

# Retrieval-augmented Generation to Improve Math Question-Answering: Trade-offs Between Groundedness and Human Preference

Zachary Levonian

December 13, 2023

[zach@digitalharbor.org](mailto:zach@digitalharbor.org)



# Quick notes

- Presenting at the NeurIPS'23 Workshop on Generative AI for Education (GAIED)
- Co-authors: **Chenglu Li**, Wangda Zhu, **Anoushka Gade**, Owen Henkel, Millie-Ellen Postle, and Wanli Xing
- Paper: [arxiv.org/abs/2310.03184](https://arxiv.org/abs/2310.03184)
- Code & data: [github.com/DigitalHarborFoundation/rag-for-math-qa](https://github.com/DigitalHarborFoundation/rag-for-math-qa)
- Demo: [llm-math-education-levon003.streamlit.app](https://llm-math-education-levon003.streamlit.app)



These slides released CC BY 4.0



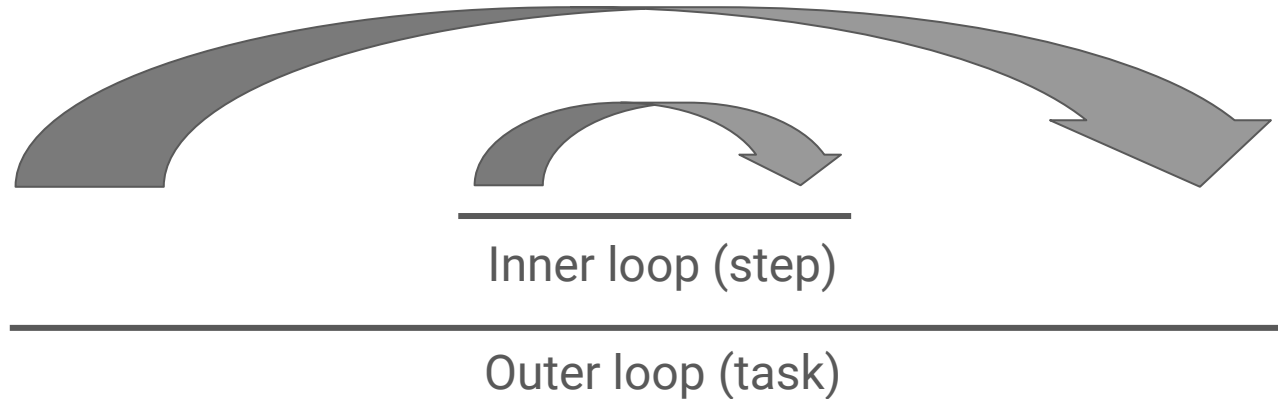
# Large language models for education

- So many applications!



# Large language models for education

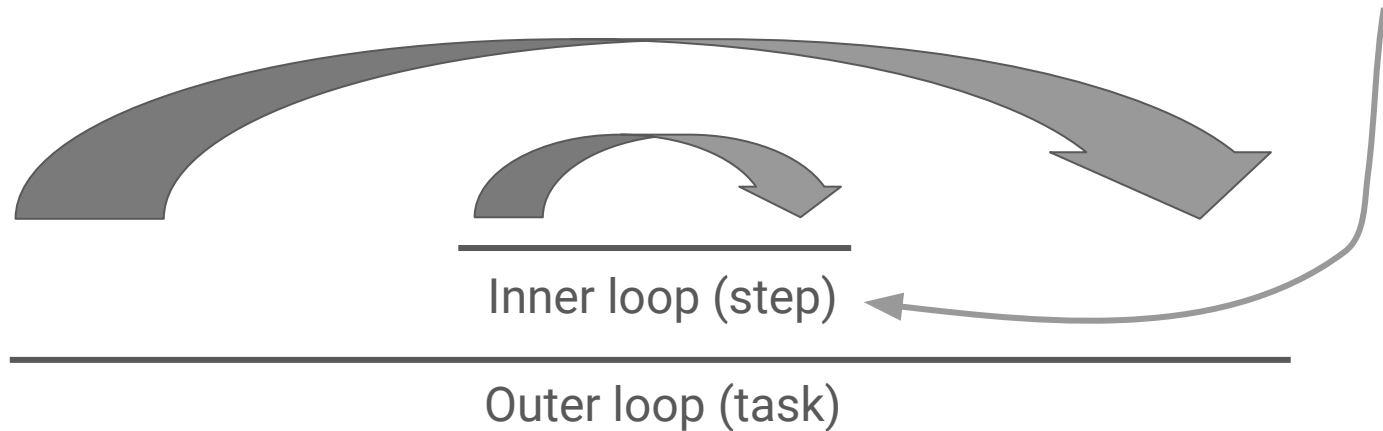
- So many applications!
- Let's improve intelligent tutoring systems



# Large language models for education

- So many applications!
- Let's improve intelligent tutoring systems

Inner loop intervention:  
Hint generation  
Procedural Q&A





# Rori: a chatbot math tutor





# Learning Engineering Virtual Institute

Doubling the rate of middle school math progress  
for low-income students by 2027.

CARNEGIE  
LEARNING

Carnegie  
Mellon  
University

I= Eedi



RIISING ACADEMY  
NETWORK



Khan Academy

UF UNIVERSITY of  
FLORIDA



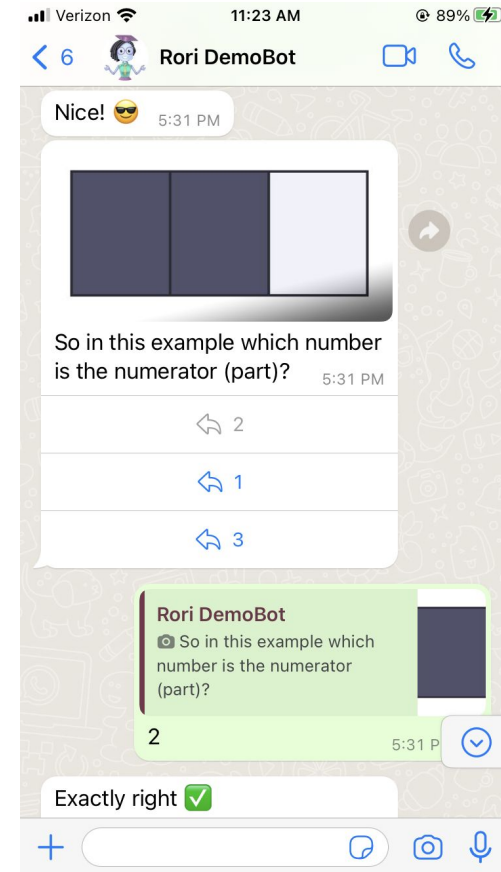
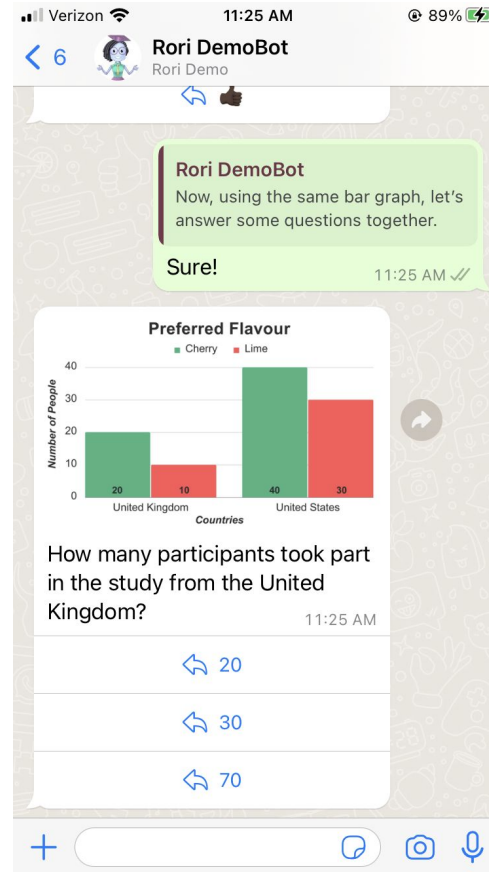
University of Colorado Boulder





