AIKLR: A Co-Attentive Infusion of Structural Knowledge into Language Models towards Legal Textual Inference

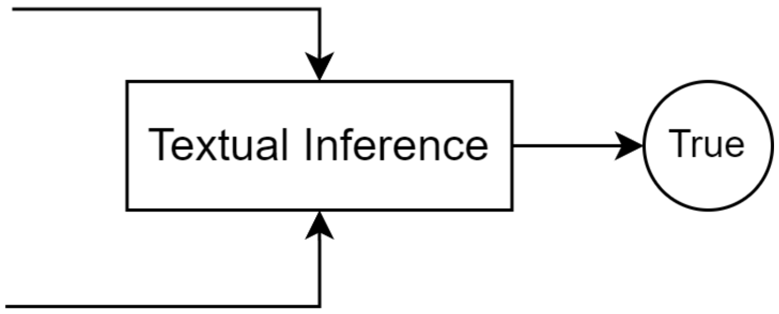
Tan-Minh Nguyen, Thi-Hai-Yen Vuong, Hoang-Trung Nguyen, Le-Minh Nguyen, Xuan-Hieu Phan

The Web Conference 2025 – Emerging World Track

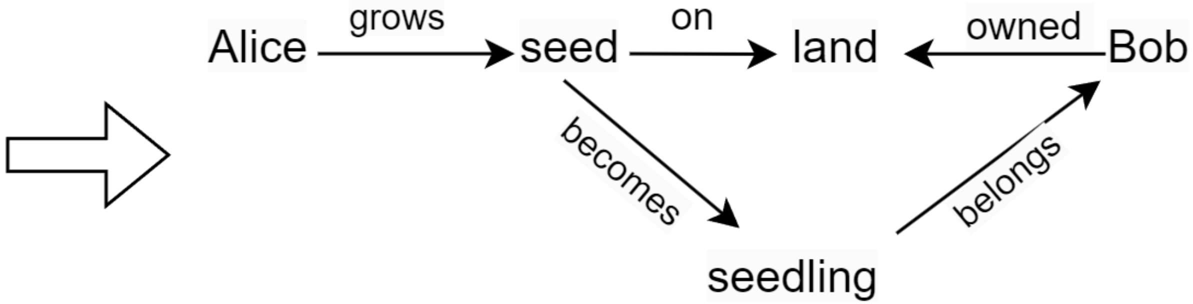
Legal Textual Inference

Hypothesis: Alice grows her seed on a land owned by **Bob**. If the seed grows to become a seedling, ownership of the seedling belongs to **Bob**.

Premise: Article 242 (Accession to Immovables) The owner of immovables ...

