Navigating Ontology Development with Large Language Models

Javad Saeedizade, Eva Blomqvist

May 26-30 in Hersonissos, Greece
ESWC 2024

LINKÖPING UNIVERSITY

ChatGPT

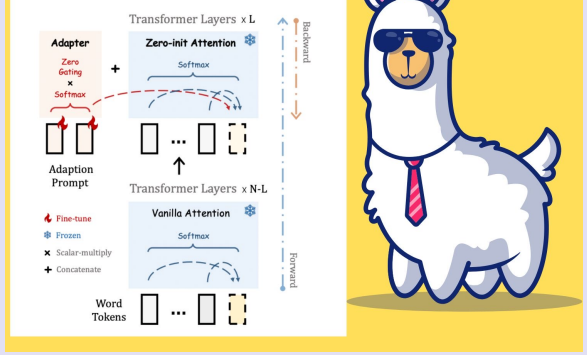
Gemini

BING CHAT

Llama 2

Meta AI

Falcon 40B LLM



Bard

Google BERT

What is ontology story

An example ontology story from the semantic web course

During each year a number of theatre festivals are held in cities around Italy. In January 2007 …

Competency questions

When and **Where** did a certain theatre festival take place?

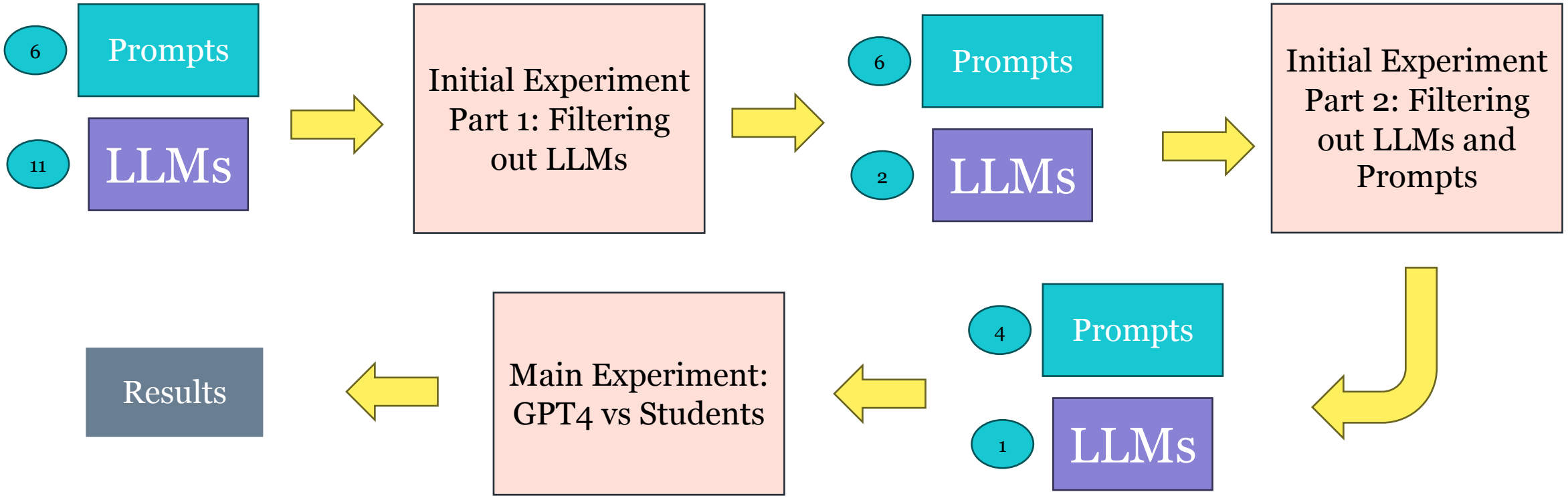
Research Questions

1- To what extent can LLMs create an ontology that meets the Ontology Requirements?

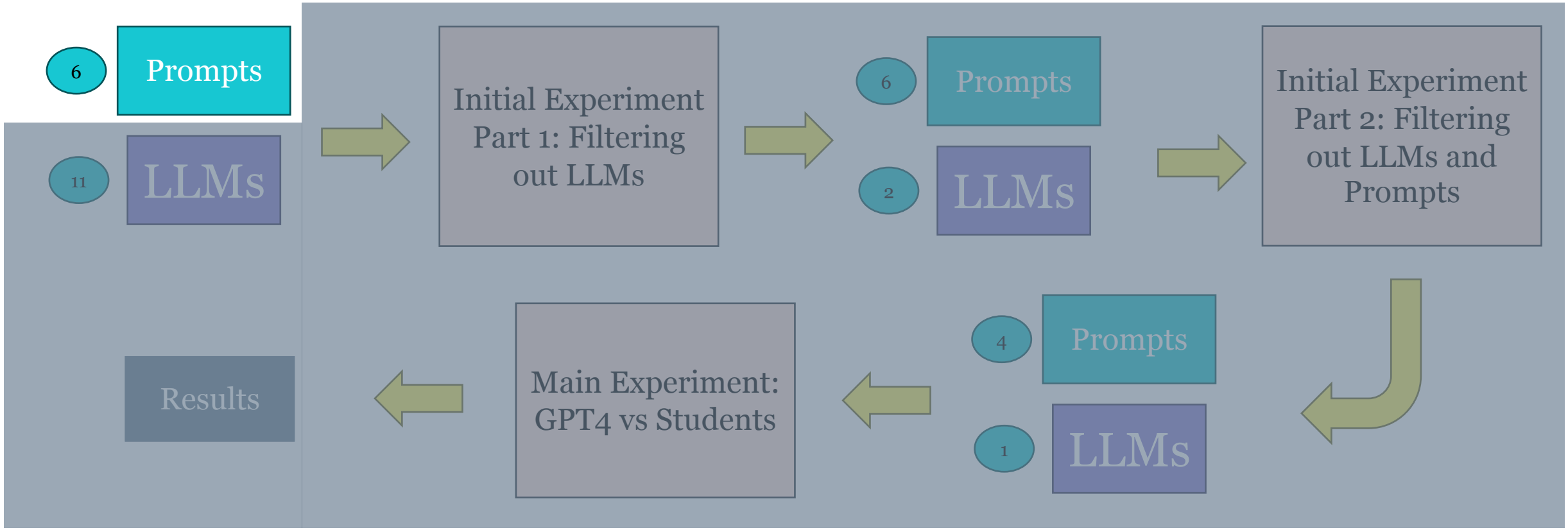
2- Which LLMs are suitable for this task?

