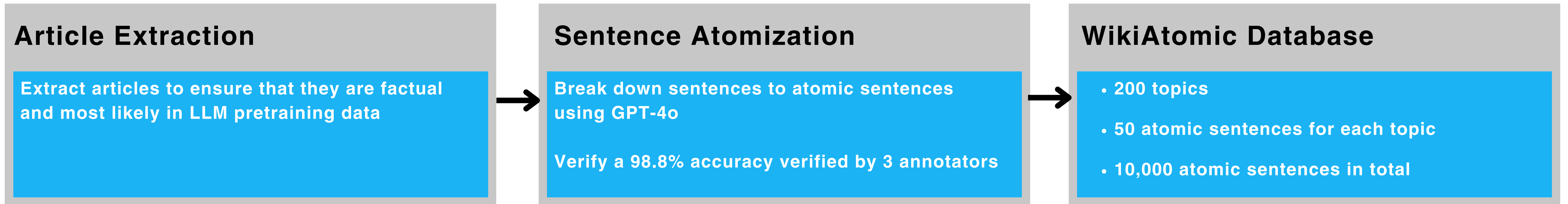


WHEN CONTEXT LEADS BUT PARAMETRIC MEMORY FOLLOWS IN LLMS

YUFEI TAO, ADAM HIATT, ERIK HAAKE, ANTONIE J. JETTER, AMEETA AGRAWAL
DEPARTMENT OF COMPUTER SCIENCE | PORTLAND STATE UNIVERSITY, USA

How do LLMs incorporate context vs. parametric knowledge in knowledge-consistent settings?

WIKIATOMIC DATASET



EXAMPLE OF CONTEXT AND OPEN-ENDED QA

Context: The Boeing 717 is a jet airliner. The Boeing 717 has two engines. The Boeing 717 is a twinjet. The Boeing 717 has two columns of seats.

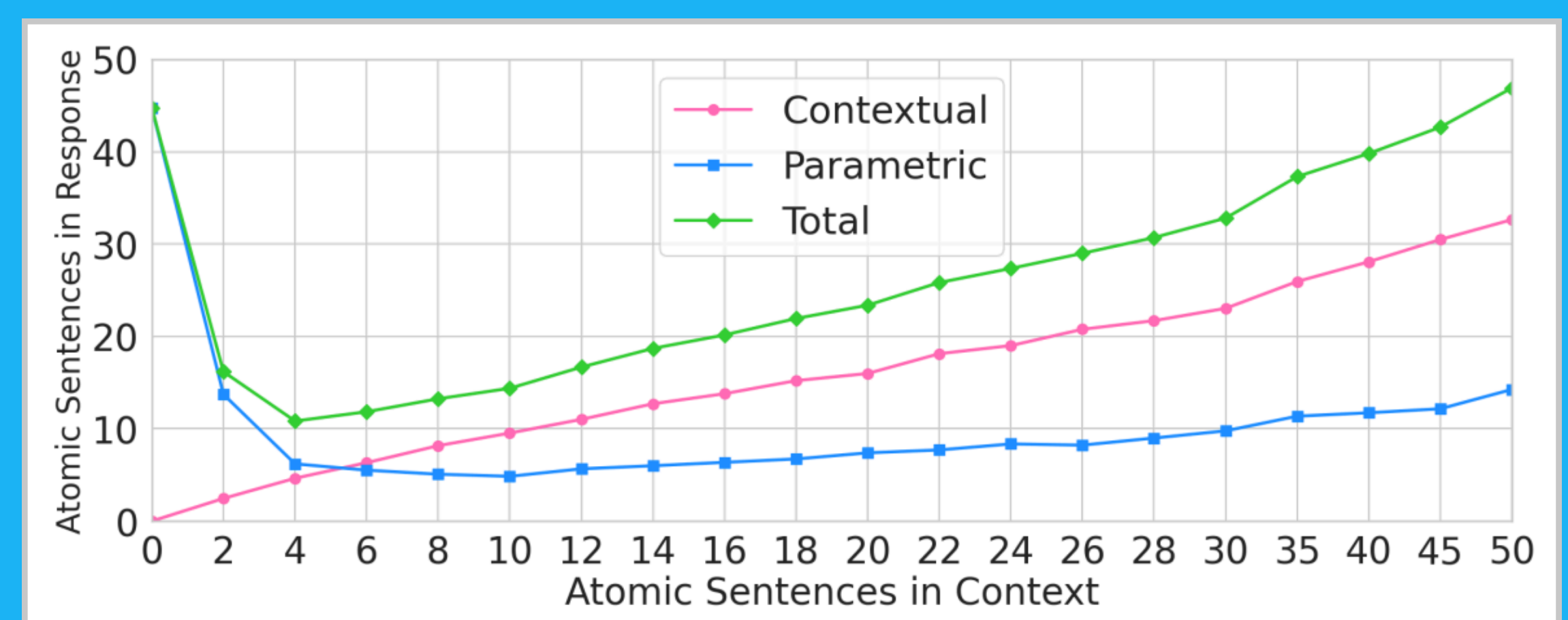
Question: With this information, tell me about Boeing 717.

