

Aligning LLMs to Answer Questions with Reasoning that Actually Helps You

Nishant Balepur

nbalepur@umd.edu

<https://nbalepur.github.io/>



UNIVERSITY OF
MARYLAND

Why do we ask questions?

Goal: Learn Something New



What does "LLM" mean?



Goal: Solve a Problem



How can I get my refund?



Goal: Receive Tailored Advice



How do I hide the fact that I'm a tourist in London?



Goal: Recall Forgotten Information



Who gave that talk with too many emojis at Imperial?



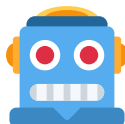
Why do we ask questions?

Goal: Learn Something New



What does “LLM” mean?

An LLM is trained on...

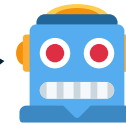


Goal: Solve a Problem



How can I get my refund?

Please hold for 3 hours

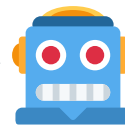


Goal: Receive Tailored Advice



How do I hide the fact that I’m a tourist in London?

You surely can’t, Nishant

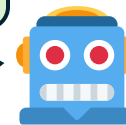


Goal: Recall Forgotten Information



Who gave that talk with too many emojis at Imperial?

You’re listening to it now



The Goal of Question Answering Research:

Building systems that answer questions and are **helpful** for these goals

What do we mean by helpfulness?

When helpfulness is discussed in NLP, it's often "vibes-based":

Our goal is not to define or prescribe what 'helpful' and 'harmless' mean ..., so for the most part we simply let our crowdworkers interpret these concepts [1]



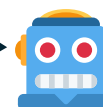
Correctness \nRightarrow Helpfulness



I want to learn how gravity works!



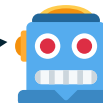
Gravity proportionally governs the attraction between objects with mass...



Correct, less helpful



Gravity is an invisible magnet that pulls large objects towards each other...



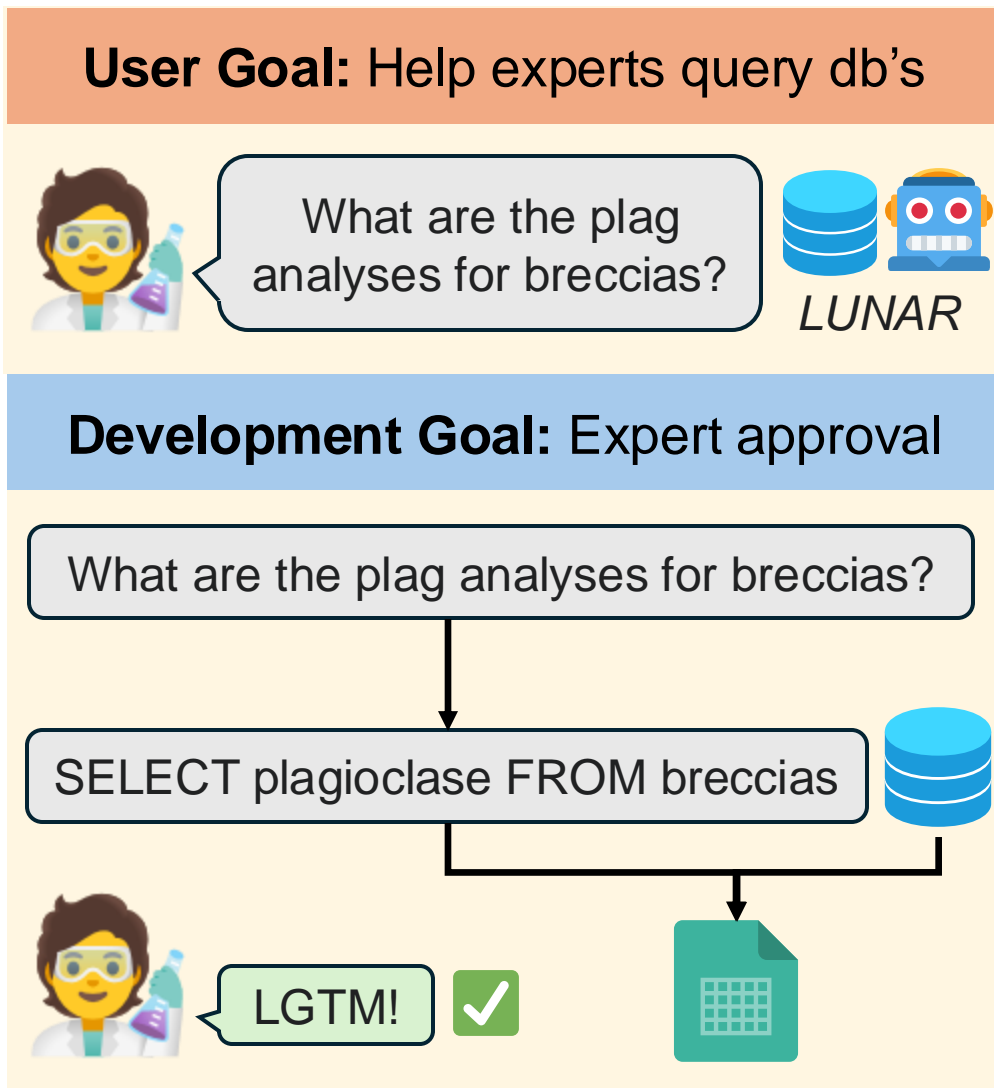
Helpful, less correct

Our definition:

*A question answering system is **helpful** if it provides responses that enable users to maximally achieve their goals*

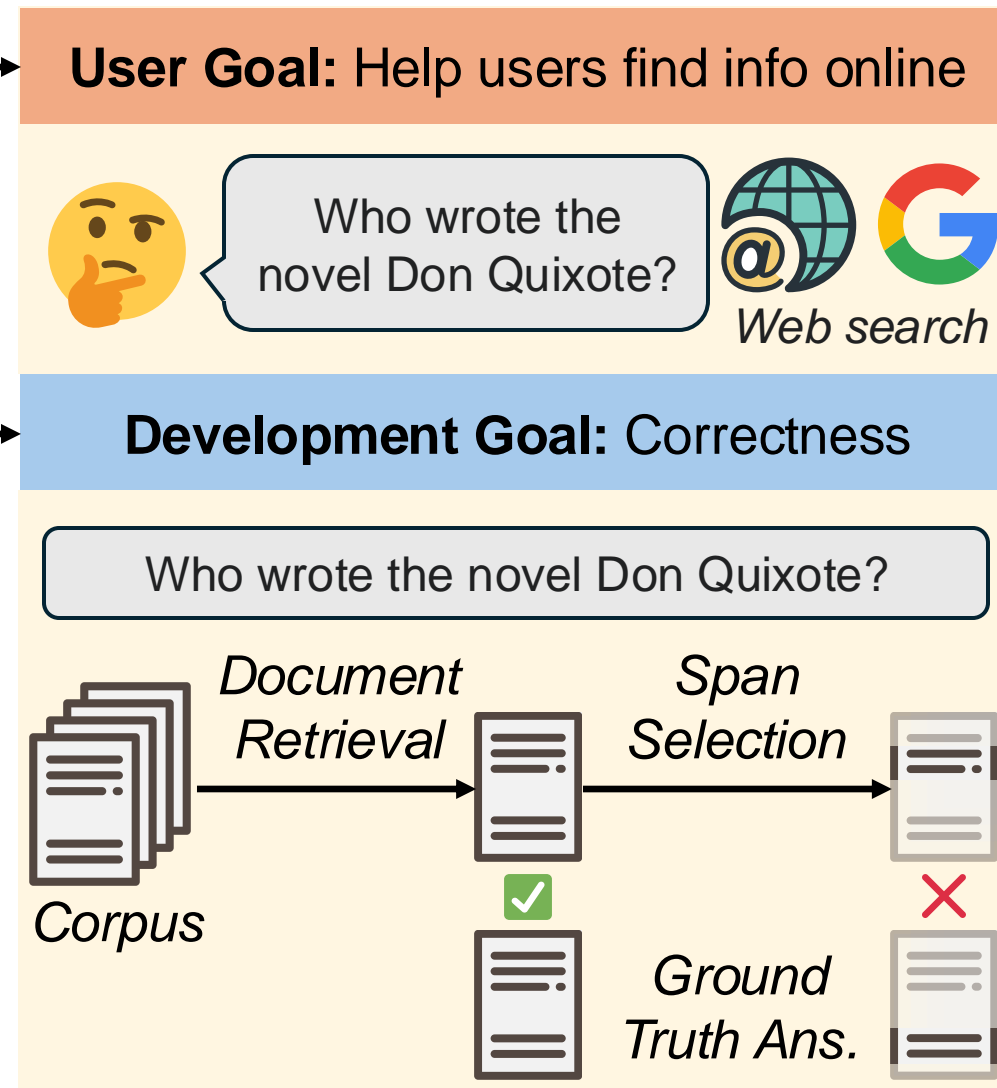
What goals did prior QA systems help users achieve?

1) Domain-Specific QA

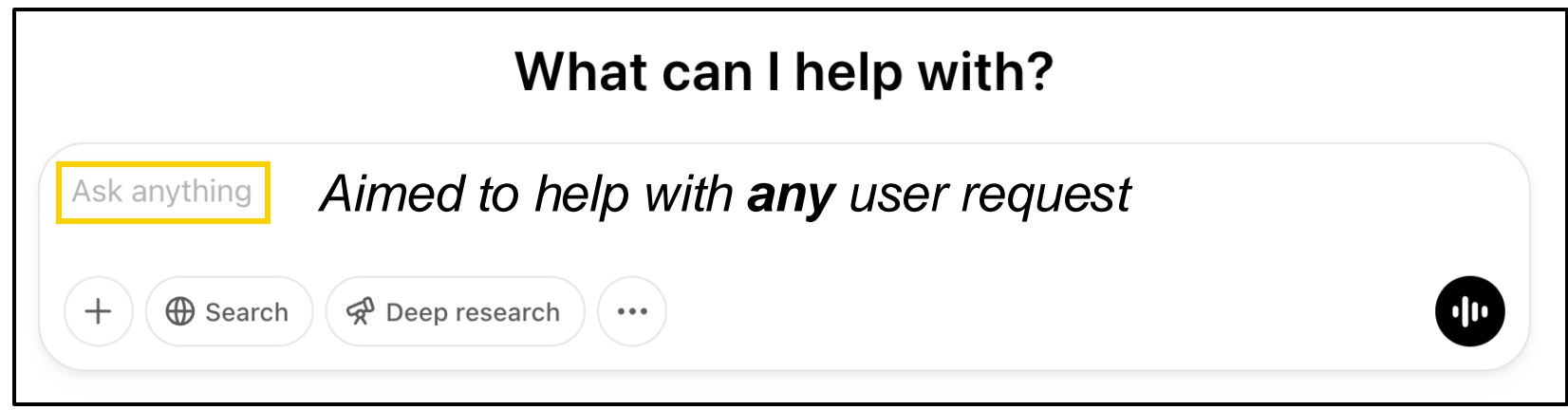


User Goal
Dev. Goals

2) Open-Domain QA



But now, we have Large Language Models



Do LLM ***user goals*** match ***development goals***?

User goals with Large Language Models

Users now ask LLMs questions with a variety of goals

Goal: Learning

What's a memorable way I can remember the capital of France?

A fun and memorable way to remember that **Paris** is the capital of France is:

◆ "A Pair (Paris) of Fancy Pants (France)"

Goal: Personalization

What's the capital of France? I'm thinking of traveling there

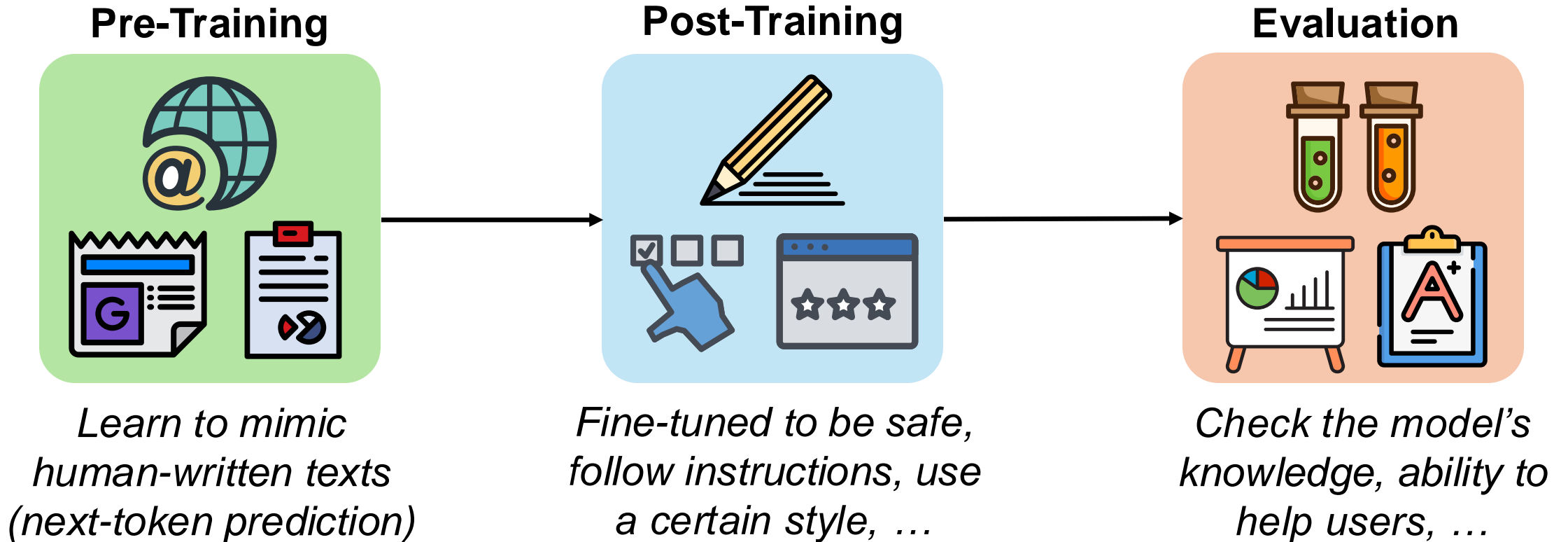
The capital of France is **Paris!** It's a beautiful city known for its iconic landmarks like the Eiffel Tower, the Louvre Museum, and Notre-Dame Cathedral. If you're planning to visit, let me know if you need recommendations on places to see, things to do, or where to eat! 😊

Users don't just expect **correct answers**, but **reasoning chains*** to help them

*AI Hype Disclaimer: "reasoning" is not meant to convey true reasoning/faithfulness, but the utility of these generations

Development goals with Large Language Models

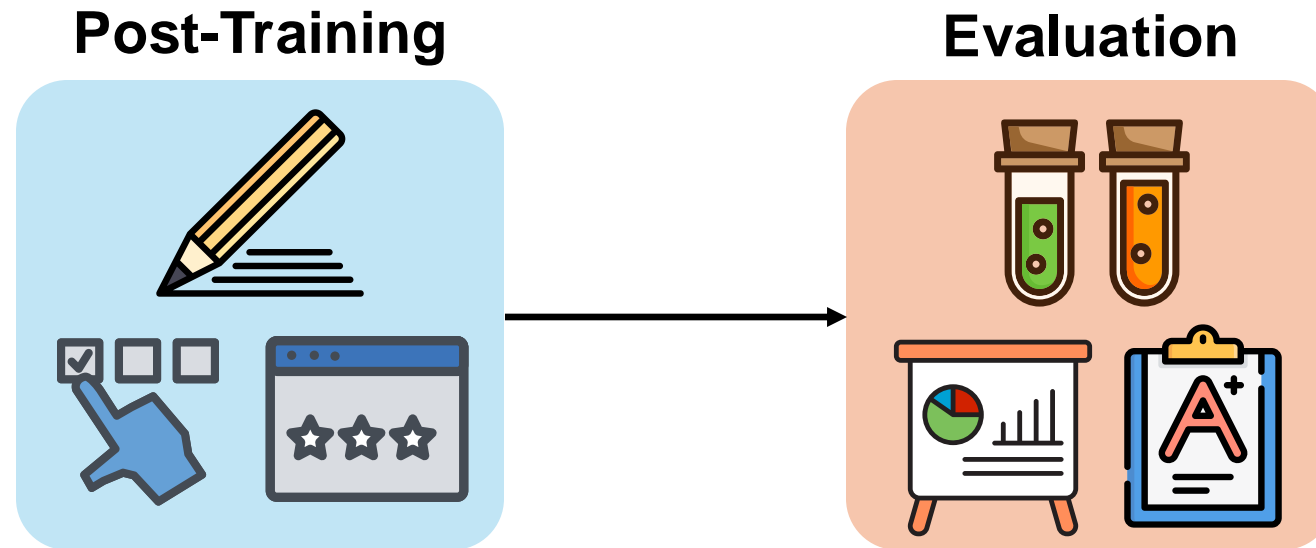
LLMs are developed to be strong text generation systems



Is LLM development aligned with helping users?

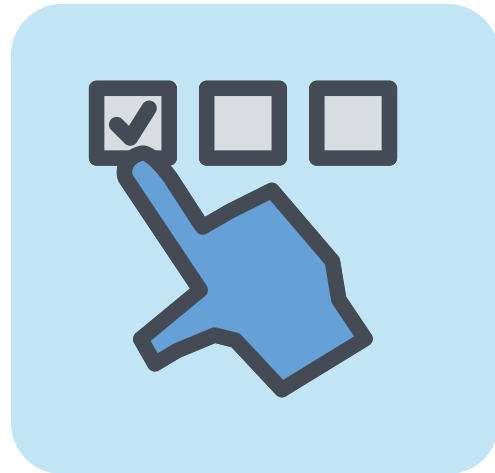


Is LLM development aligned with helping users?

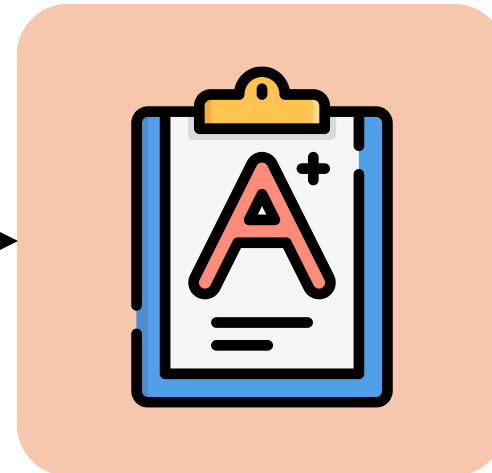


Is LLM development aligned with helping users?

Preference Training



Correctness Evaluation



Is LLM development aligned with helping users?

Preference Training



Correctness Evaluation



Is LLM development aligned with helping users?

Preference Training

Question

What's the capital of France?

Chosen

Paris is the current capital of France, known for its food...



Rejected

Starts with a "P" and rhymes with "Maris"—it's Paris!



*Trained on responses **most users** perceive to be helpful*

Correctness Evaluation

Multiple-Choice Question Answering

Question: What's the capital of France?

(A) Paris

(B) London

(C) The Moon

Answer: (A) ✓

Factual Question Answering

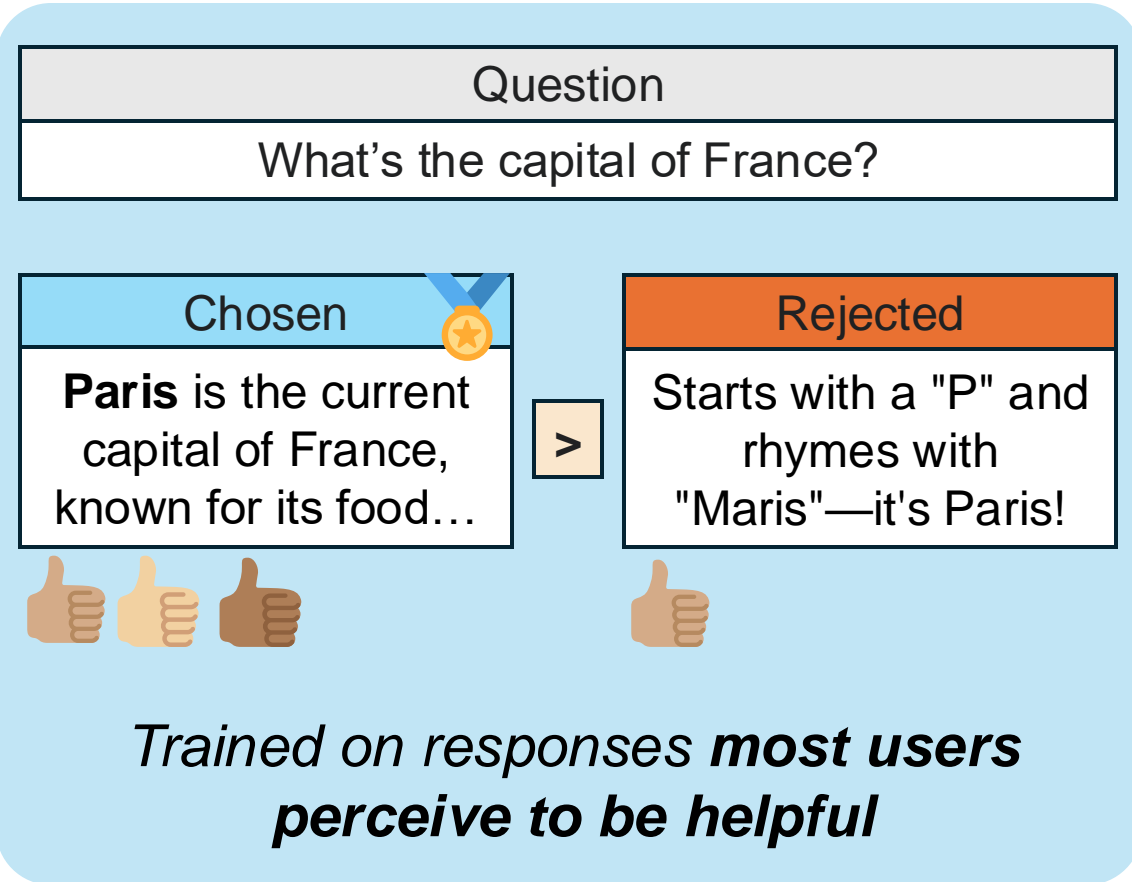
Question: What's the capital of France?

