Beyond Prompt-to-RDF: A Vision for Scalable and Explainable Graph Transformations via LLM-assisted Schema Mappings

Yannis Marketakis and Yannis Tzitzikas

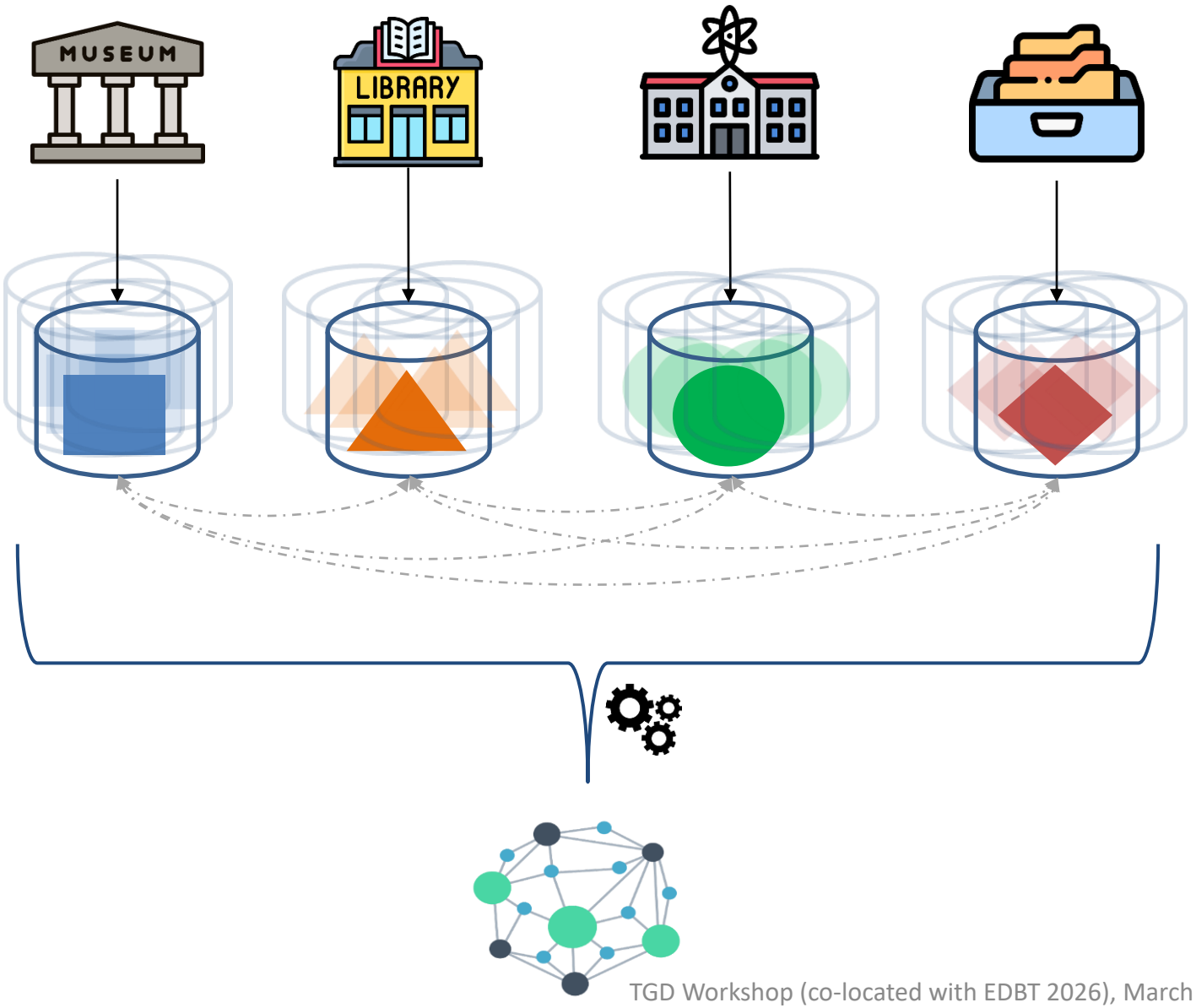
Institute of Computer Science, Foundation for Research and Technology – Hellas, Heraklion, Greece