GPT-4o Response: The Boeing 717 is a jet airliner equipped with two engines, which classifies it as a twinjet. Inside the aircraft, there are two columns of seats for passengers. The aircraft can accommodate up to 117 passengers.

Response Atomic Sentences:

- [The Boeing 717 is a jet airliner., Contextual Knowledge]
- [The Boeing 717 is equipped with two engines., Parametric Knowledge]
- ...
- [The Boeing 717 can accommodate up to 117 passengers.]

PRIORITIZING SOURCES OF KNOWLEDGE



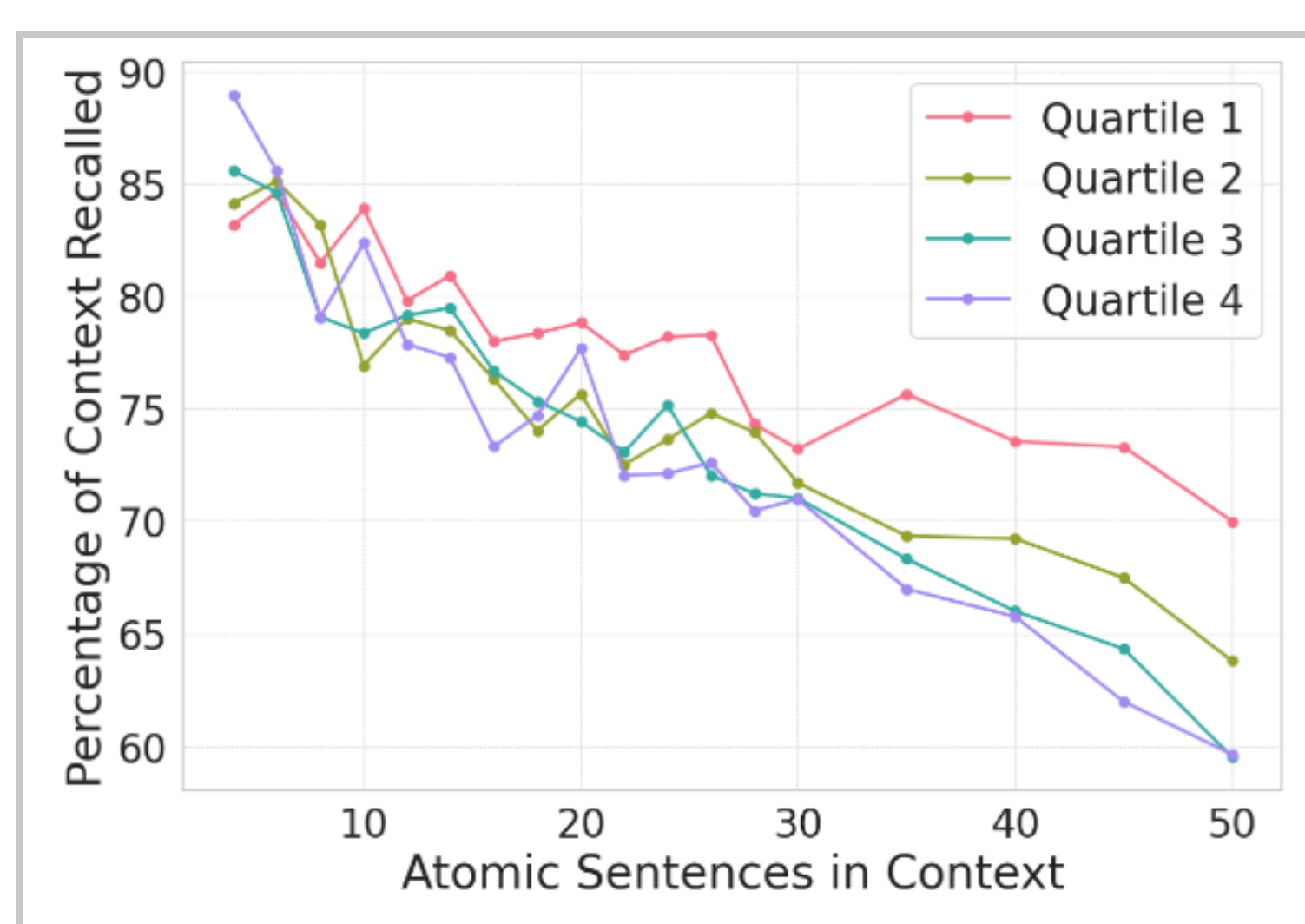
EVALUATION

- Models: GPT-4o, Claude 3 (Opus, Sonnet and Haiku), Llama 3 (70B and 8B), Mixtral 8x22b, Mistral 7B and Phi-3.
- Metric: INFUSE framework to calculate entailment (0 to 1). Threshold was set at 0.5.

FINDINGS

- All LLMs have same pattern
- All of them never utilize 100% of contexts
- All of them always provide parametric knowledge
- Parametric knowledge converges to ~30%