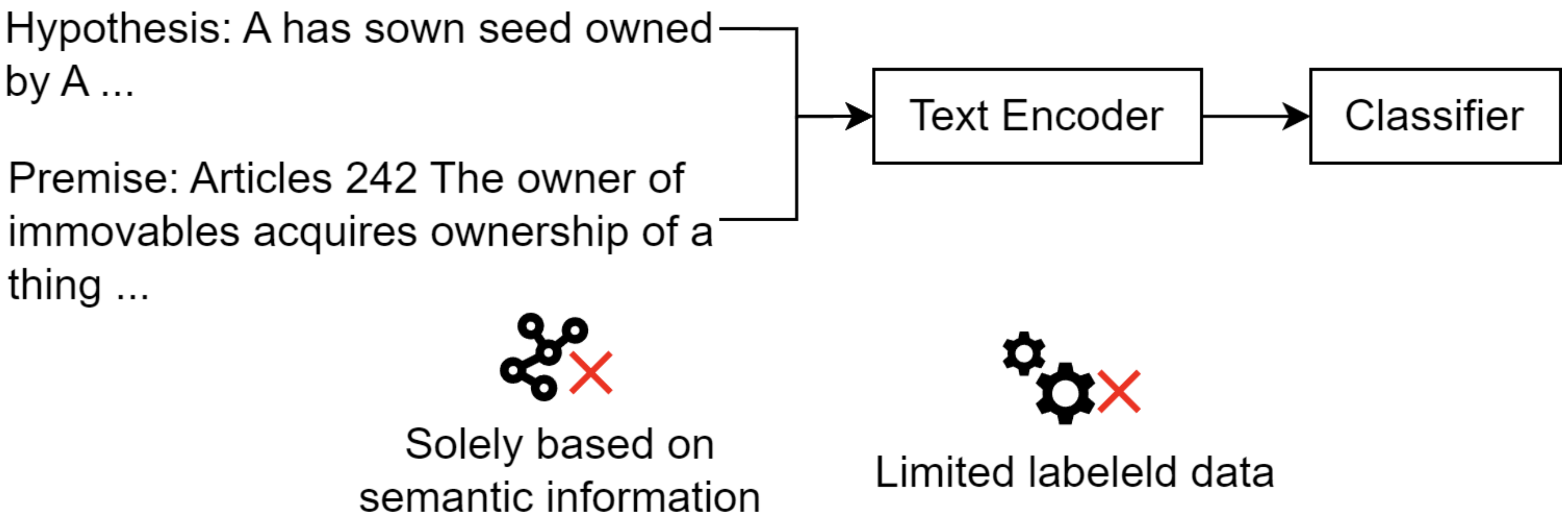


Hypothesis



Problems

- Inefficient fine-tuning of pre-trained language models (PLMs)