3- What prompting techniques are the most effective?

Structure of the presentation



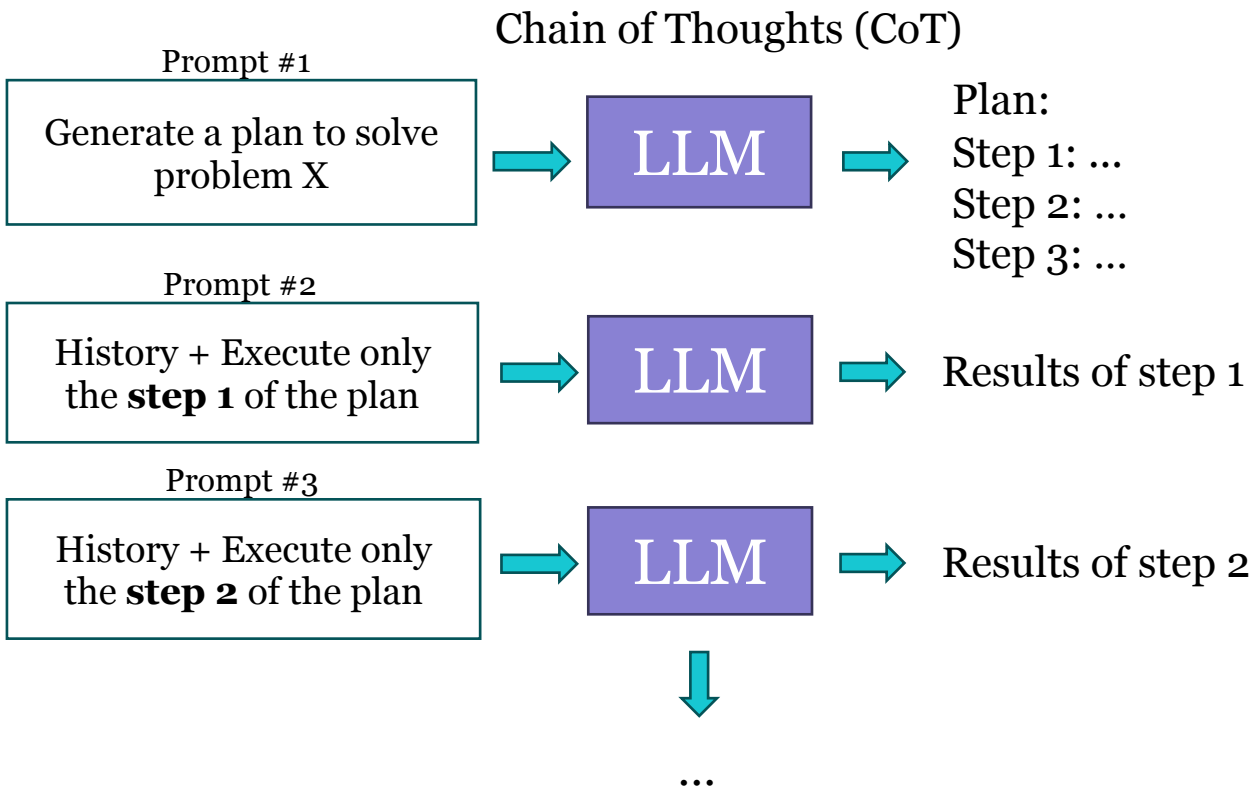
Structure of the presentation



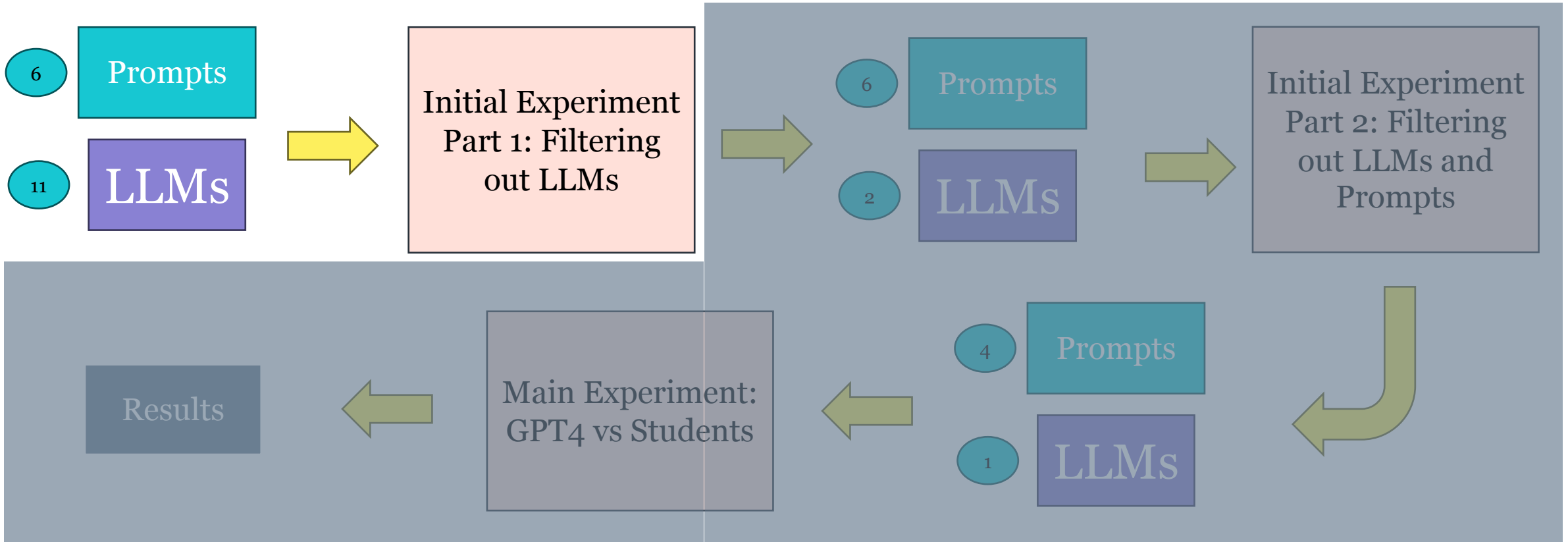
What is a prompt?

What is a prompting technique?

- Zero-Shot
- Chain of Thoughts (**CoT**)
- Chain of Thoughts Self-Consistency (**CoT-SC**)
- Sub-task Decomposed Prompting - Waterfall approach (**Waterfall**)
- Sub-task Decomposed Prompting - Competency Question by Competency Question (**CQbyCQ**)
- Graph of Thoughts (**GoT**)