Answer: Paris ✓

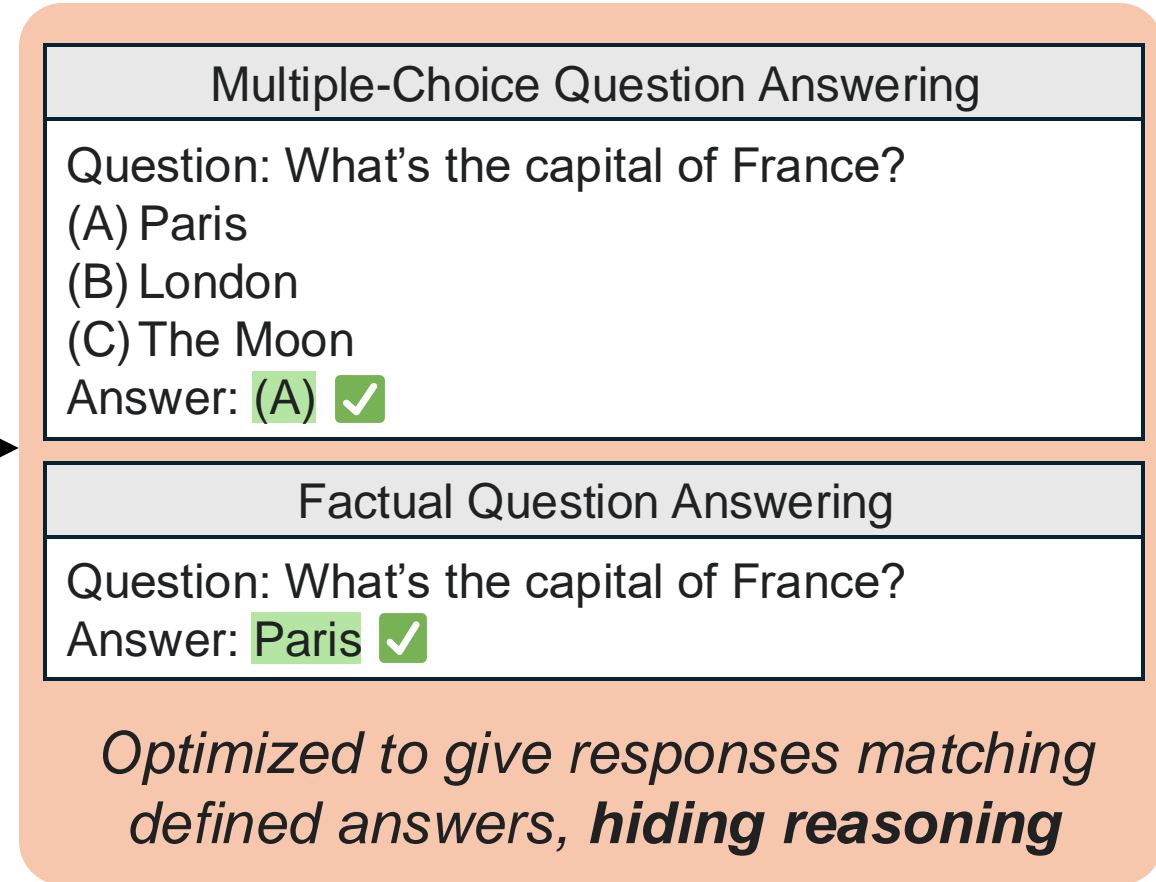
*Optimized to give responses matching defined answers, **hiding reasoning***

Is LLM development aligned with helping users? **No**

Preference Training



Correctness Evaluation

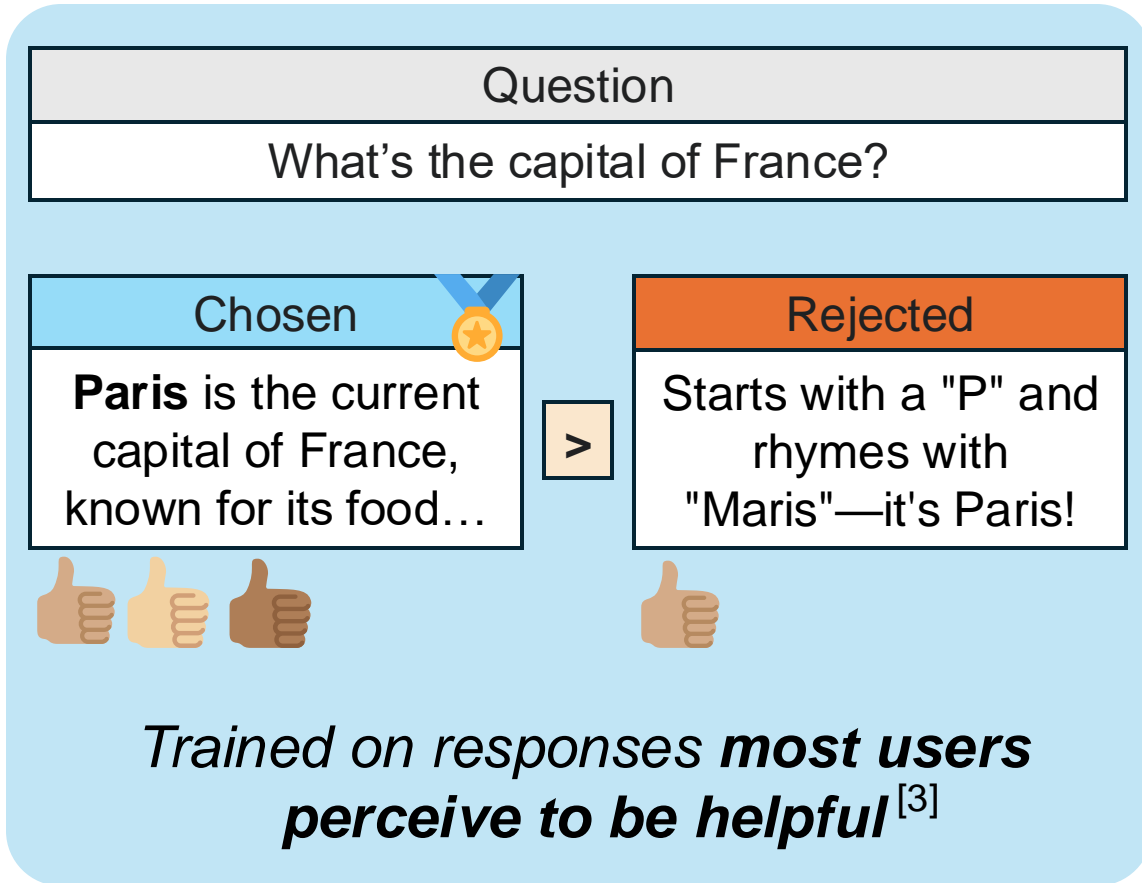


Poor proxies for true helpfulness

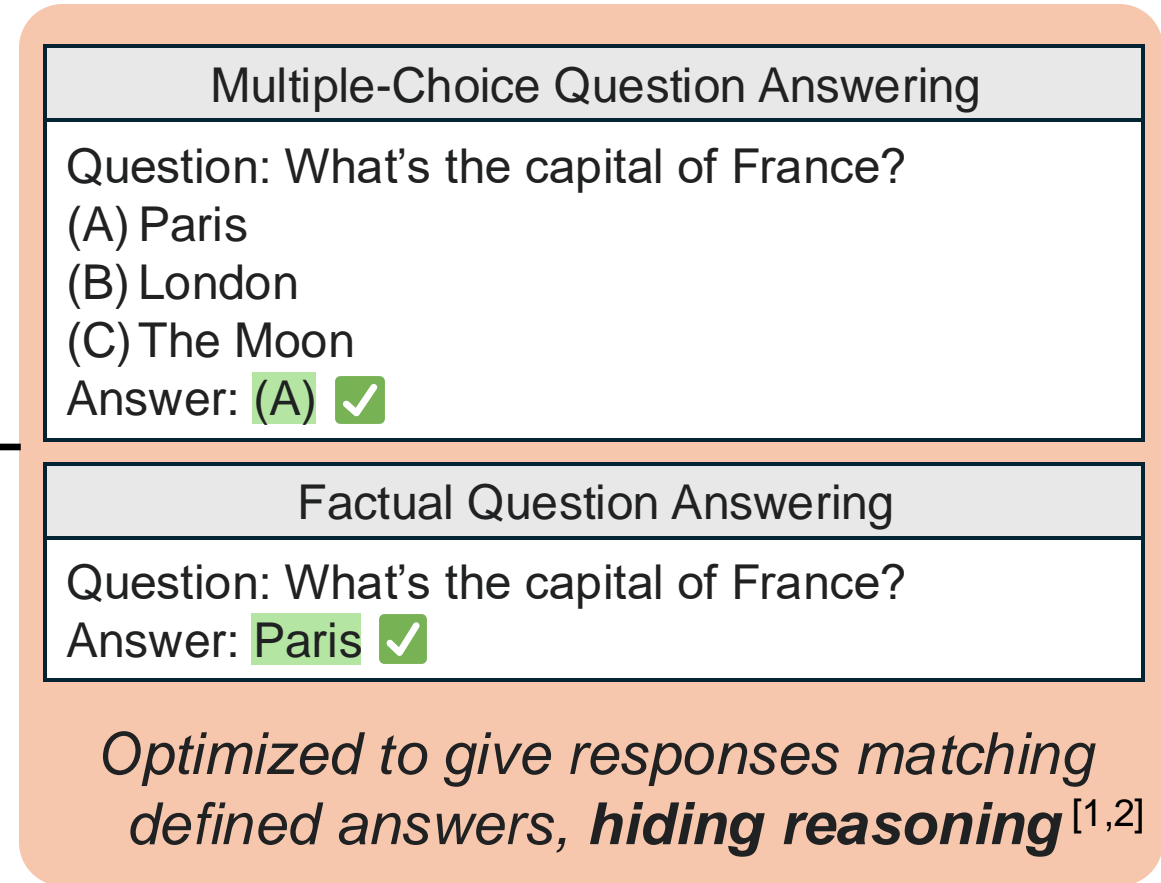


Is LLM development aligned with helping users? **No**

Preference Training



Correctness Evaluation

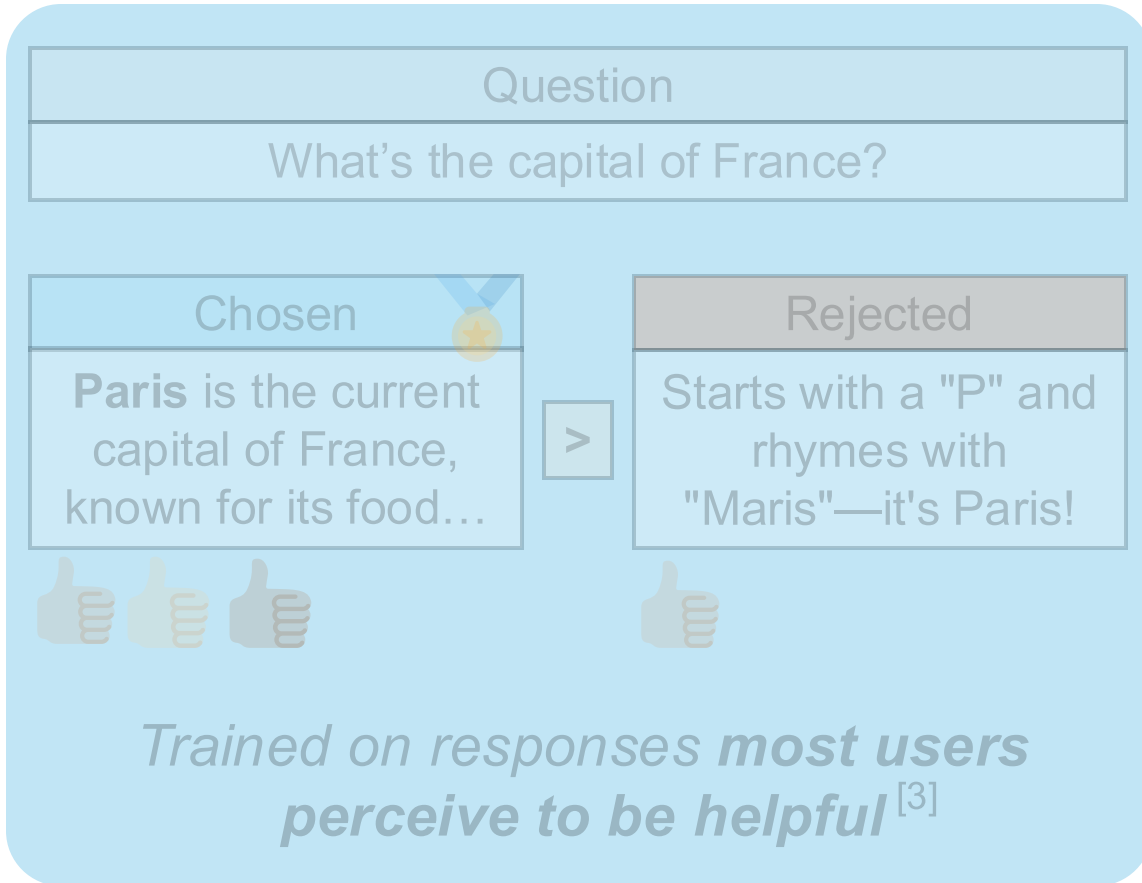


[3] A SMART Mnemonic Sounds like "Glue Tonic" (EMNLP 2024)

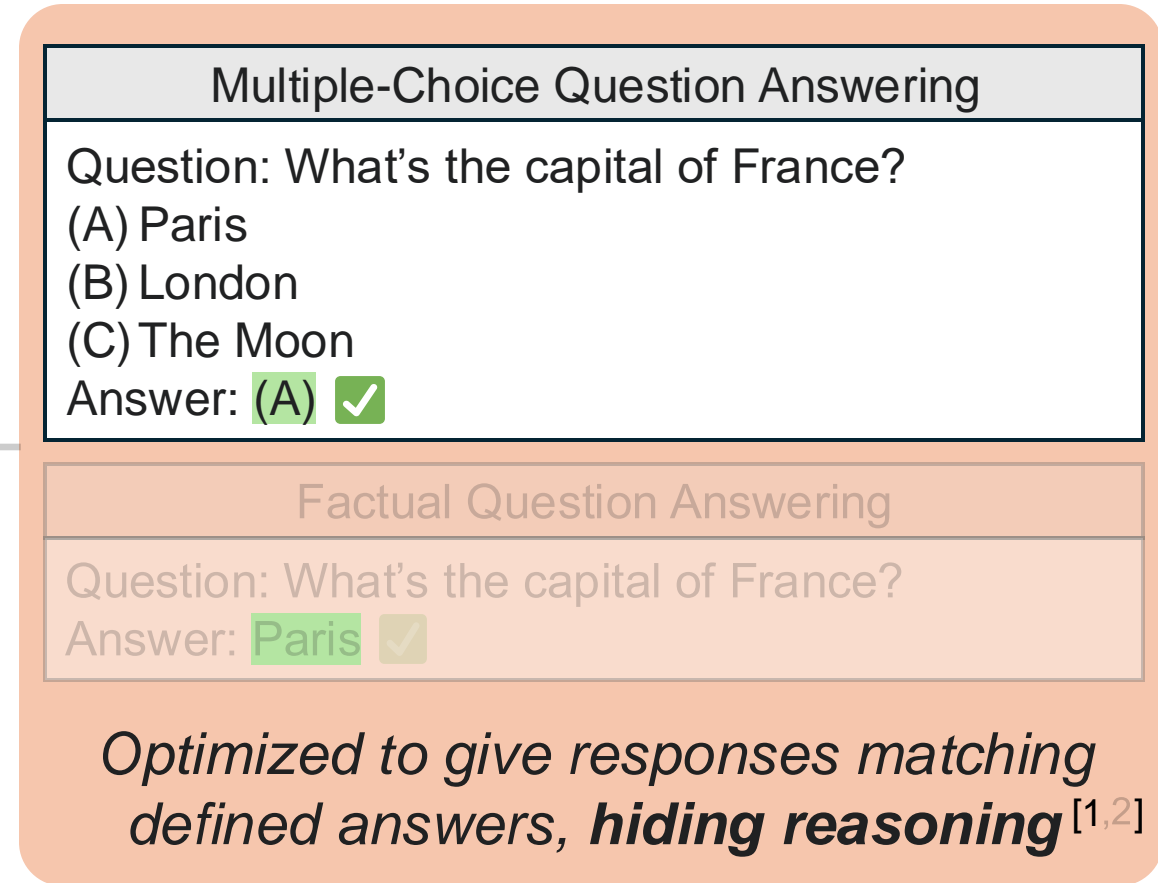
[1] It's Not Easy Being Wrong (ACL 2024, Findings)
[2] Reverse Question Answering (NAACL 2025)

Is LLM development aligned with helping users? **No**

Preference Training



Correctness Evaluation



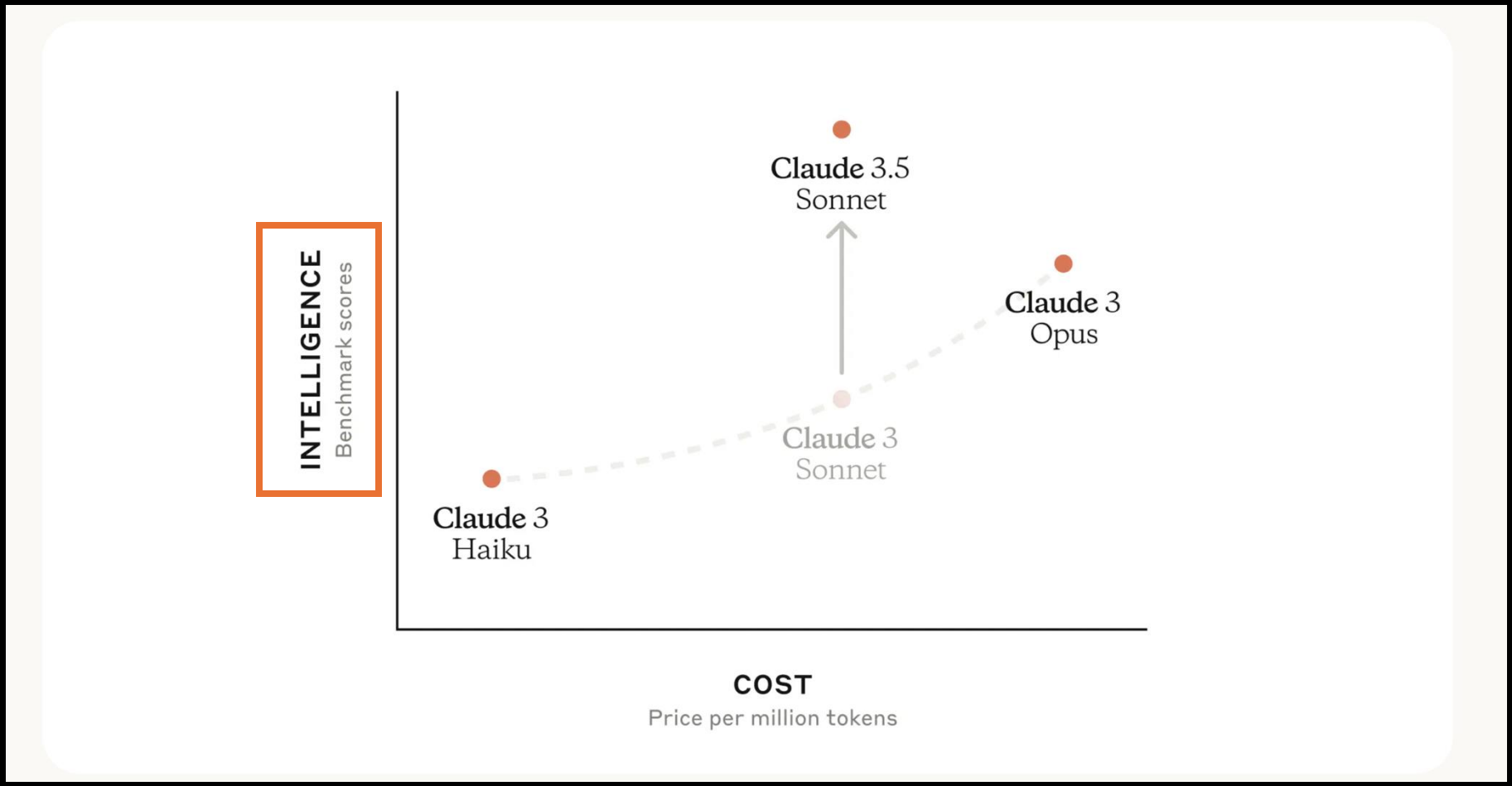
[3] A SMART Mnemonic Sounds like "Glue Tonic" (EMNLP 2024)

[1] It's Not Easy Being Wrong (ACL 2024, Findings)

[2] Reverse Question Answering (NAACL 2025)

Multiple-Choice Questions are Standard for LLM Evaluations

Simple and mirrors human testing



Type	Model
◆	MaziyarPa
☰	MaziyarPa
☰	dfurman/C
☰	MaziyarPa

MMLU...
70.03 %
68.72 %
66.80 %
66.69 %

MCQA Task Format

Given a question and set of choices, LLMs generate the letter of the correct answer

Direct Answer
Question: What's the capital of France? (A) London (B) Paris Answer: (B)

MCQA Task Format

Given a question and set of choices, LLMs generate the letter of the correct answer

Direct Answer + Chain-of-Thought ^[1, 2]
Question: What's the capital of France? (A) London (B) Paris Answer: Let's think step by step. (B)

[1] Wei et. al., Chain-of-Thought Prompting Elicits Reasoning in Large Language Model (2022)

[2] Kojima et. al., Large Language Models are Zero-Shot Reasoners (2022)

MCQA Task Format

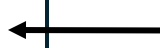
Given a question and set of choices, LLMs generate the letter of the correct answer

Direct Answer + Chain-of-Thought ^[1, 2]
Question: What's the capital of France? (A) London (B) Paris Answer: Let's think step by step. Paris is the only city in the choices that is found in France. So the correct answer is: (B)

Helpful reasoning



Boost accuracy



(more context, justifications...)

[1] Wei et. al., Chain-of-Thought Prompting Elicits Reasoning in Large Language Model (2022)

[2] Kojima et. al., Large Language Models are Zero-Shot Reasoners (2022)

A *New* MCQA Task Format

Given a question and set of choices, LLMs generate the letter of the ~~correct~~ ^{incorrect} answer

Direct Answer + Chain-of-Thought

Question: What's the capital of France?

(A) London

(B) Paris

Answer: Let's think step by step. Paris is the only city in the choices that is found in France. So the **correct** answer is: **(B)**

Process of Elimination + Chain-of-Thought

Question: What's the capital of France?

(A) London

(B) Paris

Answer: Let's think step by step.

Process of Elimination is a common reasoning strategy used by students!

Our Goal: Can LLMs do this?

Why do Process of Elimination?

Direct Answer + Chain-of-Thought

Question: What's the capital of France?
(A) London
(B) Paris
Answer: Let's think step by step. Paris is the only city in the choices that is found in France. So the correct answer is: (B)

Process of Elimination + Chain-of-Thought

Question: What's the capital of France?
(A) London
(B) Paris
Answer: Let's think step by step. London is the capital of the United Kingdom, which is not in France. So the incorrect answer is: (A)

Simple test of logical consistency in 2-choice settings

Helpful QA reasoning should be adaptable

Useful for diagnoses of exclusion, troubleshooting, ...

Implementing MCQA Strategies

Direct Answer + Chain-of-Thought

Question: What's the capital of France?
(A) London
(B) Paris
Answer:

Process of Elimination + Chain-of-Thought

Question: What's the capital of France?
(A) London
(B) Paris
Answer:

Implementing MCQA Strategies

Direct Answer + Chain-of-Thought

Task: Your goal is to pick the correct...

Question: What's the capital of France?

(A) London

(B) Paris

Answer:

Process of Elimination + Chain-of-Thought

Task: Your goal is to pick the incorrect...

Question: What's the capital of France?

(A) London

(B) Paris

Answer:

Implementing MCQA Strategies

Direct Answer + Chain-of-Thought

Task: Your goal is to pick the correct...

Question: Who wrote Don Quixote?

(A) Miguel de Cervantes ✓

(B) William Shakespeare

Answer: Cervantes wrote Don Quixote in 1605, so the correct answer is (A)

Question: What's the capital of France?

(A) London

(B) Paris

Answer:

Process of Elimination + Chain-of-Thought

Task: Your goal is to pick the incorrect...

Question: Who wrote Don Quixote?

(A) Miguel de Cervantes ✓

(B) William Shakespeare

Answer: William Shakespeare wrote plays, not novels, so the incorrect answer is (B)

Question: What's the capital of France?

(A) London

(B) Paris

Answer:

Implementing MCQA Strategies

Direct Answer + Chain-of-Thought

Task: Your goal is to pick the correct...

Question: Who wrote Don Quixote?

(A) Miguel de Cervantes

(B) William Shakespeare

Answer: Cervantes wrote Don Quixote in 1605, so the correct answer is (A)

[9 more examples...]

Question: What's the capital of France?

(A) London

(B) Paris

Answer:

Process of Elimination + Chain-of-Thought

Task: Your goal is to pick the incorrect...

Question: Who wrote Don Quixote?

(A) Miguel de Cervantes

(B) William Shakespeare

Answer: William Shakespeare wrote plays, not novels, so the incorrect answer is (B)

[9 more examples...]

Question: What's the capital of France?

(A) London

(B) Paris

Answer:

Implementing MCQA Strategies

Direct Answer + Chain-of-Thought

Task: Your goal is to pick the correct...

Question: Who wrote Don Quixote?

(A) Miguel de Cervantes

(B) William Shakespeare

Answer: Cervantes wrote Don Quixote in 1605, so the correct answer is (A)

[9 more examples...]

Question: What's the capital of France?

(A) London

(B) Paris

Answer: *Solve this MCQ with DA*

Process of Elimination + Chain-of-Thought

Task: Your goal is to pick the incorrect...

Question: Who wrote Don Quixote?

(A) Miguel de Cervantes

(B) William Shakespeare

Answer: William Shakespeare wrote plays, not novels, so the incorrect answer is (B)

[9 more examples...]