Computer Science Department, University of Crete, Heraklion, Greece

Outline

Introduction	3-5	2’
Related Works	6	1’
Prompt-to-RDF Limitations	7-8	3’
Vision: LLM-assisted Schema Mappings	9-1	4’
Grounding with X3ML	12-14	4’
Conclusion & Research Agenda	15	1’

Introduction – RDF Transformation at scale

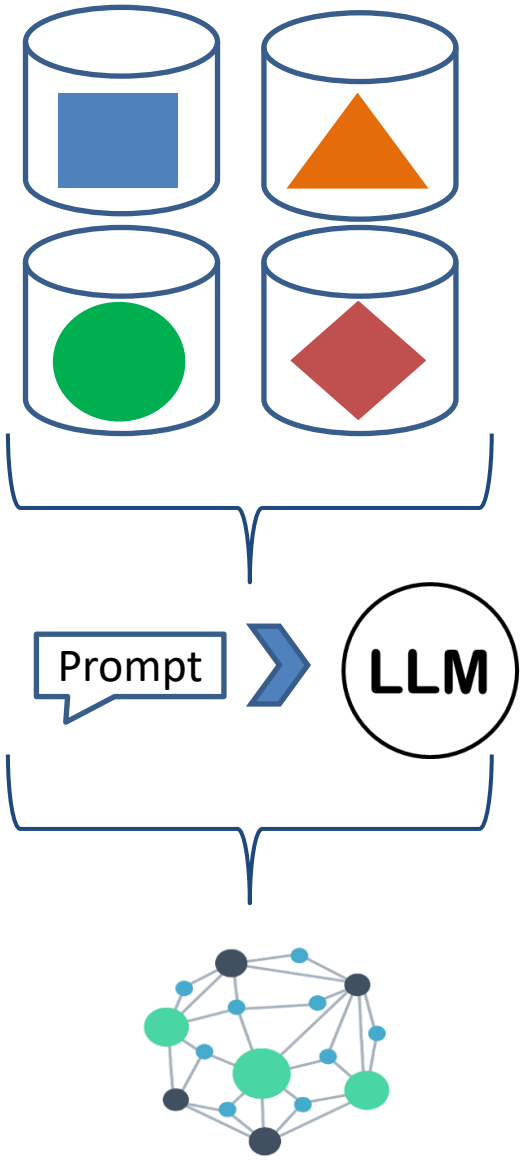


- Fragmented data landscape
- Structural and semantic heterogeneity
- Rapidly increasing data scale
- Requirements for scalable graph construction

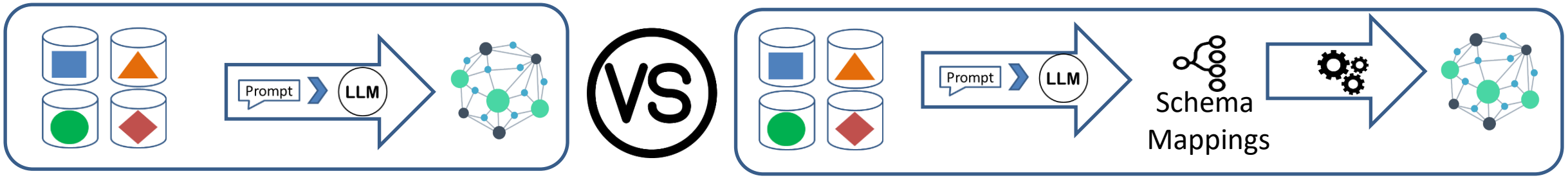
Introduction – LLMs enter the scene

- Large Language Models (LLMs) can facilitate the semantic data integration

- **How?**
  - Generation of RDF from several structured data sources
  - Instance-level construction of RDF knowledge graphs
- **Why?**
  - Strong semantic interpretation
  - Flexible handling of heterogeneous inputs
  - Relatively simple to use through prompting
  - Ideal for rapid prototyping
- **…But, what about?**
  - Scalability
  - Transparency
  - Reproducibility



