

Text simplification leads to info loss, but we can recover it through reader-centric QA.



InfoLossQA: Characterizing and Recovering Information Loss in Text Simplification

Jan Trienes, Sebastian Joseph, Jörg Schlötterer, Christin Seifert, Kyle Lo, Wei Xu, Byron C. Wallace, Junyi Jessy Li

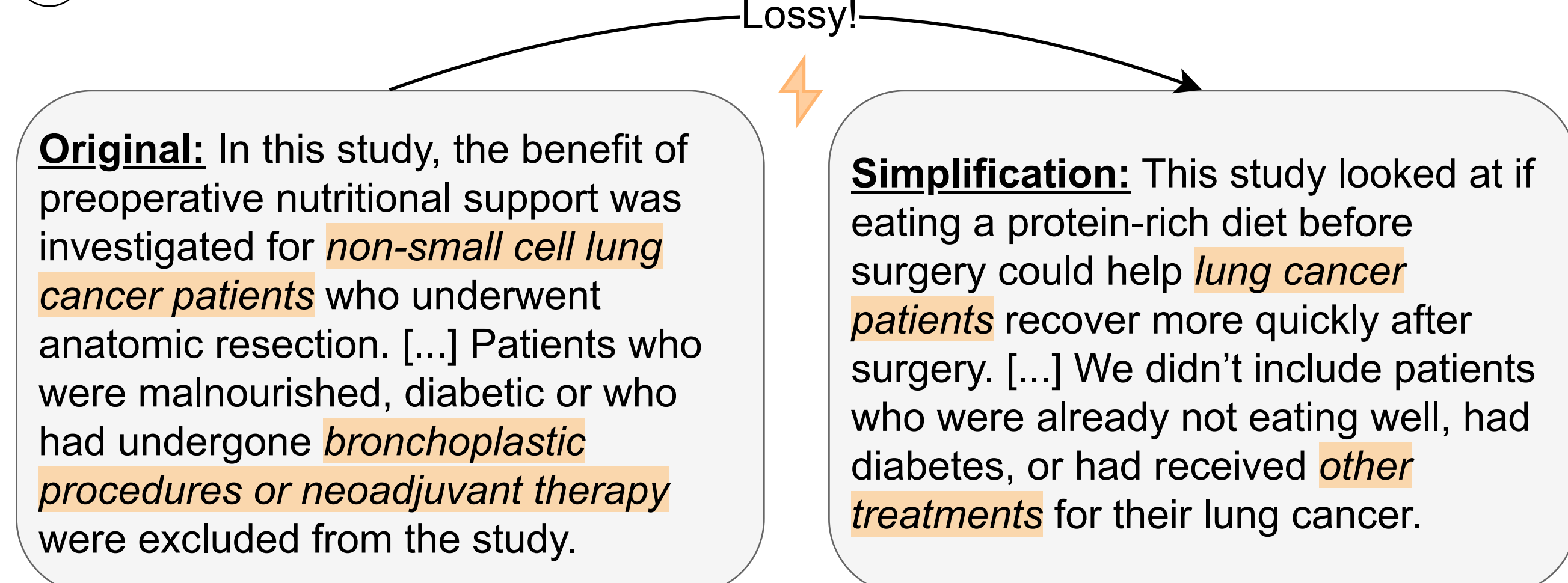
jan.trienes@uni-marburg.de