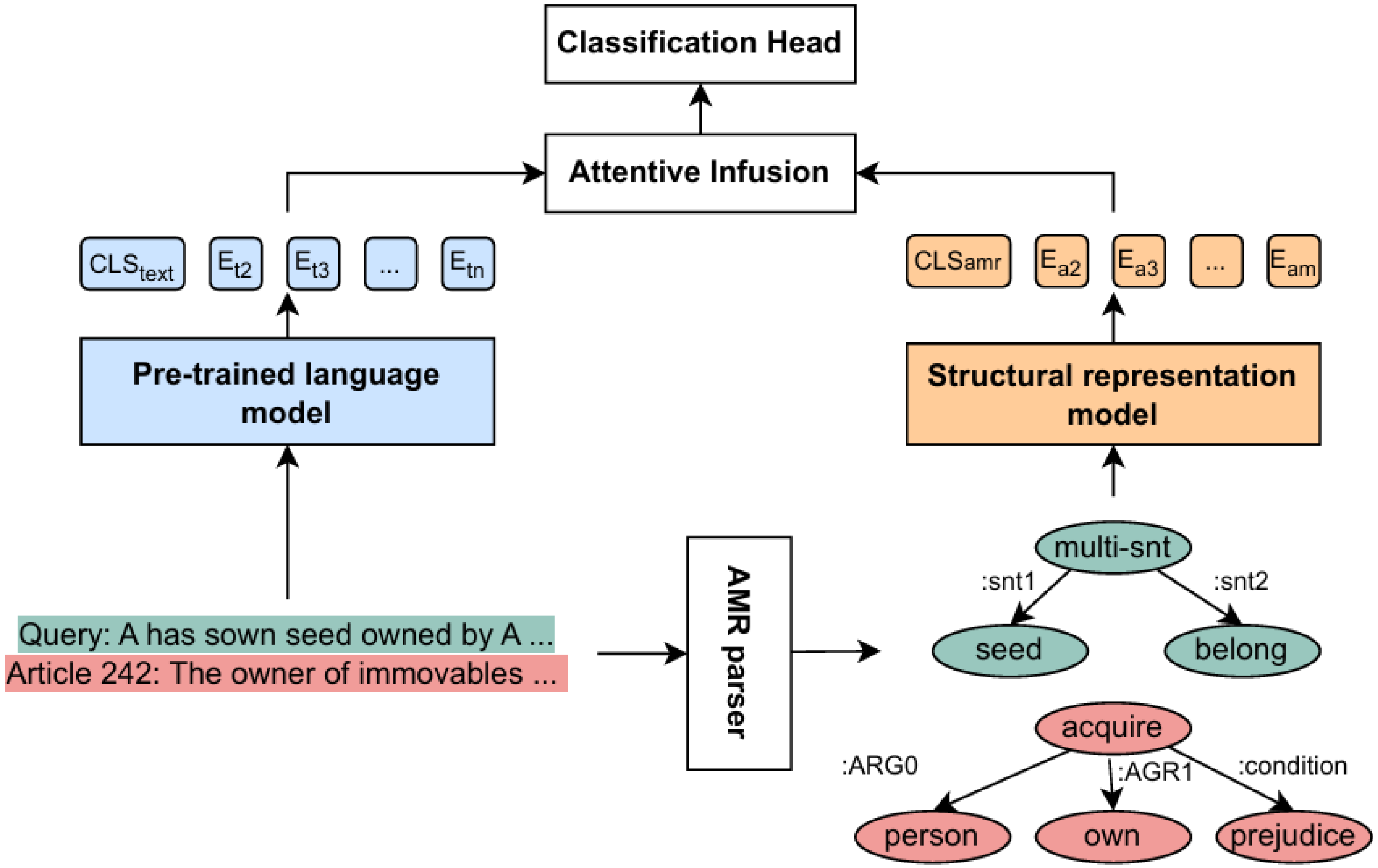
Key contributions

- We introduce **AIKLR** to improve the **inference quality of PLMs**, allow models to learn from **both structural and semantic patterns** without scaling up of model size and training data.