Introduction – Beyond Prompt-to-RDF



- Using LLMs for automating the schema mapping generation is more sustainable than direct LLM-based data transformation

- Our contributions:
  - Reframing the role of LLMs in RDF graph transformation
  - Analysis of prompt-to-RDF limitations
  - Grounded vision using X3ML-based experimentation

Related Works

Schema Mapping Frameworks

- Declarative schema mapping languages (X3ML, RML, R2RML)
- Explicit transformation logic executed by dedicated engines
- Scalable, reproducible, reusable
- Manual, expert-driven schema mapping definition

Prompt-to-RDF approaches

- Direct generation of RDF triples (SQLMorpher; Ghazzai et al., 2024; Norouzi et al., 2024)
- Prompt-to-RDF/ Prompt-to-Turtle workflows (LLM2KB, Frey et al., 2023)
- Iterative prompt refinement approaches (Carta et al., 2023)
- Instance-level graph construction

Our approach repositions LLMs as accelerators of data transformation workflows by assisting the generation of schema mappings

Prompt-to-RDF Limitations

Conceptual Limitations

No explicit mappings

- Modeling decisions hidden inside RDF triples
- No inspectable intermediate representation

Limited verification & validation

- Errors detected only after RDF generation
- Logic cannot be reviewed before execution

Debugging & maintenance

- Black-box transformation without any traceable transformation step
- Difficult to identify source of errors

Variability and Reproducibility

- Varying and inconsistent outputs across runs of prompts
- Sensitive to model versions

Transformation logic is coupled with execution

Prompt-to-RDF Limitations – cont'd

Scalability and Cost

Instance-level processing

- Requires model invocation for each record or data fragment
- Inefficient for (very) large datasets

Repeated model invocation

- Full transformation requires many subsequent LLM calls
- Costly for evolving or frequently updated data

Computational latency & cost

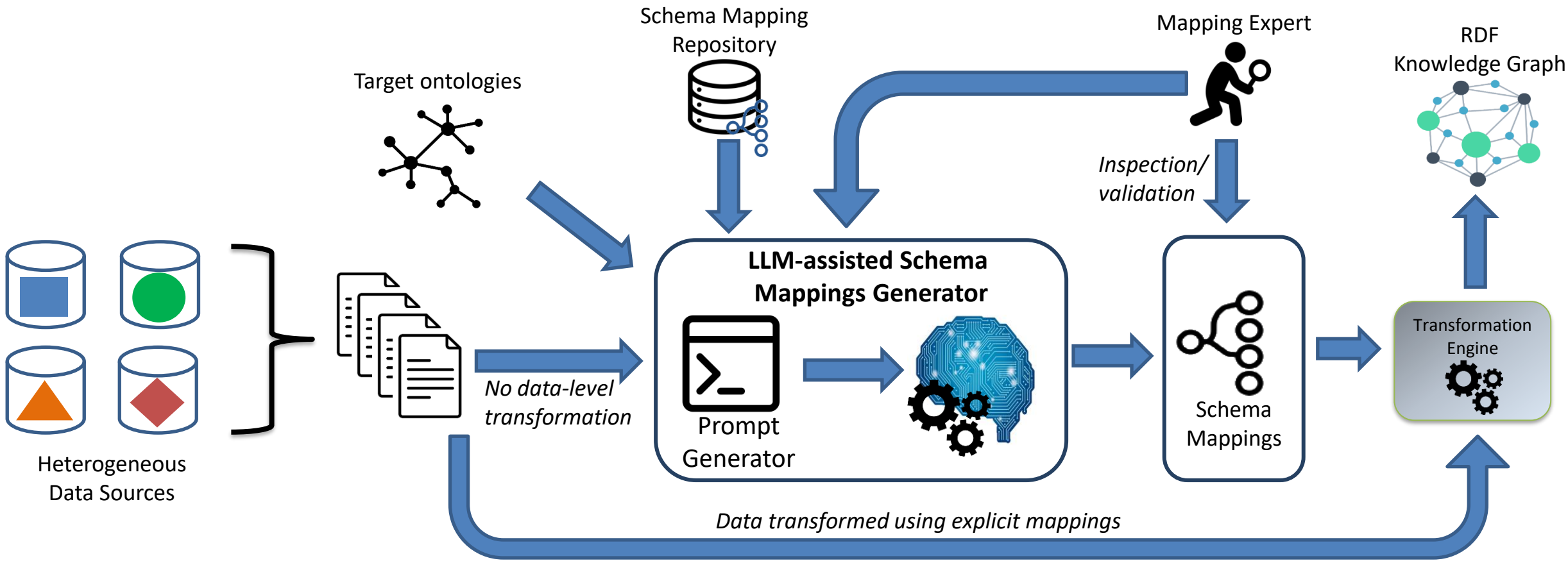
- High cost and increased processing time
- Unsuitable for continuous production pipelines

