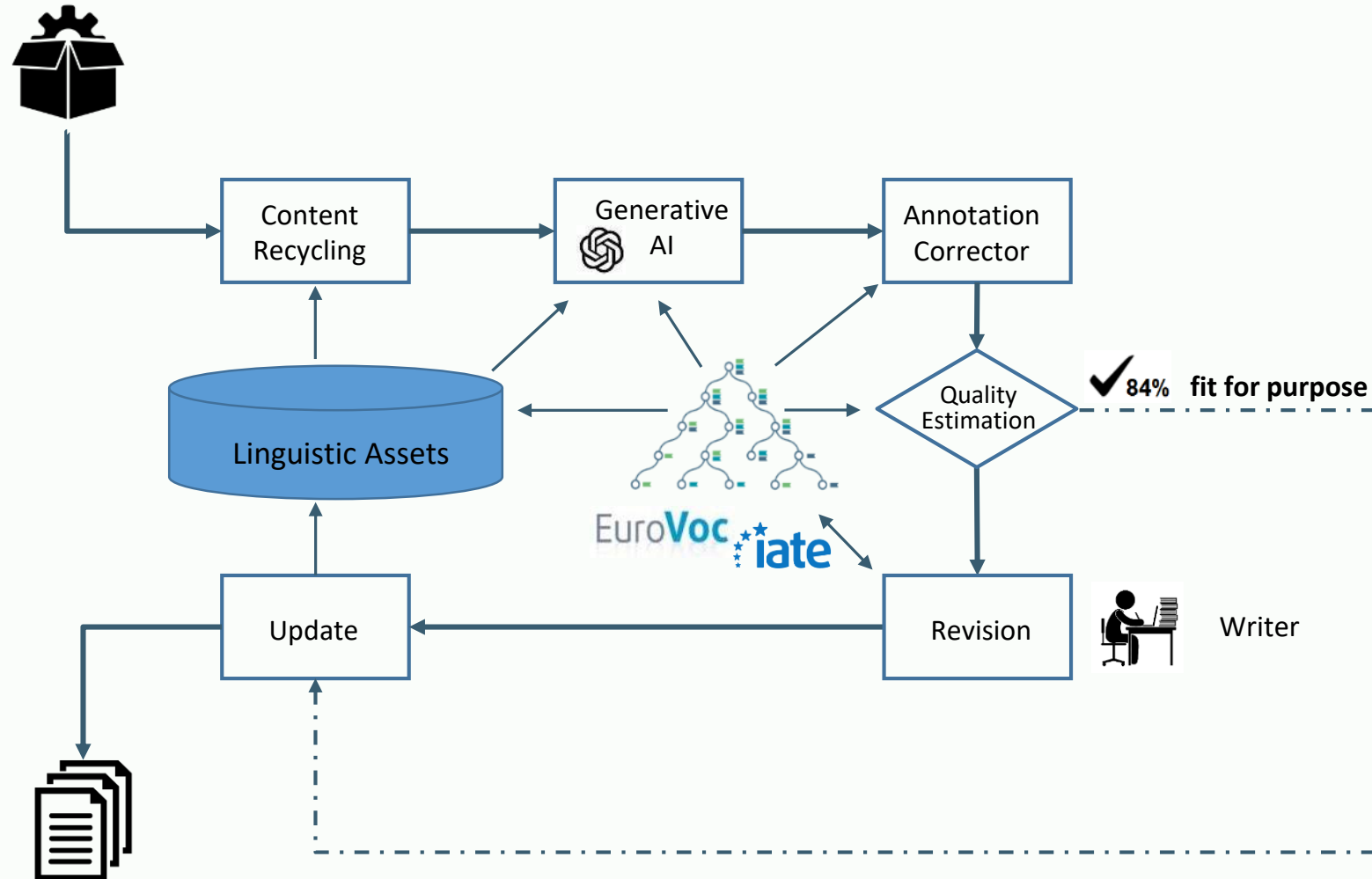
# Key Enabler Multilingual Knowledge System