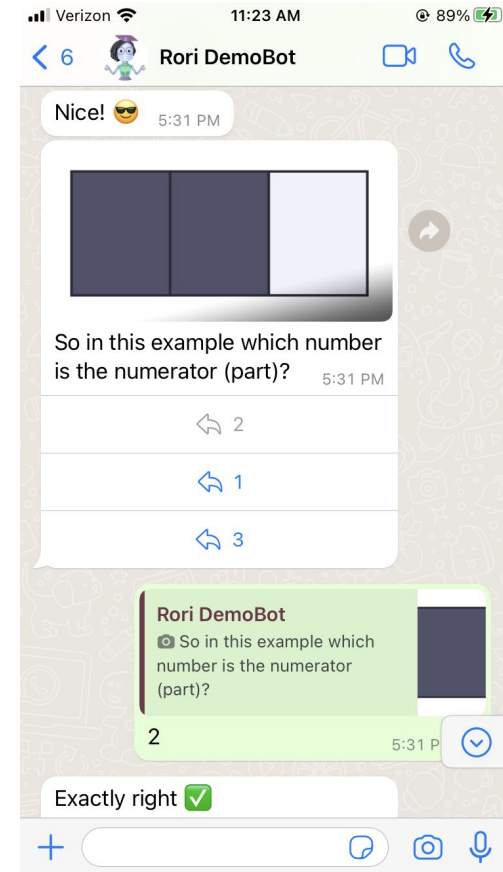
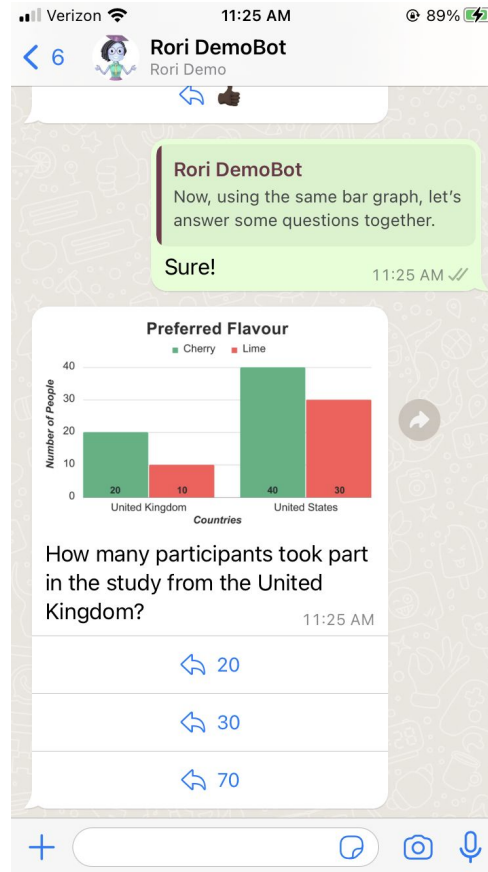
# Rori: a chatbot math tutor

- Organized into micro-lessons
- Images and text
- Multiple-choice and free response
- Inflexible!
- Lots of LLM opportunities
  - Hint generation
  - Personalization
  - Q&A

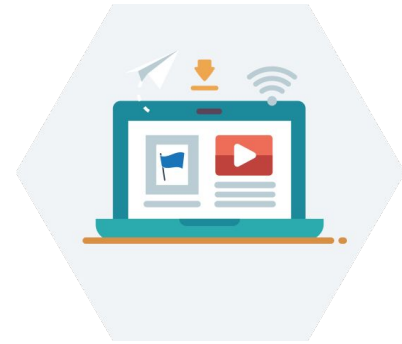
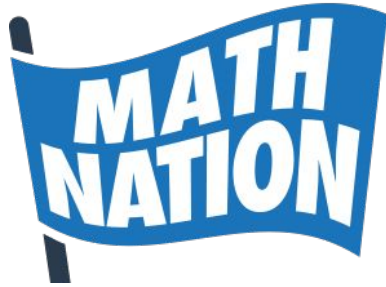


# Rori: a chatbot math tutor

- Goal: prototype conceptual Q&A
- No existing Rori data!

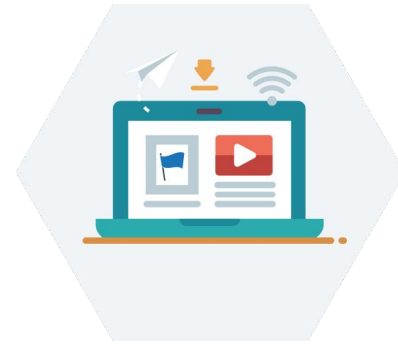


# Math Nation forums



# Math Nation forums

- 554 Math Nation threads
  - Memes
  - Homework help
  - Conceptual questions
- 51 (9%) conceptual questions



# Math Nation forums

- what is monomial
- How do I add line segments again??
- How do I multiply fractions???????
- How do you know if a number is a constant?
- I always forget the difference between commutative and associative. Does anybody know a way to make me remember?



# Math Nation forums

- what is monomial
- How do I add line segments again??
- How do I multiply fractions???????
- How do you know if a number is a constant?
- **I always forget the difference between commutative and associative. Does anybody know a way to make me remember?**



LE

**You**

I always forget the difference between commutative and associative. Does anybody know a way to make me remember?

**ChatGPT**

Of course! I can help you with that. Commutative and associative properties are important concepts in mathematics.

Here's a simple way to remember them:

- The commutative property applies to addition and multiplication. It means that the order in which you add or multiply numbers doesn't change the result. For example, in addition,  $2 + 3 = 3 + 2$  and in multiplication,  $4 \times 5 = 5 \times 4$ .
- The associative property also applies to addition and multiplication. It means that the grouping of numbers does not affect the result. For example, in addition,  $(2 + 3) + 4 = 2 + (3 + 4)$  and in multiplication,  $(4 \times 5) \times 6 = 4 \times (5 \times 6)$ .

I hope this helps you remember the difference between commutative and associative properties. Let me know if you have any follow-up questions related to math.



