

ARG2ST AT CQS-GEN 2025

Critical Questions Generation through LLMs and Usefulness-based Selection

Alan Ramponi,¹ Gaudenzia Genoni,² Sara Tonelli¹

¹Fondazione Bruno Kessler ²University of Trento

The generation of critical questions (CQs) aims to foster individuals' critical thinking and counter misinformation. CQs are *the set of inquiries that could be asked in order to judge if an argument is acceptable or fallacious* [1] and have been proven useful for identifying fallacies and evaluating argumentative essays.

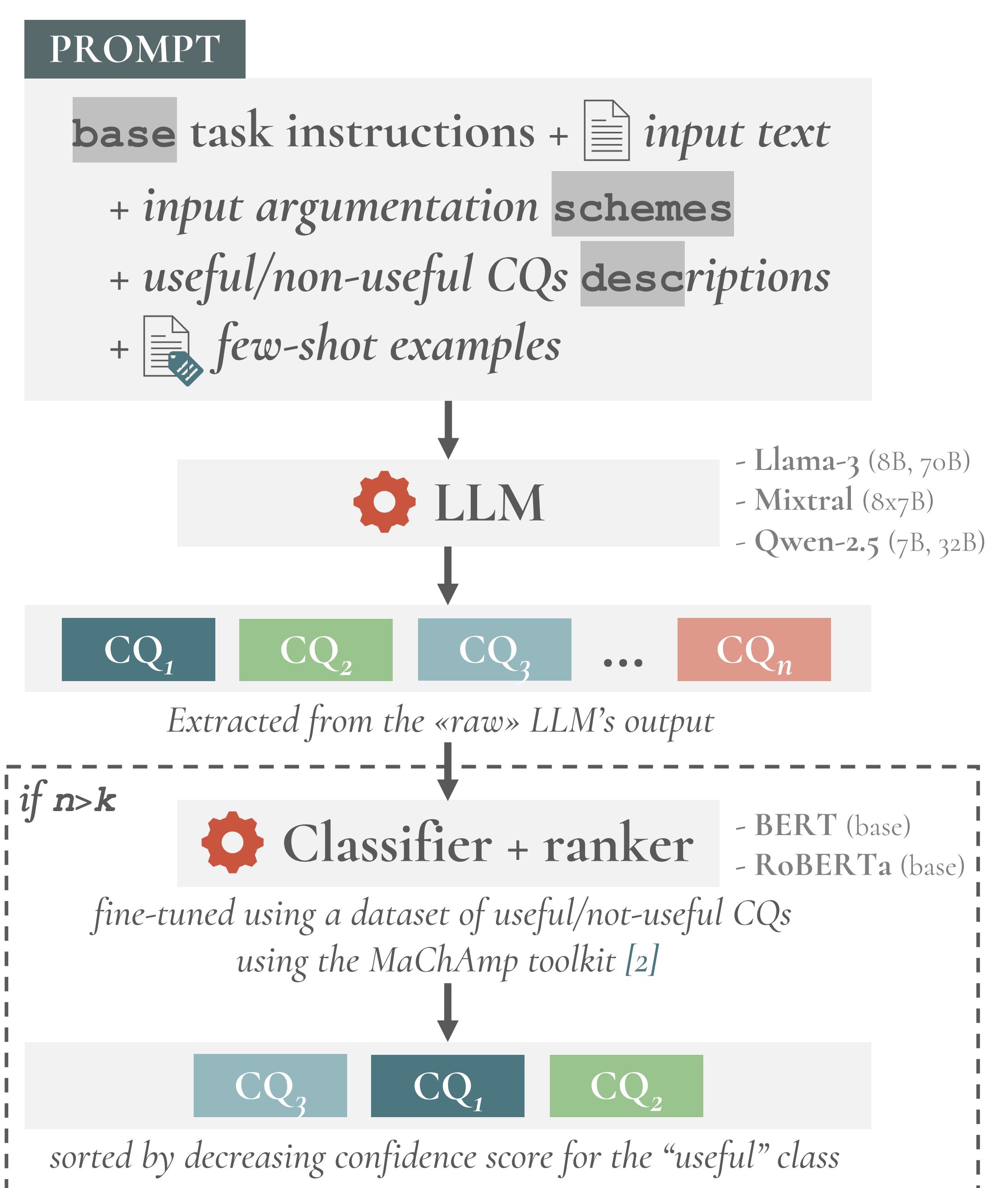
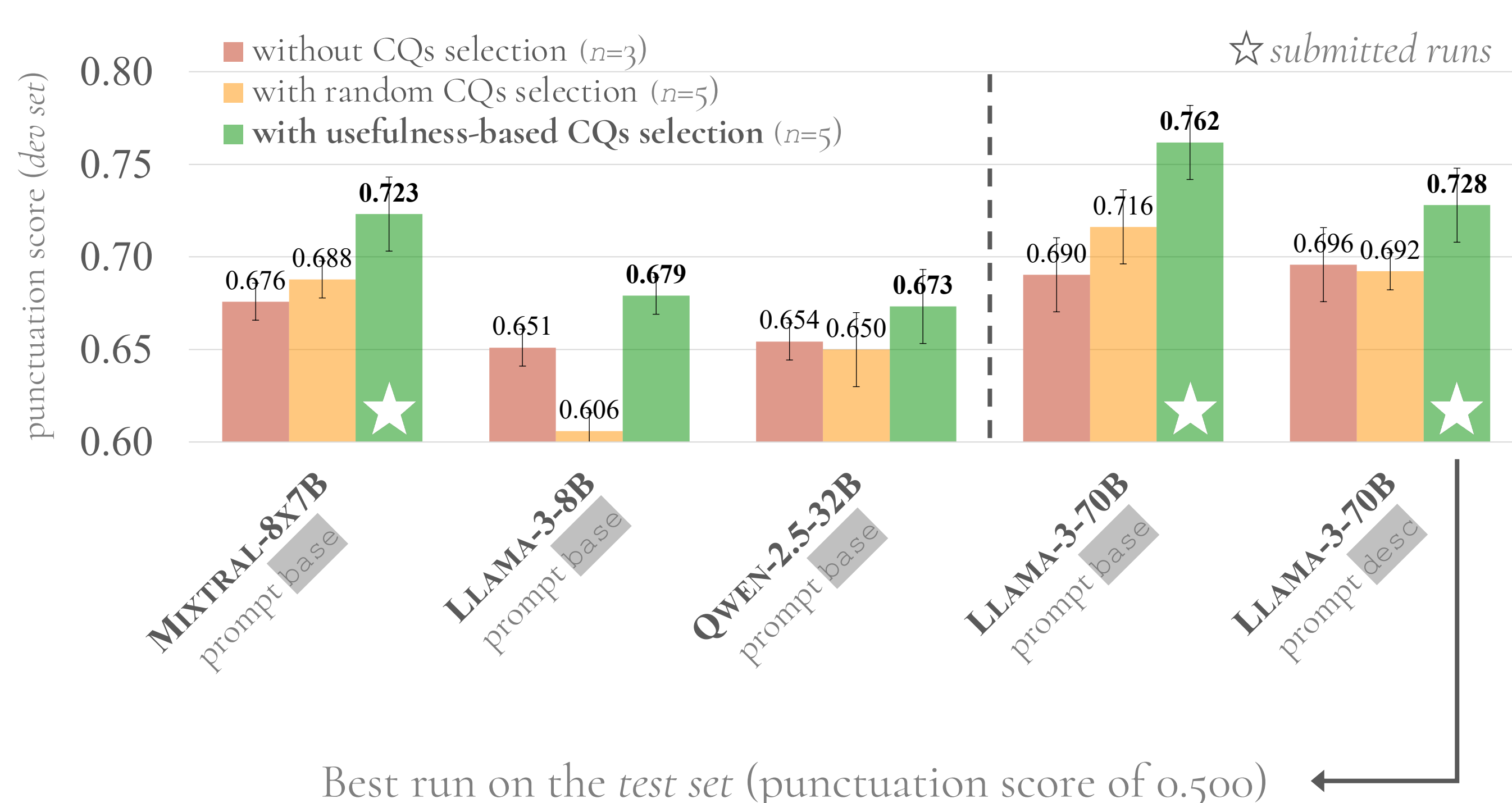
A TWO-STEP APPROACH