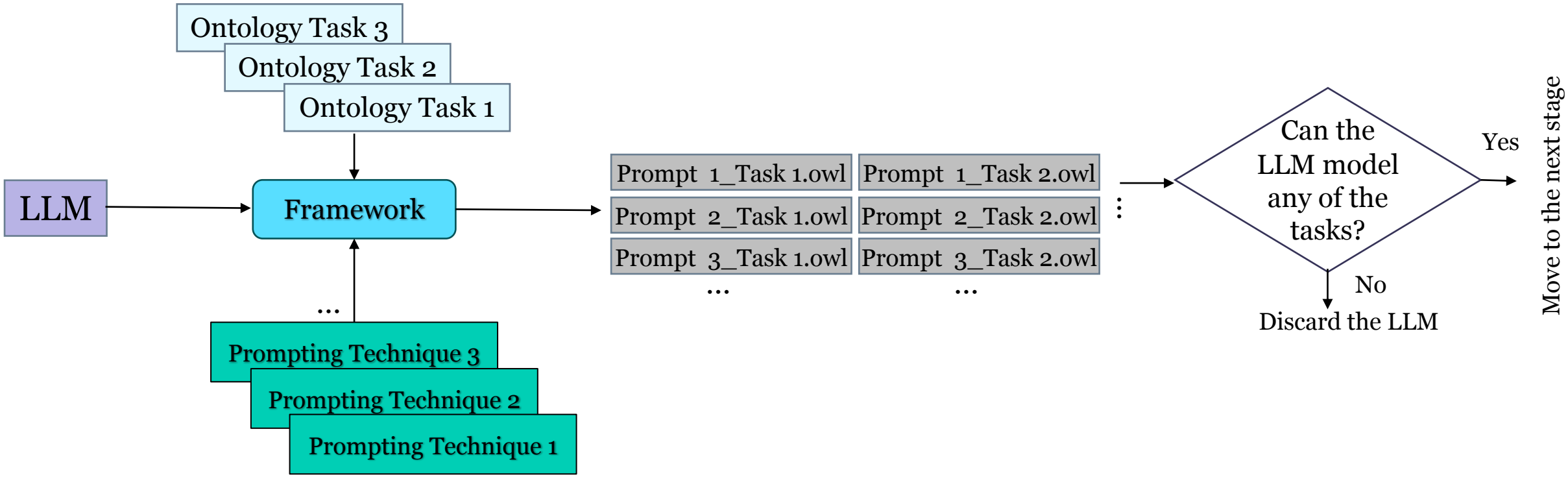


Structure of the presentation



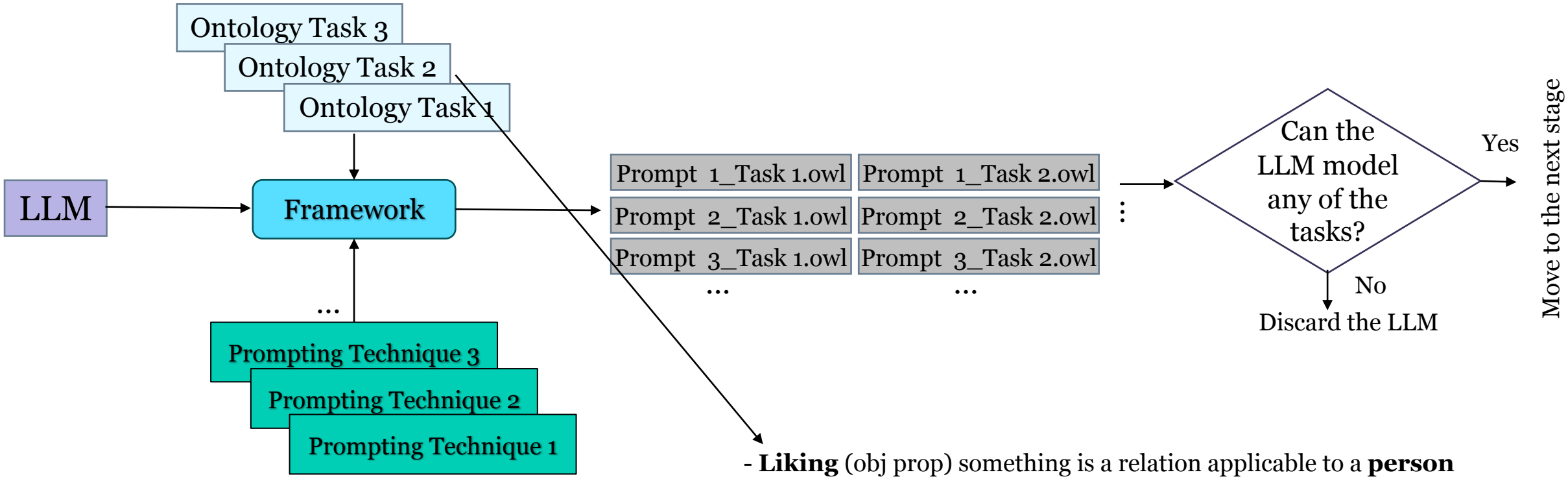
Initial Experiment: Phase 1

Which LLMs are suitable for ontology generation?



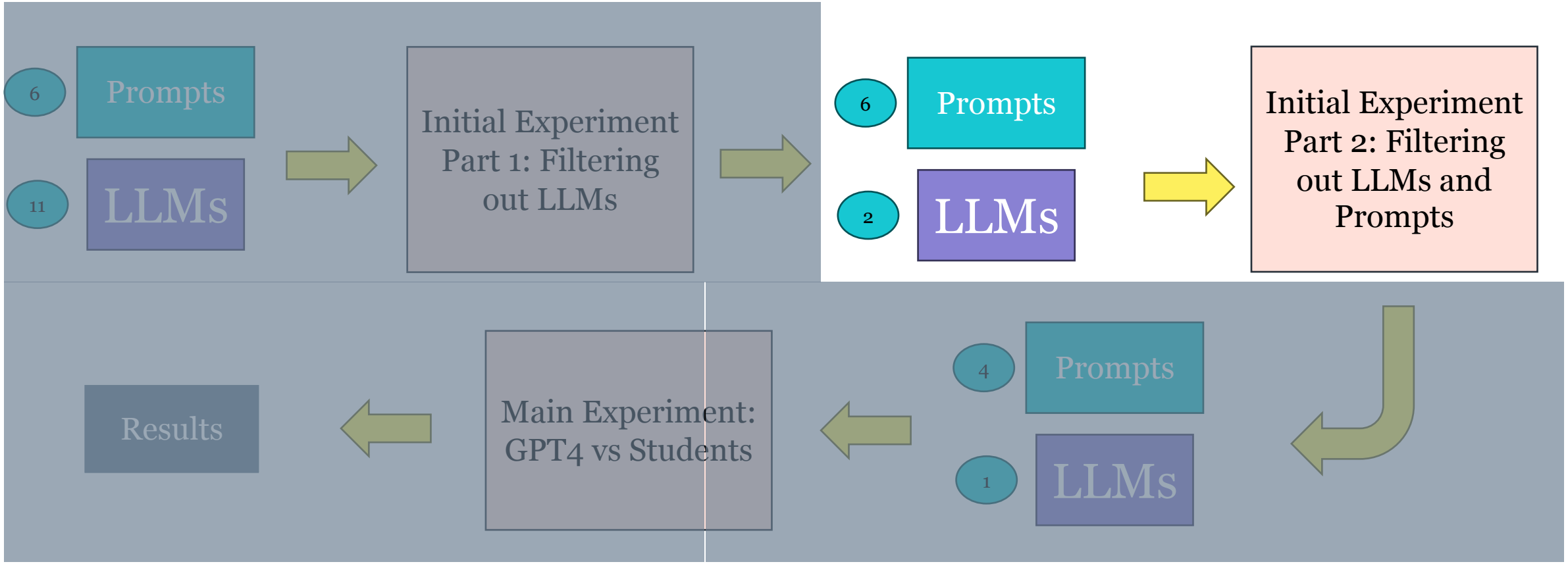
Initial Experiment: Phase 1

Which LLMs are suitable for ontology generation?