- Results on **2 benchmark** show significant performance gain on legal textual inference, achieving comparative with a number of **recent LLMs.**

AILKR Architecture

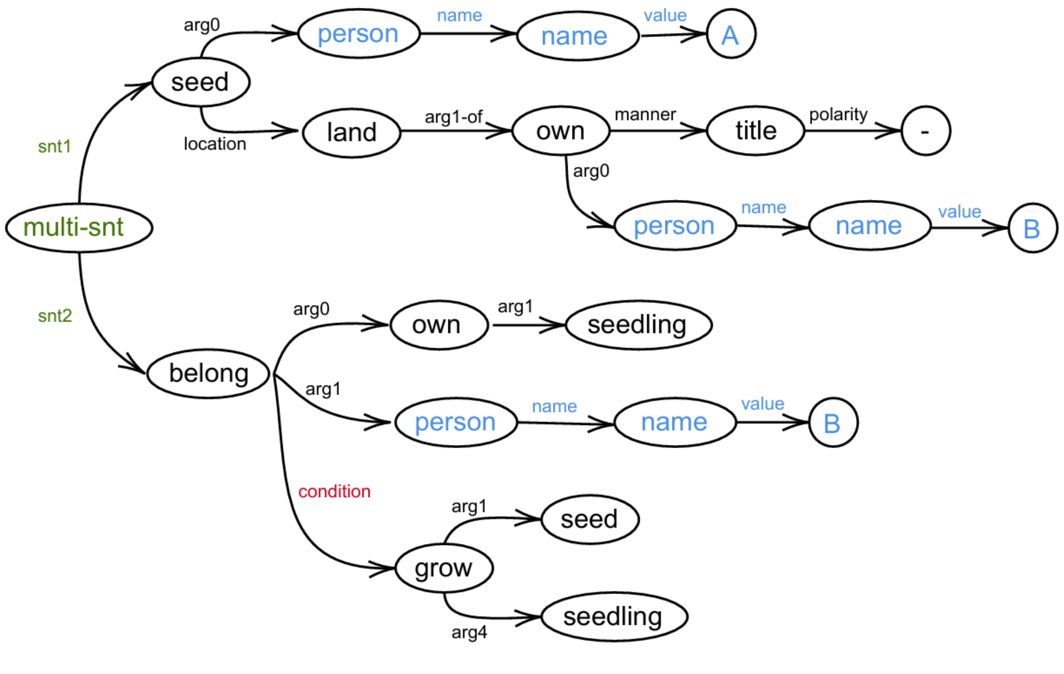
- Text Representation
- Structure Representation
- Attentive Infusion