Limited operational sustainability

- Difficult to scale to massive collections
- Not robust for long-term graph maintenance

Instance-level LLM transformation does not scale sustainably

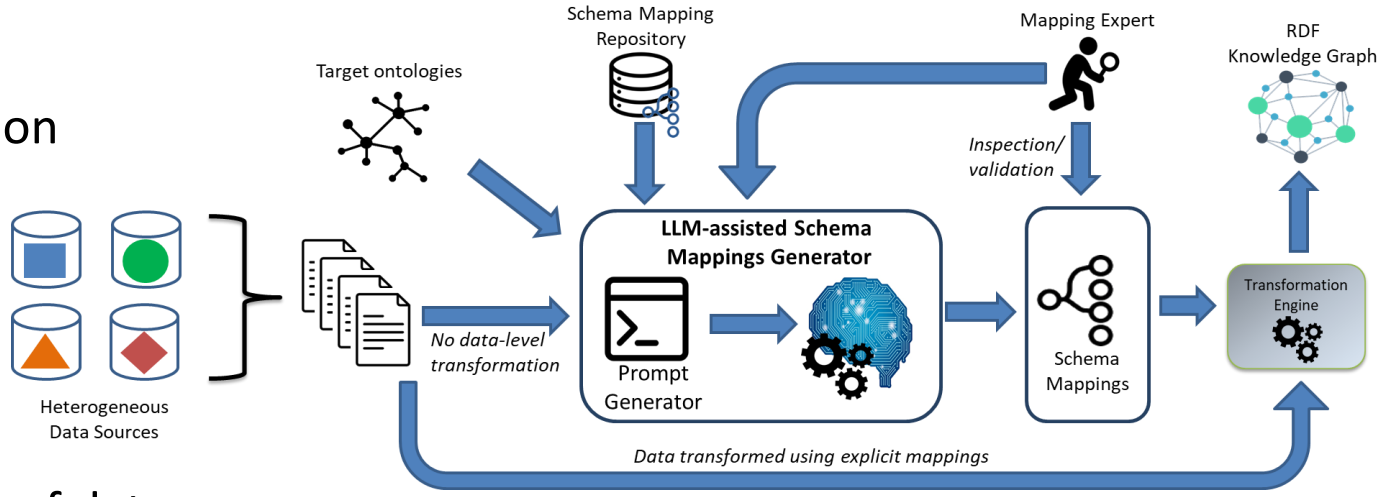
Vision: LLM-assisted Schema Mapping Generation