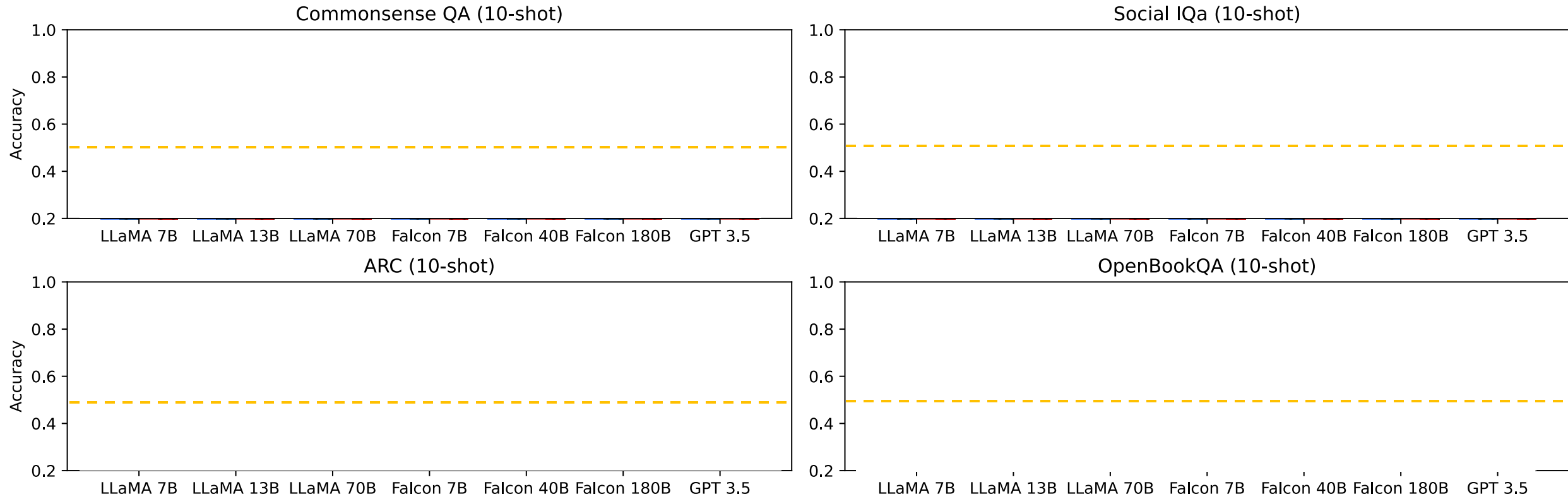
Question: What's the capital of France?

(A) London

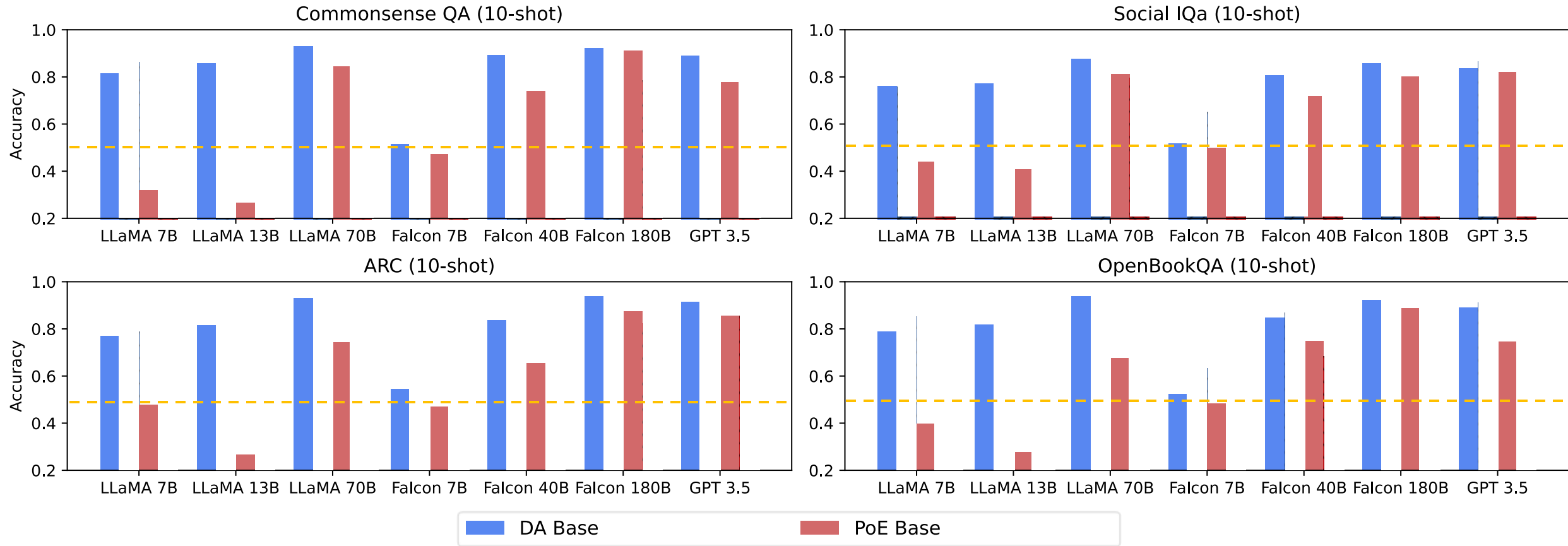
(B) Paris

Answer: *Solve this MCQ with PoE*

LLM Accuracy with Direct Answer vs Process of Elimination

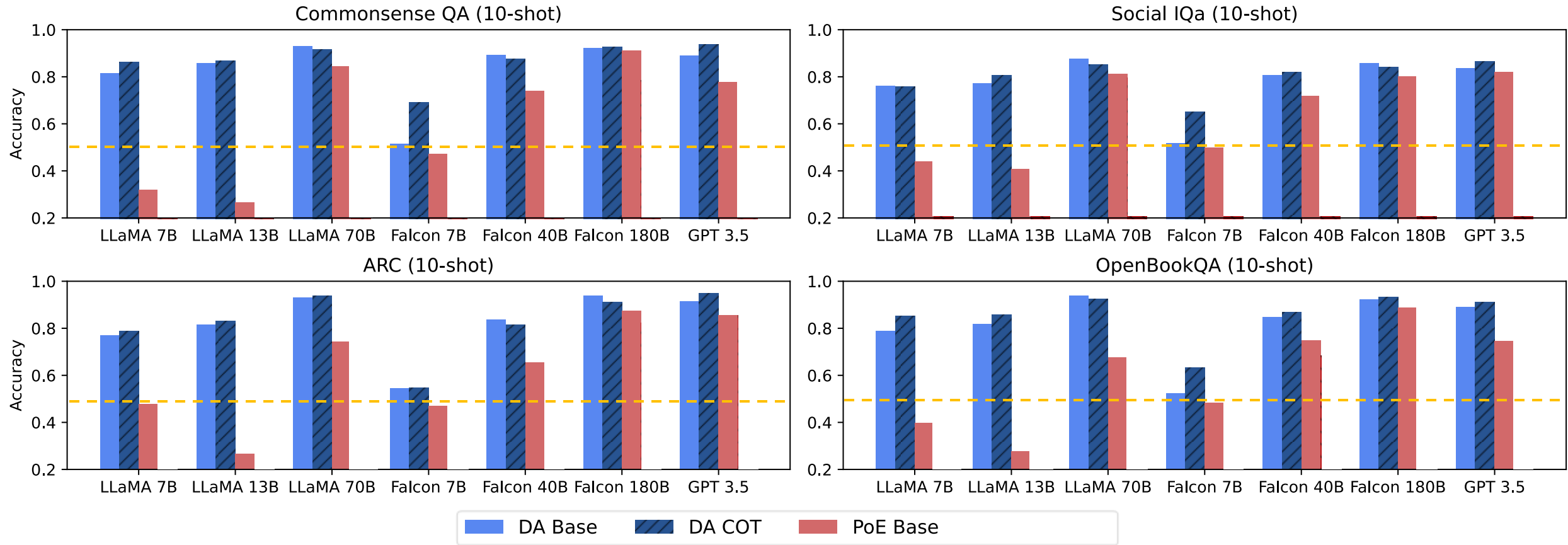


LLM Accuracy with Direct Answer vs Process of Elimination



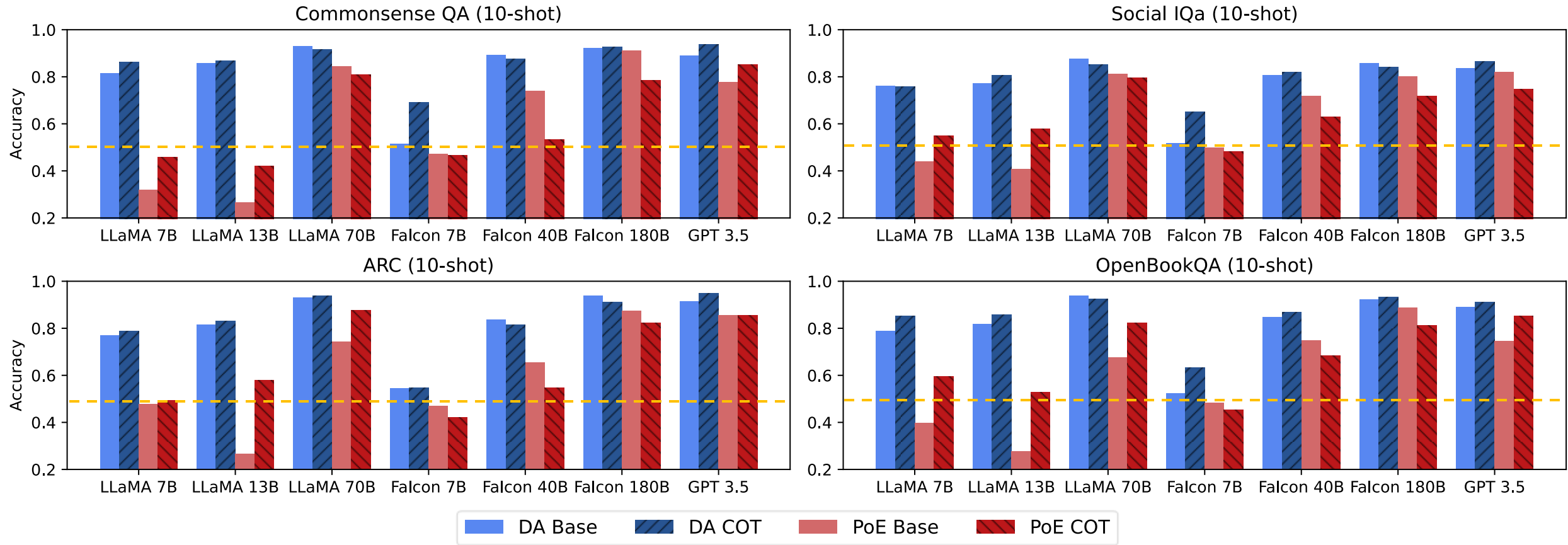
➤ LLMs struggle with eliminating incorrect answers versus picking correct ones

LLM Accuracy with Direct Answer vs Process of Elimination



- LLMs struggle with eliminating correct answers versus picking correct ones
- Chain-of-thought typically helps DA...

LLM Accuracy with Direct Answer vs Process of Elimination



➤ LLMs struggle with eliminating correct answers versus picking correct ones

➤ Chain-of-thought typically helps DA... but often **harms PoE!** 😬

Process of Elimination is Logically Inconsistent

Direct Answer

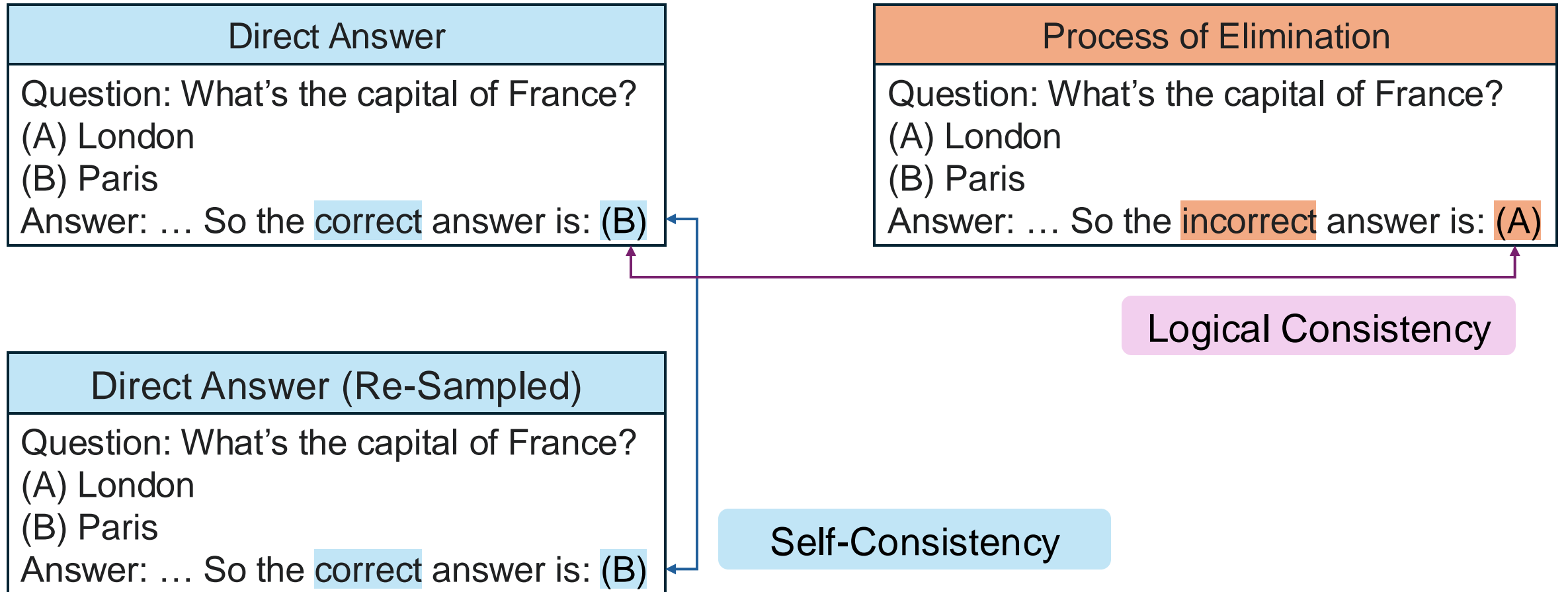
Question: What's the capital of France?
(A) London
(B) Paris
Answer: ... So the correct answer is: (B)

Process of Elimination

Question: What's the capital of France?
(A) London
(B) Paris
Answer: ... So the incorrect answer is: (A)

Logical Consistency

Process of Elimination is Logically Inconsistent



Process of Elimination is Logically Inconsistent

Model: GPT-3.5

Self-Consistency

Logical Consistency

Base

+ CoT

Base

+ CoT

Commonsense QA

97.3%

97.6%

81.3%

86.0%

Social IQA

97.0%

95.2%

80.7%

76.2%

ARC

97.5%

97.0%

86.6%

85.5%

Open Book QA

98.2%

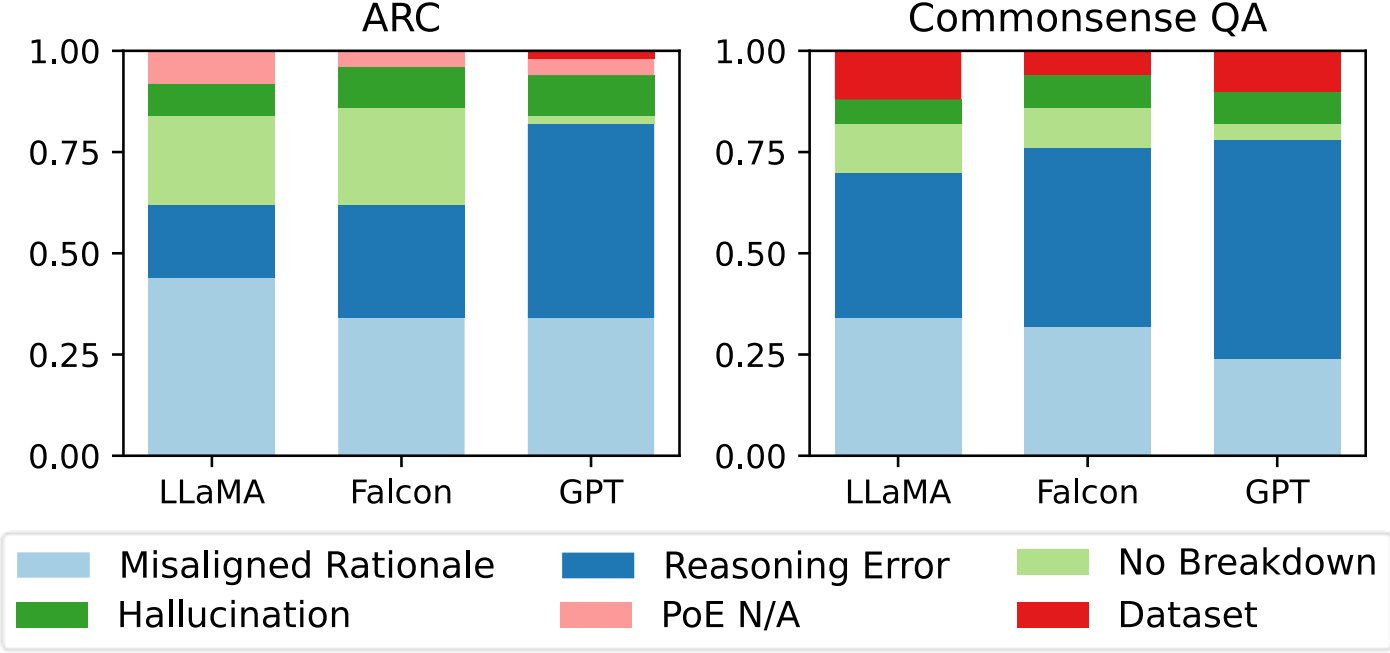
96.8%

72.3%

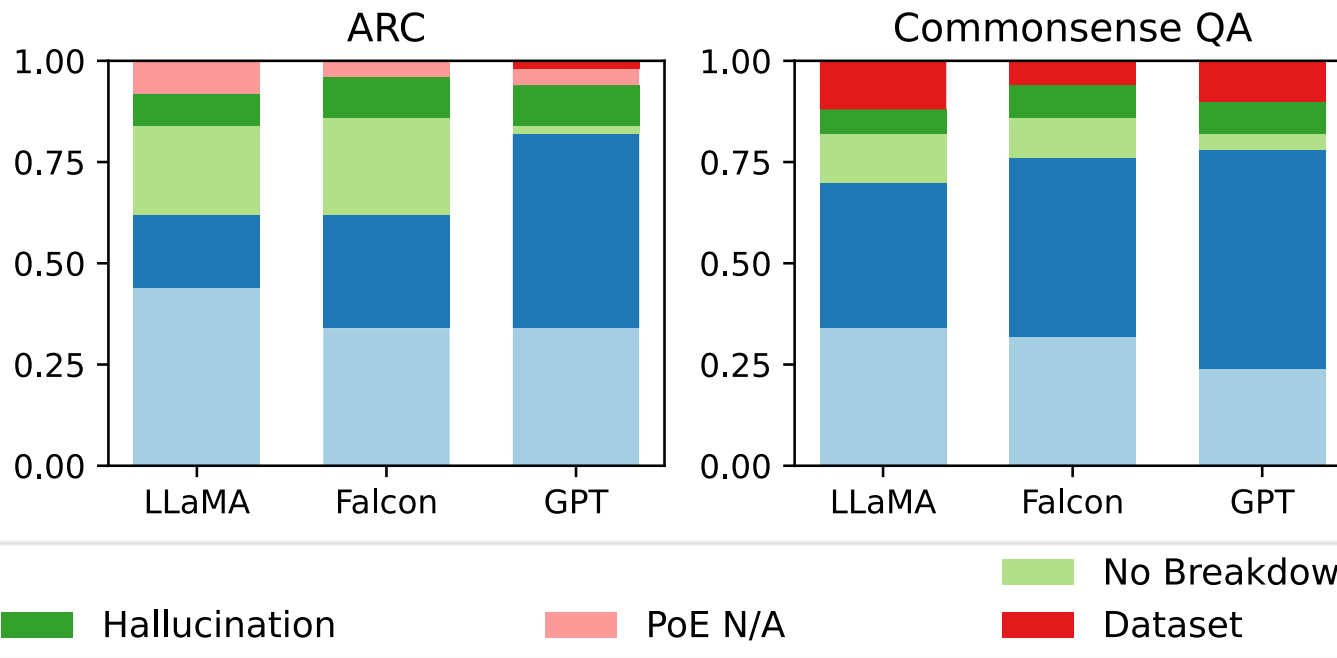
81.4%

>>>

Why does Process of Elimination Fail?



Why does Process of Elimination Fail?



Overfitting to one reasoning type?

MCQA scores != reasoning skills

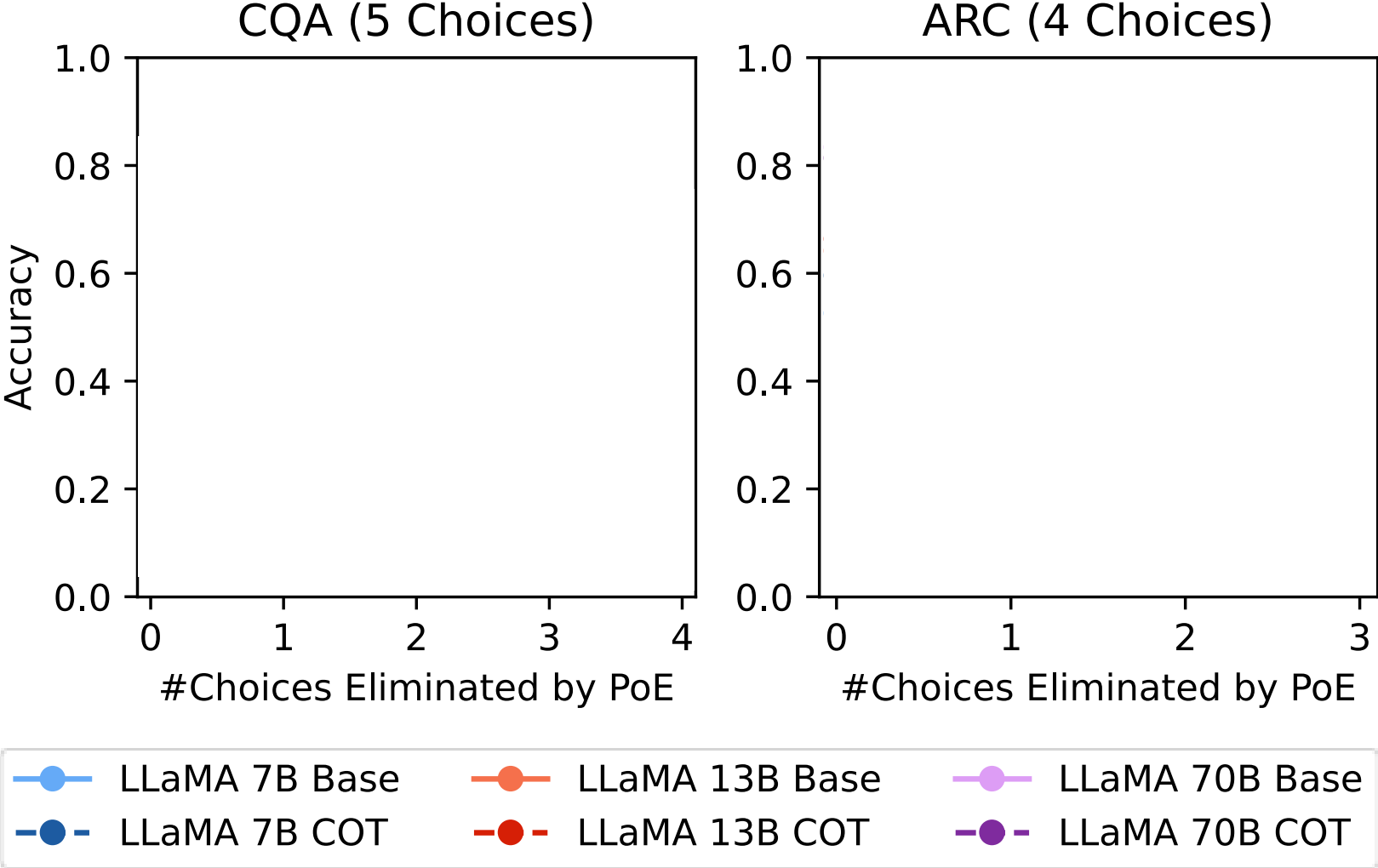
Misaligned Rationale

Reasoning Error

Which molecule does not have a carbon-nitrogen bond?
 (A) nucleic acid (B) carbohydrate

Which will retain the most energy from the sun?
 (A) sand (B) glass


Can we use Process of Elimination Iteratively?