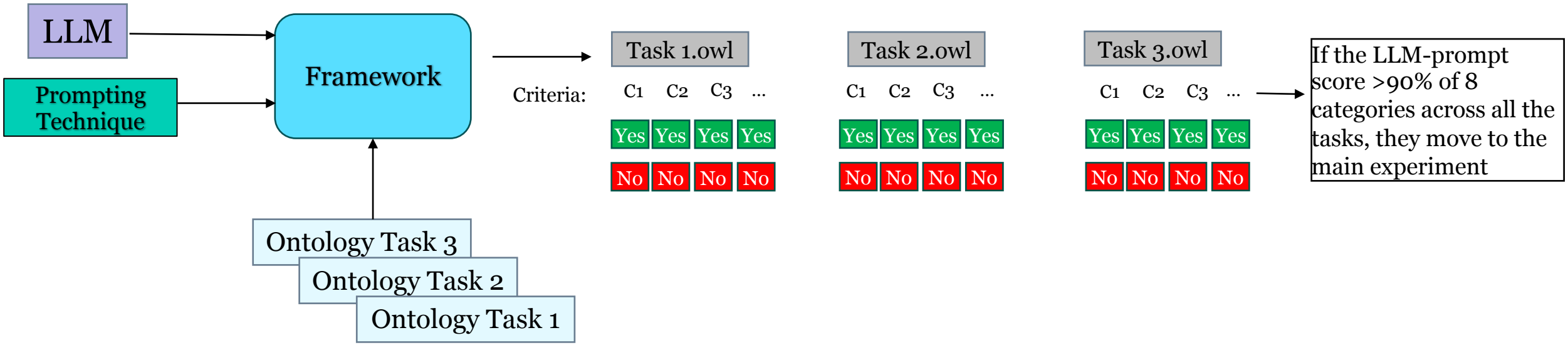
- **Liking** (obj prop) something is a relation applicable to a **person**
- **Adore** (obj prop) is only for **cars** (only cars are adored)
- Thomas adores Lena.

Structure of the presentation



Initial Experiment: Phase 2

Which pair of LLMs and prompts are suitable for ontology generation?



1: Correctness of the Turtle syntax.
2: Usage taxonomy (class hierarchy) in the ontology
3: Usage of Data Properties (if needed)
4: Presence of reification class

5: Presence of instances (if needed)
6: Establishment of domain or range restriction for properties
7: Presence of “EquivalenceClass” restriction
8: Semantic coherence of class hierarchy

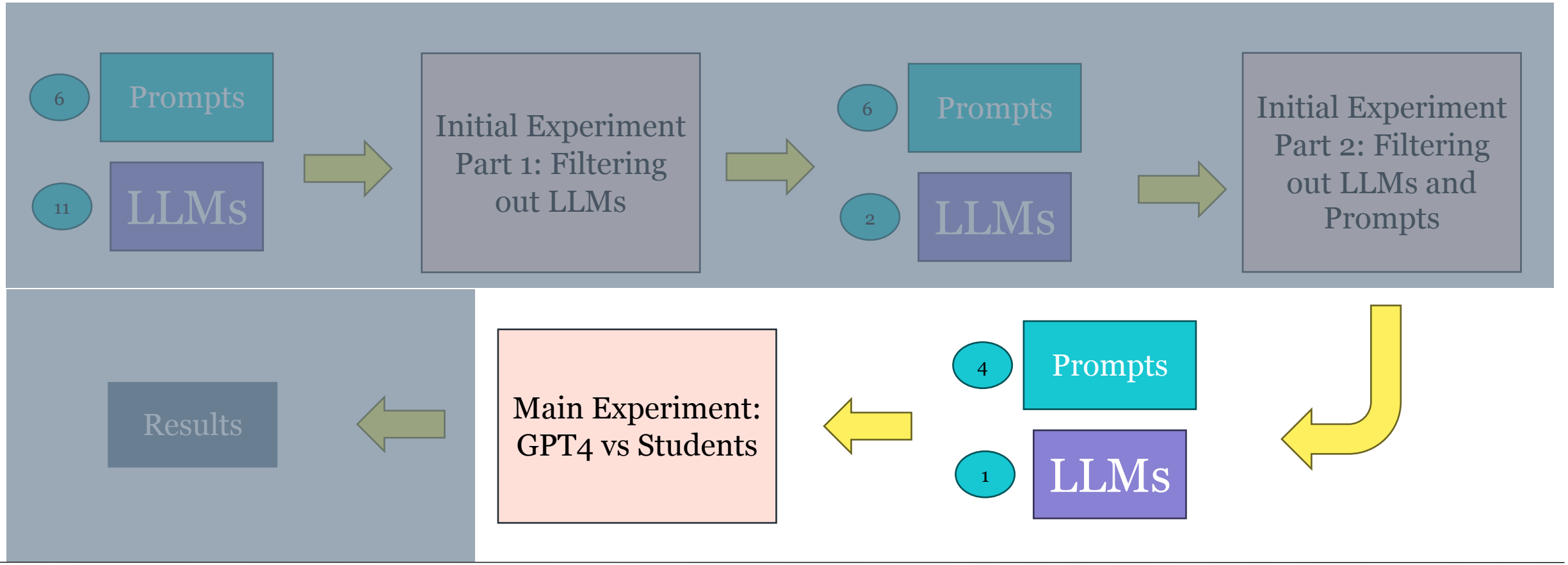
Result of Initial Experiment

LLMs: **GPT-4**

Prompting Techniques:

- **CoT** (chain of thoughts)
- **CoT-SC** (Chain of Thoughts Self-Consistency)
- **CQbyCQ** (Sub-task Decomposed Prompting - Competency Question by Competency Question)
- **GoT** (Graph of Thoughts)