Vision: LLM-assisted Schema Mapping Generation – cont'd

Key Principles

- Decoupling logic from execution
  - Mapping definition ≠ data transformation
- Explicit & inspectable schema mappings
  - Intermediate representation
  - Human validation possible
- Efficient use of LLMs
  - LLM uses a small representative subset of data
- Reuse of existing schema mappings
  - To facilitate more accurate schema mappings
- Iterative Refinement
  - Expert feedback loop



Vision: LLM-assisted Schema Mapping Generation – cont'd

How our approach addresses the limitations of prompt-to-RDF

Aspect	Prompt-to-RDF	LLM-assisted Schema Mappings
Input	Full Data	**Sample data** ✓
Transparency	Implicit	**Explicit** ✓
Reproducibility	Limited	**High** ✓
Debugging	Limited	**High** ✓
Variability	High	**Limited** ✓
Scalability	Limited	**High** ✓
Human Control	Minimal	**High** ✓

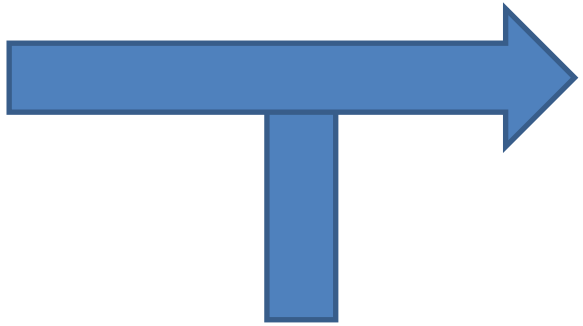
*Decoupling mapping definition from execution enables **transparency**, **reproducibility**, and **scalability***

Grounding with X3ML – Example

X3ML framework provides a declarative mapping language and a transformation engine

input

```
<record>
    <id>pr-01</id>
    <name>Yannis</name>
</record>
<record>
    <id>pr-02</id>
    <name>Mary</name>
</record>
```



```
:Yannis
   a crm:E21_Person.
   crm:P1_is_identified_by :pr_01

:Mary
   a crm:E21_Person.
   crm:P1_is_identified_by :pr_02

:pr_01
   a crm:E42_Identifier.

:pr_02
   a crm:E42_Identifier.
```