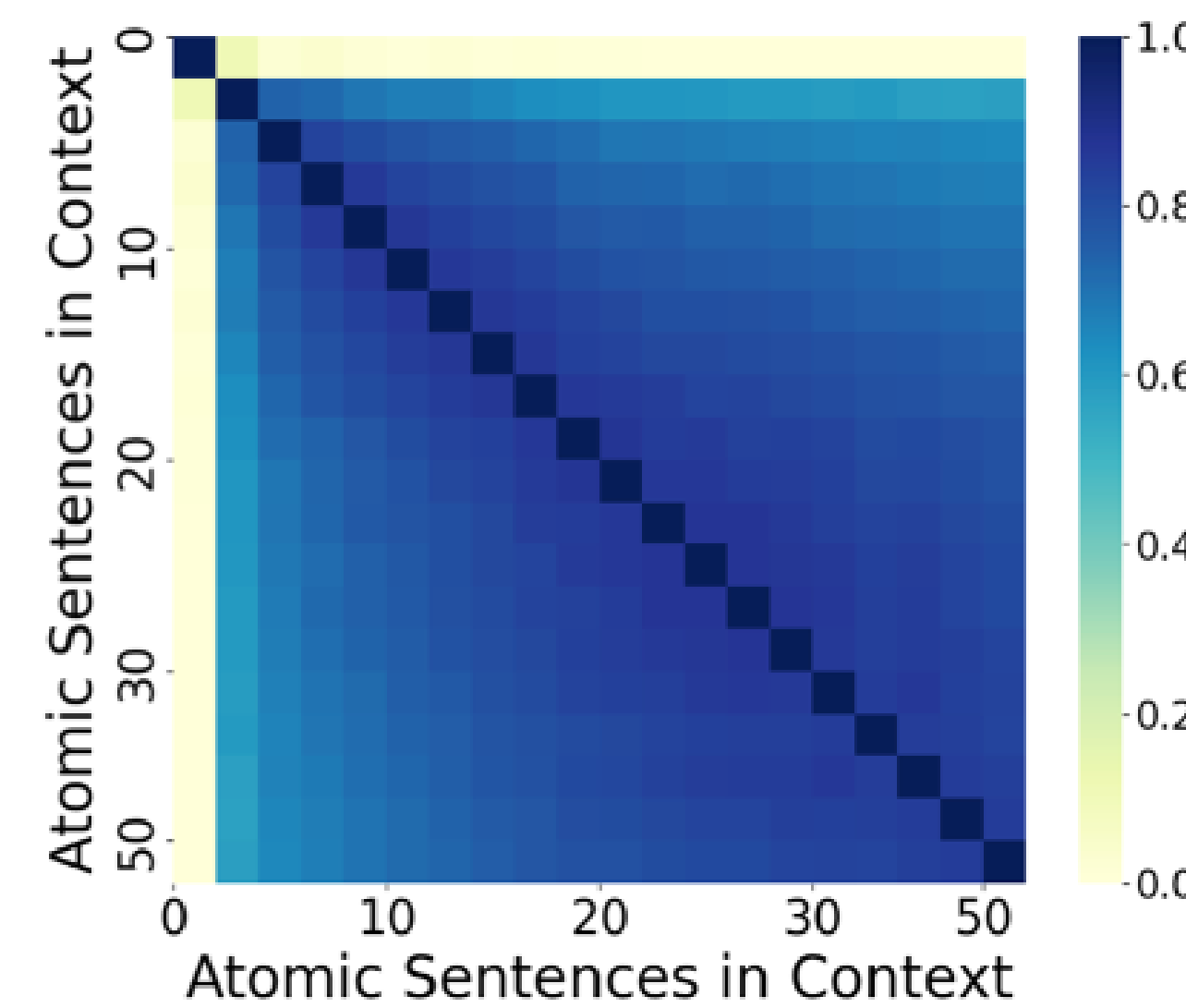
CONTEXTS POSITIONS



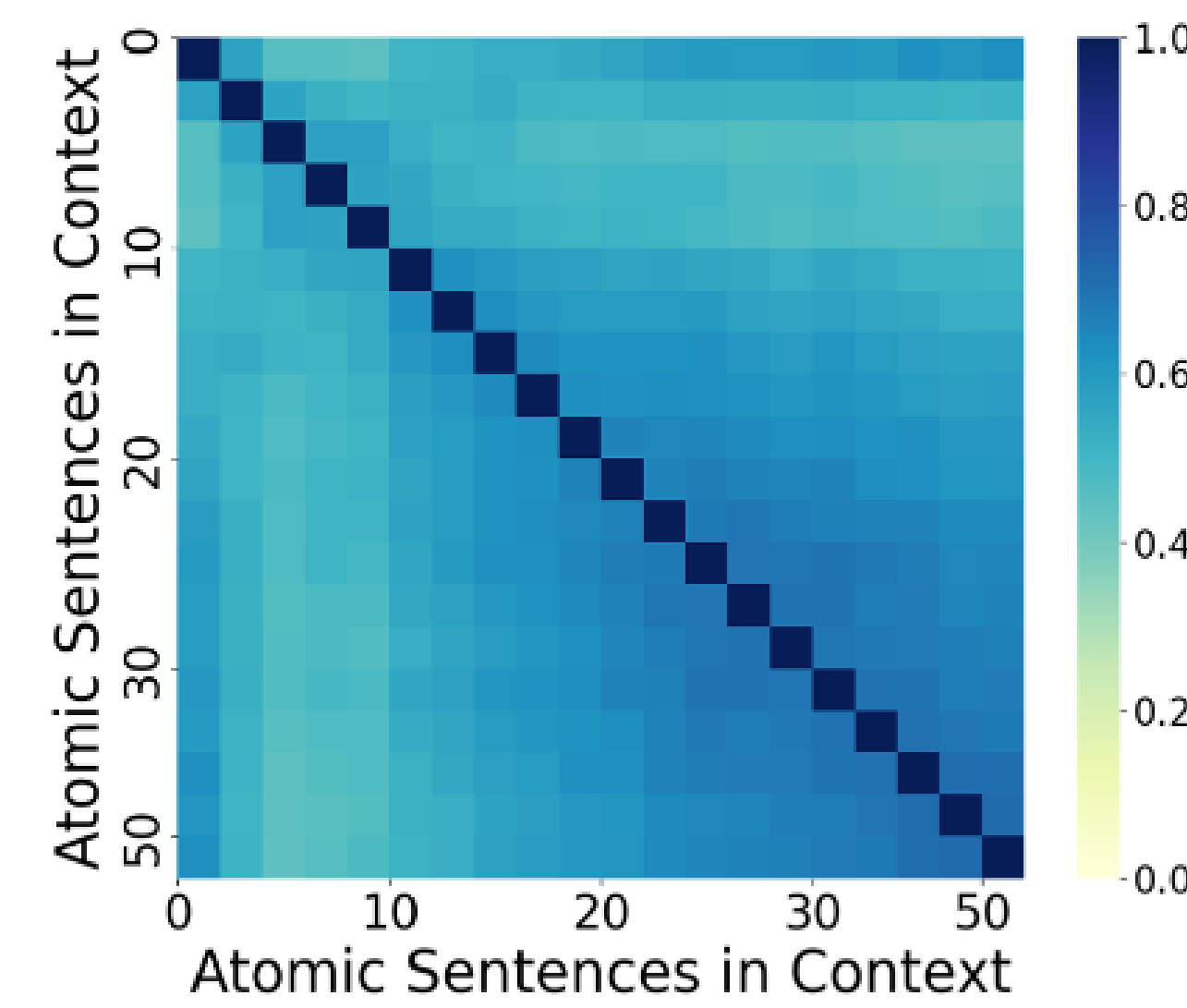
Analyzed context recalled per quartile