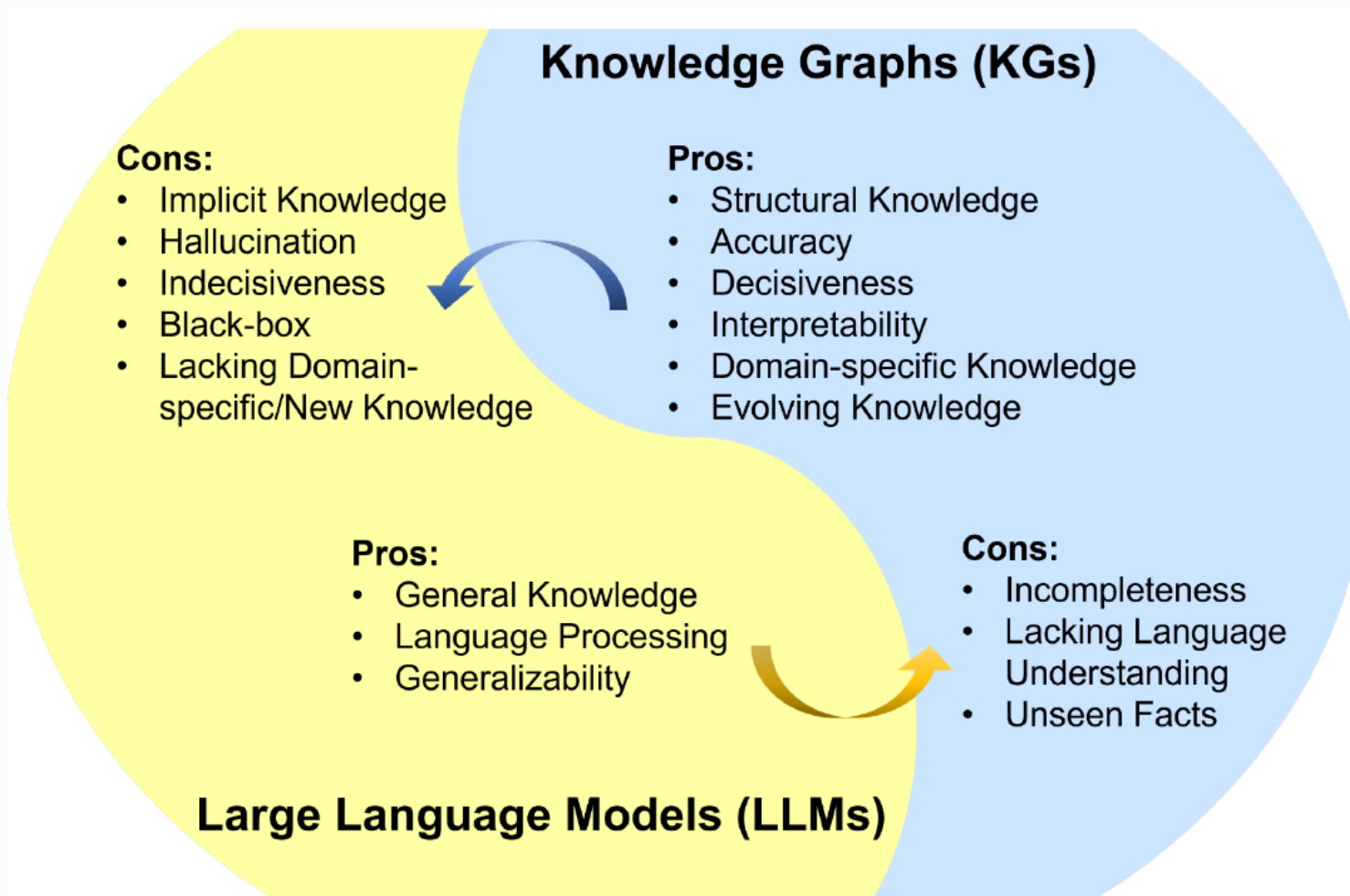


Coreon classifies linguistic assets, human resources, workflow rules, and projects in a unified system which is expandable, dynamic, and provides fallback paths.