transformed RDF

mappings

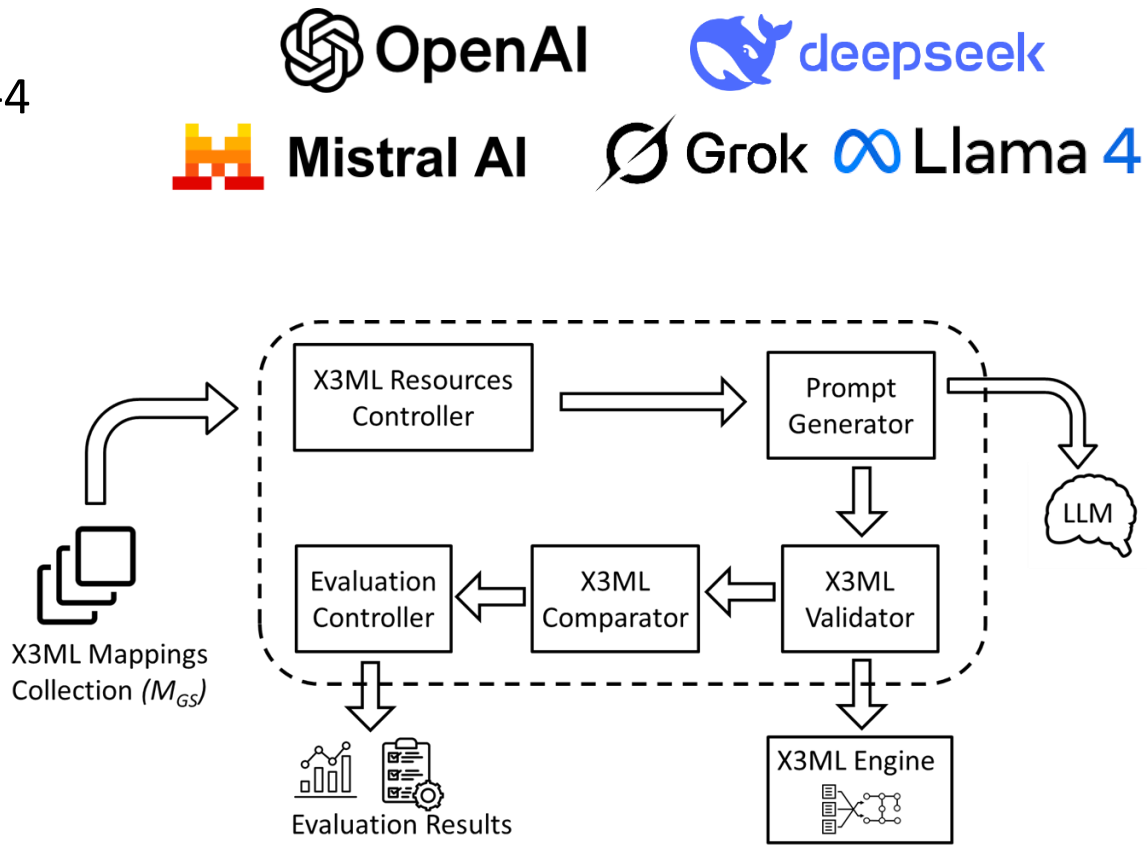
```
<x3ml><mappings><mapping>
    <domain>
      <source_node>//record</source_node>
      <target_node><entity><type>crm:E21_Person</type></entity></target_node>
    </domain>
    <link>
      <path>
        <source_relation><relation>id</relation></source_relation>

<target_relation><relationship>crm:P1_is_identified_by</relationship></target_relation>
      </path>
      <range>
        <source_node>id</source_node>
        <target_node><entity><type>crm:E42_Identifier</type></entity></target_node>
      </range>
    </link>
</mapping></mappings></x3ml>
```

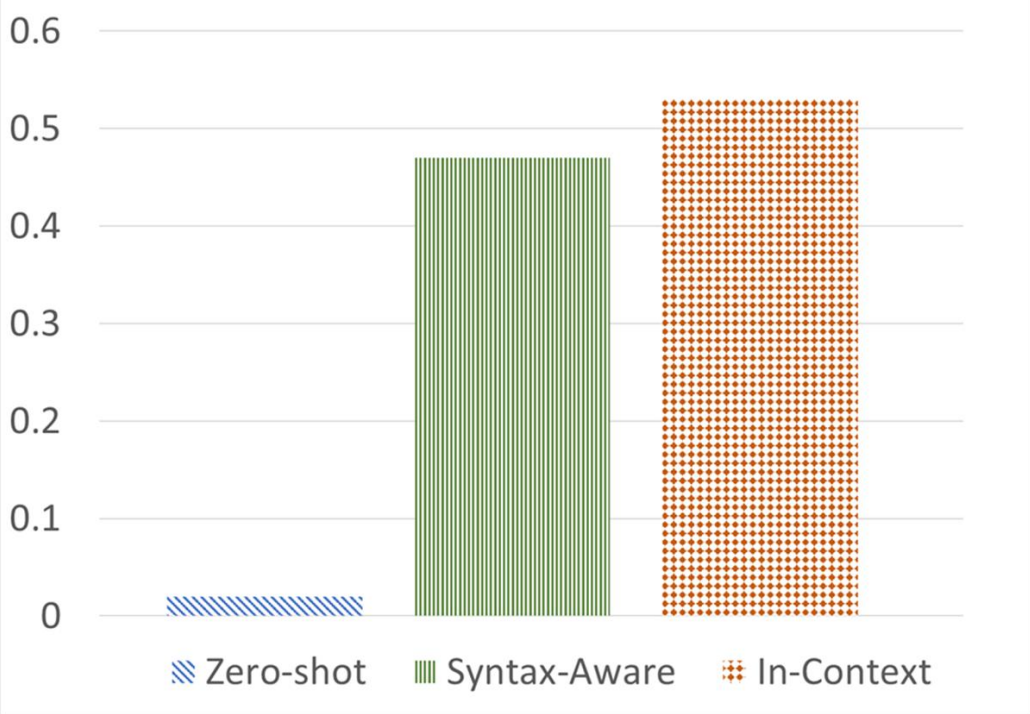
Grounding with X3ML – Experimental Setup