- Generation of candidate questions**
Prompt an LLM to obtain a raw output with n candidate CQs for an argumentative text
- Usefulness-based questions selection**
Apply a *useful/not-useful* classifier to the n candidate CQs, rank them by decreasing “usefulness”, and keep the top $k=3$

FINDINGS AND RESULTS

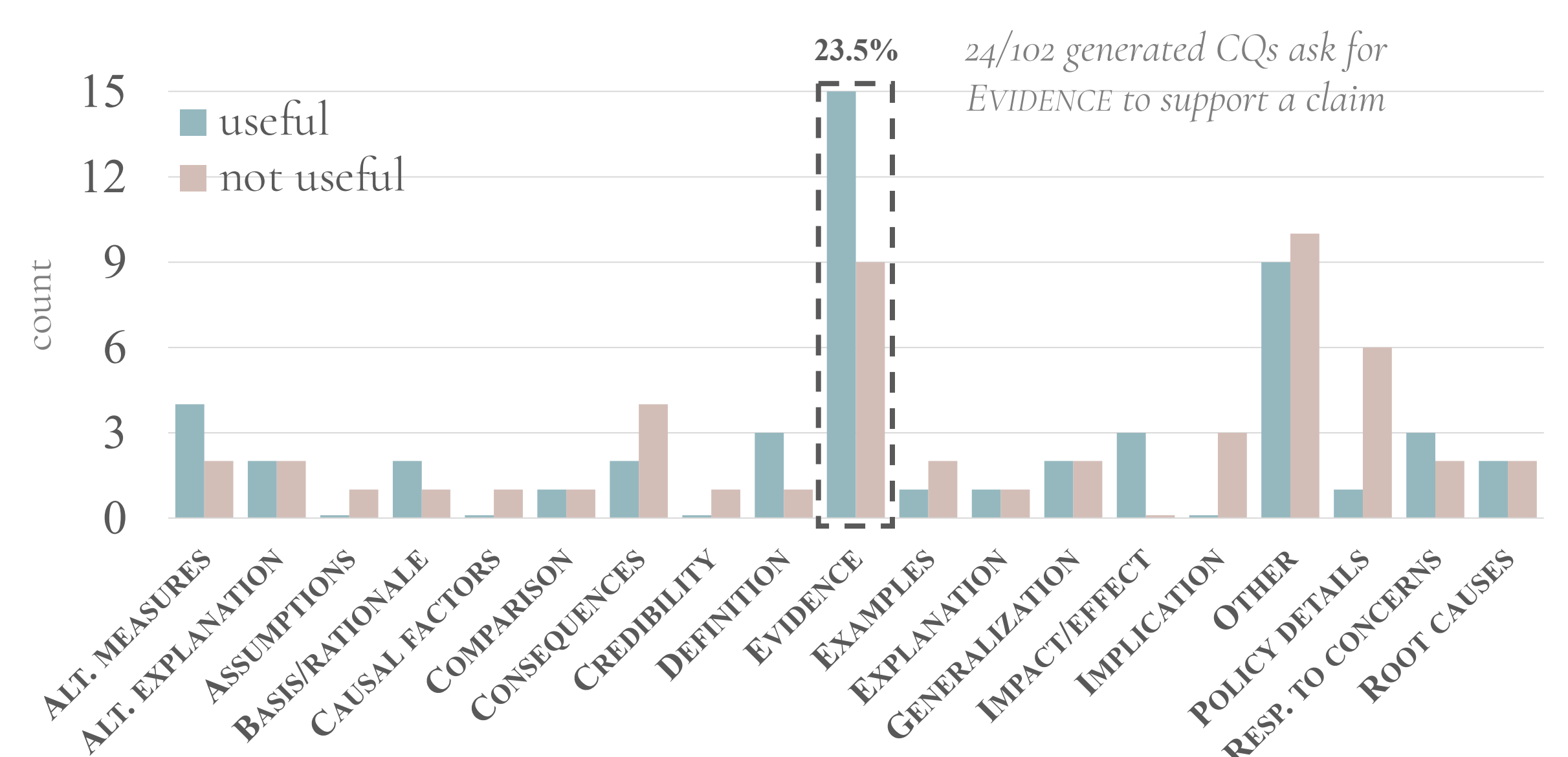
- LLMs & prompts:** Llama-3 8B & Mixtral outperform Qwen2.5 7B across all prompts. The base prompt provides the best performance followed by desc
- Few-shot settings:** large performance degradation!
- Classifier:** best performance with BERT (0.756 F_{1M}) fine-tuned with a mix of gold & synthetic CQs

✓ **Usefulness-based CQs selection consistently improves the performance across the models!**



QUALITATIVE ANALYSIS

Manual classification of CQs by the best run according to their argumentative gap type based on argumentative function, structure, and semantics



References:

- [1] Calvo Figueras & Agerri, 2024. “Critical Questions Generation: Motivation and Challenges”. CoNLL.
[2] van der Goot, Üstün, Ramponi, Sharaf, & Plank, 2021. “Massive Choice, Ample Tasks (MaChAmp): A Toolkit for Multi-task Learning in NLP”. EACL.