Structure of the presentation



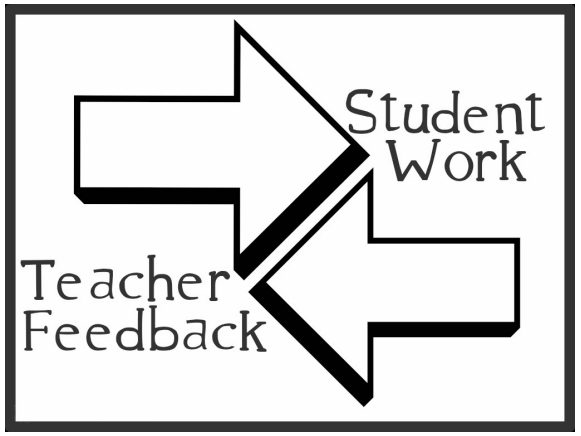
Baseline

Baseline: Master's students who took a **semantic web course**

Course: **3** ontology tasks/stories, **15 competency** questions each

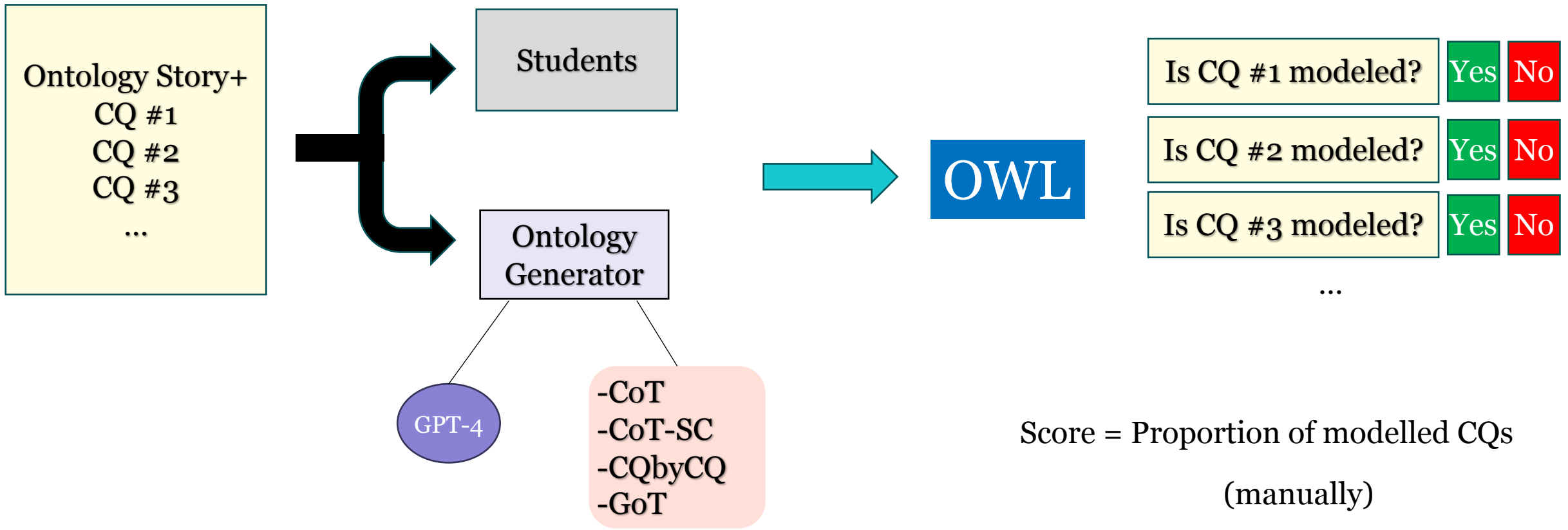
10 groups, and **2 students** on average in each group