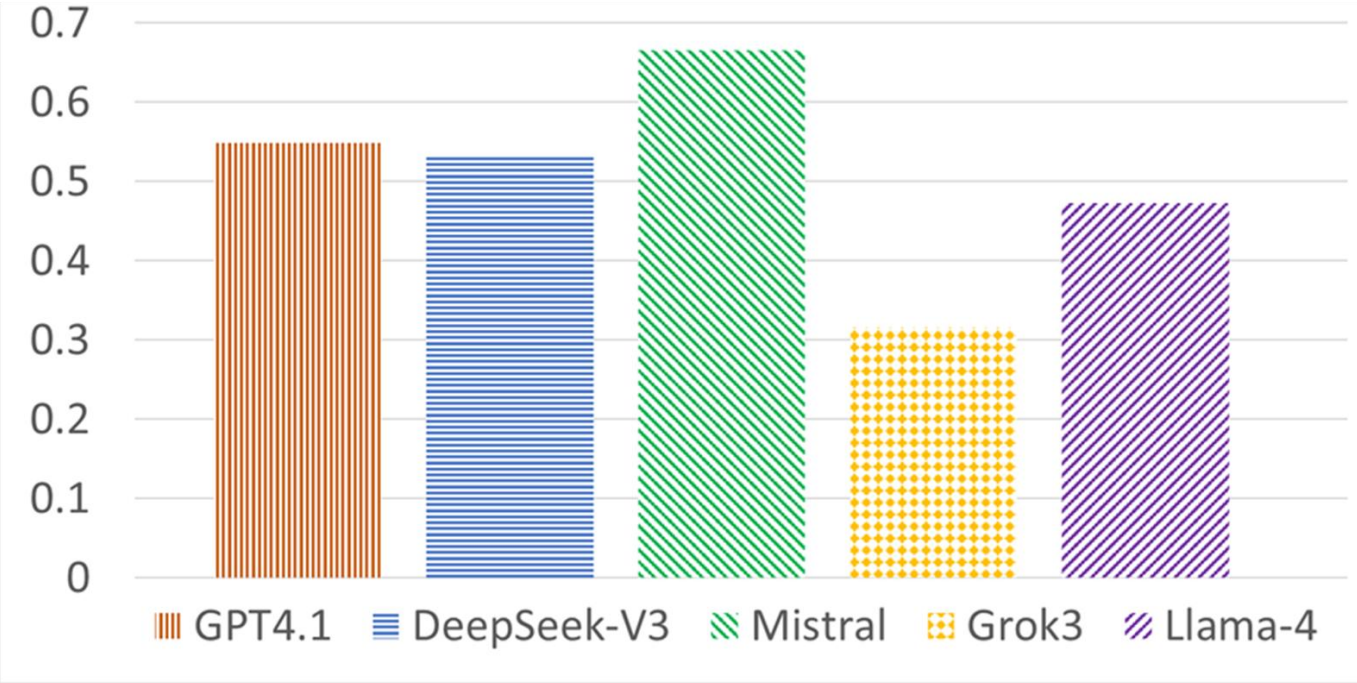
In (Marketakis et al., 2026) we experiment with different LLMs and prompting methods

- **Five** different LLMs
  - GPT 4.1, DeepSeek-V3, Mistral, Grok-3, Llamma-4
- **Three** prompting techniques
  - **Zero-shot** with input:
    - source data
  - **Syntax-aware** with input:
    - source data
    - X3ML structure
  - **In-Context** with input
    - source data
    - relevant X3ML schema mapping definitions



