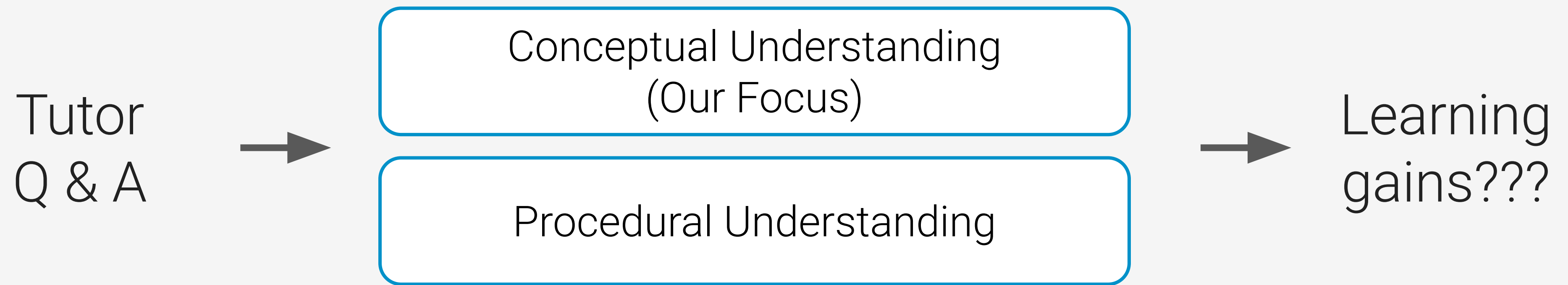


# Retrieval-augmented Generation to Improve Math Question-Answering

## Trade-offs Between Groundedness and Human Preference

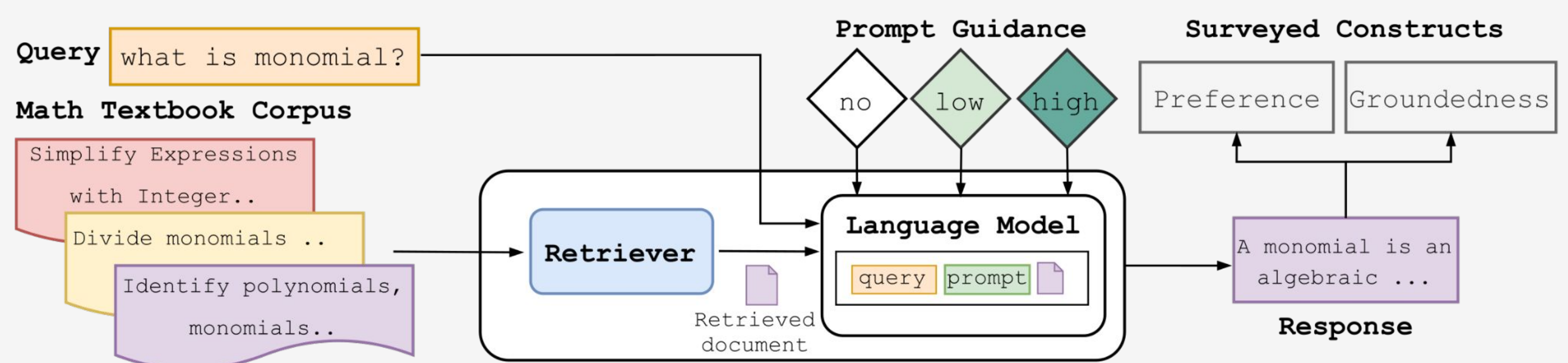
Zachary Levonian, Chenglu Li, Wangda Zhu, Anoushka Gade, Owen Henkel, Millie-Ellen Postle, Wanli Xing  
Presenter: Zach (Digital Harbor Foundation) <zach@digitalharbor.org> @zwlevonian@hci.social

### Large Language Models for Math Question-Answering



### Retrieval-augmented Generation for Correctness & Groundedness

- Implemented a RAG system
- Asked GPT-3.5 student questions from Math Nation
- Retrieved textbook sections from OpenStax (prealgebra)
- Evaluated via survey
- Code & data available!