# Correctness is not enough

- Grade level
- Curricular alignment
- Cultural relevance





# The automated student Q&A continuum™

LLMs



Conventional  
information  
retrieval



# The automated student Q&A continuum™

LLMs



Conventional  
information  
retrieval

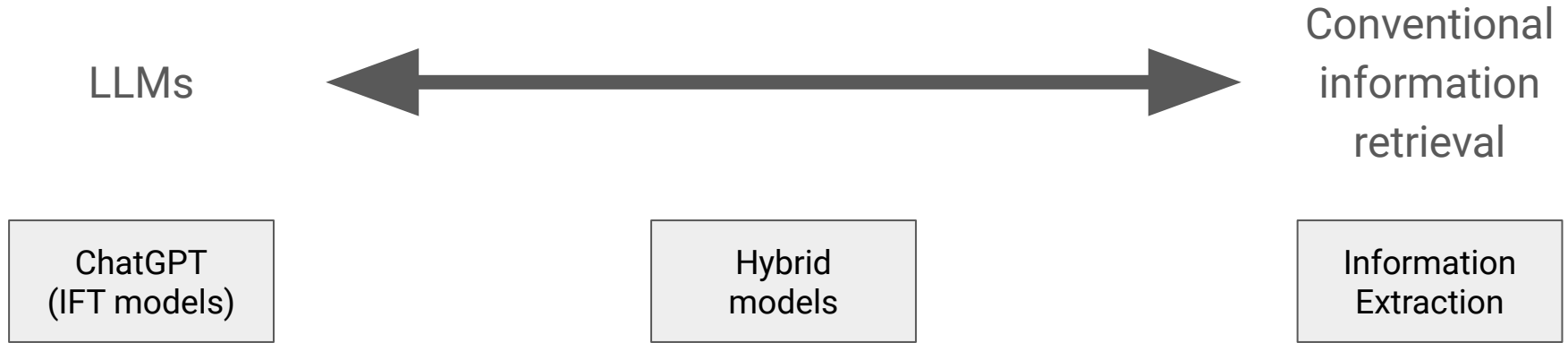
Information  
Extraction



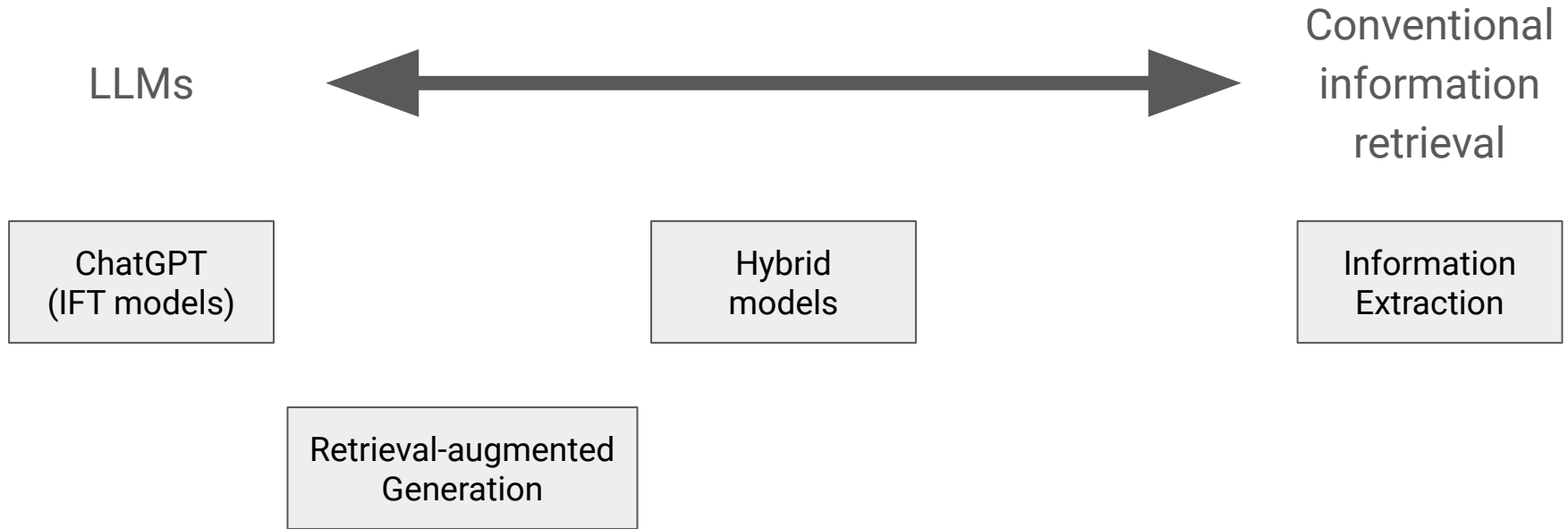
# The automated student Q&A continuum™



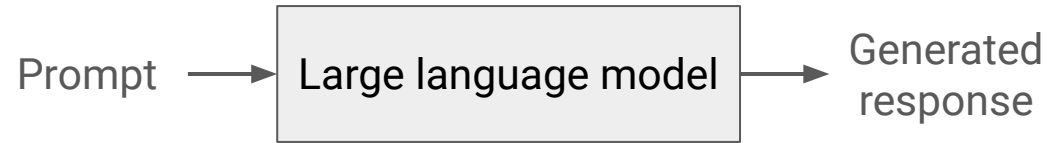
# The automated student Q&A continuum™



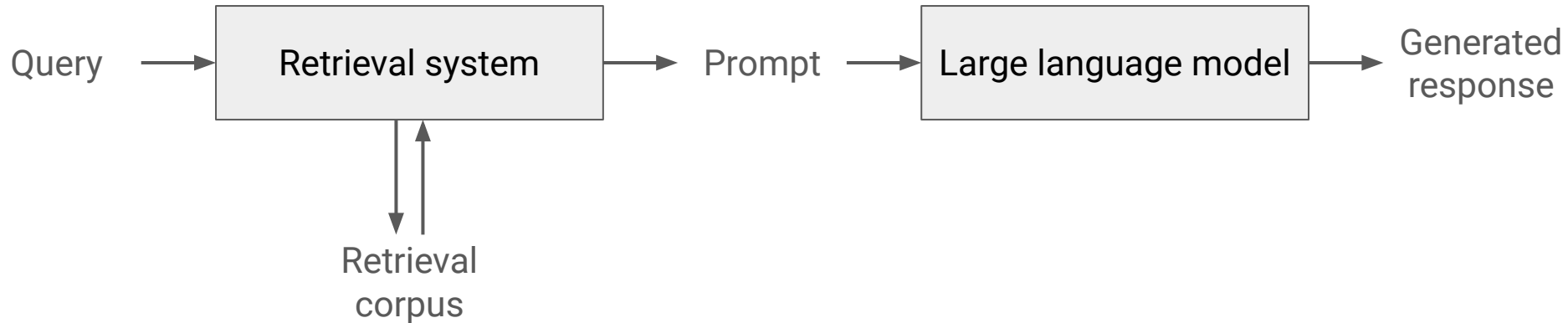
# The automated student Q&A continuum™



# Retrieval-augmented generation



# Retrieval-augmented generation



# Retrieval-augmented generation



**Luca Soldaini**   
@soldni

Pet peeve: "RAG" used as a stand-in for any LLM+retrieval 

11:25 AM · Oct 10, 2023 · **6,949** Views





# Retrieval-Enhanced Machine Learning

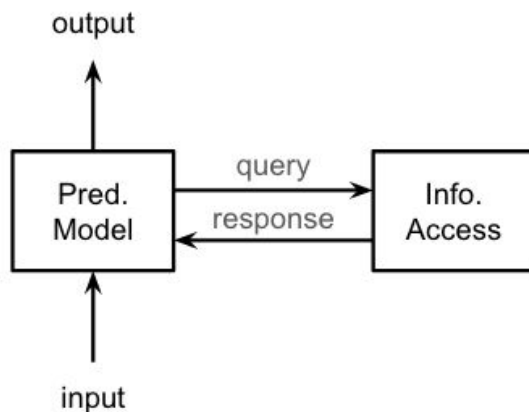
Hamed Zamani\*  
University of Massachusetts Amherst  
zamani@cs.umass.edu