Structural Representation

- Abstract Meaning Representation
- Directed acyclic graph
  - Node: concept
  - Edge: relation
- Encoded without losing any adjacency information

Sentences: A has sown seed owned by A on land owned by B without title. If the seed grows to become a seedling, ownership of the seedling belongs to B.



Attentive Infusion

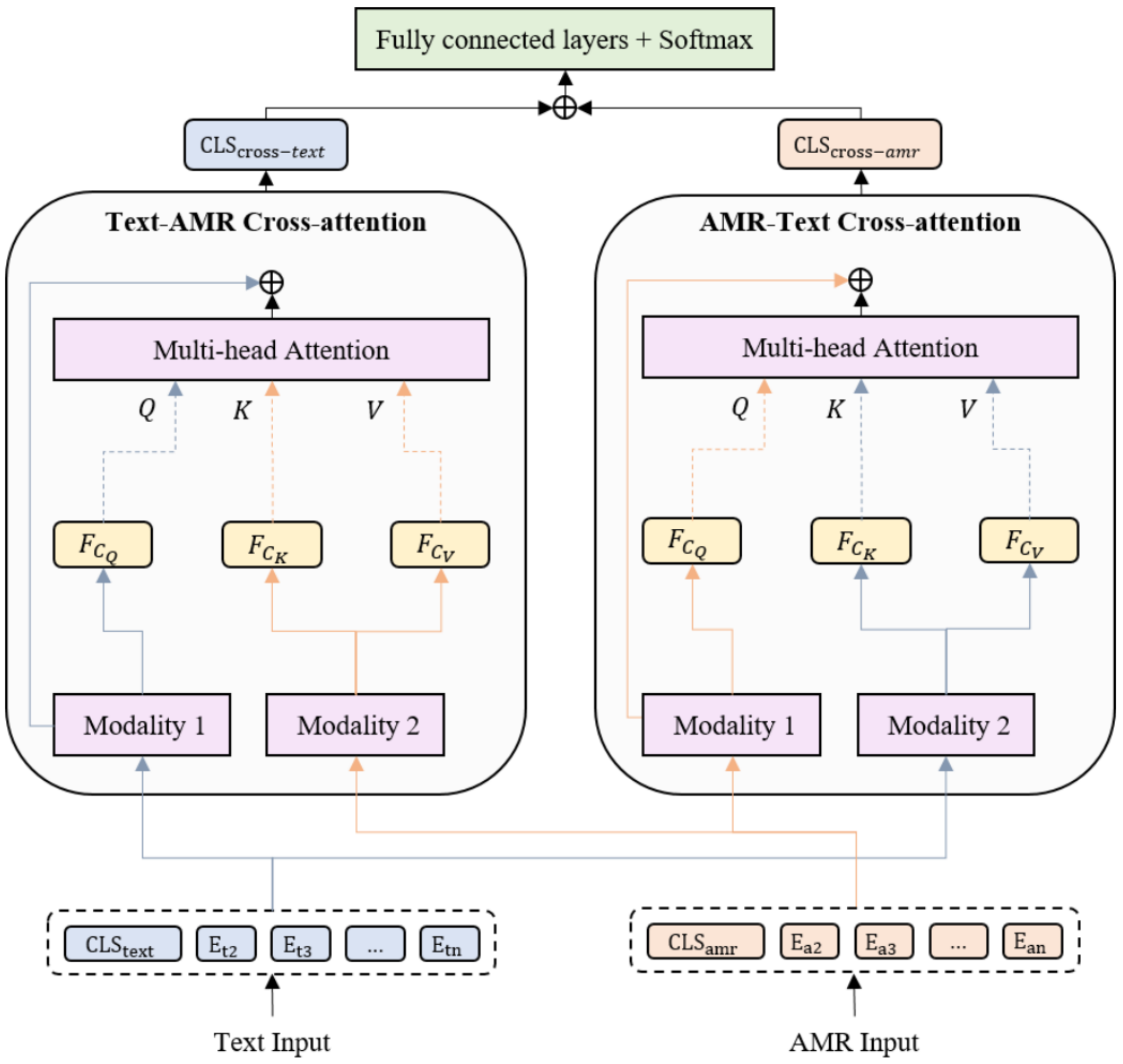
Fuse semantic and structure information

- Co-attentive block
- Residual connection

- Cross Entropy loss

$$\mathcal{L}_{CE} = -\sum_{i=1}^{N}\sum_{c=1}^{2} w_c \log \frac{\exp(x_{i,c})}{\sum_{j=1}^{2}\exp(x_{i,j})}$$

x is the input, y is the target, N spans the minibatch



Experiments

- Dataset
  - COLIEE 2023
  - COLIEE 2024
  - Input: {Statement, Articles}
  - Output: {Yes, No}