As context length increased, the model focused on the first quartile

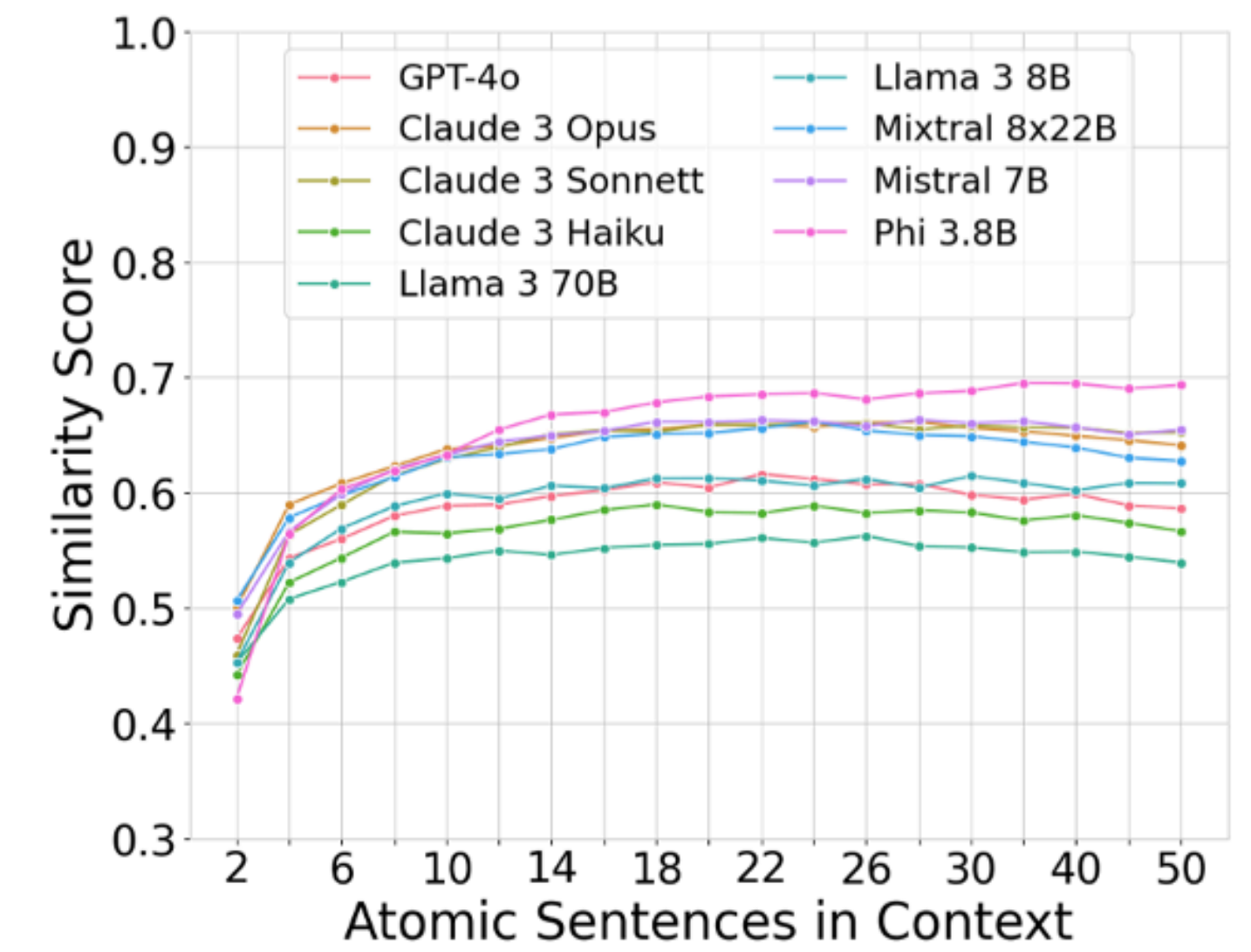
KNOWLEDGE SIMILARITIES