Fernando Diaz\*  
Google Research  
diazf@acm.org

Mostafa Dehghani  
Google Research  
dehghani@google.com

Donald Metzler  
Google Research  
metzler@google.com

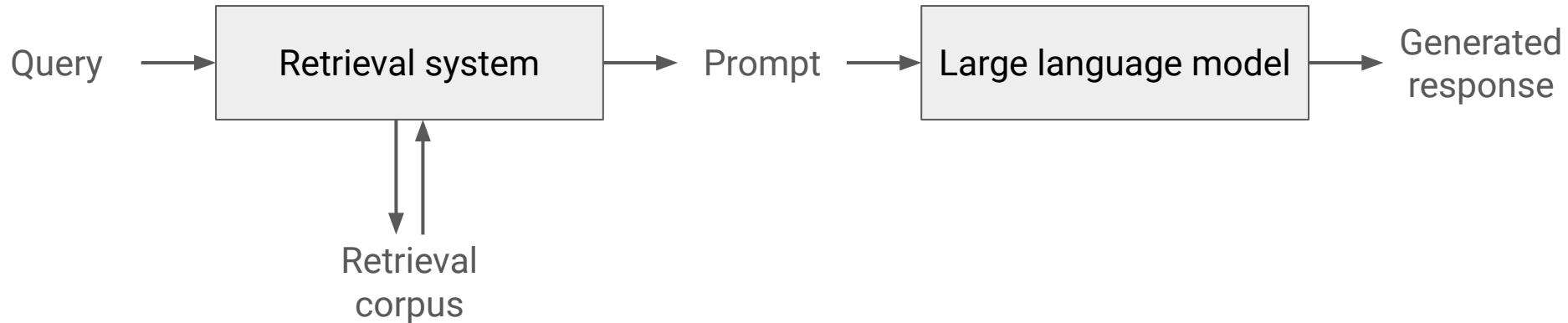
Michael Bendersky  
Google Research  
bemike@google.com



(a) Cat 1: Retrieval-only



# Retrieval-augmented generation



# Prompt Engineering

You are going to act as a mathematics tutor for a 13 year old student who is in grade 8 or 9 and lives in Ghana.

You will be encouraging and factual.

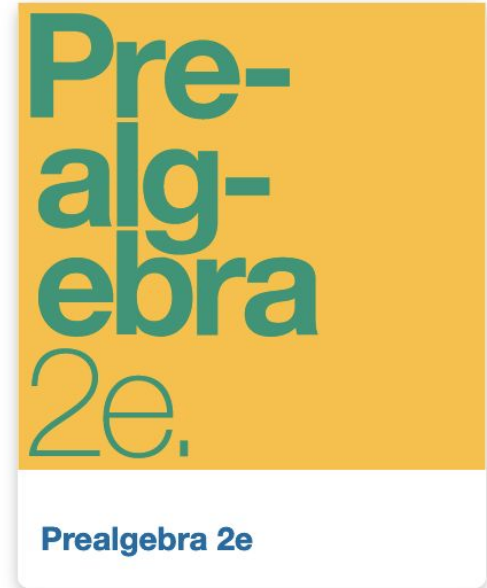
Prefer simple, short responses.

If the student says something inappropriate or off topic you will say you can only focus on mathematics and ask them if they have any math-related follow-up questions.

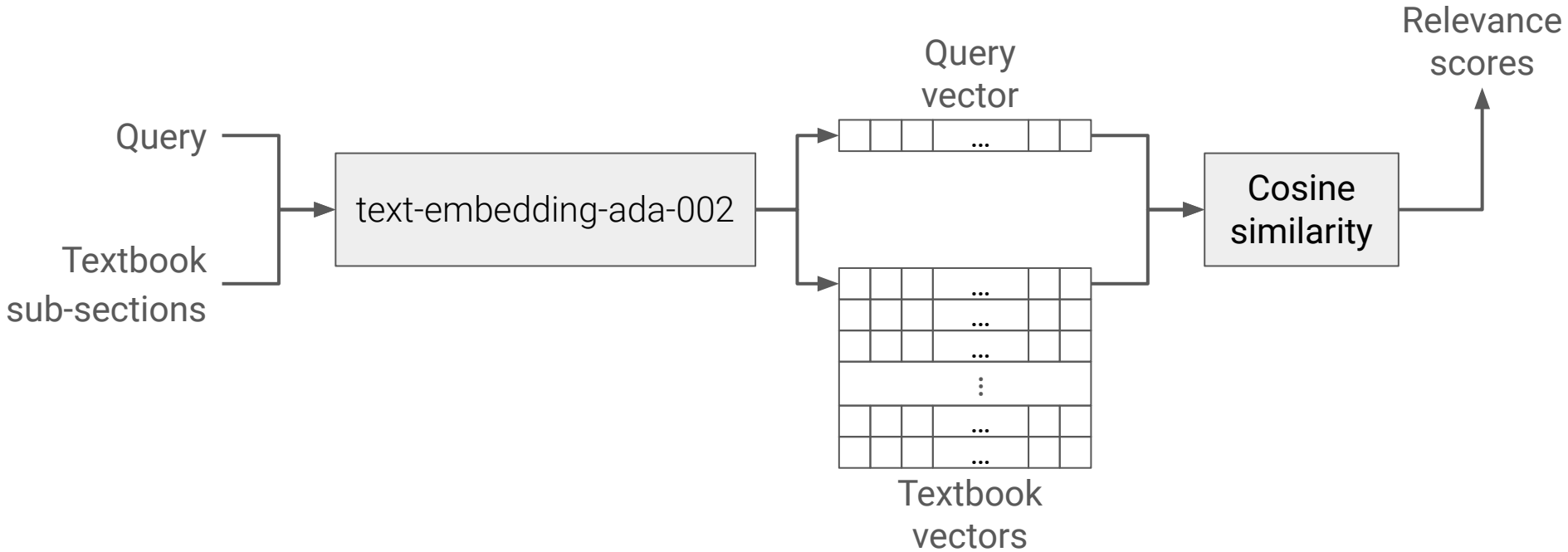


# Retrieval corpus: OpenStax textbook

- Trusted
- Available CC BY 4.0
- Structured (by section)



# Retrieval system



# Parent retrieval

I always forget the difference between commutative and associative. Does anybody know a way to make me remember?