Conclusion: How will this help us build helpful QA systems?

Overall:

- LLMs struggle to eliminate incorrect answers
- This is an issue stemming from logical inconsistencies
- Models may overfit on certain reasoning types (e.g. correctness)

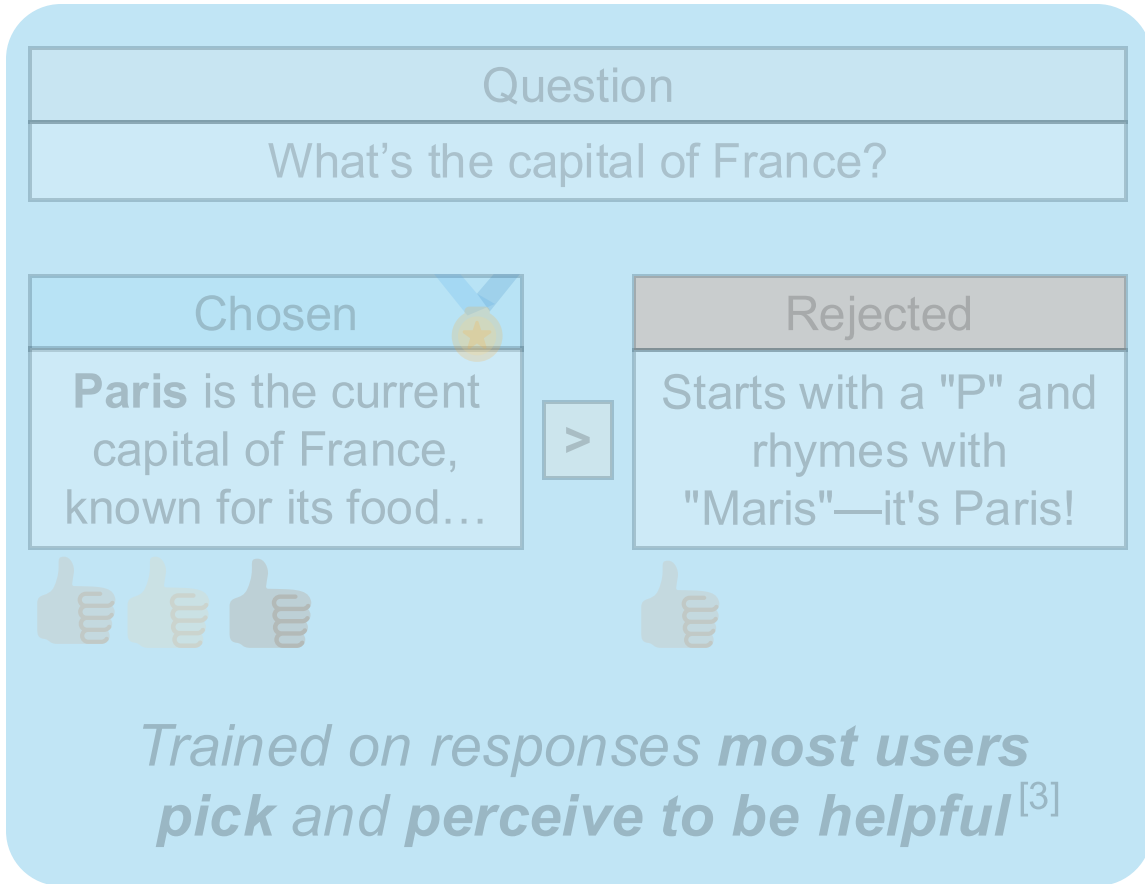
 MCQA correctness cannot evaluate reasoning chains^[1]

 LLMs fail to give adaptable reasoning chains for answers

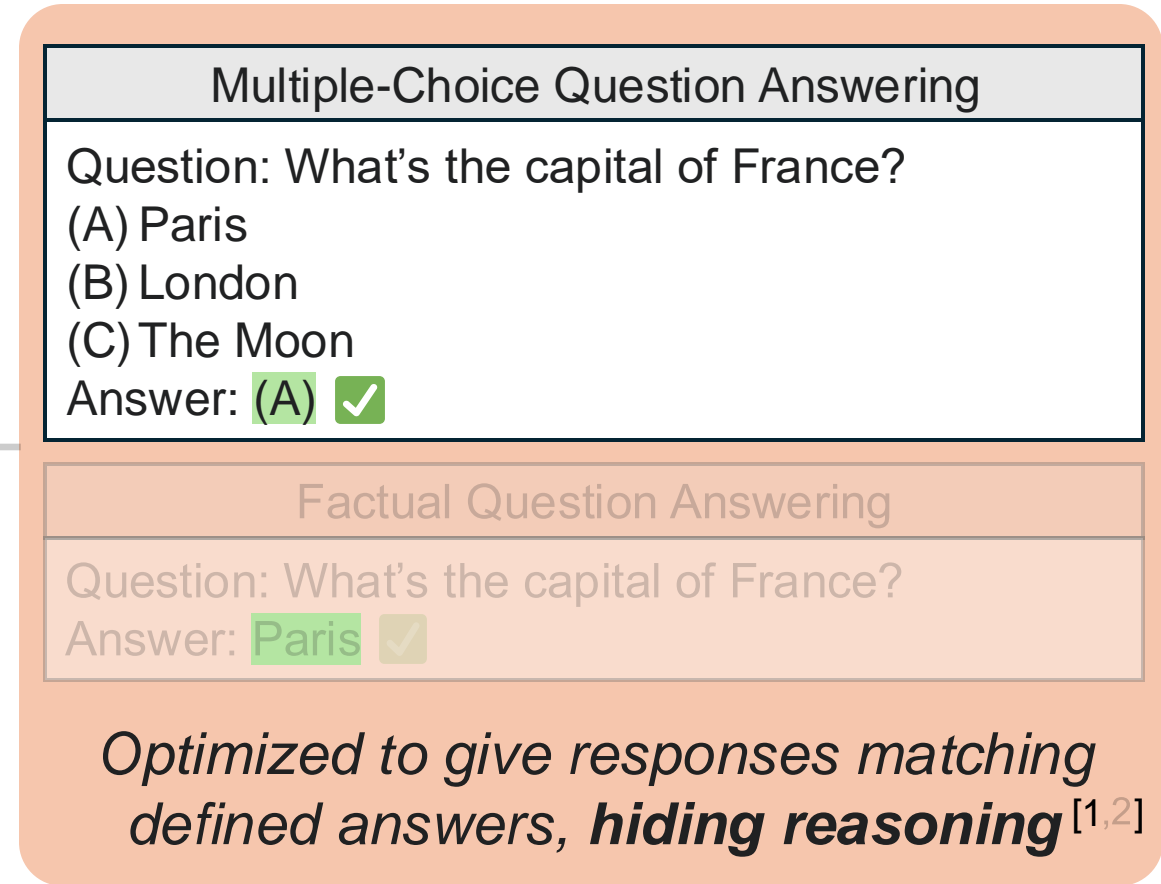
[1] Which of These Best Describes Multiple Choice Evaluation with LLMs? A) Forced B) Flawed C) Fixable D) All of the Above

Is LLM development aligned with helping users? **No**

Preference Training



Correctness Evaluation



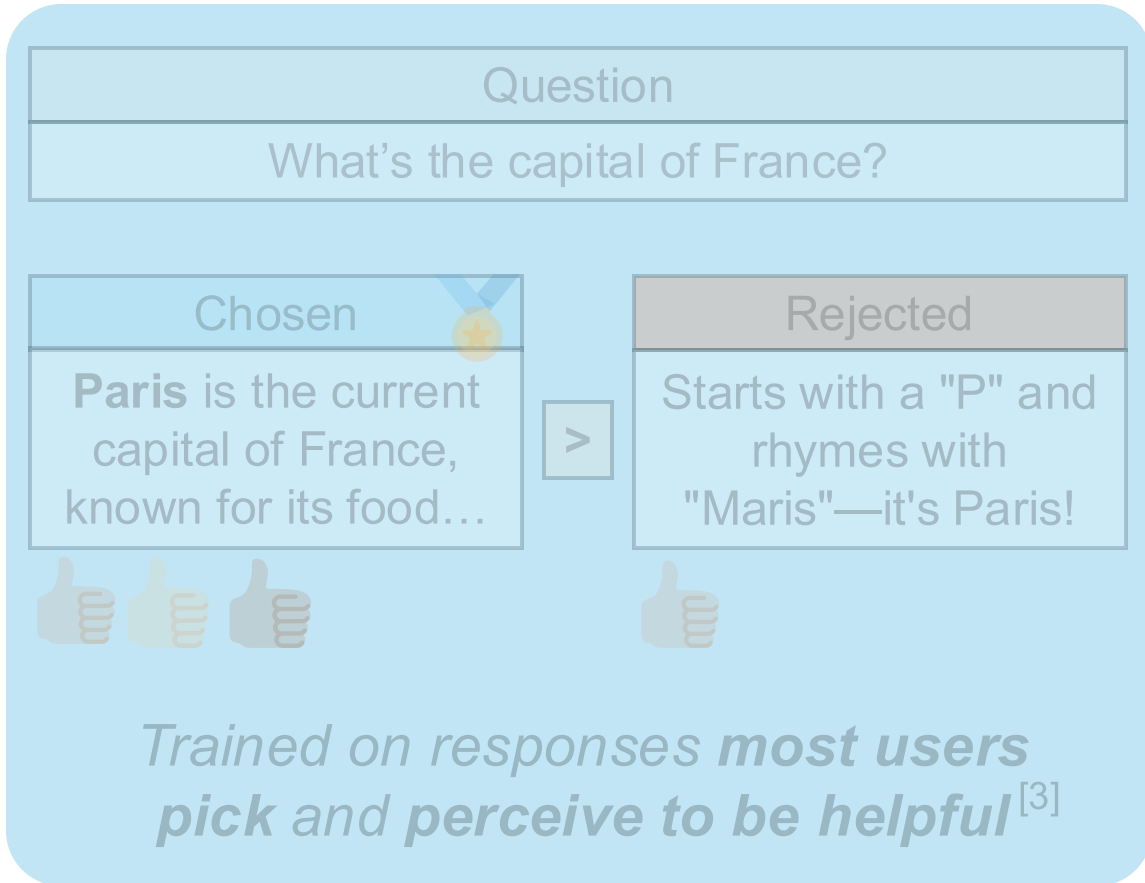
[3] A SMART Mnemonic Sounds like "Glue Tonic" (EMNLP 2024)

[1] It's Not Easy Being Wrong (ACL 2024, Findings)

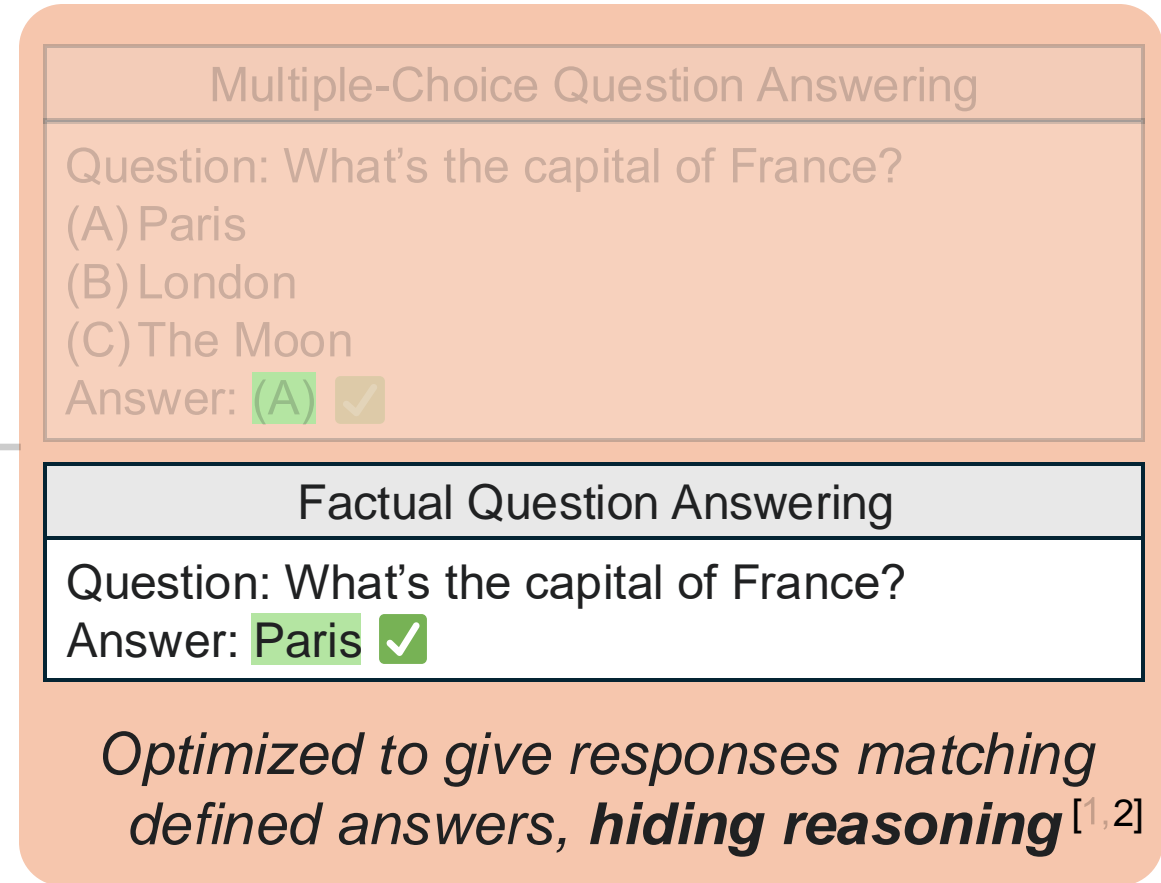
[2] Reverse Question Answering (NAACL 2025)

Is LLM development aligned with helping users? **No**

Preference Training



Correctness Evaluation



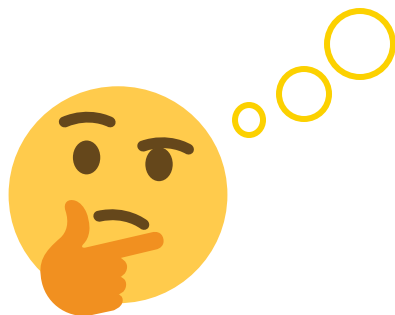
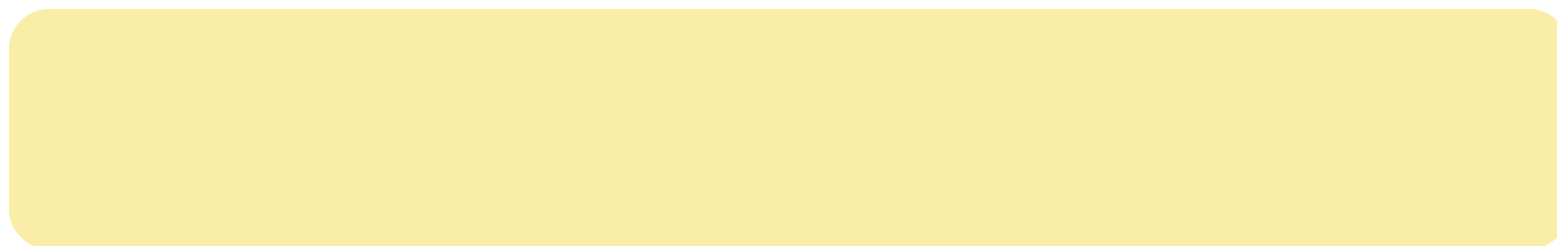
[3] A SMART Mnemonic Sounds like "Glue Tonic" (EMNLP 2024)

[1] It's Not Easy Being Wrong (ACL 2024, Findings)

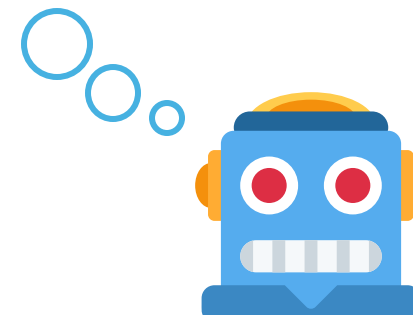
[2] Reverse Question Answering (NAACL 2025)

What abilities does (factual) question answering measure?

Question Answering
Question: What's the nationality of the author of Don Quixote? Answer: Spanish



This reasoning process is **deductive**:
Reaching *the* output conclusion (answer)
based on input premises (question)



But what about other reasoning types?

Deductive

Deriving conclusions based on premises

Inductive

Generalizing from previous observations

Abductive

Providing explanations for a given observation

But what about other reasoning types?

Abductive

Providing explanations for a given observation
by reasoning over many possible explanations

Often neglected in
QA, but important!

What's a memorable way I can
remember the capital of France?

A fun and memorable way to remember that **Paris** is the capital
of France is:

◆ "A Pair (Paris) of Fancy Pants (France)"

Memorably explaining facts

What's the capital of France? I'm
thinking of traveling there

The capital of France is **Paris!** It's a beautiful city known for
its iconic landmarks like the Eiffel Tower, the Louvre
Museum, and Notre-Dame Cathedral. If you're planning to
visit, let me know if you need recommendations on places
to see, things to do, or where to eat! 😊

Personalizing to subjective queries

How can we test abduction in question answering?

Question Answering

Task: Answer the question “What’s the nationality of Don Quixote’s author?”

Answer: Spanish

How can we test abduction in *reverse* question answering?

Question Answering

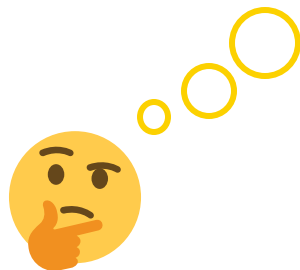
Task: Answer the question “What’s the nationality of Don Quixote’s author?”
Answer: Spanish

For a question,
deduce the
correct answer

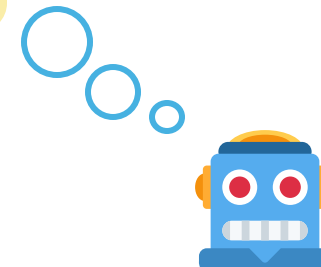
Reverse Question Answering

Task: Give me a question with the answer “Spanish”
Question: What is the official language of Spain?

For an answer,
adduce any
valid question



Our Goal:
Compare LLM abilities on QA versus RQA



Dataset Construction

Numerical Entities

Number
Question: What is 26 times 4? Answer: 104

Number + Text
Question: When did Pope Hormisdas die? Answer: 523 AD


Textual Entities

Easy Fact
Question: Who painted Stary Night? Answer: Vincent Van Gogh

Hard Fact
Question: What is Paola Uccello's last painting? Answer: The Hunt in the Forest

Question Answering

Question: What is 26 times 4
Answer: 103  104 <i>Gold answer</i>

 Accuracy metric

Reverse Question Answering

Answer: 104
Question: What is 100 + 4?

GPT
4o

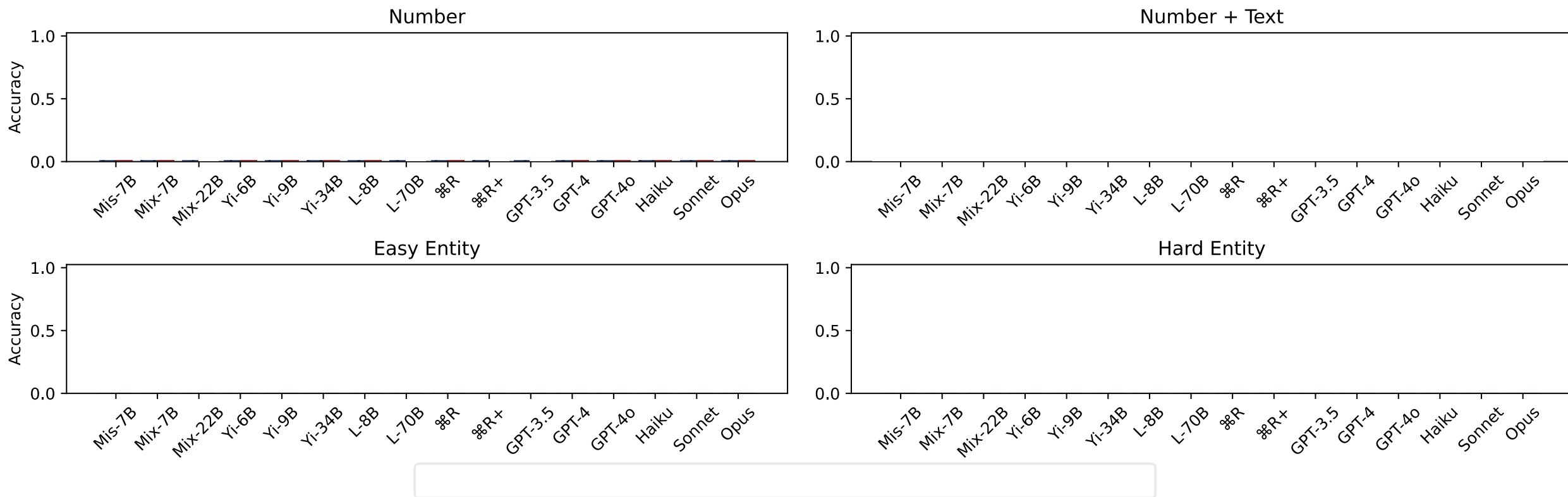


Does 104 answer "What is 100+4?"

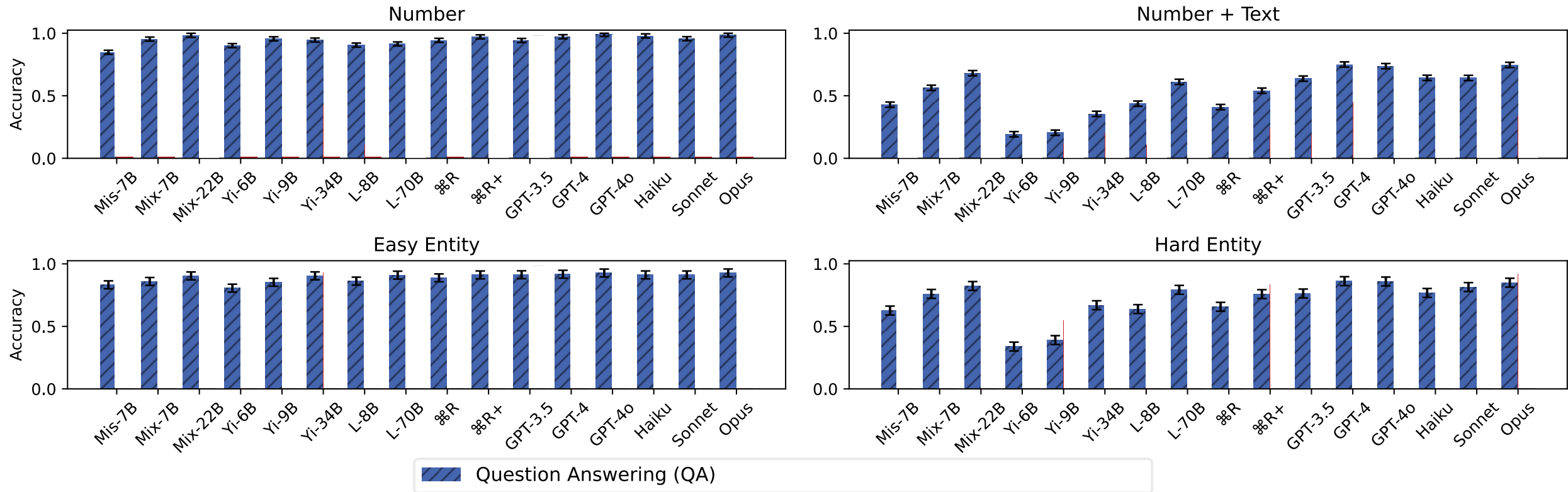


Accuracy metric (90% human agree.)