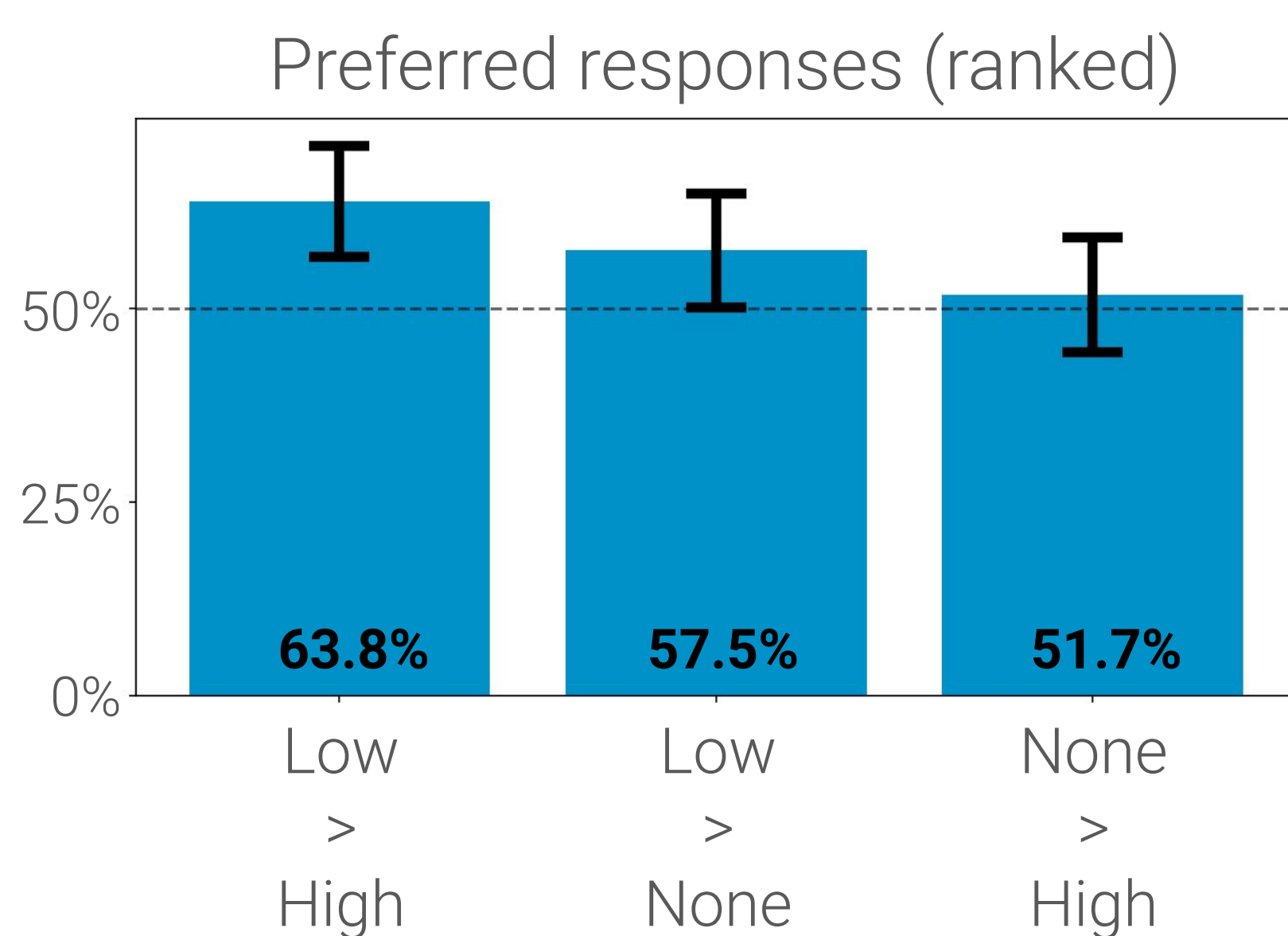


[github.com/DigitalHarborFoundation/rag-for-math-qa](https://github.com/DigitalHarborFoundation/rag-for-math-qa)

**What we found:** Humans prefer LLM responses to students' conceptual math questions when created with retrieval-augmented generation and "just the right amount" of prompting guidance.

### Survey Results

- Prompt guidance affects groundedness
- Raters preferred low guidance ( $n=144$ )
- Too much guidance is possible!



### Rori: a chatbot math tutor

- Accessed via WhatsApp
- Used in classrooms and at home
- Mostly in West Africa, esp. Ghana



## Learning Engineering Virtual Institute



RISING  
ACADEMIES

DIGITAL HARBOR FOUNDATION



NeurIPS'23 Workshop  
Generative AI for Education (GAIED)  
December 15, 2023

Partners: University of Utah, University of Florida, University of Oxford