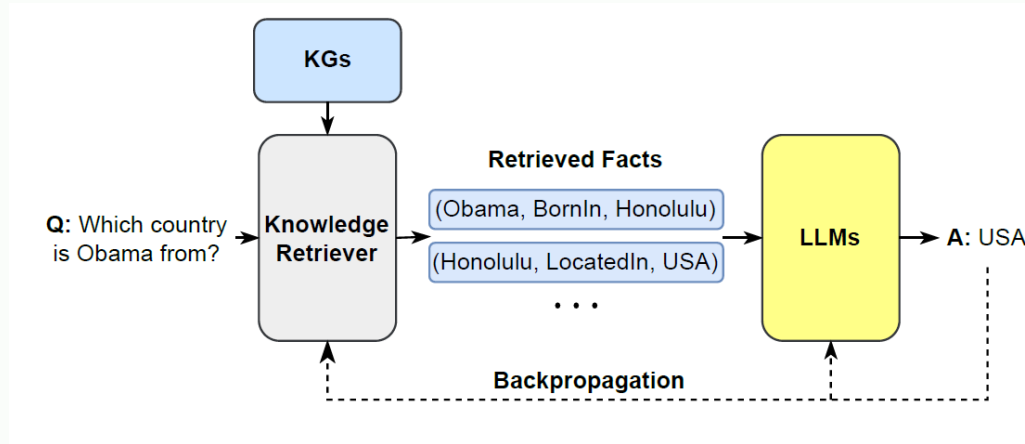


# Content Factory

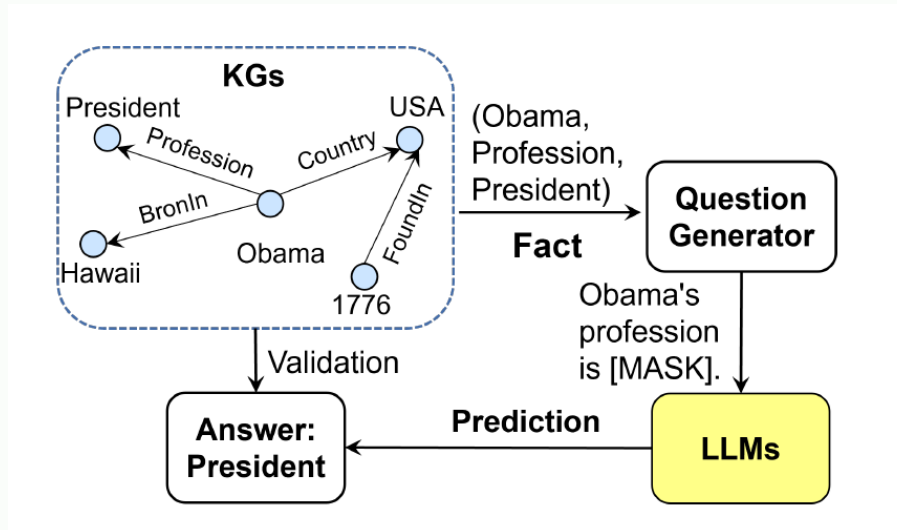




# Hallucination Detection



Add facts stored in the knowledge graph to the input



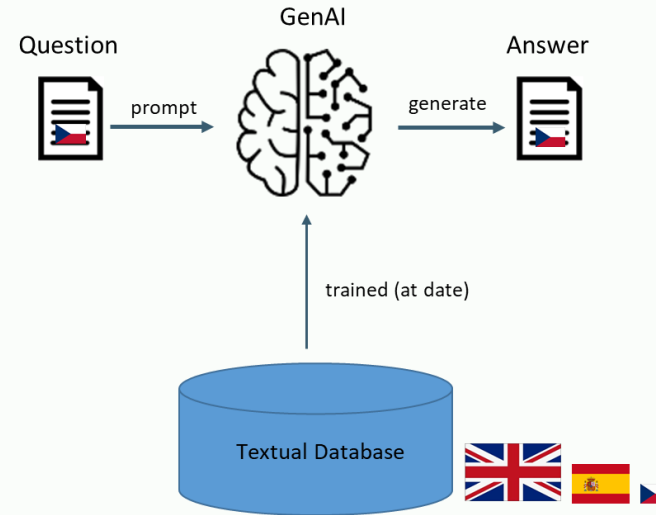
Validate the output of the LLM with the facts stored in the KG.