## ▼ 7 The Properties of Real Numbers

Introduction to the Properties of Real Numbers

7.1 Rational and Irrational Numbers

**7.2 Commutative and Associative Properties**

Highest relevance sub-section: “Simplify Expressions Using the Commutative and Associative Properties”



# Prompt Engineering + Retrieval

You are going to act as a mathematics tutor for a 13 year old student who is in grade 8 or 9 and lives in Ghana.

You will be encouraging and factual.

Prefer simple, short responses.

If the student says something inappropriate or off topic you will say you can only focus on mathematics and ask them if they have any math-related follow-up questions.

Use examples and language from the section below to format your response:

===

Use the Commutative and Associative Properties: Think about adding two numbers, such as 5 and 3...

===



# Which is better?

Of course! I can help you with that. Commutative and associative properties are important concepts in mathematics.

Here's a simple way to remember them:

- The commutative property applies to addition and multiplication. It means that the order in which you add or multiply numbers doesn't change the result. For example, in addition,  $2 + 3 = 3 + 2$  and in multiplication,  $4 \times 5 = 5 \times 4$ .

- The associative property also applies to addition and multiplication. It means that the grouping of numbers does not affect the result. For example, in addition,  $(2 + 3) + 4 = 2 + (3 + 4)$  and in multiplication,  $(4 \times 5) \times 6 = 4 \times (5 \times 6)$ .

I hope this helps you remember the difference between commutative and associative properties. Let me know if you have any follow-up questions related to math.

Yes, I can help you remember the difference between commutative and associative properties in mathematics.

The commutative property has to do with order. For addition and multiplication, it means that changing the order of the numbers does not change the result. For example, in addition,  $5 + 3$  is the same as  $3 + 5$ . In multiplication,  $5 \times 3$  is the same as  $3 \times 5$ .

On the other hand, the associative property has to do with grouping. For addition and multiplication, it means that changing the grouping of the numbers does not change the result. For example, in addition,  $(2 + 3) + 4$  is the same as  $2 + (3 + 4)$ . In multiplication,  $(2 \times 3) \times 4$  is the same as  $2 \times (3 \times 4)$ .

To remember the difference, here's a helpful way:

- Commutative starts with "c" just like "change", so it helps to remember that the commutative property is about changing the order.

- Associative starts with "a" just like "associate", so it helps to remember that the associative property is about grouping or associating numbers together.

I hope this makes it easier for you to remember the difference between commutative and associative properties! Let me know if you have any more questions.

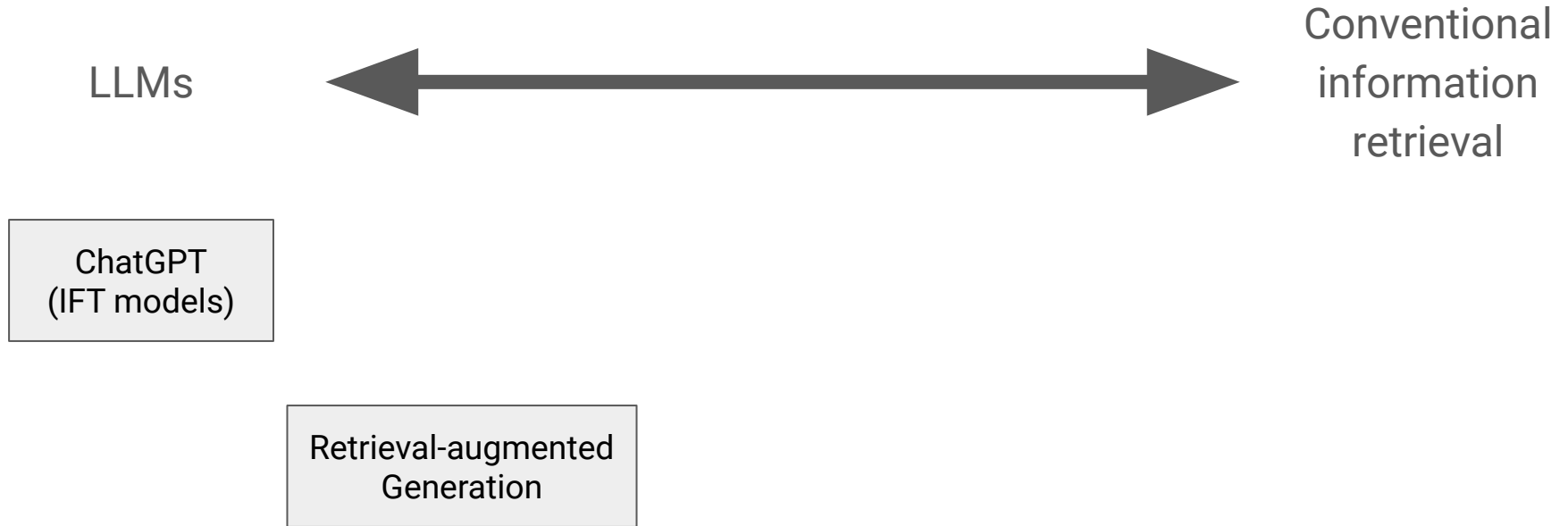




# The automated student Q&A continuum™



# Groundedness



# Knowledge F1

- Turn **retrieved text** and **generated response** into bags of words
- Compute F1:
  - Retrieved text is the ground truth
  - Generated response is the prediction



# Knowledge F1++

- Turn **retrieved text** and **generated response** into bags of words
- Compute F1:
  - Retrieved text is the ground truth
  - Generated response is the prediction

## **The Dangers of trusting Stochastic Parrots: Faithfulness and Trust in Open-domain Conversational Question Answering**

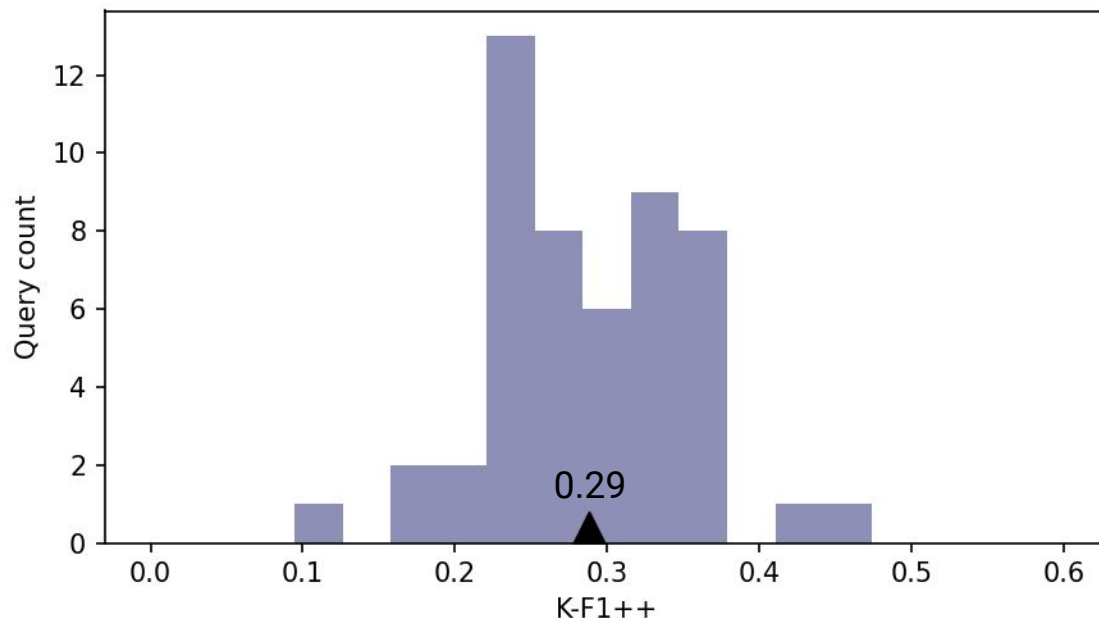
**Sabrina Chiesurin\* Dimitris Dimakopoulos\* Marco Antonio Sobrevilla Cabezudo  
Arash Eshghi Ioannis Papaioannou Verena Rieser† Ioannis Konstas**