Grounding with X3ML – Experimental Setup





Large Language Models can be a powerful accelerator for schema mapping and Knowledge Graph construction

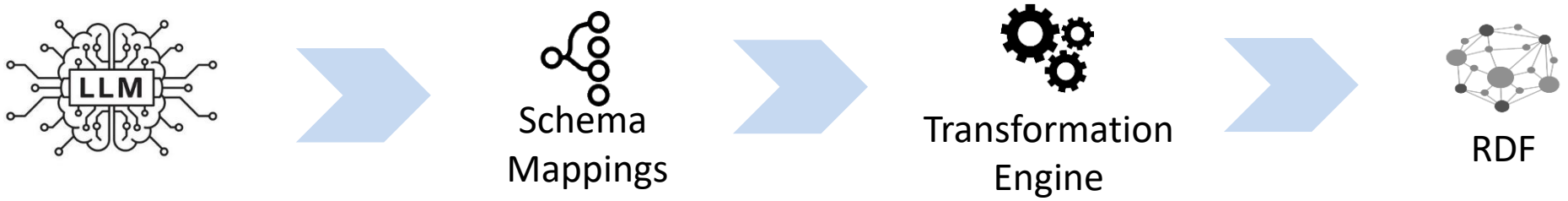
Conclusions & Research Agenda

- **Conclusion**
  - Direct prompt-to-RDF approaches have limitations
  - Explicit schema mappings are essential for transparency, reproducibility, and validation
  - Schema mappings are usually manual, time-consuming and expertise-intensive
  - LLMs should assist in mapping generation, towards accelerating it
- **Research agenda**
  - Improving the LLM-assisted mapping generation
  - Retrieval-enhanced mapping assistance
  - Hybrid human-LLM workflows
  - Benchmarking and evaluation frameworks

References

- X3ML-related
  - Y. Marketakis, M. Lintanff-Castel, Y. Tzitzikas, Using llms to automate schema mappings for rdf knowledge graphs construction, in: Poster in the 41st ACM/SIGAPP Symposium on Applied Computing, Thessalloniki, Greece, March 2026.
  - Y. Marketakis, N. Minadakis, H. Kondylakis, K. Konsolaki, G. Samaritakis, M. Theodoridou, M. Doerr, X3ml mapping framework for information integration in cultural heritage and beyond, International Journal on Digital Libraries 18 (2017) 301–319.
- Related Prompt-to-RDF Works
  - A. Nayak, H. P. Timmapathini, Llm2kb: Constructing knowledge bases using instruction tuned context aware large language models, international semantic web conference (iswc) workshop on knowledge base construction from pre-trained language models, 2023.
  - A. Sharma, X. Li, H. Guan, G. Sun, L. Zhang, L. Wang, K. Wu, et al., Automatic data transformation using large language model-an experimental study on building energy data, 2023 ieee international conference on big data (bigdata) (2023) 1824–1834, 2023.
  - S. Ghazzai, D. Grigori, B. Benatallah, R. Rebai, Harnessing gpt for data transformation tasks, 2024 ieee international conference on web services (icws) (2024) 1329–1334, 2024.
  - S. S. Norouzi, A. Barua, A. Christou, N. Gautam, A. Eells, P. Hitzler, C. Shimizu, Ontology population usingllms, 2024. doi:10.48550/arXiv.2411.01612.
  - S. Carta, A. Giuliani, L. Piano, A. S. Podda, L. Pompianu, S. G. Tiddia, Iterative zero-shot llm prompting for knowledge graph construction, 2023. doi:10.48550/arXiv.2307.01128
  - J. Frey, L. P. Meyer, N. Arndt, F. Brei, K. Bulert, Benchmarking the abilities of large language models for rdf knowledge graph creation and comprehension: How well do llms speak turtle?, 2023. doi:10.48550/arXiv.2309.17122.

Thank you for your attention



“keep the mapping explicit”