# The Million-Dollar Question



How do we deploy  
Generative AI  
to make better use of our  
vast textual databases?

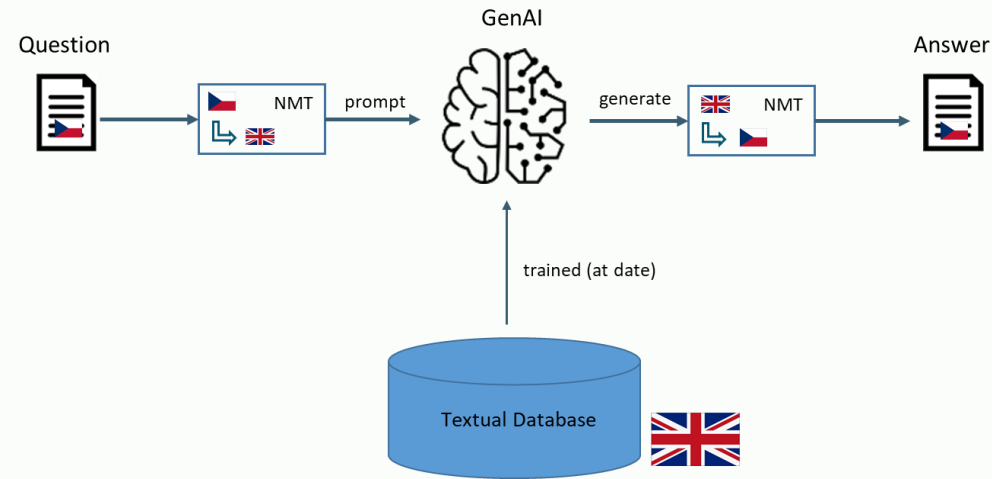
# Train Language



- For most languages there's too little content to properly train an LLM.
- Not enough text to train LLM for general language skills.
- Effort multiplies by language to be supported.
- No cross-language leverage.
- Inconsistent performance per language.