Alana AI

hello@alanaai.com



# Knowledge F1++



# Prompt Engineering + Retrieval: Prompt 1

You are going to act as a mathematics tutor for a 13 year old student who is in grade 8 or 9 and lives in Ghana.

You will be encouraging and factual.

Prefer simple, short responses.

If the student says something inappropriate or off topic you will say you can only focus on mathematics and ask them if they have any math-related follow-up questions.

Use examples and language from the section below to format your response:

===

Use the Commutative and Associative Properties: Think about adding two numbers, such as 5 and 3...

===



# Prompt Engineering + Retrieval: Prompt 2

You are going to act as a mathematics tutor for a 13 year old student who is in grade 8 or 9 and lives in Ghana.

You will be encouraging and factual.

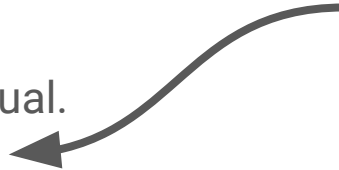
Prefer simple, short responses.

If the student says something inappropriate or off topic you will say you can only focus on mathematics and ask them if they have any math-related follow-up questions.

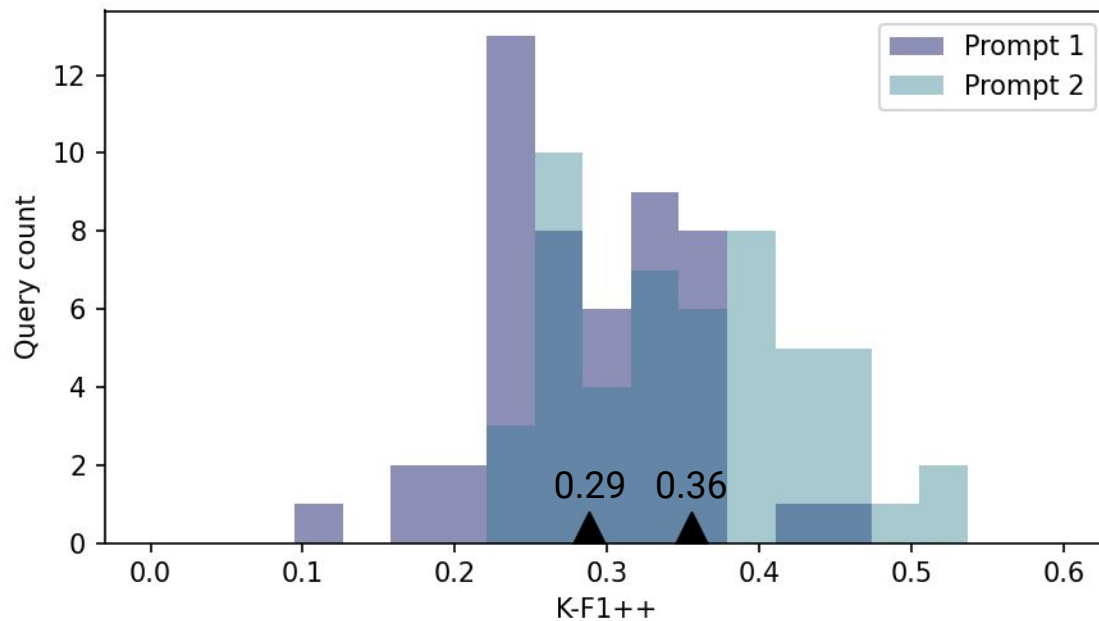
**Reference content from this textbook section in your response:**

Use the Commutative and Associative Properties: Think about adding two numbers, such as 5 and 3...

End your response by relating the question to an example or definition in the textbook.