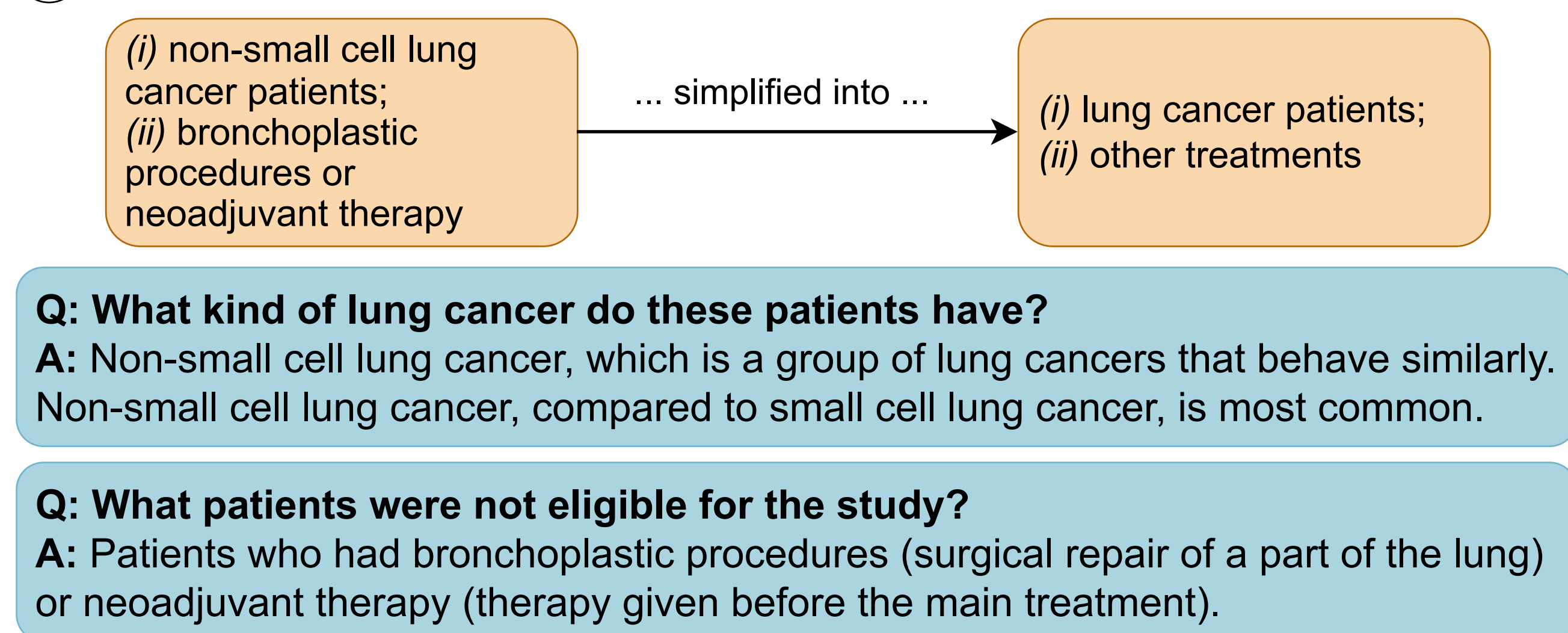
MOTIVATION

- Text simplification is inherently lossy
- This may reduce comprehension and factuality
- Q: How can we characterize info loss, and help lay readers to intuitively recover it?**

1 Text simplification

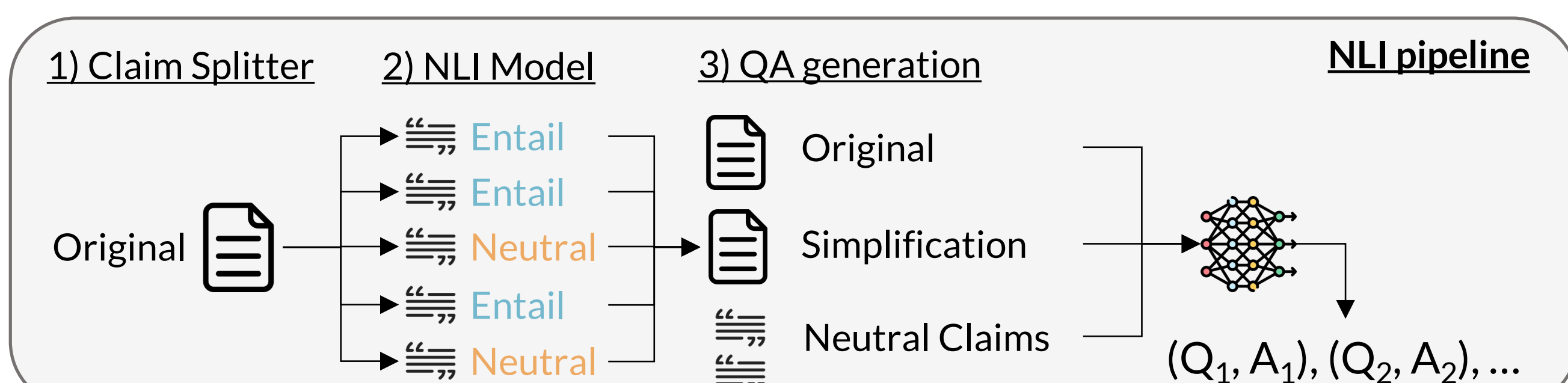
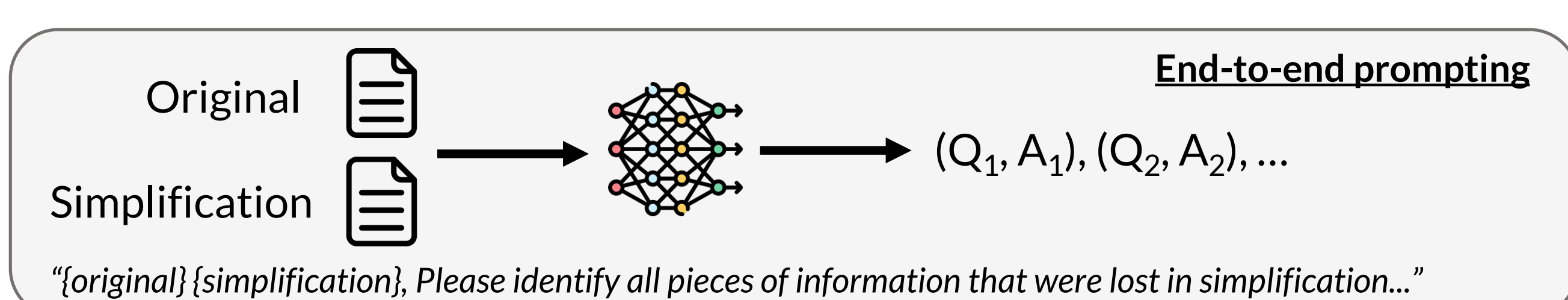