Local vs Local: Moderately similar in smaller contexts but became increasingly similar with larger contexts

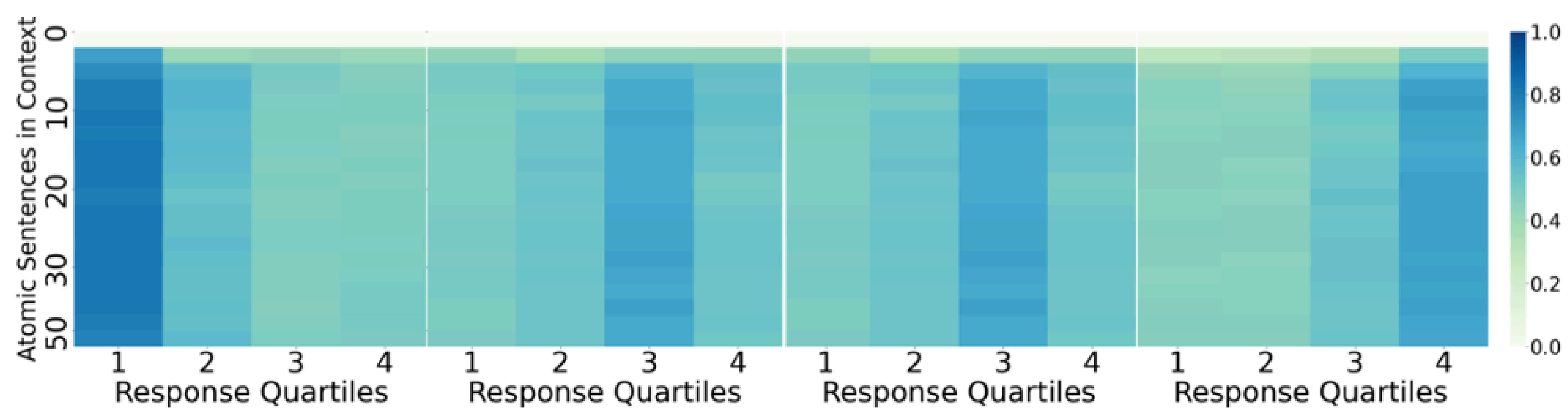


Global vs Global: More contexts -> theme more aligned, suggesting LLMs draws from small pools of knowledge