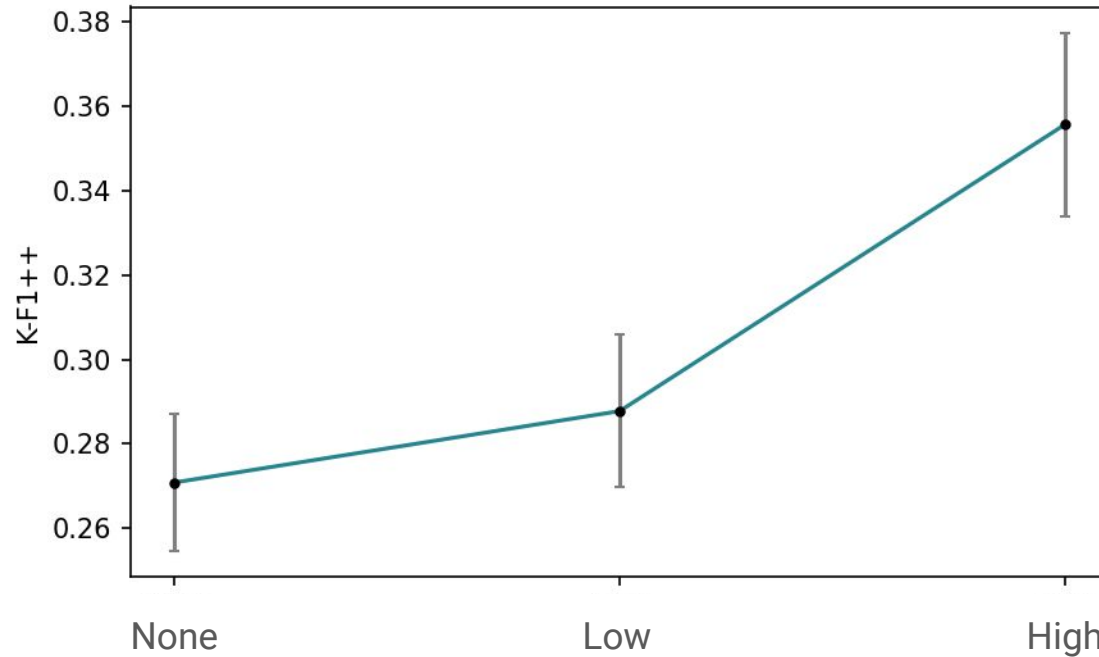


# Knowledge F1++

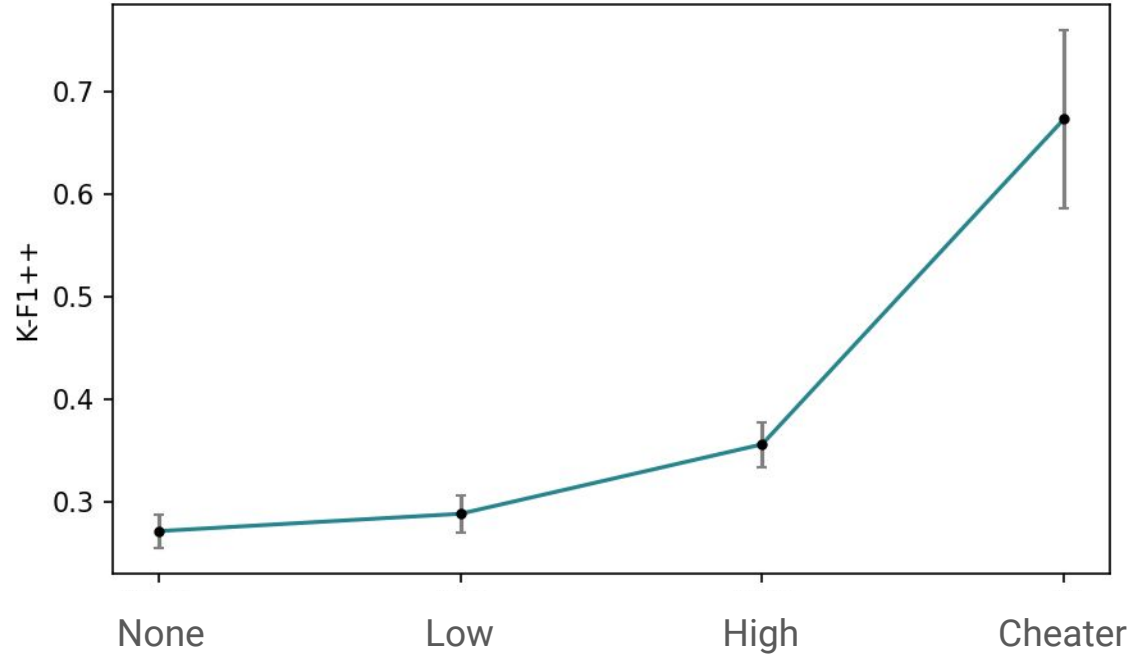




# Knowledge F1++



# Knowledge F1++



# Too much groundedness?

- Unclear how much we should care about groundedness
- Might have a trade-off with other quality metrics
- In the future: faithfulness might be a good solution

## **The Dangers of trusting Stochastic Parrots: Faithfulness and Trust in Open-domain Conversational Question Answering**

**Sabrina Chiesurin\* Dimitris Dimakopoulos\* Marco Antonio Sobrevilla Cabezudo  
Arash Eshghi Ioannis Papaioannou Verena Rieser† Ioannis Konstas**

Alana AI

hello@alanaai.com

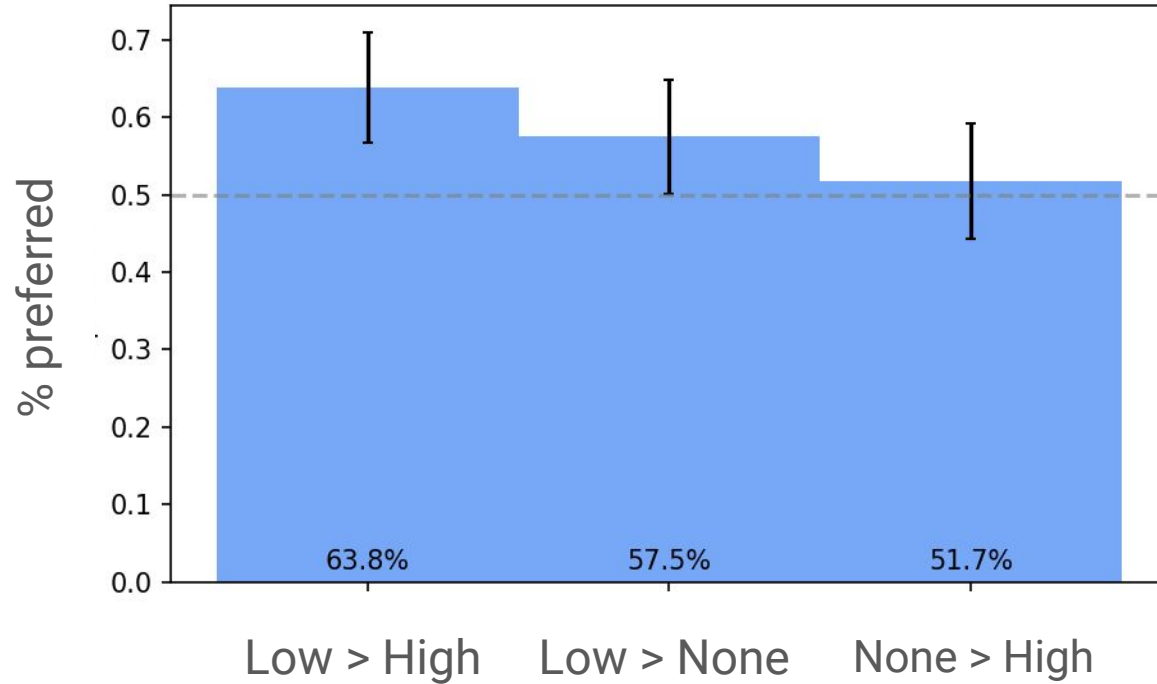


# Human annotation

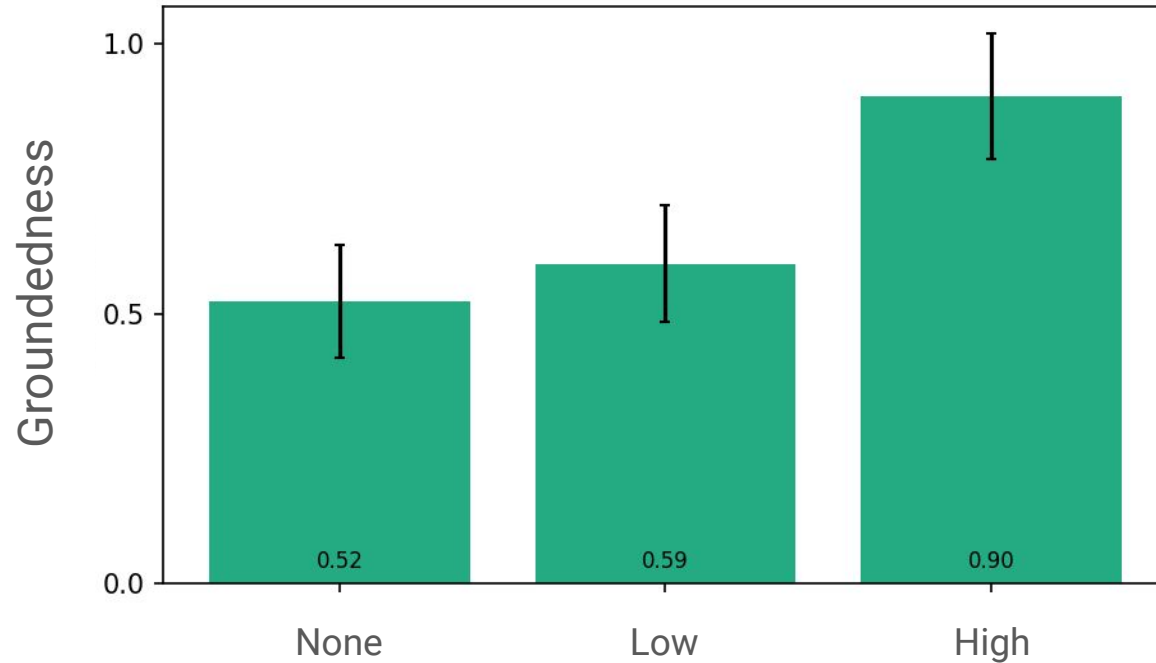
- 9 educators and researchers
- Three conditions:
  - None (No retrieval-augmented generation)
  - Low guidance
  - High guidance
- Two constructs:
  - Preference (ranking)
  - Groundedness (3-level ordinal scale)