Reverse QA is largely weaker than QA on Numbers

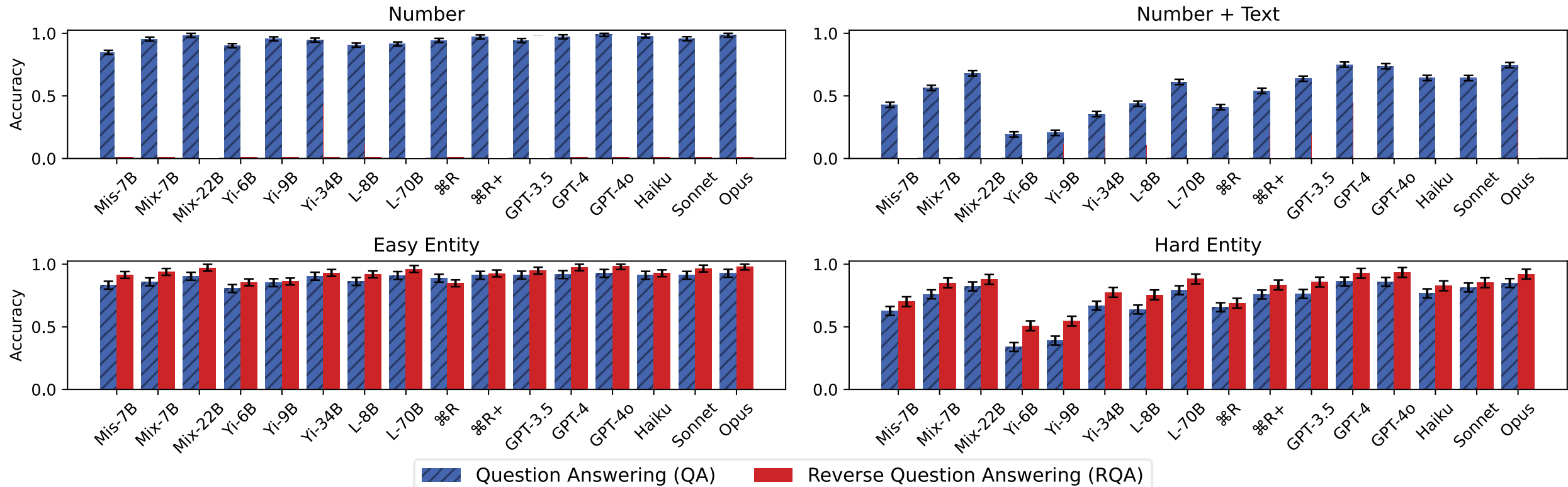


Reverse QA is largely weaker than QA on Numbers



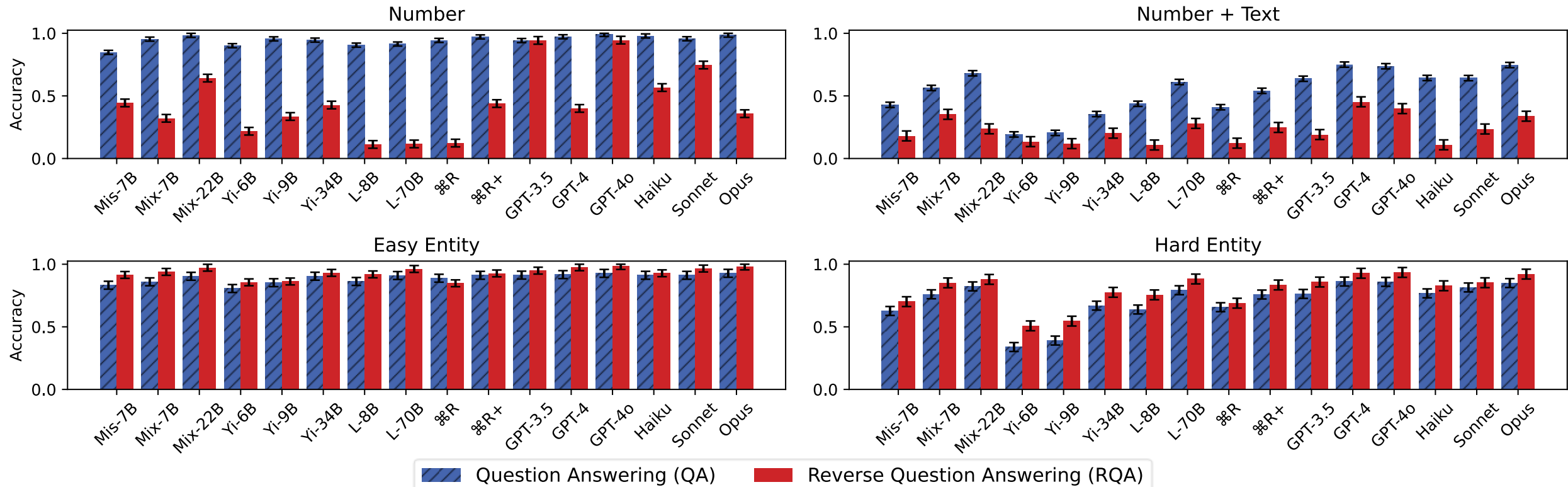
➤ LLMs appear fairly accurate in QA/deduction

Reverse QA is largely weaker than QA on Numbers



➤ LLMs appear fairly accurate in QA/deduction and textual RQA/abduction

Reverse QA is largely weaker than QA on Numbers



- LLMs appear fairly accurate in QA/deduction and textual RQA/abduction
- But are significantly weaker at numerical RQA/abduction!

Can LLMs answer their own invalid questions?

Reverse Question Answering

Give me a question with the answer “488”

Question:

Question Answering

Question:

Answer:

Can LLMs answer their own invalid questions?

Reverse Question Answering

Give me a question with the answer “488”

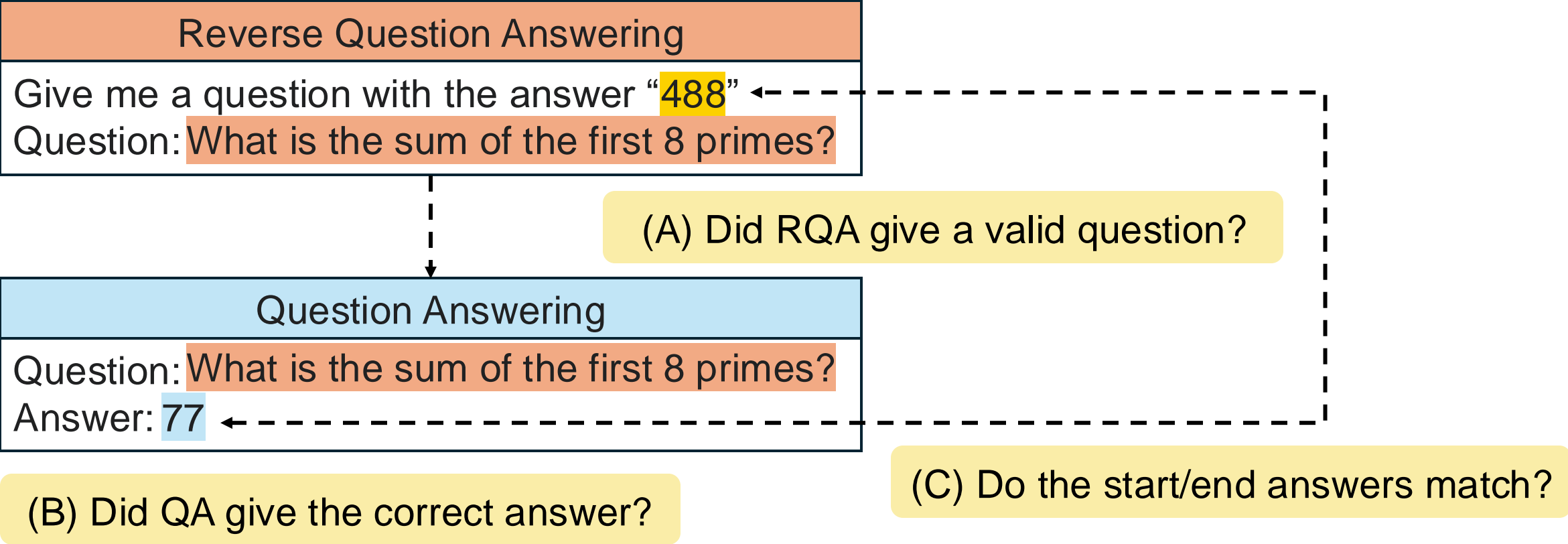
Question: What is the sum of the first 8 primes?

Question Answering

Question:

Answer:

Can LLMs answer their own invalid questions?



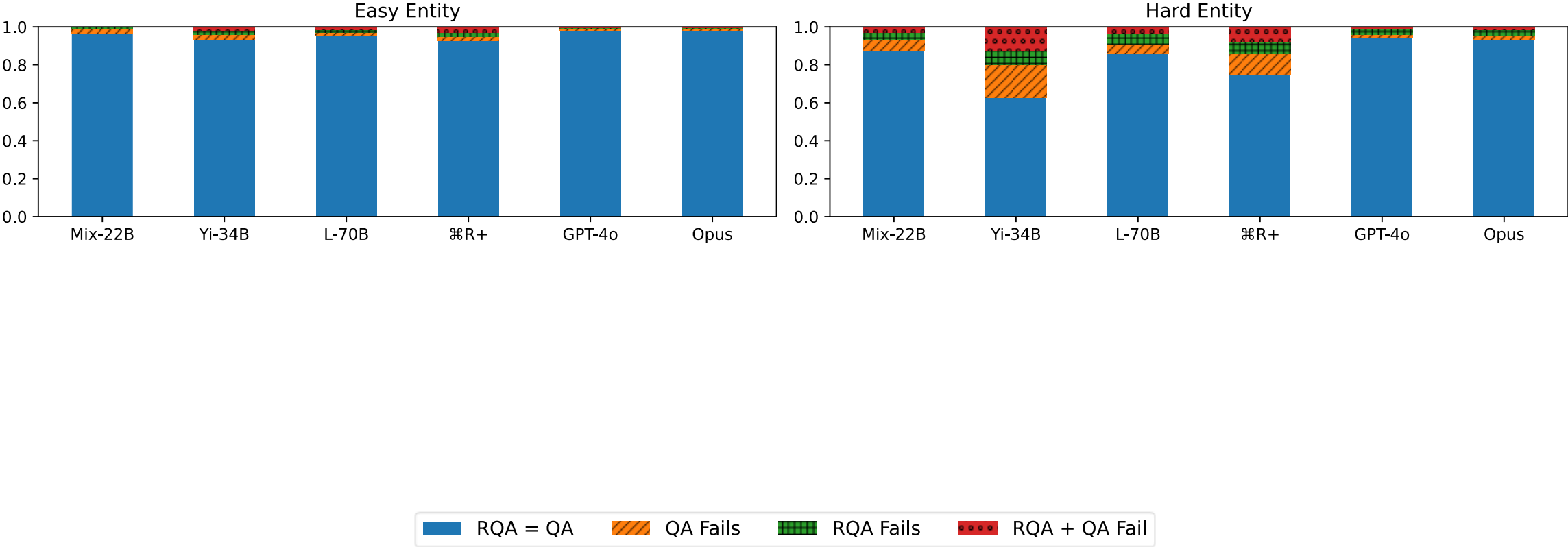
Can LLMs answer their own invalid questions?

The questions form another logical consistency check!

	Consistent	QA Fails	RQA Fails	Both Fail
(A) Did RQA give a valid question?	Yes	Yes	No	No
(B) Did QA give the correct answer?	Yes	No	Yes	No
(C) Do the start/end answers match?	Yes	No	No	No

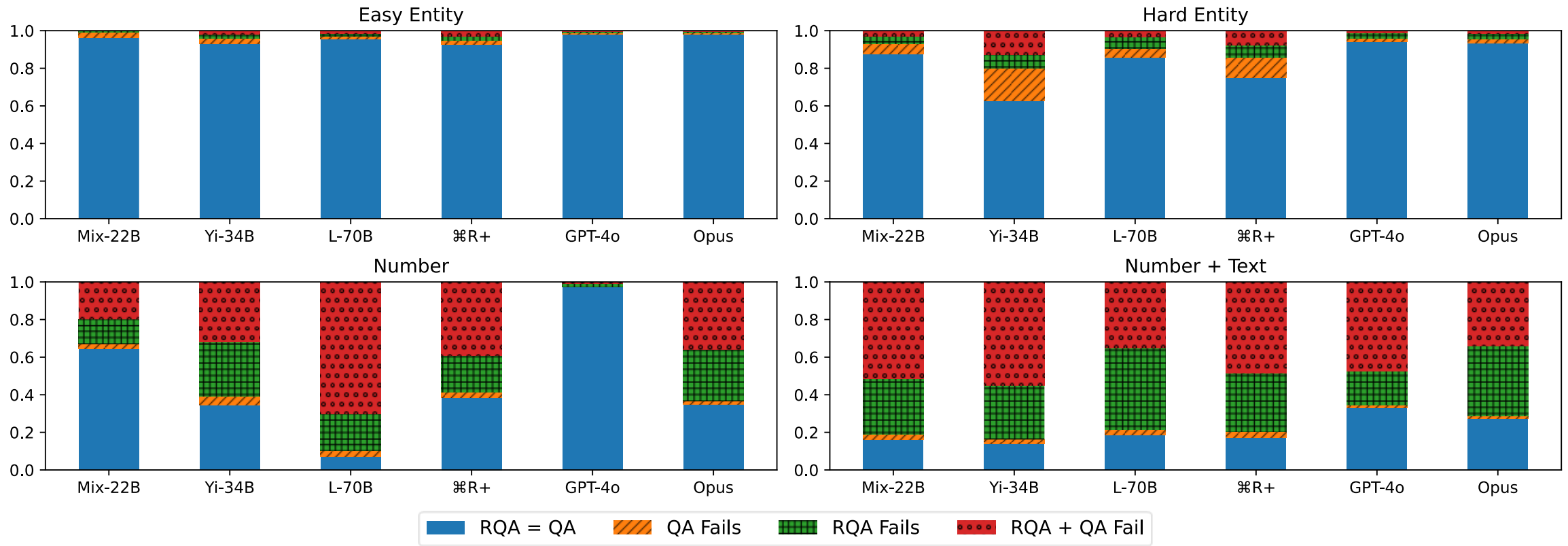
Can LLMs answer their own invalid questions? Sometimes!

Can LLMs answer their own invalid questions? Sometimes!



➤ Highly consistent on textual answers

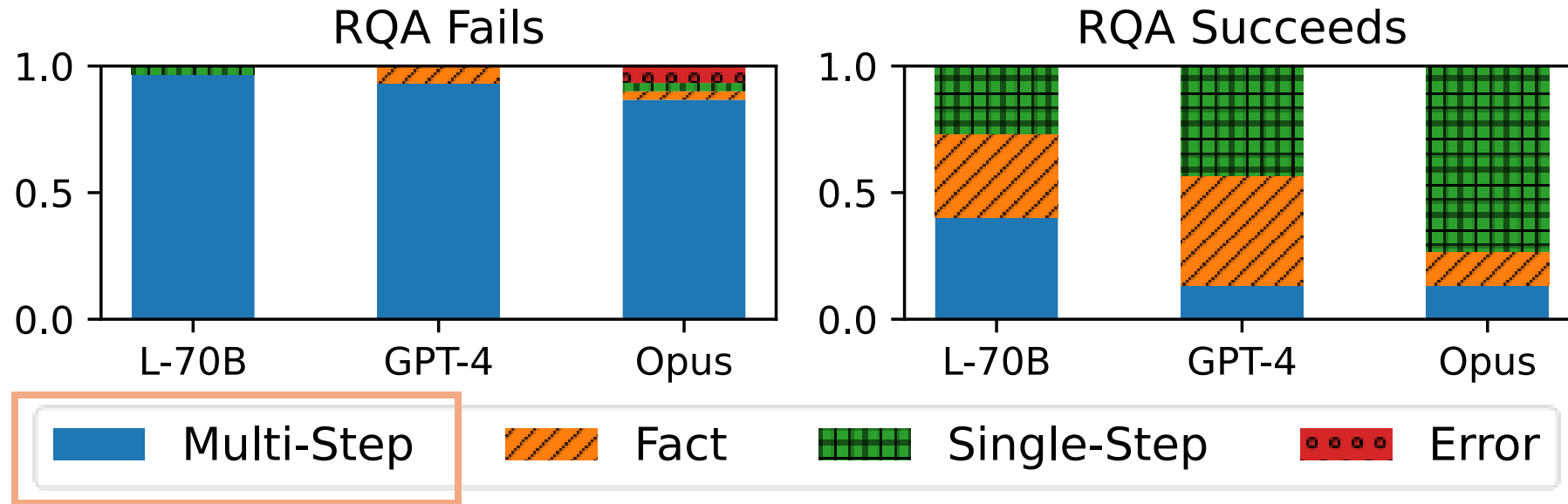
Can LLMs answer their own invalid questions? Sometimes!



- Highly consistent on textual answers
- But on numerical answers, QA rarely fails alone => LLMs detect their own question errors!
- Not just due to knowledge gaps (*How Language Model Hallucinations Can Snowball*)

When might RQA *specifically* fail?

We analyze questions for numbers when RQA fails and categorize them:



Generate a question for “437”
Question: What is the sum of the numbers of legs of a group of 23 cats, 12 humans, and 1 spider?

Generate a question for “756”
Question: What is the sum of the numbers from 1 to 27, inclusive?

Look helpful, but are in fact bogus

When might RQA *specifically* fail?

We speculate: could this be due to preference training?

Looks helpful, but isn't (complexity bias)^[1]

Prompt

Generate a question for "756"

Response 1

Question: What is the sum of the numbers from 1 to 27, inclusive?



Response 2

Question: What is $755 + 1$?



[1] Language Models Learn to Mislead Humans via RLHF

Conclusion: Again, how will this help us build helpful QA systems?

Overall:

- LLMs struggle to generate accurate questions for numerical entities
- Not just due to knowledge gaps, as models can solve their own questions
- Models can give questions that *appear* helpful, but are actually faulty



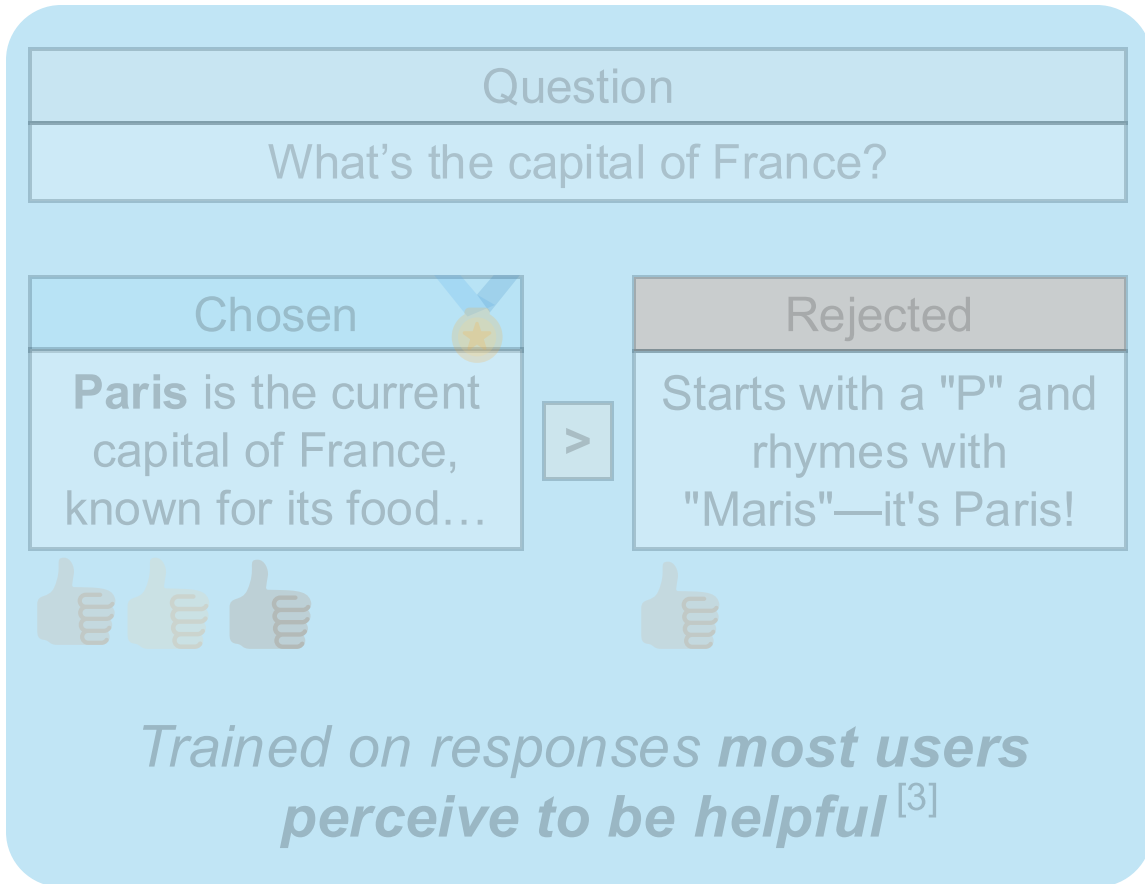
QA correctness **again** cannot evaluate reasoning chains



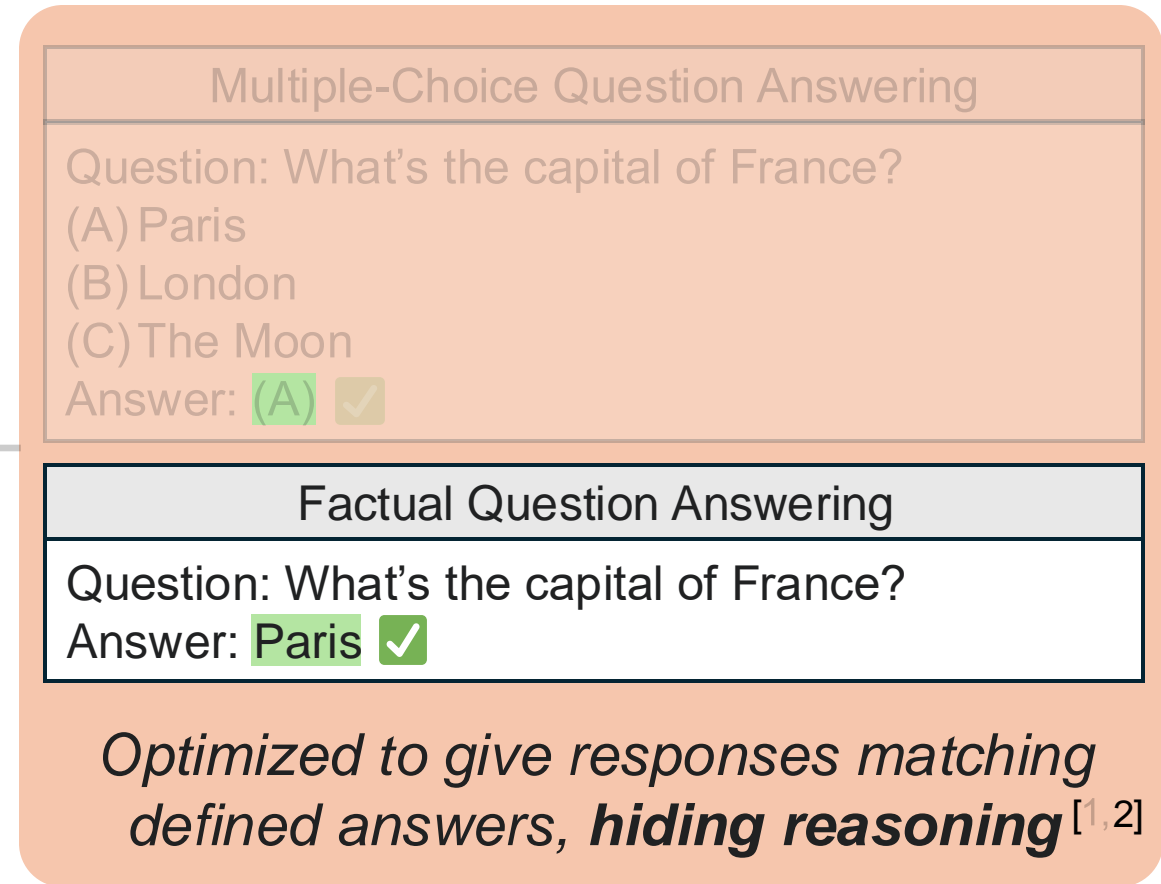
LLMs can generate responses that solely **appear** helpful