# Double MT

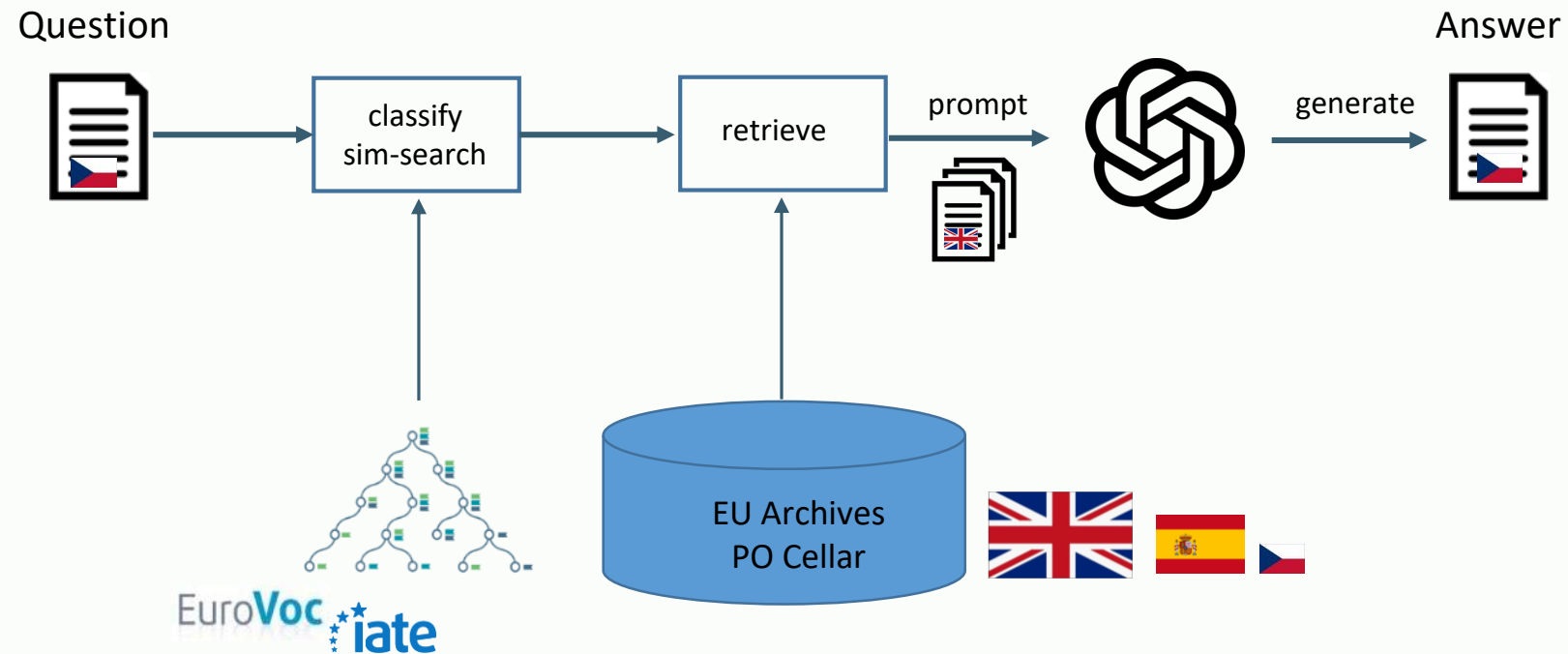


- Expensive to train and maintain an LLM, even for one language.
- LLMs only know what they have been trained with.
- DIY LLMs are not on par with leading general LLMs.
- NMT errors/inaccuracies of the two translation steps multiply.
- Inconsistent performance per language.