Students submitted each modelling on average **3 times** to pass the lab

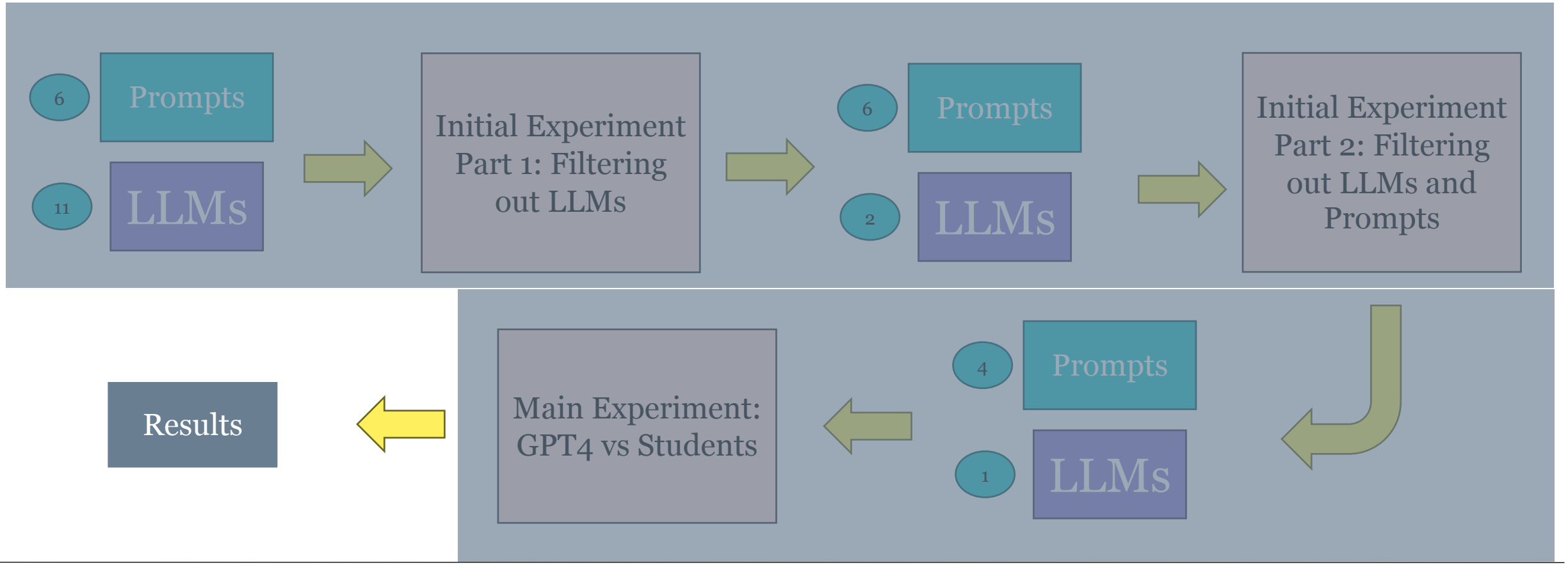


The baseline is the **first** and the **last student submissions** of the course.