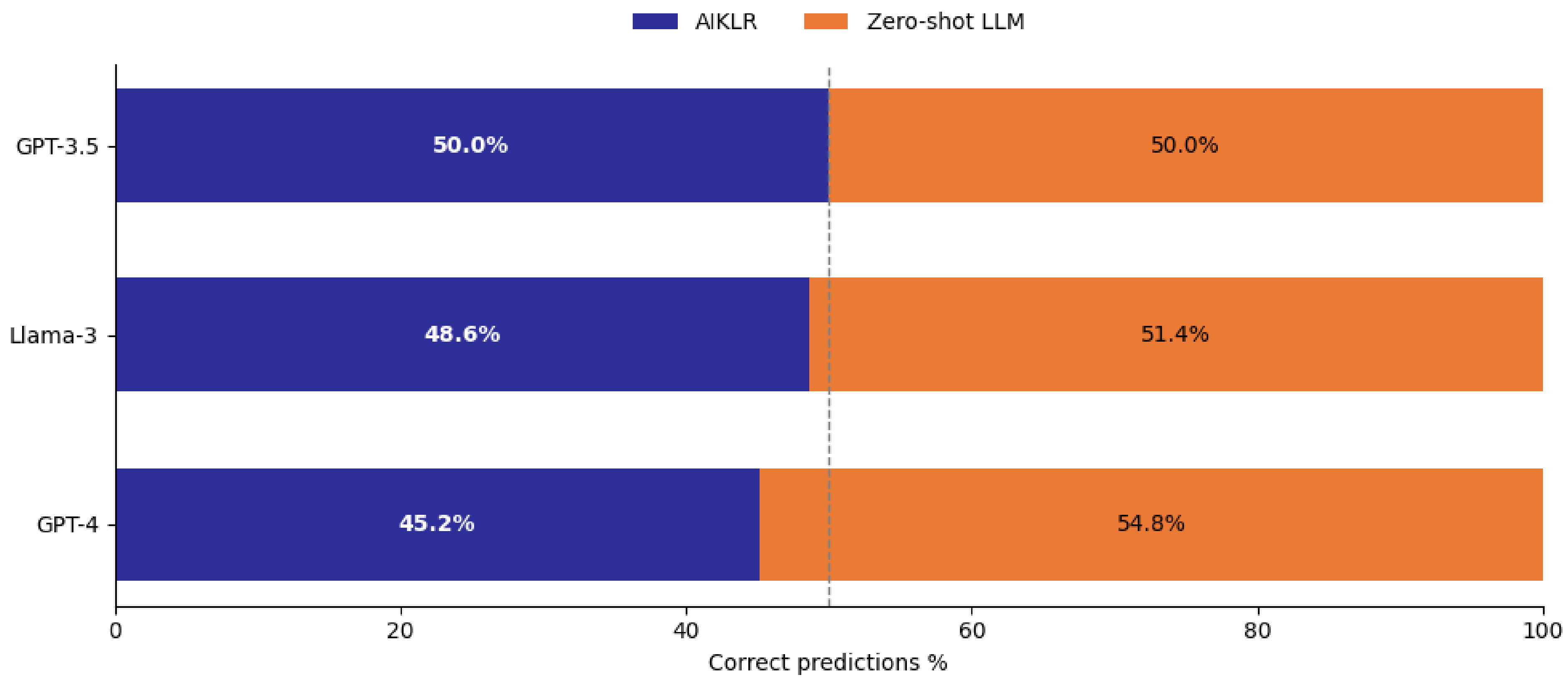
- Metric
  - Accuracy

Attribute	COLIEE'23		COLIEE'24	
	Train	Test	Train	Test
# queries	996	101	1097	109
# relevant articles per query				
Maximum	6	2	6	2
Average	1.27	1.29	1.28	1.19
# tokens per query				
Minimum	13	25	13	16
Maximum	248	130	248	166
Average	62.21	65.28	62.49	67.91

Main Results

Model	Size	Accuracy	
		COLIEE'23	COLIEE'24
mBERT (baseline)	110M	0.6040	0.6880
LEGALBERT	110M	0.5841 (-3.29% pt)	0.6880 (+0.00% pt)
LEGALRoBERTa	125M	0.6040 (+0.00% pt)	0.6789 (-1.32% pt)
GPT-3.5	175B	0.7029 (+16.38% pt)	0.7706 (+12.00% pt)
Llama-3	70B	**0.7920 (+31.13% pt)**	0.7798 (+13.34% pt)
GPT-4	1.7T	0.7722 (+27.86% pt)	**0.7981 (+16.00% pt)**
AIKLR (Ours)	296M	0.6534 (+8.18% pt)	0.7706 (+12.00% pt)
Ensemble (Ours)	-	0.7129 (+18.03% pt)	**0.7981 (+16.00% pt)**

Compared to LLMs



Compare to others in COLIEE'24

Team	Brief Description	Model size	Accuracy
CAPTAIN	Fine-tuning FlanT5-xxl with data augmentation	11B	**0.8257**
JNLP	Ensembling multiple LLMs (Qwen-14B, Mixtral8x-7B, FlanT5, Flan-Alpaca)	14B	0.8165
UA	Prompt selection, zero-shot prompting FlanT5-xxl, voting ensemble	11B	0.7981
AMHR	Mixture of expert models, FlanT5-xxl	11B	0.7706
HI	Zero-shot inference Flan-Alpaca-GPT4-xl, preprocessing by GPT-3.5	3B	0.7523
NOWJ	Legal prompting Panda-7B-v0.1-GPTQ	7B	0.7523
OVGU+	Fine-tuning DeBERTa-base, Word Overlap, Contradiction Word Artefacts	184M	0.7064
KIS	LLM tuning, CoT interpretability, rule-based ensemble	10B	0.6972
MIG*		-	0.6330
Our methods			
AIKLR		296M	0.7706
Ensemble	Weighted ensemble	(2) 296M (2) 110M	0.7981

Conclusion

- We introduce **AIKLR** to improve the **inference quality**, allow models to learn from **both structural and semantic patterns** without scaling up of model size and training data.
- Results on **2 benchmark** show significant performance gain on legal textual inference, achieving comparative with a number of **recent LLMs.**
- Future work focuses on extending **AIKLR** on other tasks in Legal NLP