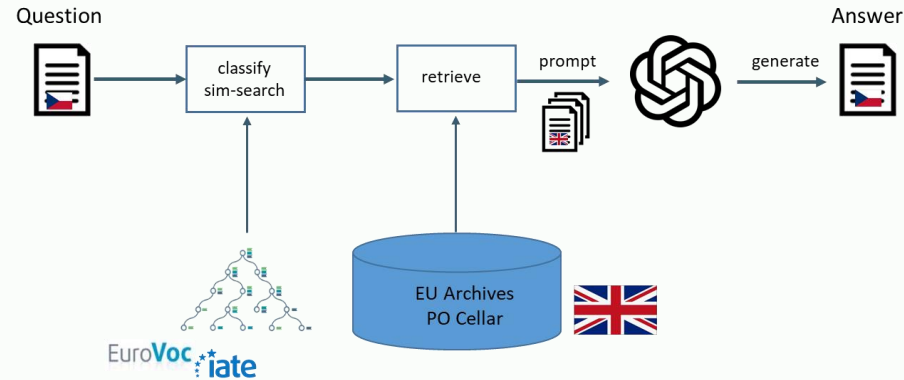
# Train, tune, or prompt



# „Multilingual RAG“ Approach

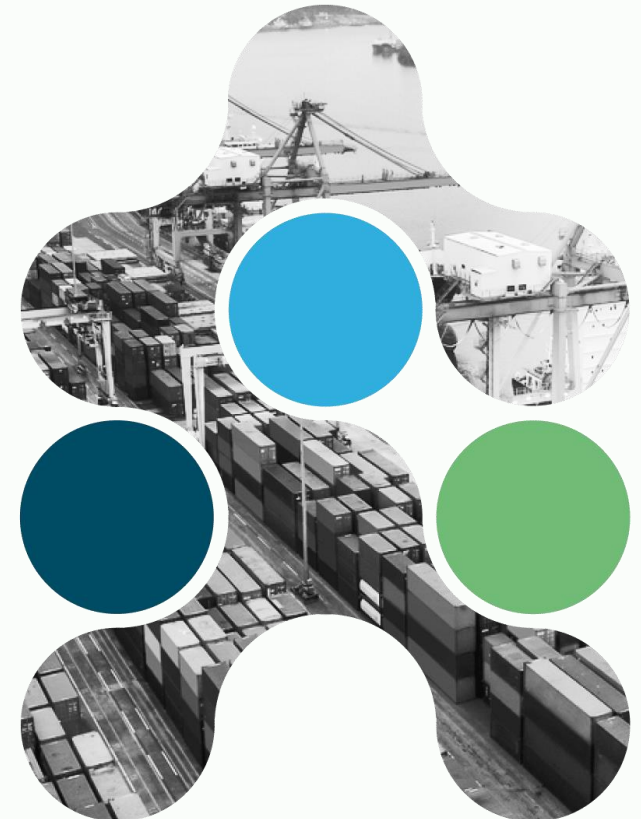
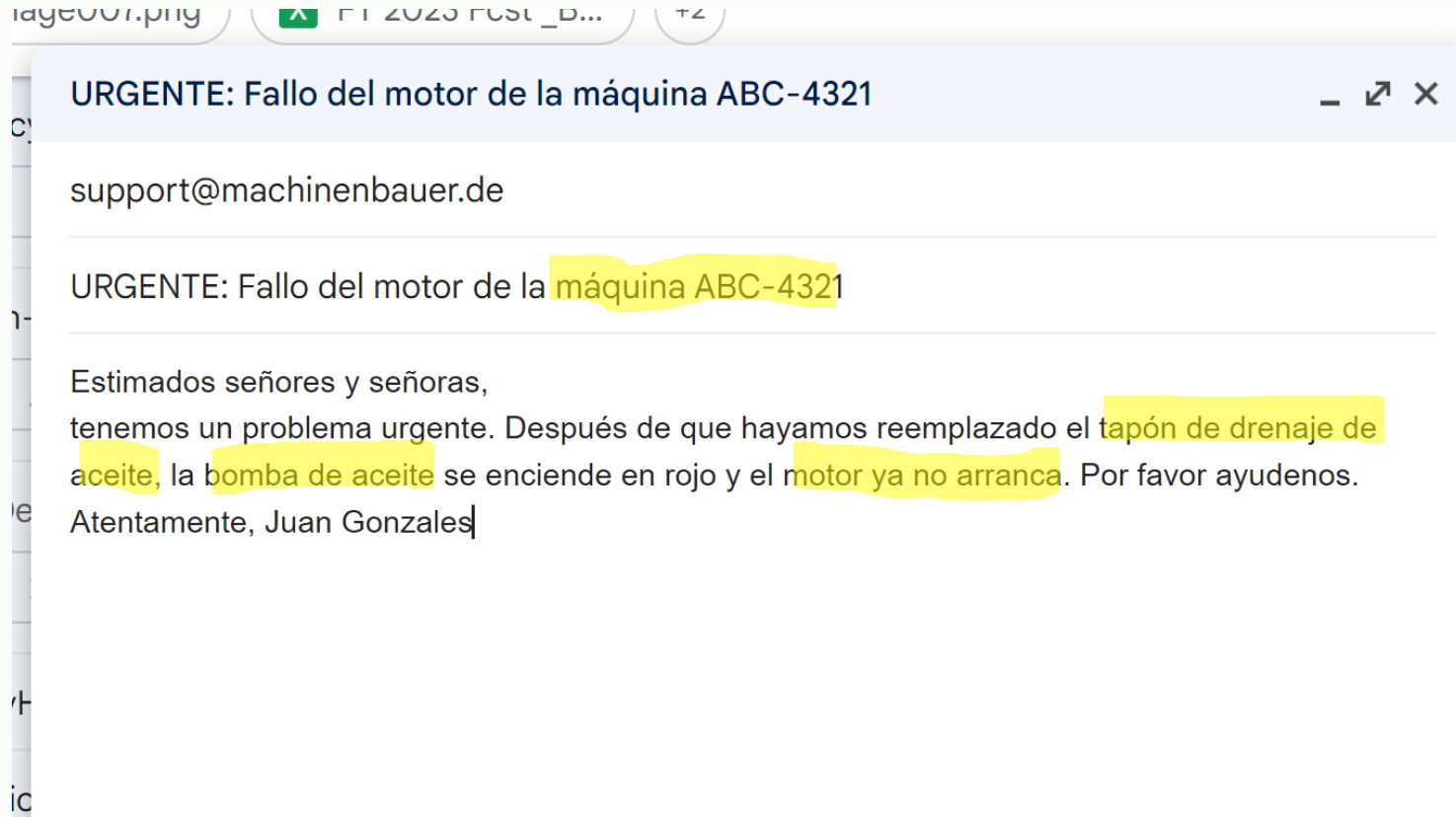


# M-RAG



- ✓ No need to train LLM.
- ✓ Data is always up-to-date.
- ✓ 3<sup>rd</sup> party LLMs costs probably lower than to train/host own LLM.
- ✓ Target language generation produces better quality than Double MT.
- ✓ M-RAG stack always SOA by using latest and greatest LLM.