Evaluation Criteria



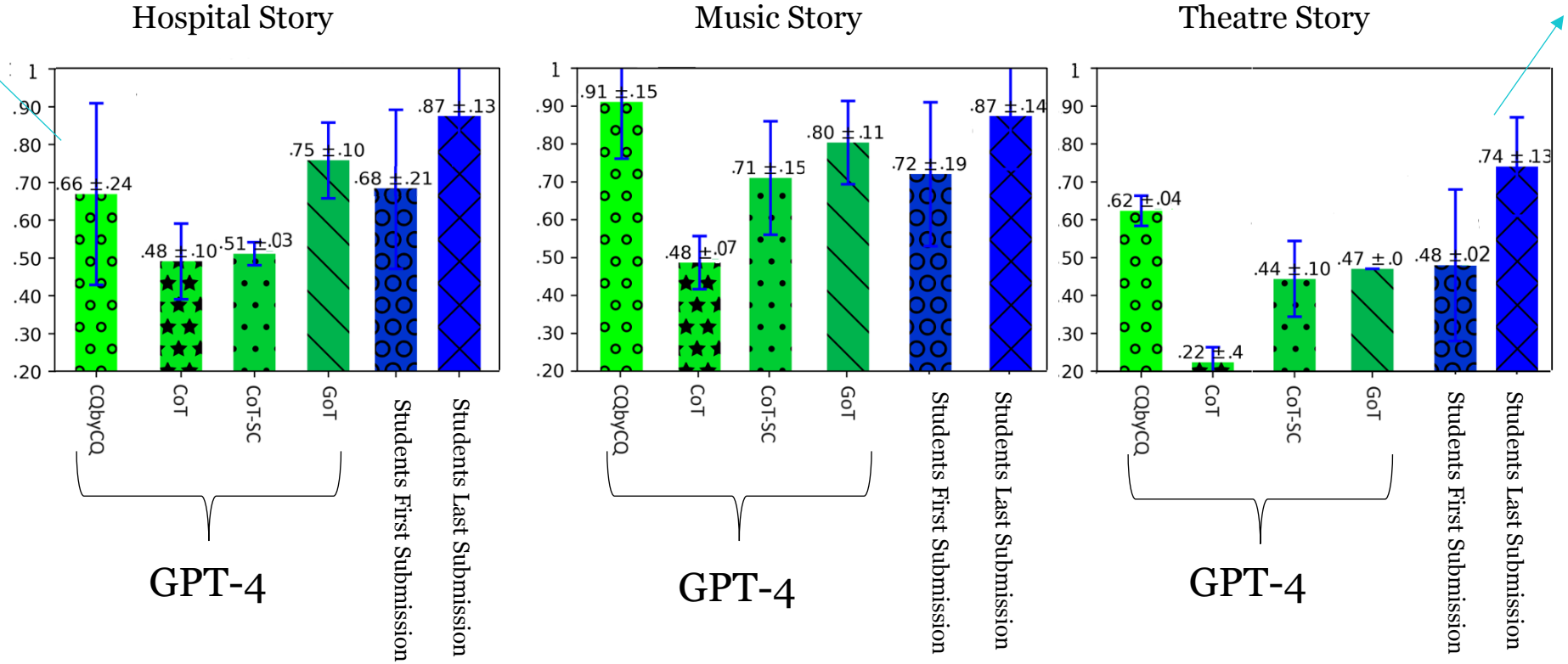
Structure of the presentation



Results

Std: we ran models 3 times

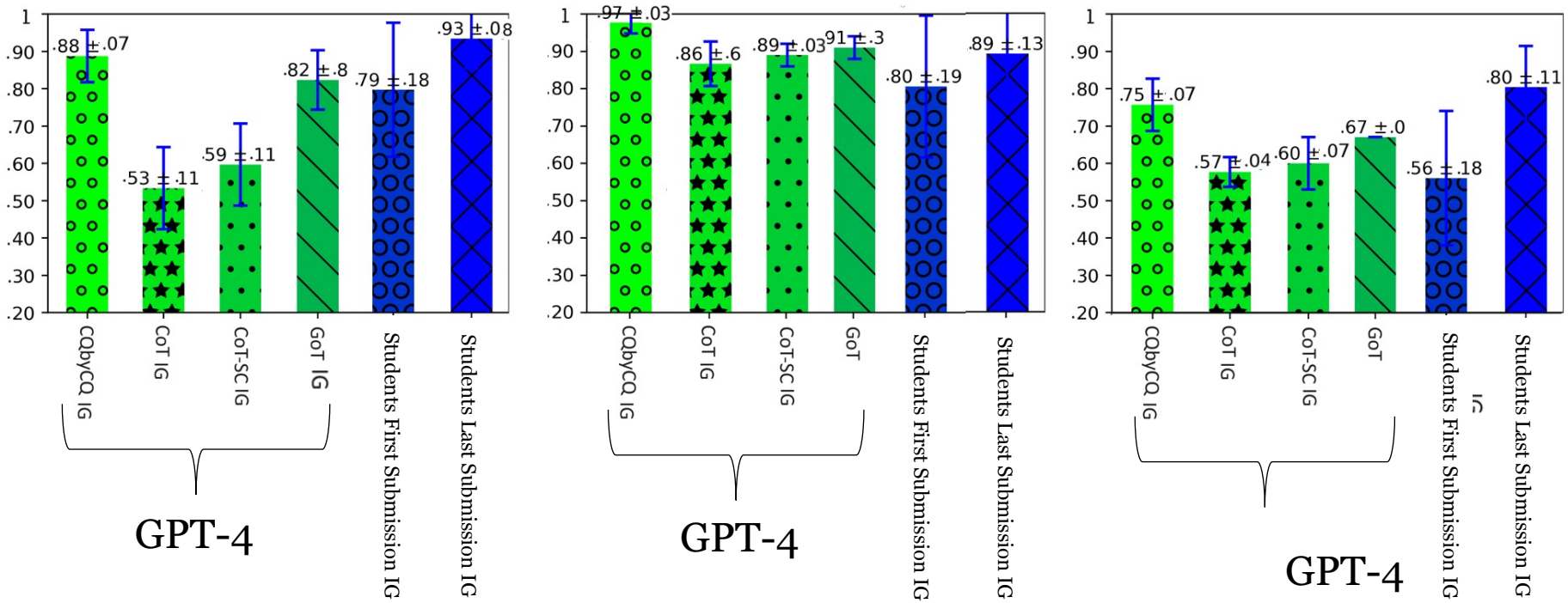
Std of students' submissions



CQbyCQ	GoT	Stu F S	Stu L S
0.73	0.67	0.62	0.82

Results with a relax criteria

missing one single data/object property



CQbyCQ	GoT	Stu F S	Stu L S
0.86	0.80	0.71	0.87

Evaluation: types of CQs

	Theatre Story				Music Story				Hospital Story			
	DP	OP	Reif.	Rest.	DP	OP	Reif.	Rest.	DP	OP	Reif.	Rest.
CQbyCQ	**.93**	**.78**	.55	.33	**1.0**	**.94**	**1.0**	**1.0**	**1.0**	.94	**1.0**	0
CoT	.80	.50	.22	**1.0**	.80	.78	**1.0**	**1.0**	.53	.50	.44	**1.0**
CoT-SC	**.93**	.55	0	**1.0**	.80	.89	**1.0**	**1.0**	.53	.66	.66	.33
GoT	.86	.61	.44	**1.0**	.80	**.94**	**1.0**	**1.0**	.73	**1.0**	.66	.66
Students First Submission	.64	.63	.43	.10	.84	.83	.87	.30	.78	.85	.73	.60
Students Last Submission	.92	**.78**	**.73**	.50	.94	.92	.90	.50	.94	**1.0**	.90	.60

Simpler
- DP: Simple data property modeling
- OP: Simple object property modeling

More Complex
- Reif.: Reifications
- Rest.: Restrictions on classes or properties

Evaluation: types of CQs

	Theatre Story				Music Story				Hospital Story			
	DP	OP	Reif.	Rest.	DP	OP	Reif.	Rest.	DP	OP	Reif.	Rest.