Is LLM development aligned with helping users? **No**

Preference Training



Correctness Evaluation



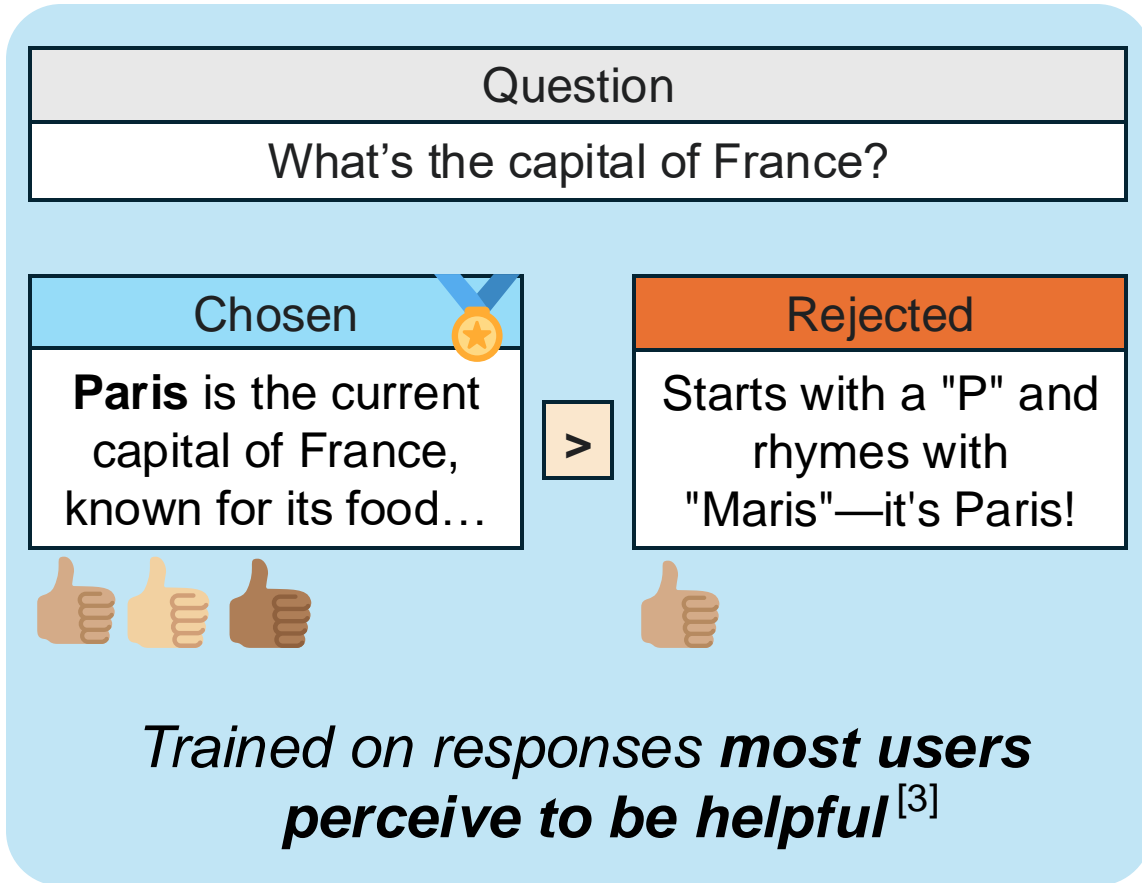
[3] A SMART Mnemonic Sounds like "Glue Tonic" (EMNLP 2024)

[1] It's Not Easy Being Wrong (ACL 2024, Findings)

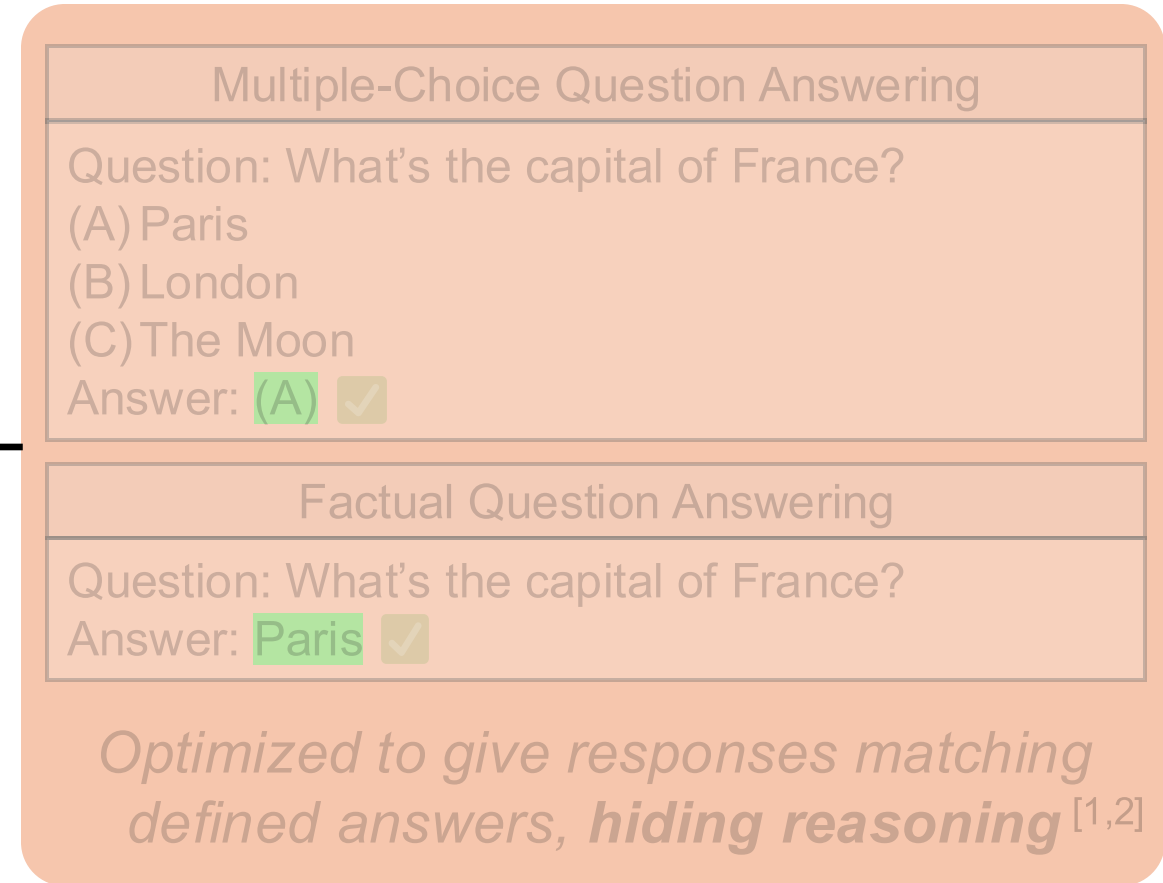
[2] Reverse Question Answering (NAACL 2025)

Is LLM development aligned with helping users? **No**

Preference Training



Correctness Evaluation



[3] A SMART Mnemonic Sounds like "Glue Tonic" (EMNLP 2024)

[1] It's Not Easy Being Wrong (ACL 2024, Findings)

[2] Reverse Question Answering (NAACL 2025)

Recall: How can QA systems help users achieve their goals?

Goal: Learn Something New



What does "LLM" mean?



Goal: Solve a Problem



How can I get my refund?



Goal: Receive Tailored Advice



How do I hide the fact that I'm a tourist in London?



Goal: Recall Forgotten Information





Who gave that talk with too many emojis at Imperial?



Recall: How can QA systems help users achieve their goals?



Students are expected to learn 1000+ vocab terms for the GRE 


 johnnyboyblablublublu · 6y ago ·

I would say electronic apps suffice for decent GRE score. I used three apps- magoosh flash cards, magoosh vocabulary, and galvanize (I would highly recommend this, it will also help understand the meaning of words).


Generally, aim for around new 1000-1200 words in your vocabulary. More than that just felt like dumping a lot of words in a lot less time.

↑ 5 ↓ Reply ...

You may need to know 1,000+ GRE vocab words to be ready for whatever could come your way on test day. Even if you already have a broad vocabulary, you probably will still need to learn at least a few hundred vocab words over the course of your GRE test prep. Dec 29, 2023

 TTP GRE Blog
<https://gre.blog.targettestprep.com/how-to-learn-voca...>

[How to Learn Vocabulary for GRE Verbal - TTP GRE Blog](#)

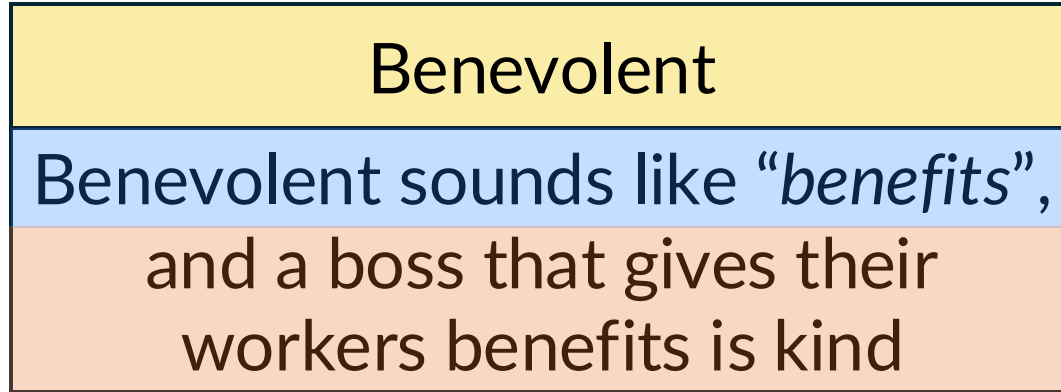
 **ToomuchLes** ...
Posted August 29, 2013

Hey all!

So.. I've been studying GRE words for nearly 2 months now, and as of right now, I've memorized Kaplan GRE Vocab Flashcards (500), Manhattan Prep GRE 500 Essential Words, and Barrows 2nd Edition GRE Words (500). In total, with synonyms and additional words on the comment section, I probably memorized more than 2,500 GRE Words (minimum), and when I mean 'memorized' I mean I can recite definition, synonyms and everything without a pause. As you can see, Im very proud of such an achievement lol.

How can systems make studying vocab more effective/engaging?

*Mnemonics must be correct **and** memorable*



Let's use **mnemonic devices!**

1) Link term to a **simpler** and **similarly sounding** keyword

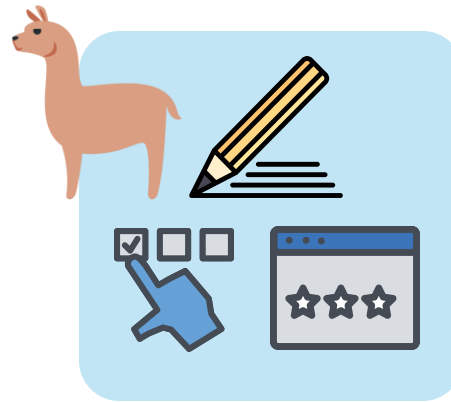
2) **Explain the link** between the keyword + term

Our Goal: Build a mnemonic generator that helps students learn vocabulary

Designing **SMART**: A Keyword Mnemonic Generator

Student **M**nemonic **A**lignment to aid the **R**ecall of **T**erms

Post-Training
with LLaMA-2 70B



Supervised
Fine-Tuning



Diverse Student
Preference Collection



Bayesian
Modeling



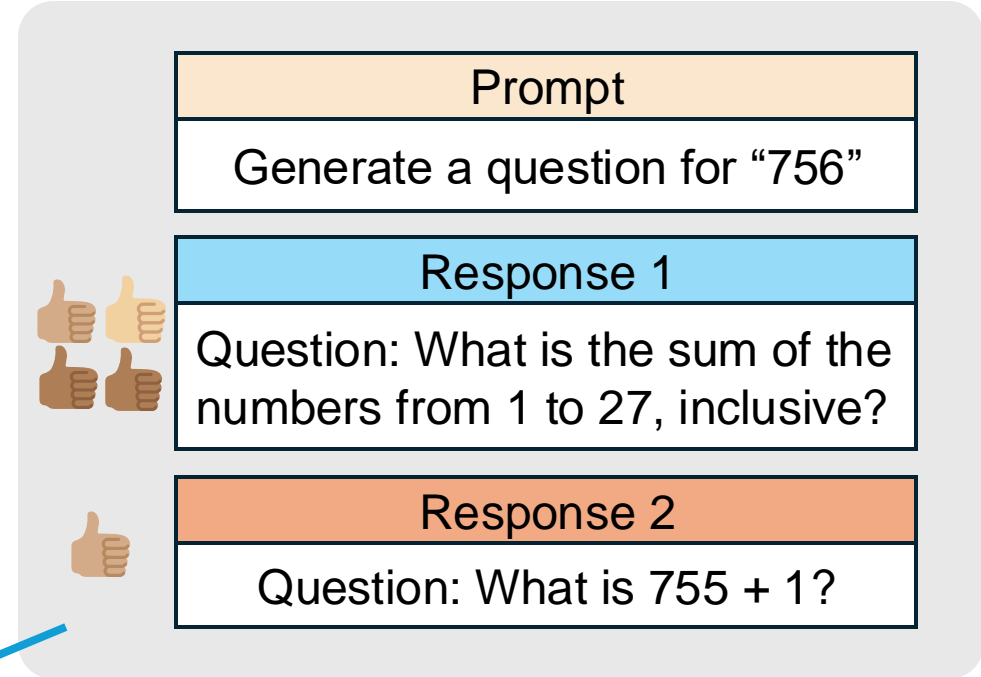
Direct Preference
Optimization

Designing **SMART**: A Keyword Mnemonic Generator

Student **M**nemonic **A**lignment to aid the **R**ecall of **T**erms

Can we confirm this intuition from RQA?

Post-Training
with LLaMA-2 70B



Supervised
Fine-Tuning



Diverse Student
Preference Collection



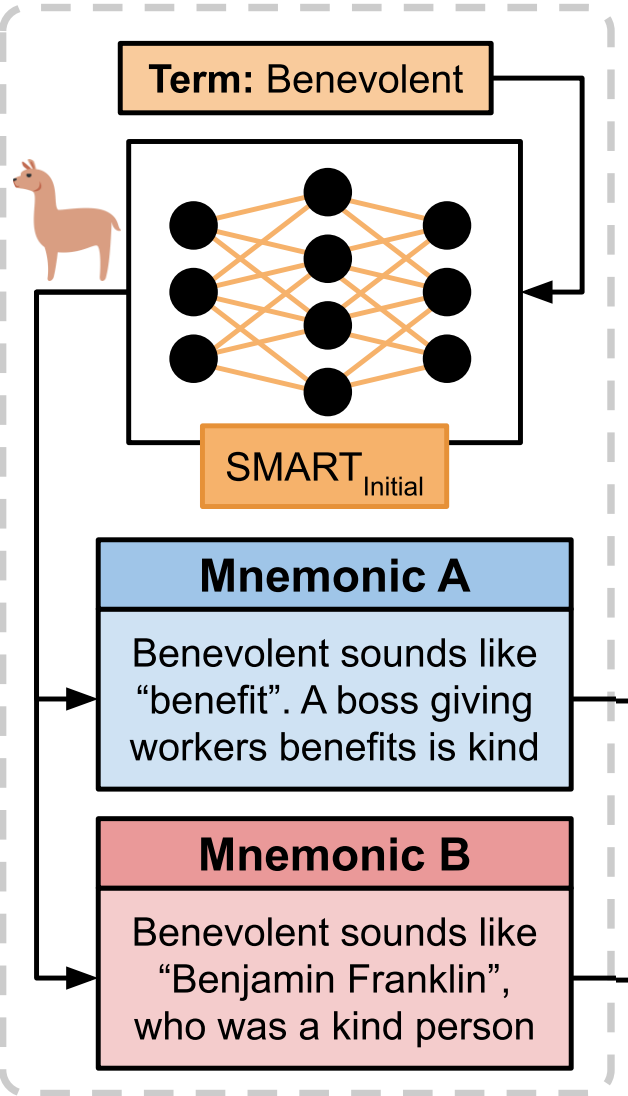
Bayesian
Modeling



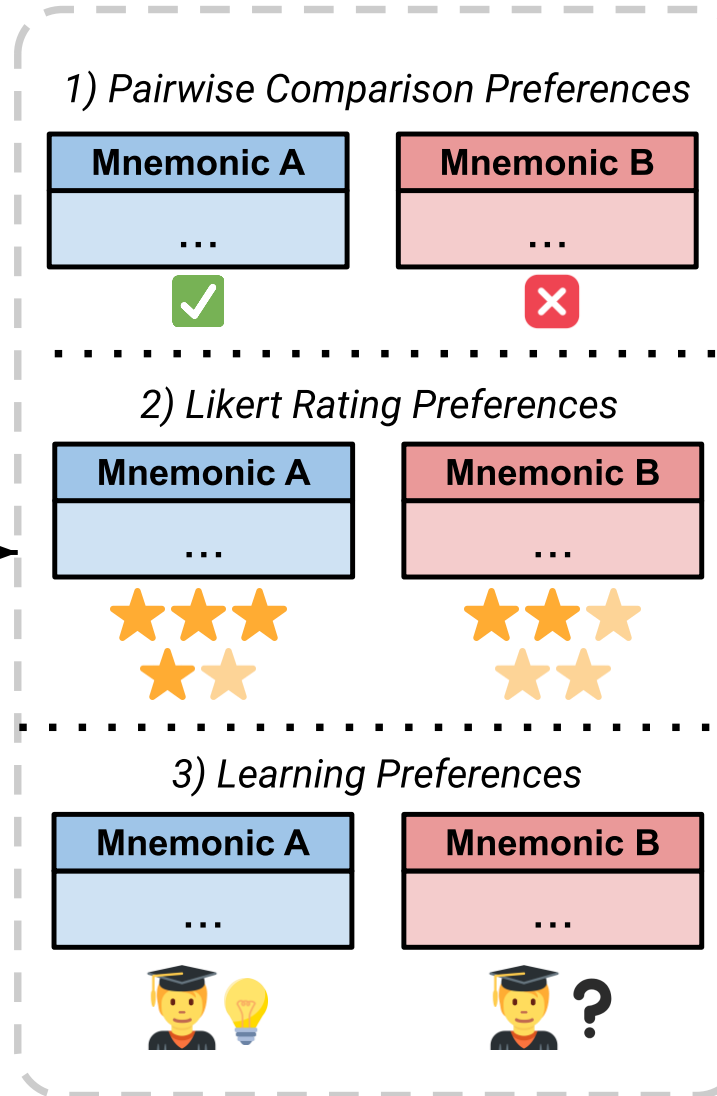
Direct Preference
Optimization

Designing **SMART**: A Keyword Mnemonic Generator

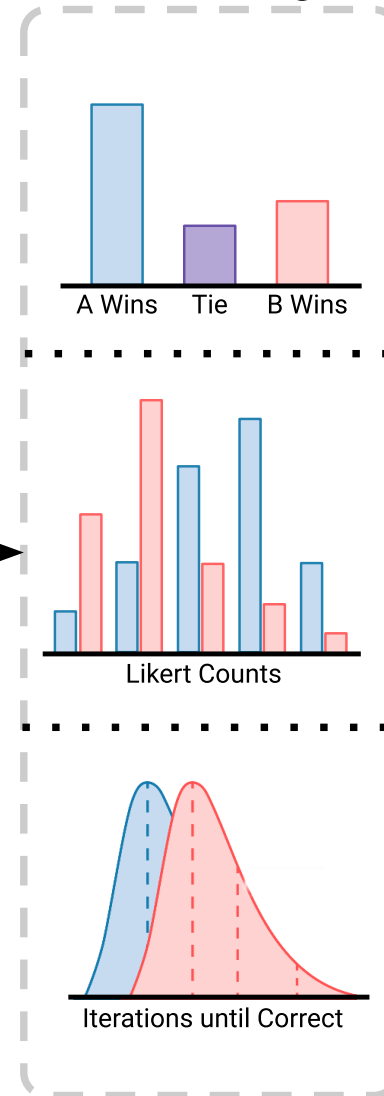
Supervised Fine-Tuning



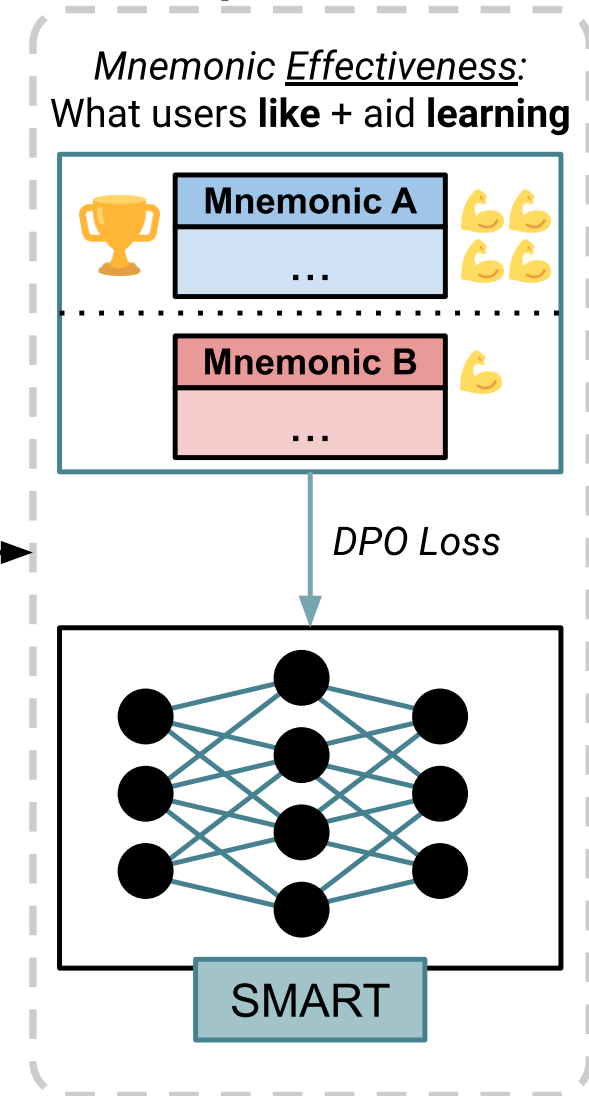
Diverse Student Preference Collection



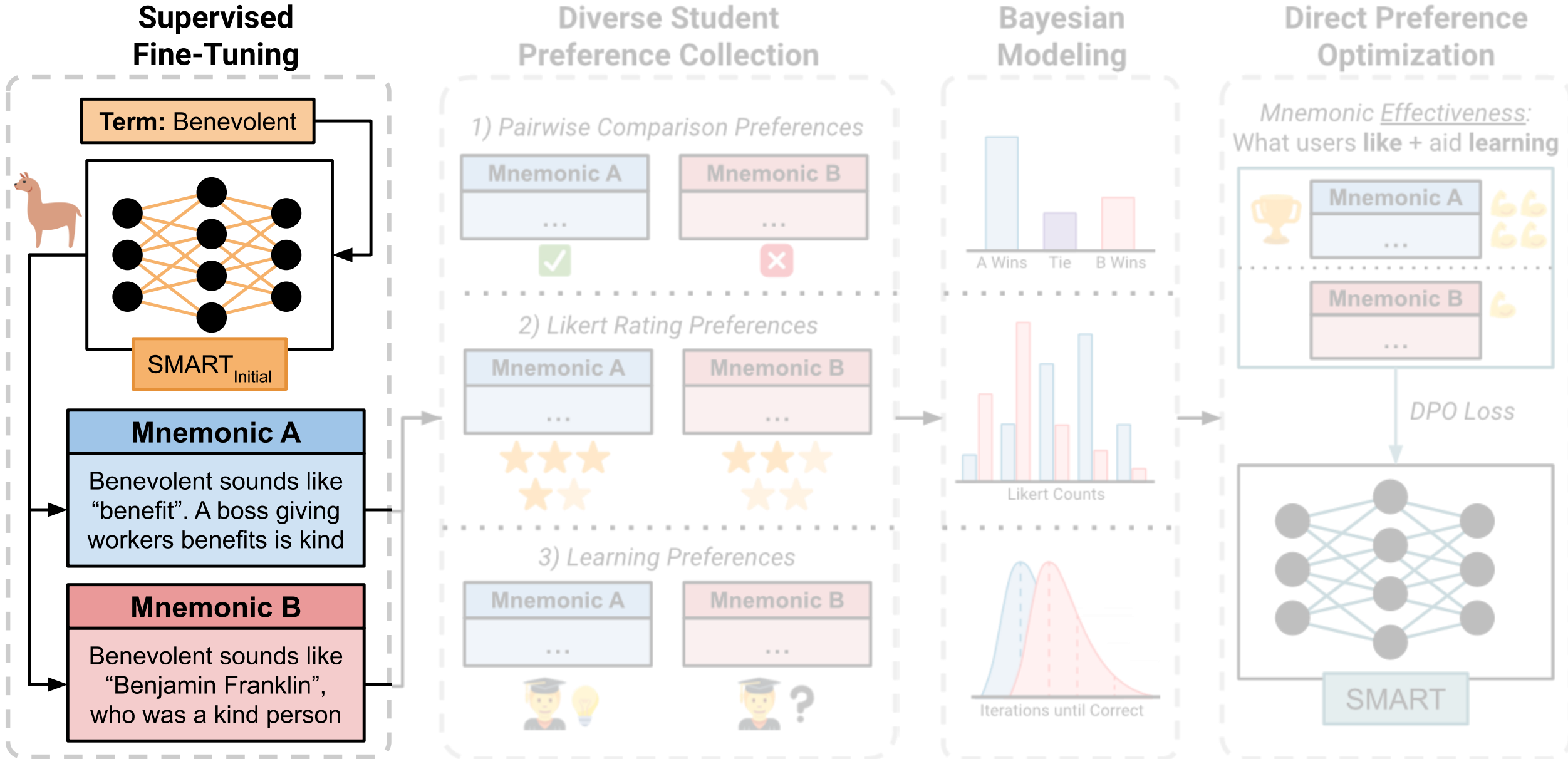
Bayesian Modeling



Direct Preference Optimization



Designing **SMART**: A Keyword Mnemonic Generator



Fine-tuning an initial mnemonic generator

Input GRE Term →

obdurate

word of the day

obdurate - Dictionary definition and meaning for word **obdurate**

Definition

(adj) stubbornly persistent in wrongdoing

💡 **Mnemonics (Memory Aids) for obdurate**

Output Mnemonics →

OB+DURA(RELATE IT TO WORD DURABILITY)..SO anything which has durability..has the resisting power, and something which can resist, is stubborn.