Local vs Global:
Small Contexts: Complementary knowledge
Larger contexts: Shared concepts

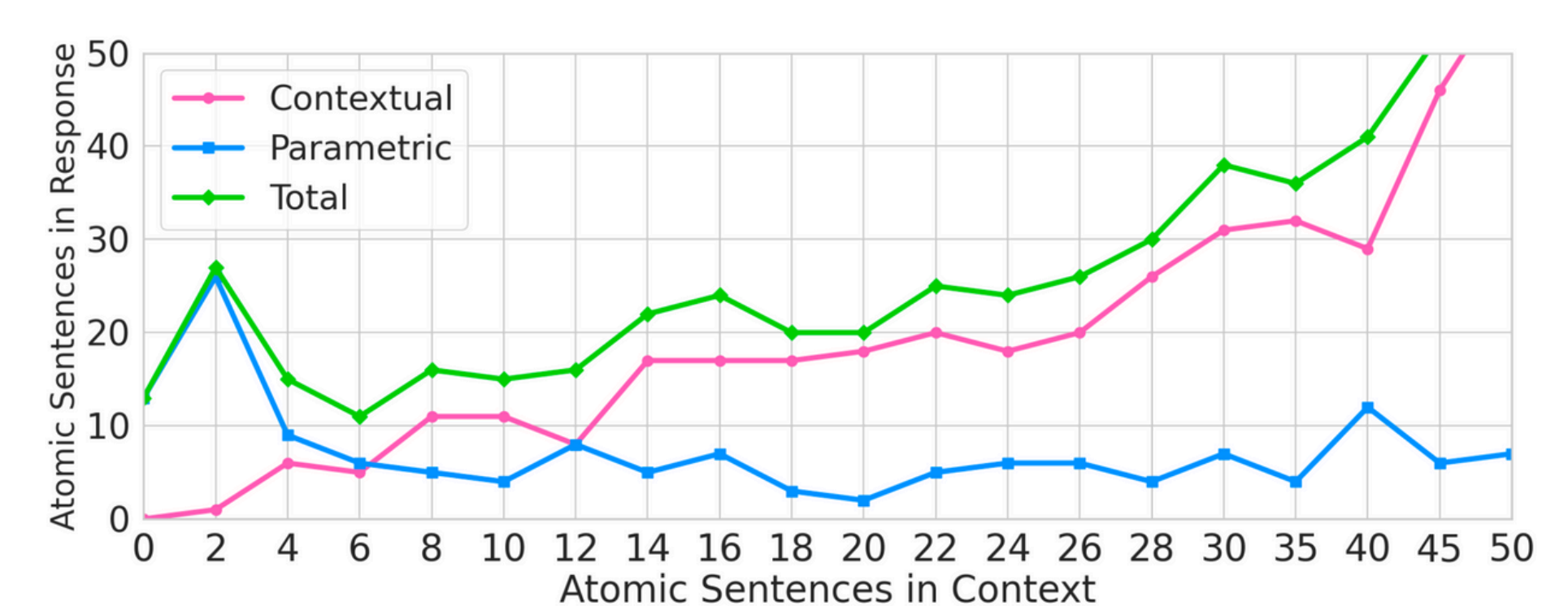
CONTEXTS POSITIONS



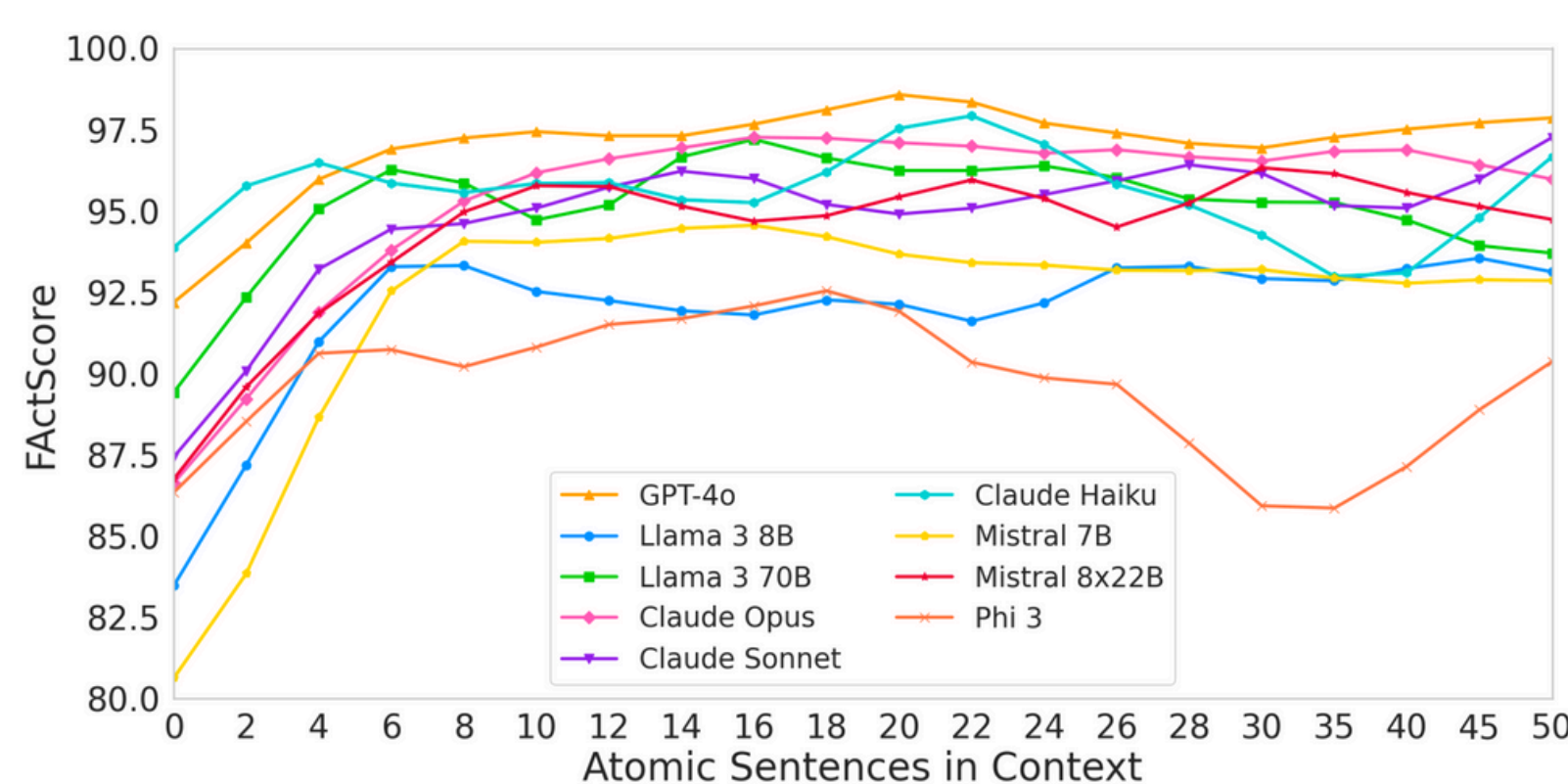
First context quartile -> first and adjacent response quartile.

Model prefers to match the positions of data

NEW/UNSEEN KNOWLEDGE



HALLUCINATION



More contexts or bigger the model, leads to less hallucination across all models

FURTHER ANALYSES

- Prompt Sensitivity
- Disregard Ambiguous Sentences

ACKNOWLEDGEMENT

• This research was supported by the National Science Foundation grant SAI-P 2228783.