CQbyCQ	**.93**	**.78**	.55	.33	**1.0**	**.94**	**1.0**	**1.0**	**1.0**	.94	**1.0**	0
CoT	.80	.50	.22	**1.0**	.80	.78	**1.0**	**1.0**	.53	.50	.44	**1.0**
CoT-SC	**.93**	.55	0	**1.0**	.80	.89	**1.0**	**1.0**	.53	.66	.66	.33
GoT	.86	.61	.44	**1.0**	.80	**.94**	**1.0**	**1.0**	.73	**1.0**	.66	.66
Students First Submission	.64	.63	.43	.10	.84	.83	.87	.30	.78	.85	.73	.60
Students Last Submission	.92	**.78**	**.73**	.50	.94	.92	.90	.50	.94	**1.0**	.90	.60

Simpler
- DP: Simple data property modeling
- OP: Simple object property modeling

More Complex
- Reif.: Reifications
- Rest.: Restrictions on classes or properties

Conclusion

RQ1- To what extent can LLMs create an ontology that meets the Ontology Requirements?

It is possible to generate ontology to meet ontology requirements to the same quality as students generating ontology.

RQ2- Which LLMs are suitable for this task?
RQ3- What prompting techniques are the most effective?

The CQbyCQ prompting using GPT-4 outperformed the first students' submissions.

Students' last lab submissions achieved the highest scores, but the performance of GPT-4 with CQbyCQ prompting was closer to this than the students' first submissions.

Thank you for attending

Any questions?

Code and the paper