2 Identify information loss and recover it in reader-focused QA (this paper)

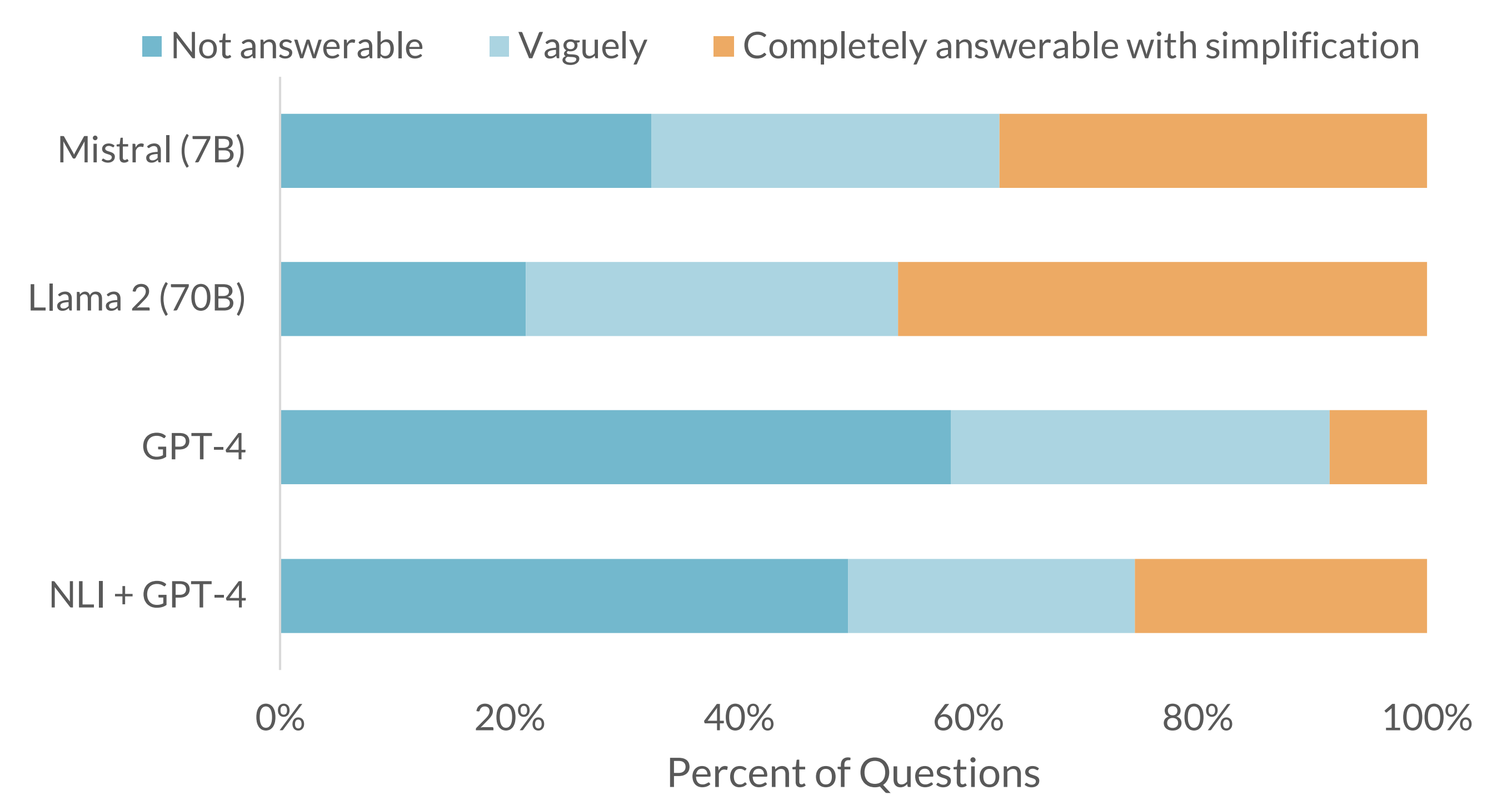


METHODS

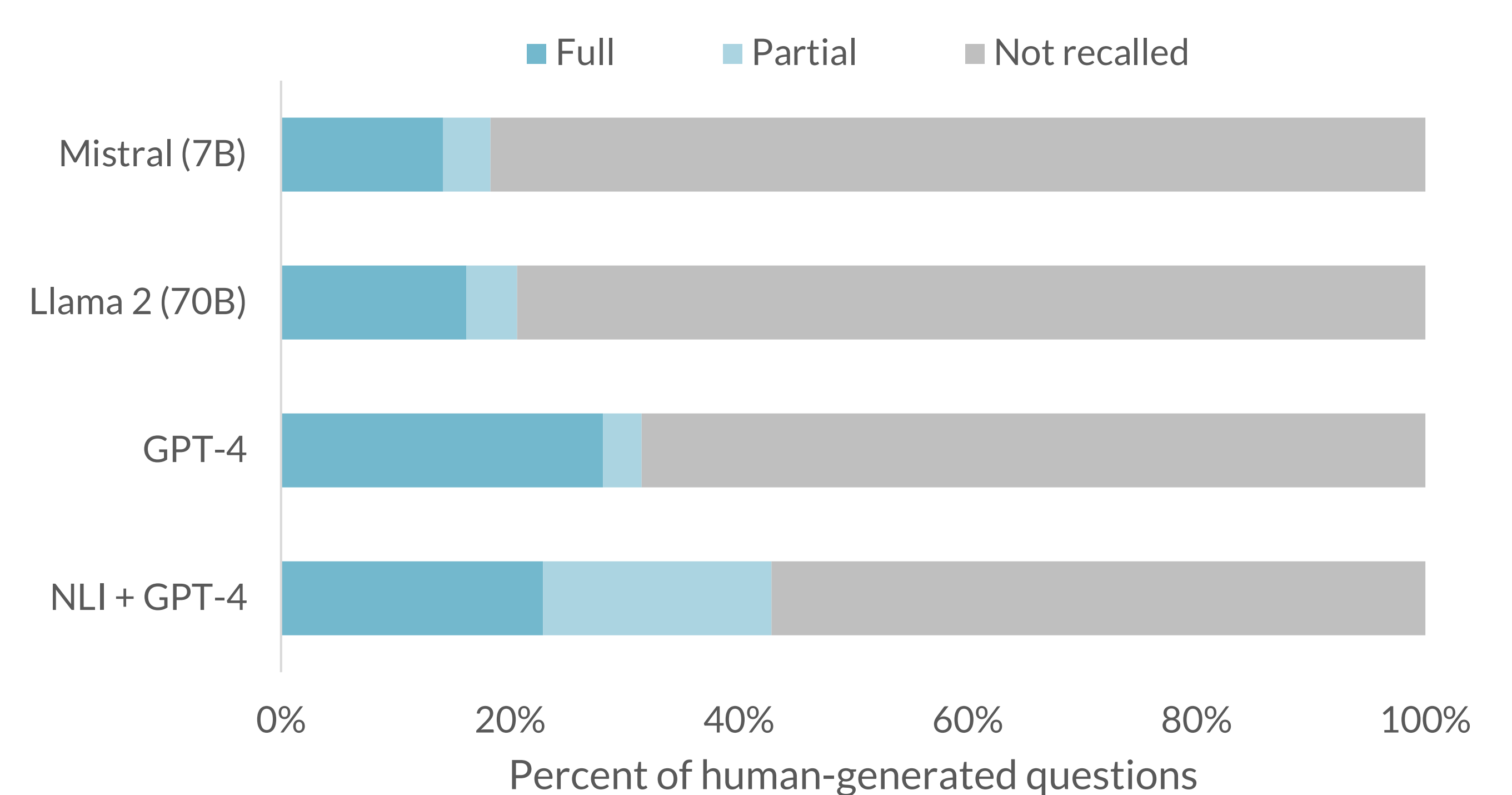
- Task: generate QAs that reveal missing info!**
- New data: annotated info loss in GPT-4 simplifications (104 docs, 1000 QAs total)
- Experiments: LLM prompting and NLI pipeline



RESULT (a). Current methods cannot reliably identify information loss. Between 10-45% of generated questions are not addressing info loss.



RESULT (b). Humans and models have different standards at what constitutes information loss.



RESULT (c). Rationalization is difficult. Models often fail to extract spans in the simple text that were oversimplified.