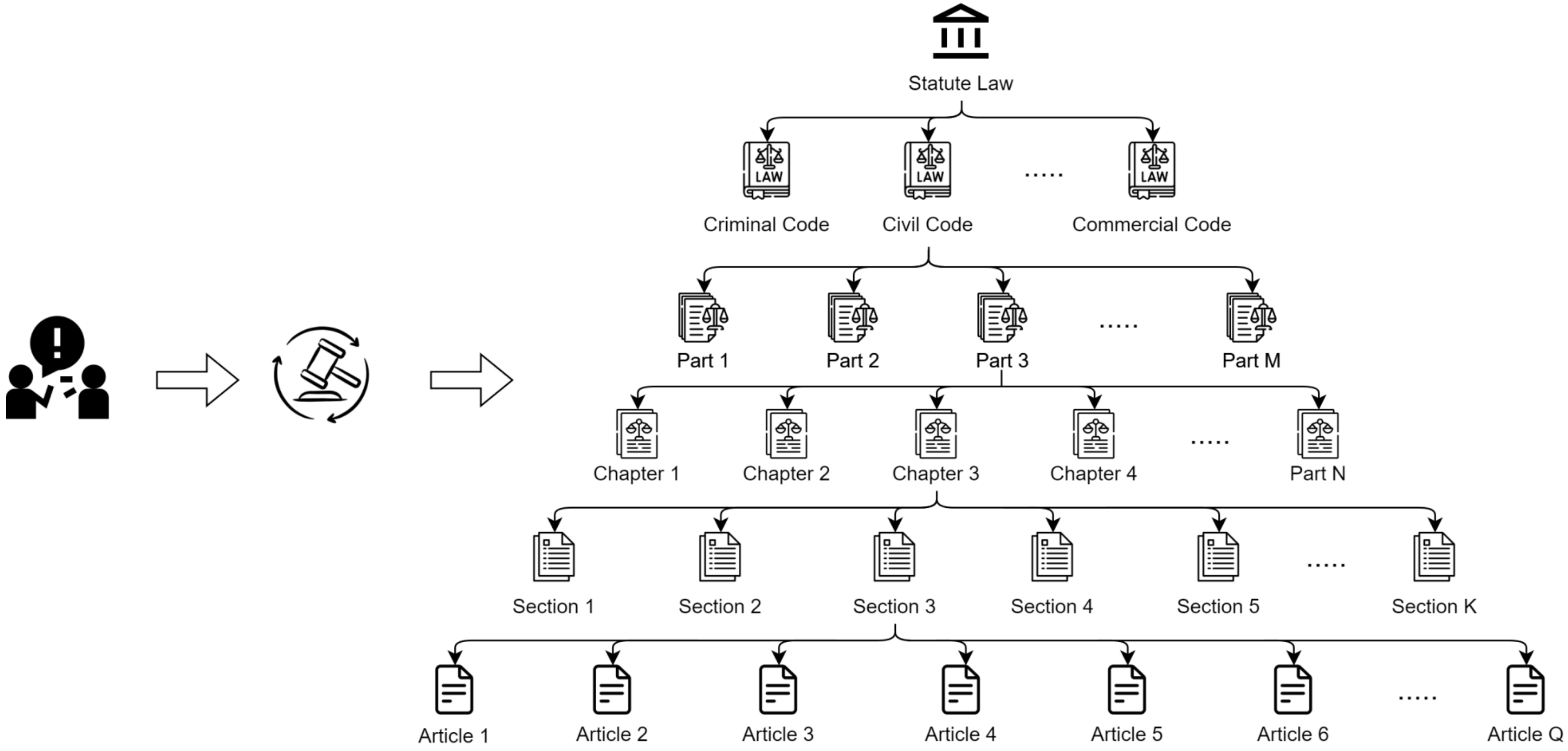
Thank you for listening
Q&A

minhnt@jaist.ac.jp

Ablation Study

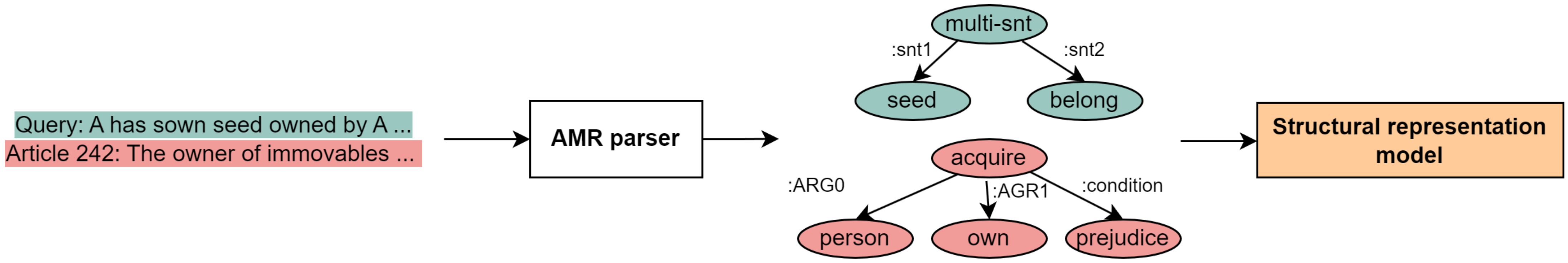
Model	AMR parser	AMR encoder	Accuracy	
			COLIEE'23	COLIEE'24
mBERT (baseline)			0.6040	0.6880
AMR-only	GSP	AMRSim	0.5544	0.5229
	SPRING	AMRSim	0.5346	0.5963
	SPRING	AMRBART	0.5247	0.5596
Shallow	GSP	AMRSim	0.6336	0.6880
	SPRING	AMRSim	0.6436	0.7339
	SPRING	AMRBART	0.6139	0.6880
AIKLR	GSP	AMRSim	0.6435	0.7431
	SPRING	AMRSim	0.6534	0.7706
	SPRING	AMRBART	0.6436	0.7114