134 👍

9 👎



OB+DU+RATE relate it as AB(OB) RATE DoUble ho gaya,and the seller was very STUBBORN about that

16 👍

13 👎



Powered by [Mnemonic Dictionary](#)
durate -duration There will be a period(duration) in everyone's life when we will be very obdurate(i.e, stubborn)..

4 👍

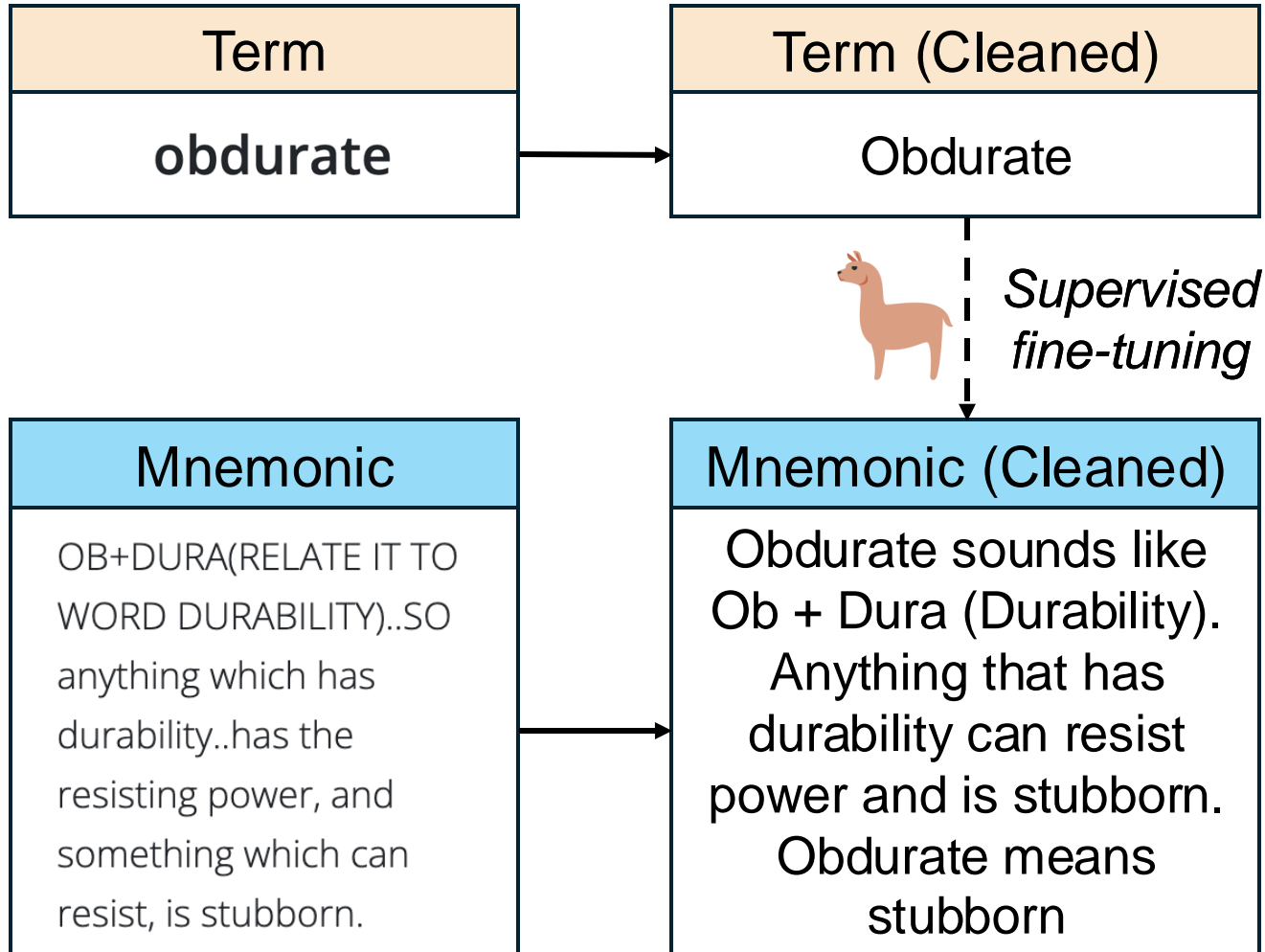
11 👎



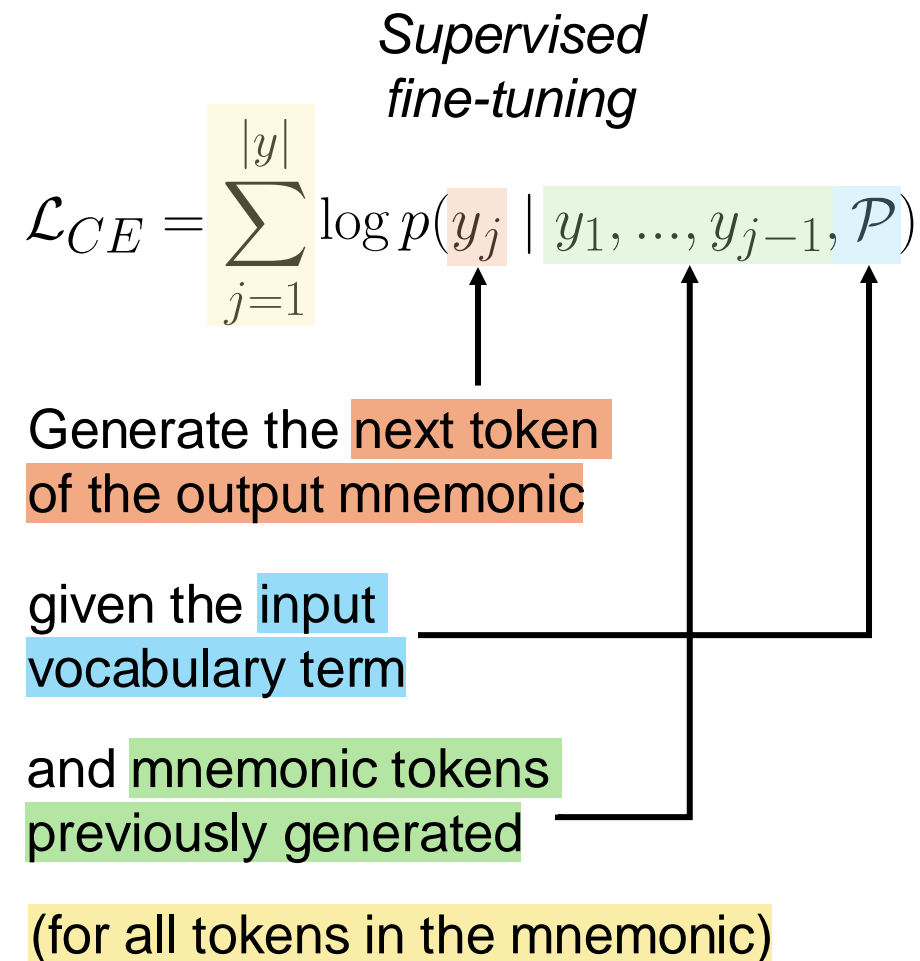
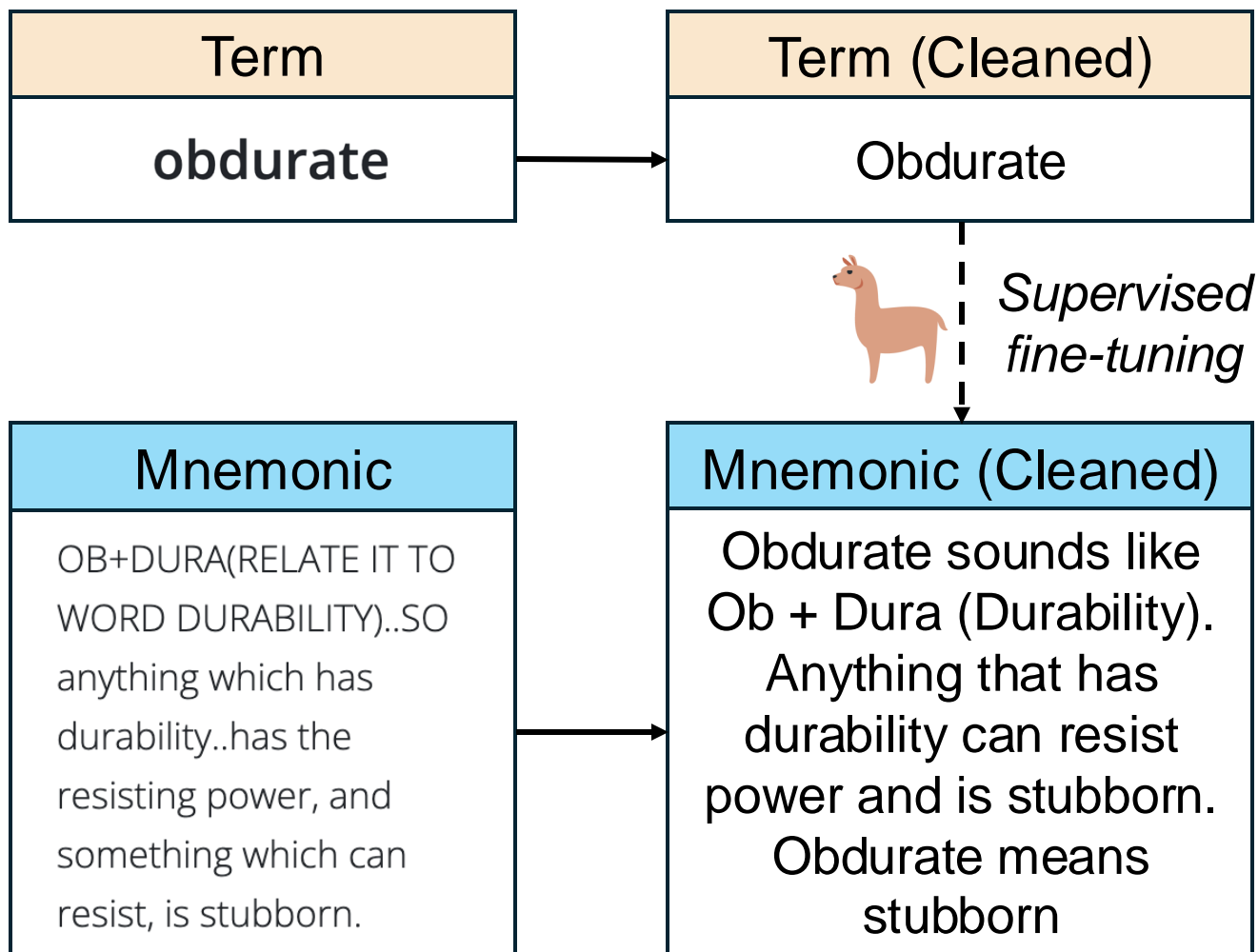
Upvotes finds **high-quality** mnemonics

<https://mnemonicdictionary.com/>

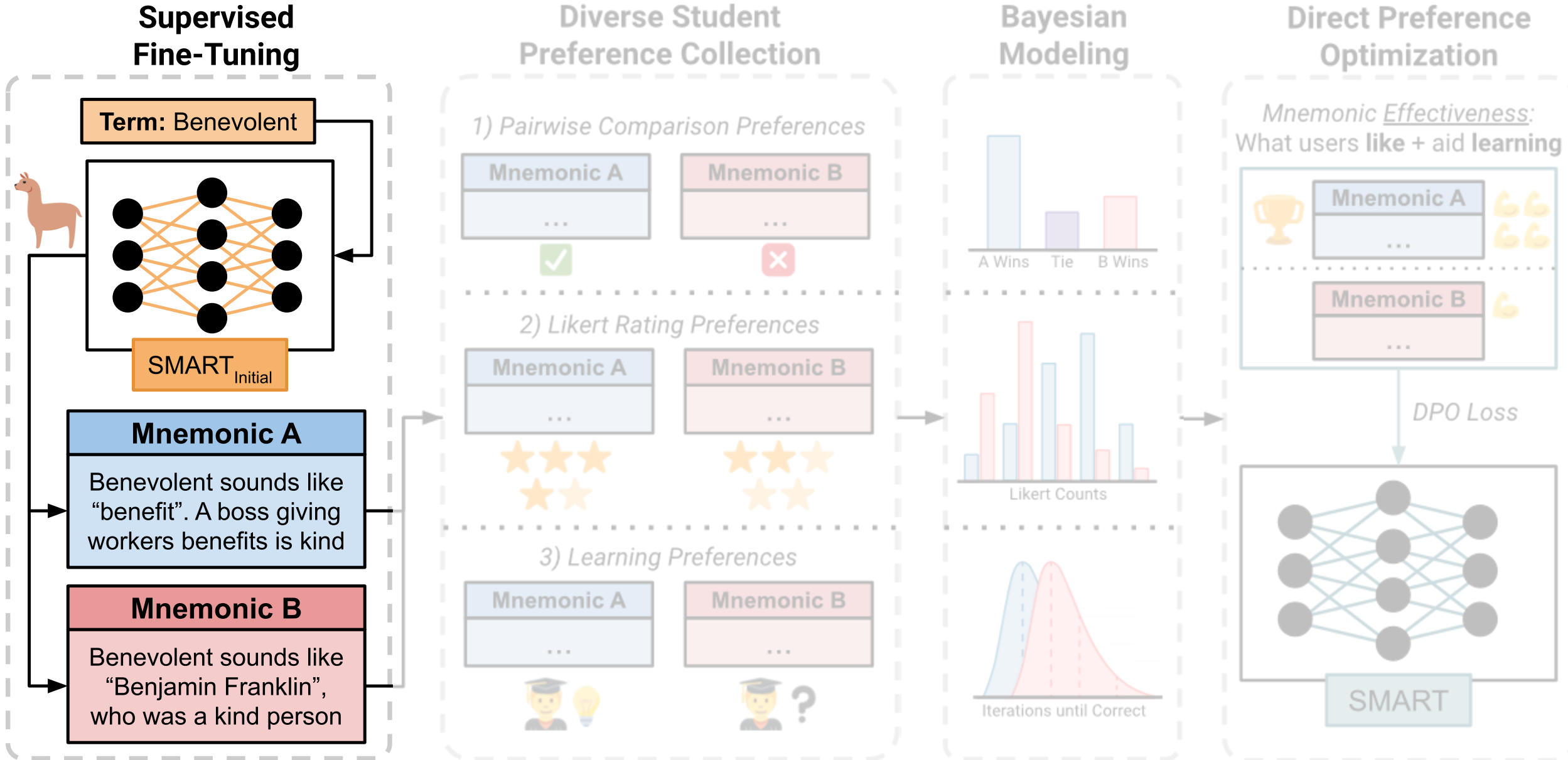
Fine-tuning an initial mnemonic generator



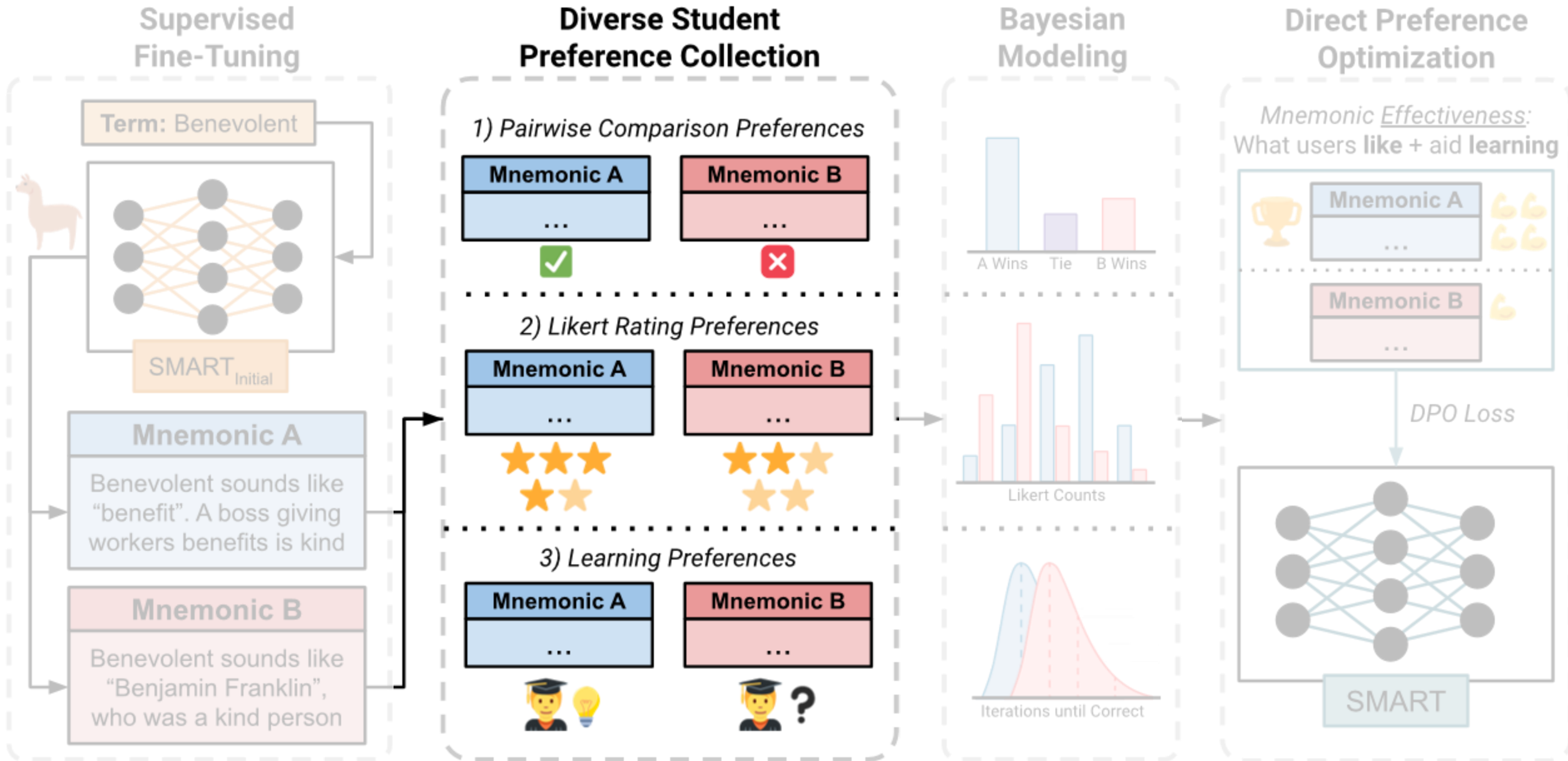
Fine-tuning an initial mnemonic generator



Designing **SMART**: A Keyword Mnemonic Generator



Designing **SMART**: A Keyword Mnemonic Generator



Collecting student feedback on mnemonics

47 learners study with mnemonics from the initial SMART model in a flashcard app:


Collecting student feedback on mnemonics

47 learners study with mnemonics from the initial SMART model in a flashcard app: [1]

Category: GRE—What's the definition?

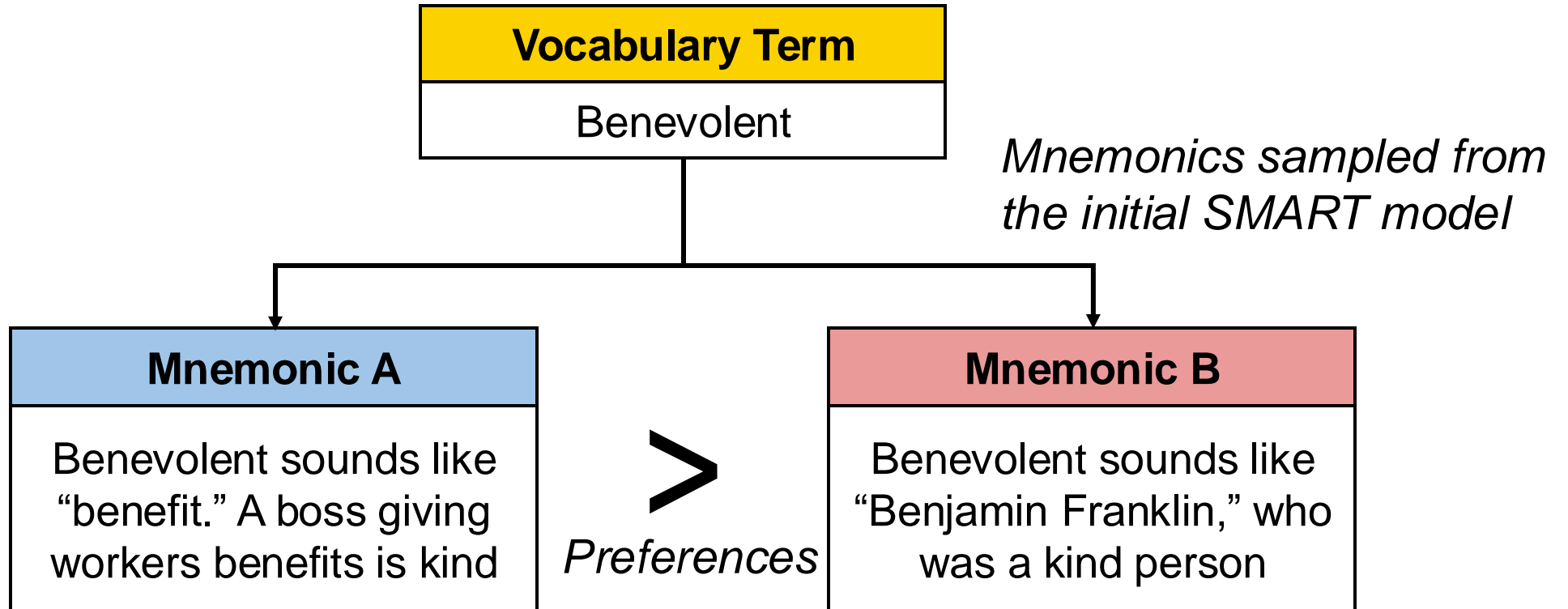
Compliant

SHOW ANSWER (ENTER) DON'T KNOW (SHIFT+ENTER)

 No, compliant means to agree with others and obey rules

[1] KARL: Knowledge-Aware Retrieval and Representations aid Retention and Learning in Students

Goal: Teach **SMART** which mnemonics are **helpful** via preferences



There are many ways to figure out which mnemonics are more/less helpful

1) Expressed Preferences: What users *think* is helpful

Standard way of collecting preference data

Pairwise
Comparisons

Which mnemonic do you think would help you learn better?