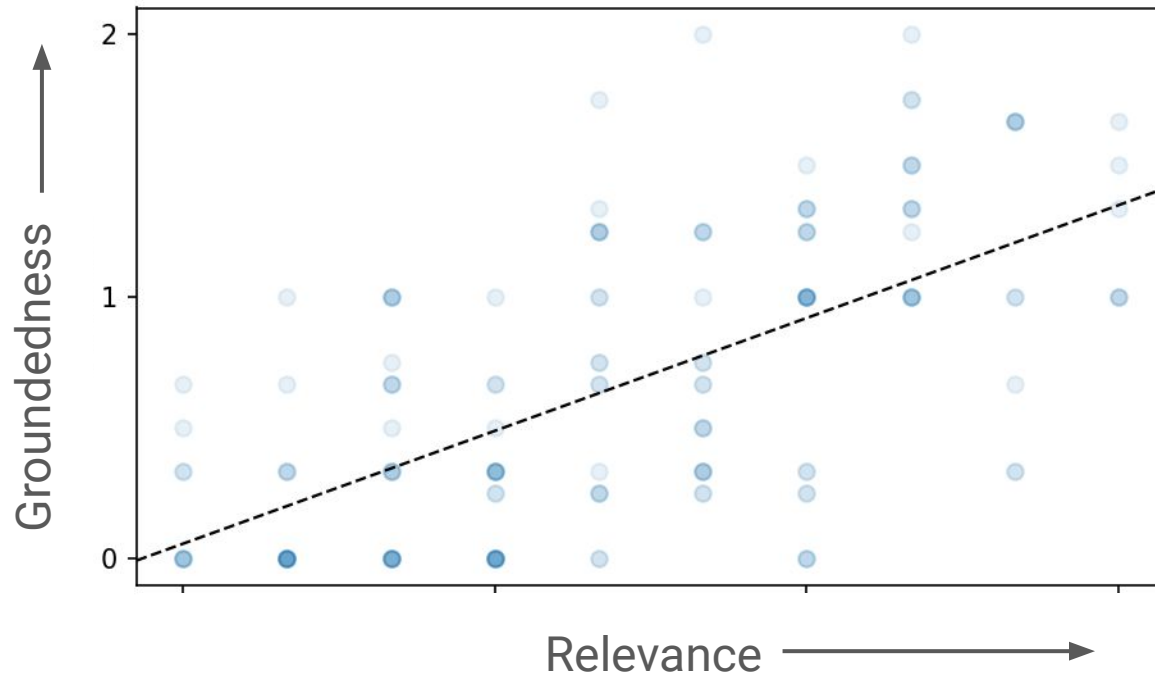
# Preference



# Perceived groundedness



# Groundedness vs Relevance



# Large Language Models Can Be Easily Distracted by Irrelevant Context

*Freda Shi, Xinyun Chen, Kanishka Misra, Nathan Scales, David Dohan, Ed H. Chi, Nathanael Schärli, Denny Zhou* Proceedings of the 40th International Conference on Machine Learning, PMLR 202:31210-31227, 2023.





# What next?

- Testing with students
- Design question: how to best integrate with Rori?
- Evaluation question: repeatable processes and metrics



# First Tragedy, then Parse: History Repeats Itself in the New Era of Large Language Models

**Naomi Saphra**

Kempner Institute at Harvard University  
nsaphra@fas.harvard.edu

**Eve Fleisig**

University of California - Berkeley  
efleisig@berkeley.edu

**Kyunghyun Cho**

New York University & Genentech  
kyunghyun.cho@nyu.edu

**Adam Lopez**

University of Edinburgh  
alopez@inf.ed.ac.uk



**Manaal Faruqui**

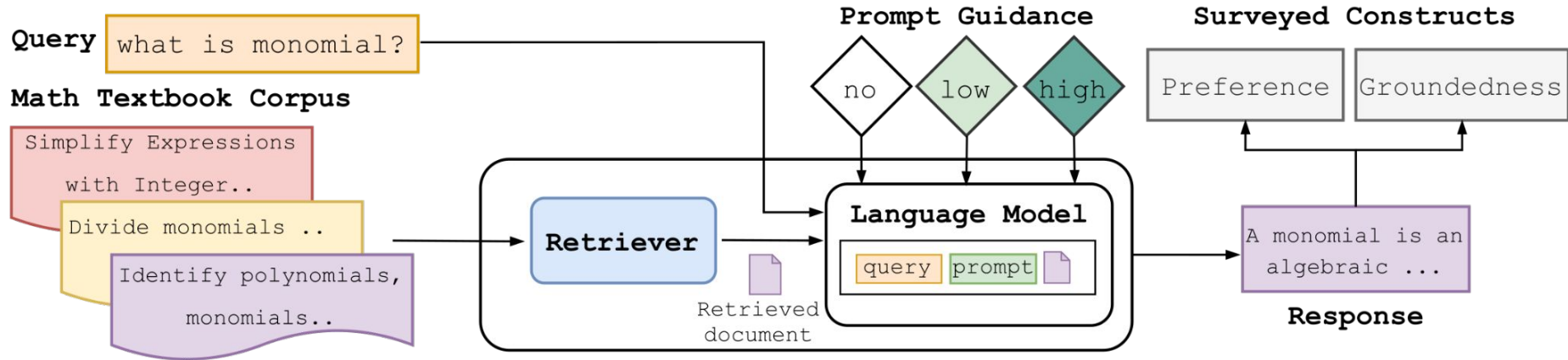
@manaalfar

...

PhD students: can you please solve the problem of long text evaluation?  
It is one of the biggest bottlenecks in the quality iteration of LLMs. Which  
response is more creative? safer? more factual?



# Research summary



# 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

We suggest: use retrieval-augmented generation and prompt engineering!  
But: don't use *too* much prompt guidance.

Code and data available!

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

@zwlevonian@hci.social

