

DOCUMENT-LEVEL CLAIM EXTRACTION AND DECONTEXTUALIZATION FOR FACT CHECKING

XPOLIC05, XMAKAI00, XKLEIN19

RESEARCH SUMMARY



Core Problem

The Challenge of Automated Fact-Checking for Czech
 This project adapts the English LOKI system to decompose Czech texts into verifiable "atomic claims" for fact-checking. We used machine translation and tested strategies like few-shot prompting and Chain of Thought.

OUR APPROACH

Adapted LOKI for Czech

Used local LLMs (Ollama)

Created 2 Czech datasets

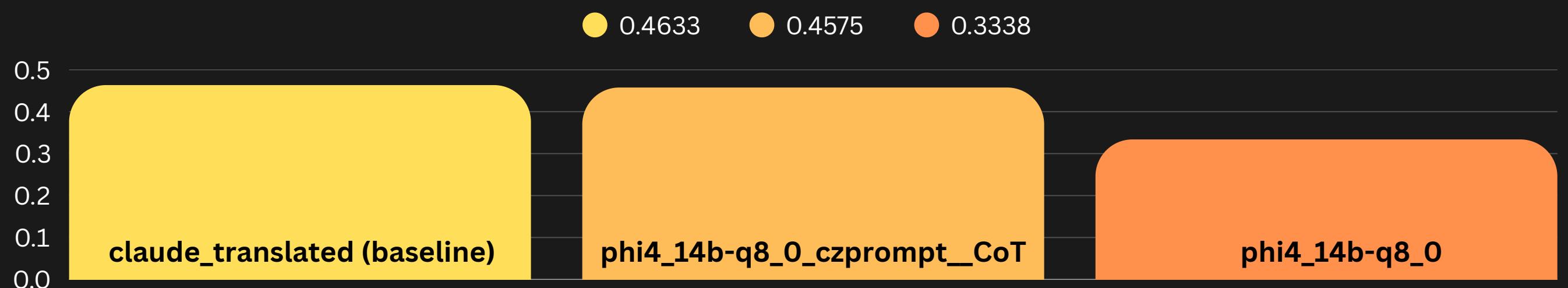
Evaluated with ROUGE-L, Edit Distance

Key Findings

Local models (esp. quantized Phi-4) can match/outperform larger LLMs with proper prompting

Quantized Phi-4 with CoT neared Claude's performance on Factcheck-Bench

On custom Czech comments, quantized Phi-4 was a top performer



Conclusion

Atomic claim decomposition for Czech is feasible with adapted LLMs. Local models offer a strong alternative to larger ones, with performance heavily influenced by prompt design and translation strategy.

More Information:

https://github.com/lmakaiova/KNN_FactChecking