# Use Case Global Customer Support





# Multilingual LLM Prompt with Relevant Context

Feed text of similar tickets (DE) to ChatGPT

Prompt ChatGPT with customer problem (ES).

RE

Q: Auch nach dem Ölwechsel leuchtet weiterhin die rote Ölwarnlampe. Es tropft kein Öl heraus.  
A: Der Kunde soll den Ölfilter überprüfen, wahrscheinlich ist der verstopft. Ein neuer Filter kann unter der Teilenummer 372623-OF bestellt werden.

Wenn die

RE

Assume the role of a service technician working for [REDACTED]. Reply in German and Spanish to the following Spanish technical enquiry based on the previously provided support tickets and mention the spare parts to be ordered: "Despues de que hayamos reemplazado el tapón de drenaje de aceite, la bomba de aceite se enciende en rojo y el motor ya no arranca". Provide only the solution not any greetings or closing text.

In Spanish:

Primero verifica el nivel y el estado del aceite. Una luz roja en la bomba de aceite puede sugerir una posible caída de presión en el sistema de aceite. También es posible que el filtro de aceite esté obstruido y necesite ser reemplazado. Puedes ordenar un nuevo filtro de aceite con el número de parte 372623-OF. Asegúrate de que todas las conexiones y mangueras estén correctamente y firmemente instaladas.

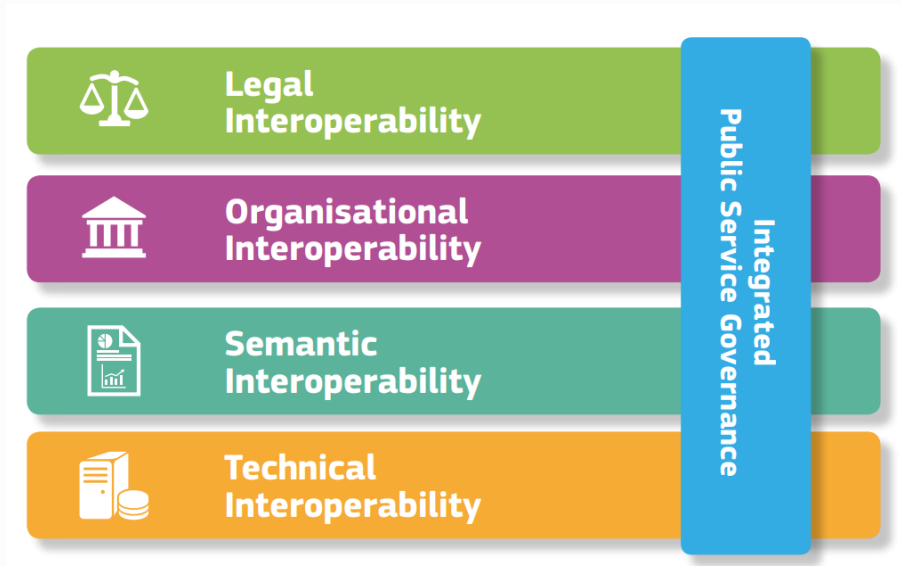
In German:

Überprüfen Sie zuerst den Ölstand und den Zustand des Öls. Eine rote Anzeige auf der Ölpumpe kann auf einen möglichen Druckabfall im Ölsystem hinweisen. Es ist auch möglich, dass der Ölfilter verstopft ist und ausgetauscht werden muss. Sie können einen neuen Ölfilter mit der Ersatzteilnummer 372623-OF bestellen. Stellen Sie sicher, dass alle Anschlüsse und Schläuche korrekt und fest montiert sind.

Instant automatic reply in Spanish  
Includes spare part number

Verify and edit the answer later (DE).