Preliminaries – Statute law system



Structural Representation

- The AMR graph is encoded without losing any adjacency information

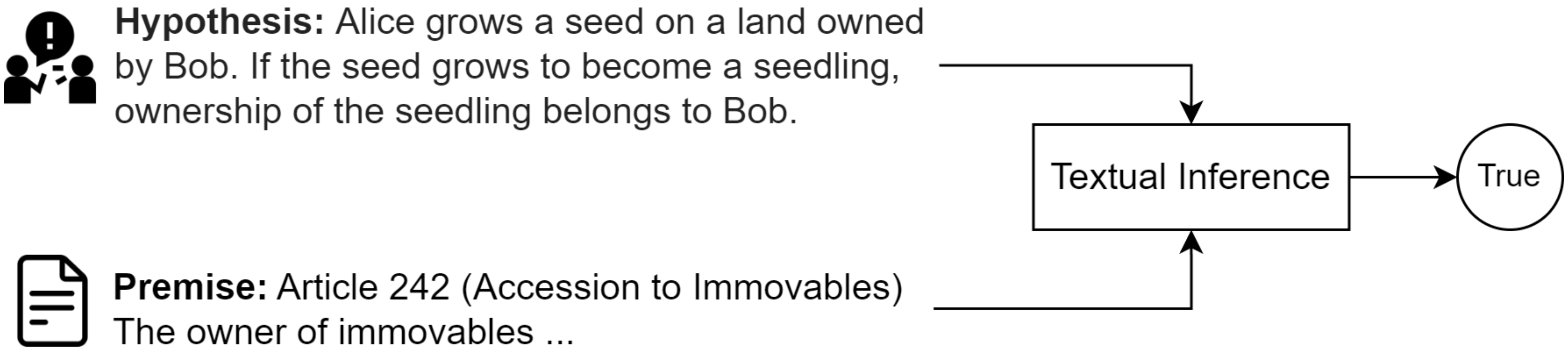


Problems

- Legal Textual Inference

"Does premise P support the hypothesis H?"

where P is legal articles, H is a legal statement (query).



Problems

- Complex relations between multiple entities

Question: "A has sown seed owned by A on land owned by B without title. If the seed grows to become a seedling, ownership of the seedling belongs to B."