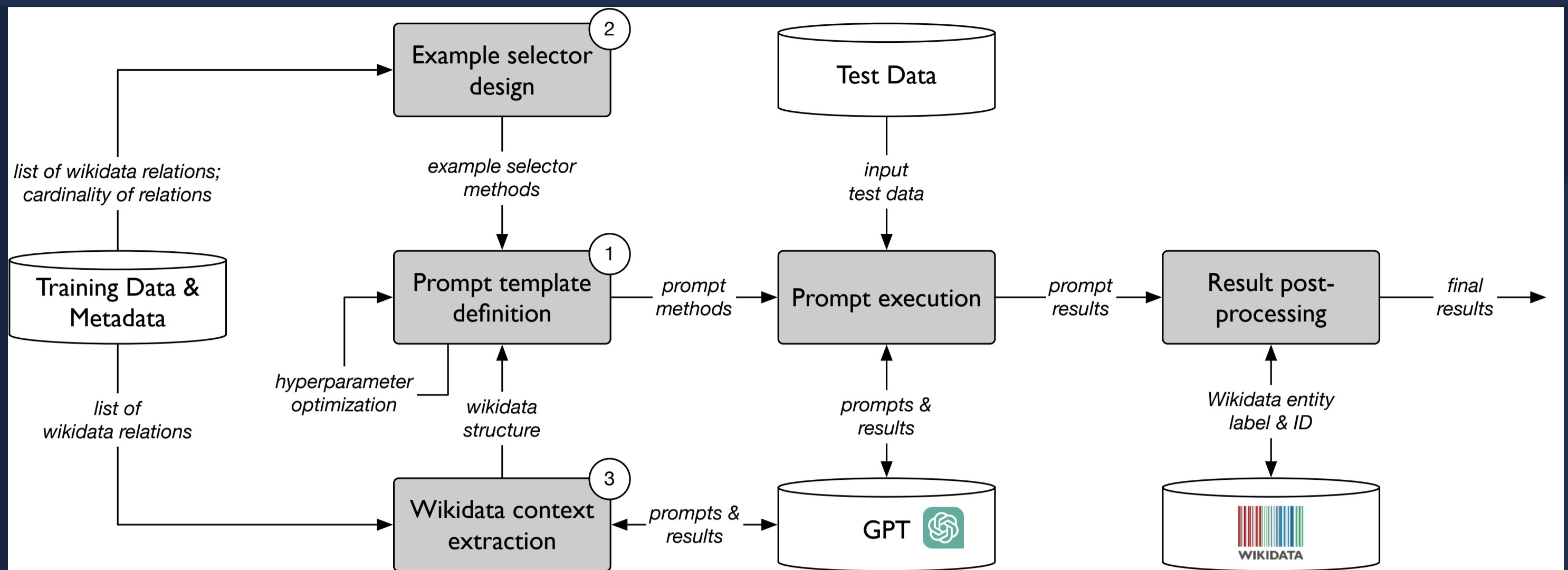# Knowledge-centric Prompt Composition for Knowledge Base Construction from Pre-trained Language Models

Xue Li, Anthony Hughes, Majlinda Llugiqi , Fina Polat , Paul Groth and Fajar J. Ekaputra

Thames' System Overview: Incorporating WikiData information and rule-based example selector in our prompts help with prompting performance.

🏆 2nd for Track 2

1. Introduction

- Given the input subject-entity (s) and relation (r), the task is to predict all the correct object-entities ({o1, o2, ..., ok}) using LM probing.
- E.g. (Matt Damon, person has number of children, ?).
- We participated for Track 2, applying in-context learning for GPT-3.5 and GPT4.

2. Methods

- Knowledge-enriched prompts: we prompt GPT-3.5 to get WikiData information about the objects and subjects such as the type, WikiData IDs and the explanation of the relation.
- We compared two example selectors: rule-based selector that accounts for the number of possible objects; and the similarity-based selector that is based on vector similarities.

3. Results

- We scored 2nd for Track 2.
- Rule-based selector outperforms similarity-based selector for 4 points.
- GPT-4 outperforms GPT-3.5 for ~9 points.

4. Discussion

- Contextual relevance matters in In-Context Learning.
- Incorporation of cardinality related rules help with the performance.
- GPT-4 still suffers from hallucination for certain relation types.

```
Your task is to predict objects based on the given subject and relation.
- Given Subject: ('AT&T','Q35476')
- Subject Type: 'organization'
- Object Type: 'organization'
- Relation: 'CompanyHasParentOrganisation'
- Relation Wikidata ID: 'P749'
- Relation Label (Wikidata): 'parent organization'
- Relation Explanation (Wikidata): 'This property is used to indicate
  the parent organization of a company.'
==>
Predicted Objects:
```

An example Wikidata context usage within the prompt.

Model	Selector	Precision	Recall	F1
GPT-3.5	Similarity-based	0.5595	0.6154	0.5484
	Rule-based	0.6105	0.6492	0.5863
GPT-4	Rule-based	**0.7128**	**0.6894**	**0.6744**

*Results for each of the presented prompt selection methodologies, and for each model utilized in the experiments. Highest scores are in **bold**.*

```
prefix = """
Imagine you are emulating Wikidata's knowledge.
Your task is to predict objects based on the given subject and relation.
Below are some examples for your reference: """
example_formatter_template = """..."""
suffix = """
End of examples. Now, it's your turn. Please only give correct answers.
The answers shall not contain duplicates and the number of answers
shall be between {} to {}. :
- Given Subject: ('{}','{}')
- Subject Type: '{}'
- Object Type: '{}'
- Relation: '{}'
- Relation Wikidata ID: '{}'
- Relation Label (Wikidata): '{}'
- Relation Explanation (Wikidata): '{}'
==>
Predicted Objects:
"""
```
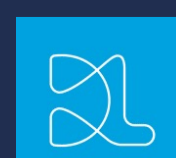
Prefix and Suffix for the prompt.

UNIVERSITY OF AMSTERDAM

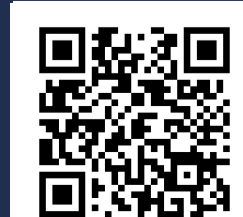TU WIEN — TECHNISCHE UNIVERSITÄT WIEN

Xue Li
corresponding author
✉ [x.li3@uva.nl]

Scan the QR code to check our github repo.