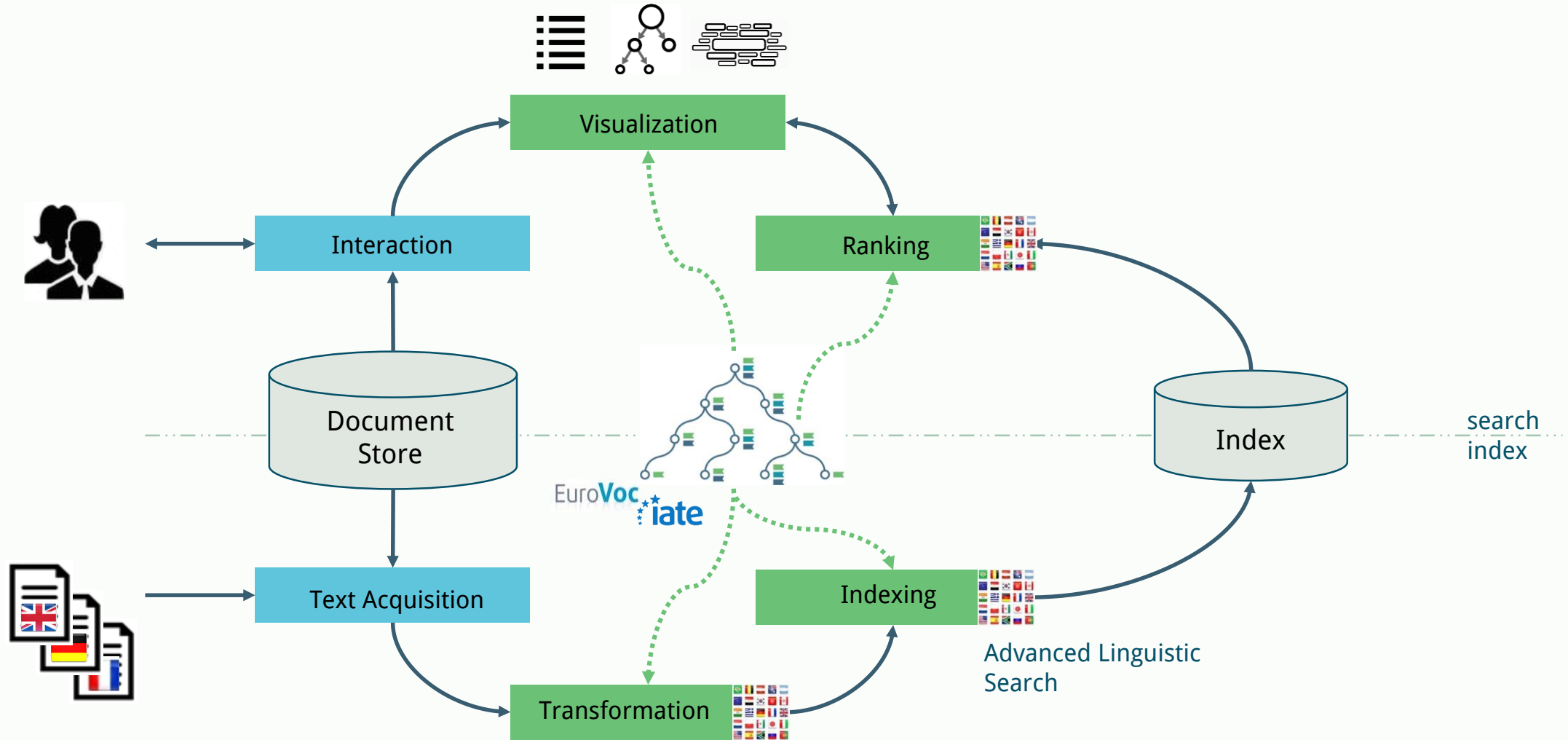
# Interoperability Act



Semantic interoperability ensures that the precise format and meaning of exchanged data and information is preserved and understood throughout exchanges between parties, in other words ‘what is sent is what is understood’.

European public services can potentially be used by anyone in any Member State. So multilingualism needs to be carefully considered when designing them.

# Cross-Language Search



# Benefits



- Information infrastructure for unstructured data
- Language agnostic
- Precise, dynamic classification
- Automates content creation and translation
- Cross-language semantic search
- Enables Generative AI
- Delivers the semantic layer of the EIF