Mnemonic A ([)

Compliant sounds like “complain”. If you complain, you are likely to follow the rules. Hence, compliant means willing to follow rules or requests.

Mnemonic B (])

Compliant sounds like “compliment”. When someone compliments you, they are agreeing with you, which is similar to being compliant.

SKIP (ENTER) EQUAL (SHIFTEENTER)

Likert
Ratings

Compliant sounds like “compliment”. When someone compliments you, they are agreeing with you, which is similar to being compliant.

Give Feedback (Optional) ⓘ

★ ★ ★ ★ ★

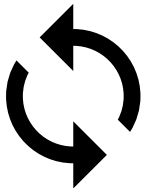
2) Observed Preferences: What *truly* helps user goals (learning)

Benevolent



Mnemonic
Benevolent sounds like “benefit.” A boss giving workers benefits is kind

Idea: **More helpful** mnemonics need **less iterations** to study



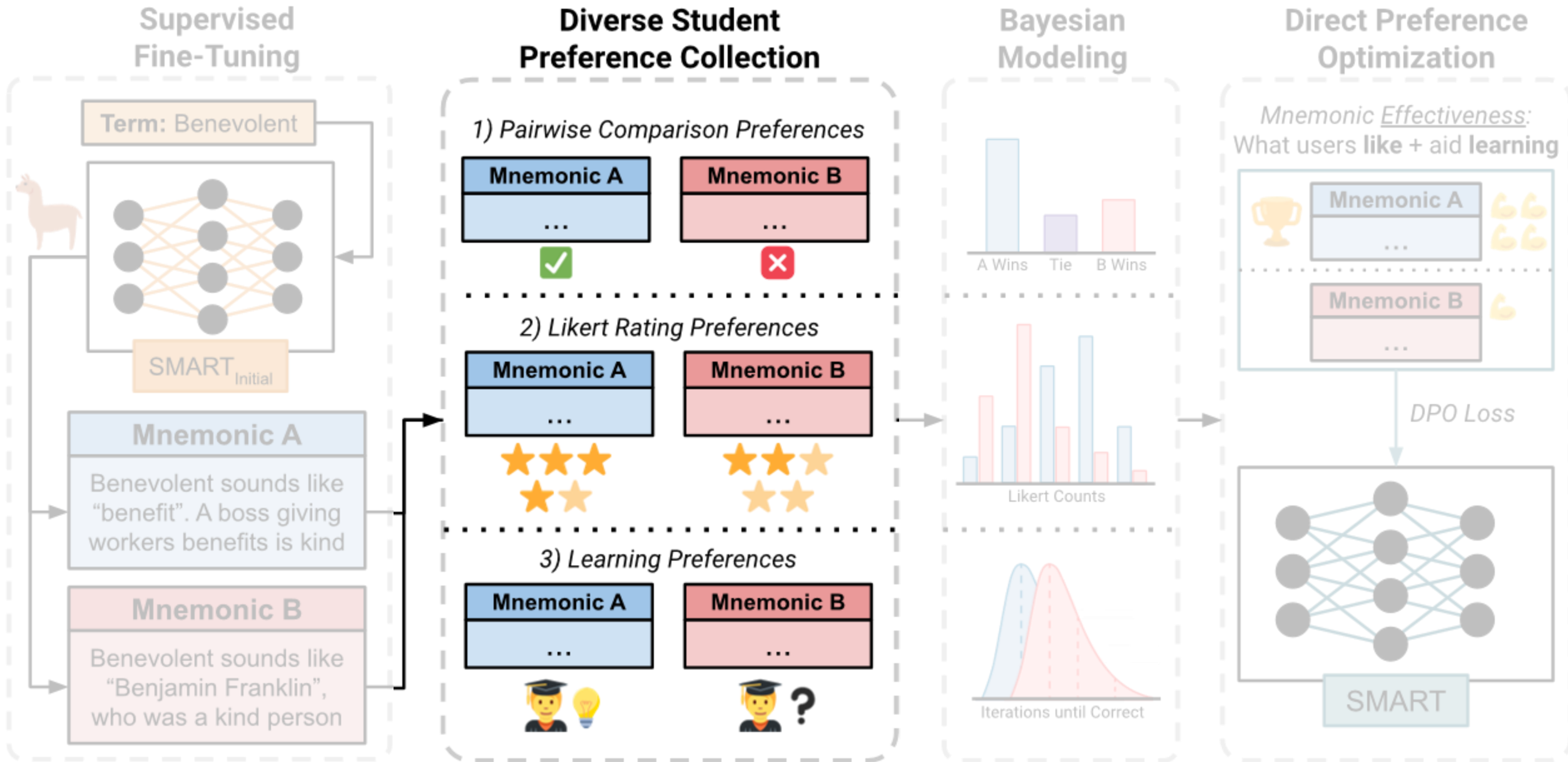
*Keep Studying
N times*

Short-Term Learning
(downstream goal)

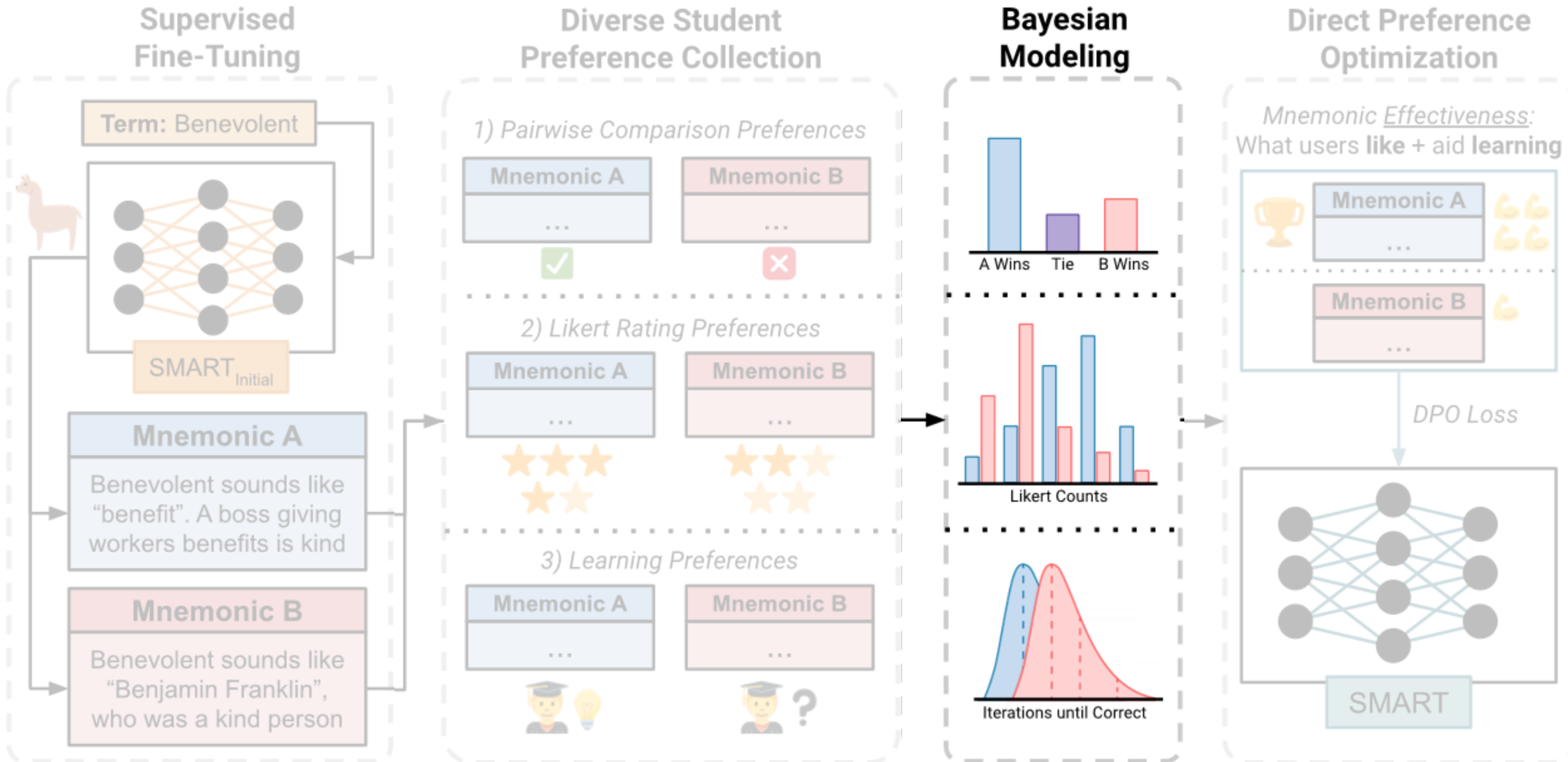


Benevolent means “well meaning and kind”

Designing **SMART**: A Keyword Mnemonic Generator



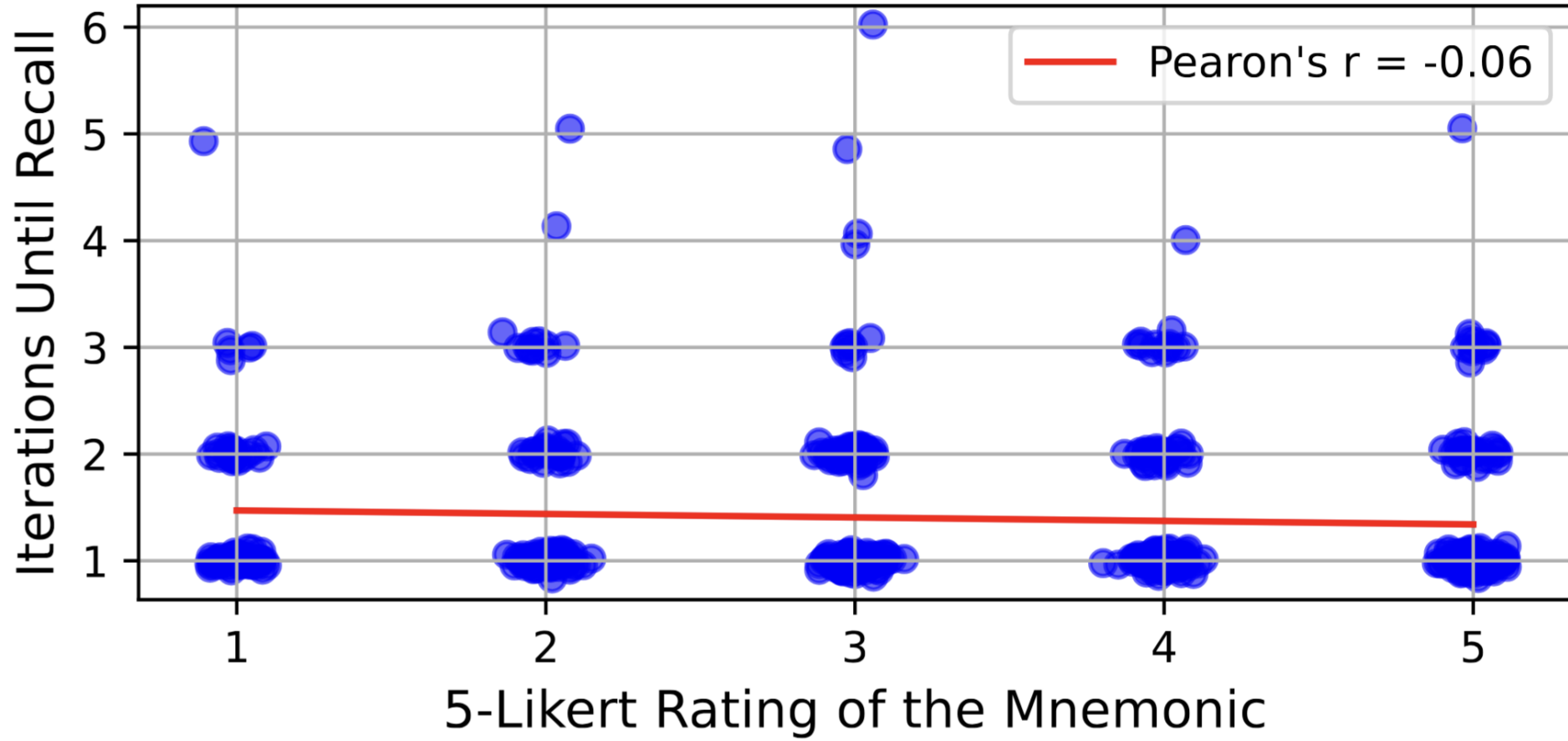
Designing **SMART**: A Keyword Mnemonic Generator



Students cannot predict what best helps them learn!



Students cannot predict what best helps them learn!



So which preferences should we use?



Expressed Preferences

Won't give us mnemonics that will help students learn



Observed Preferences

Won't give us mnemonics that students will want to see

Harmful Mnemonic for Obtuse

Obtuse sounds like "abuse". If you abuse someone, they may not understand the situation, just like an obtuse person who is slow to understand something

*But **never picked** in pairwise comparisons*




*Students **learned** in 1 iteration*

Both of these factors are important to generate **helpful** mnemonics

So we'll train SMART on both of these preferences!

Learning mnemonic **helpfulness**

A helpful mnemonic will:

-  Be chosen more in pairwise comparisons
-  Receive higher Likert ratings
-  Help students learn better

We combine these signals with
Hierarchical Bayesian Modeling

Learning mnemonic **helpfulness**: Hierarchical Bayesian Modeling

A **helpful** mnemonic will:

$$\theta_{A,i}, \theta_{B,i} \sim \text{Beta}(1, 1) \leftarrow \text{Latent "helpfulness" parameter}$$

 Be chosen more in **pairwise comparisons**

 Receive higher **Likert ratings**

$$\tau \sim \text{Beta}(1, 1),$$
$$p_i^{pair} = \frac{[p_{A,i}^{pair}; p_{B,i}^{pair}; \tau]}{p_{A,i}^{pair} + p_{B,i}^{pair} + \tau}, \leftarrow \text{Bradley-Terry with Ties}$$
$$C_i \sim \text{Multinomial}(n, p_i^{pair})$$

Predict rating distribution

$$\alpha_{rate}, \beta_{rate} \sim \text{Normal}(0, 1)^5,$$
$$p_{A,i}^{rate} = \sigma(\alpha_{rate} \cdot \theta_{A,i} + \beta_{rate}),$$
$$\mathcal{R}_{A,i} \sim \text{Multinomial}(\Sigma \mathcal{R}_{A,i}, p_{A,i}^{rate}),$$

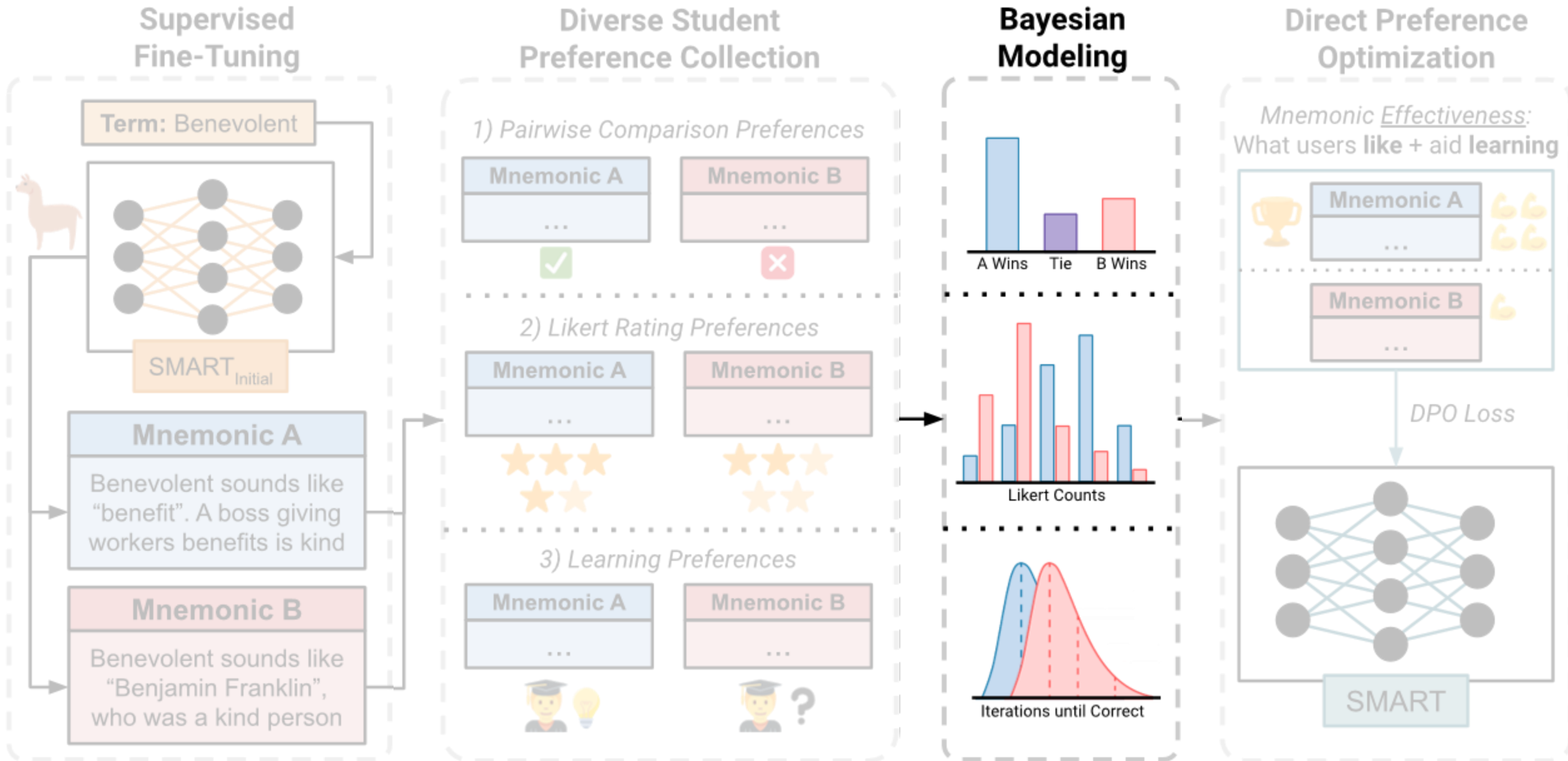
 Help students **learn better**

$$\alpha_{learn}, \beta_{learn} \sim \text{Normal}(0, 1),$$

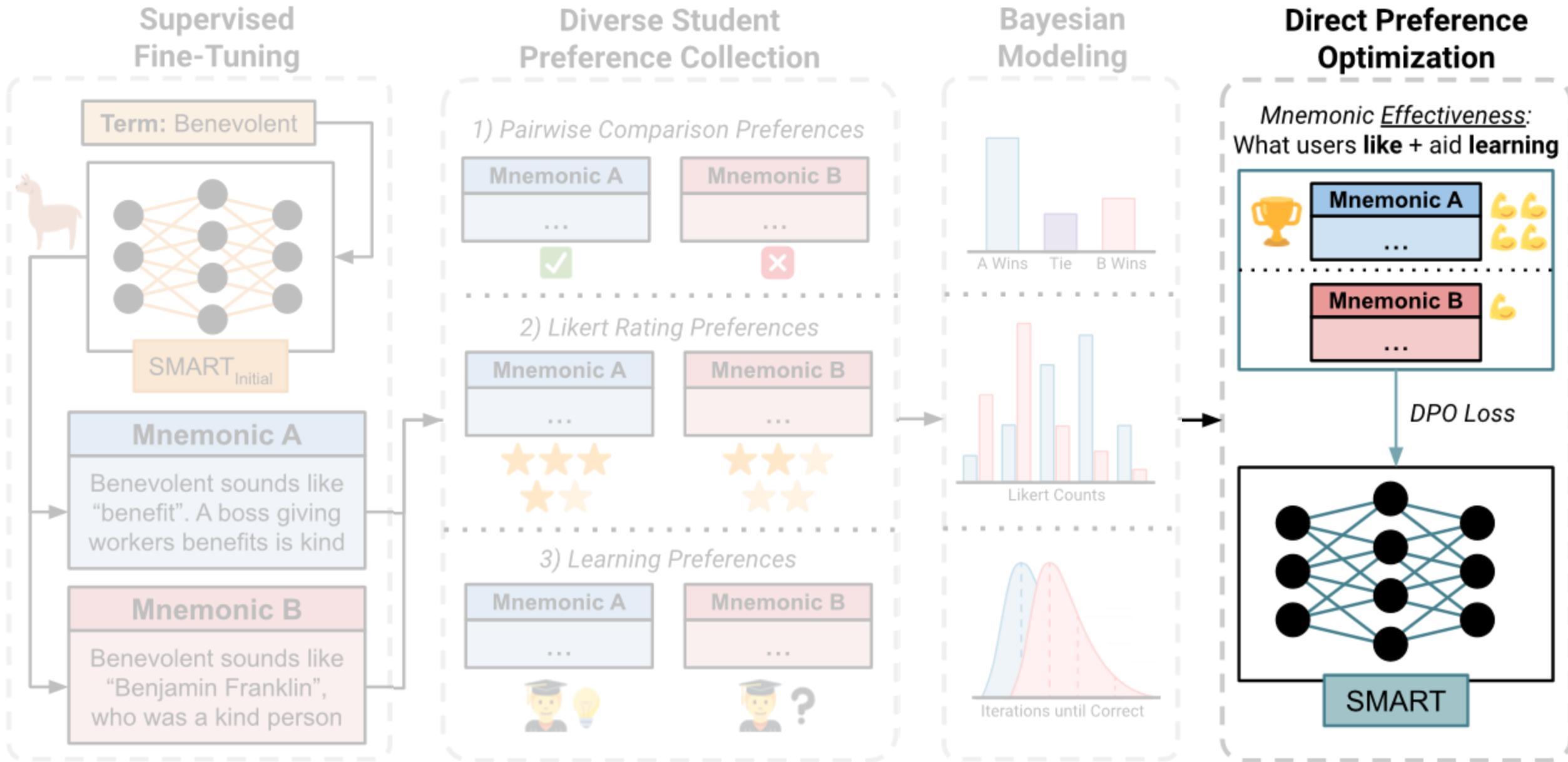
$$p_{A,i}^{learn} = \sigma(\alpha_{learn} \cdot \theta_{A,i} + \beta_{learn}), \leftarrow \text{Time taken until correct answer}$$

$$t_j \sim \text{Geometric}(p_{A,i}^{learn}),$$

Designing **SMART**: A Keyword Mnemonic Generator



Designing **SMART**: A Keyword Mnemonic Generator



Training SMART with Direct Preference Optimization (DPO)^[2]

Popular way of training on preference datasets

$$\mathcal{L}_{dpo} = -\mathbb{E}_{x, y_w, y_l \sim \mathcal{D}} \left[\ln \sigma \left(\beta \ln \frac{\pi(y_w|x)}{\pi_0(y_w|x)} - \beta \ln \frac{\pi(y_l|x)}{\pi_0(y_l|x)} \right) \right]$$

Looks **like** the chosen response

Looks **unlike** the rejected response

Rejected Response: **Less Helpful** Mnemonic

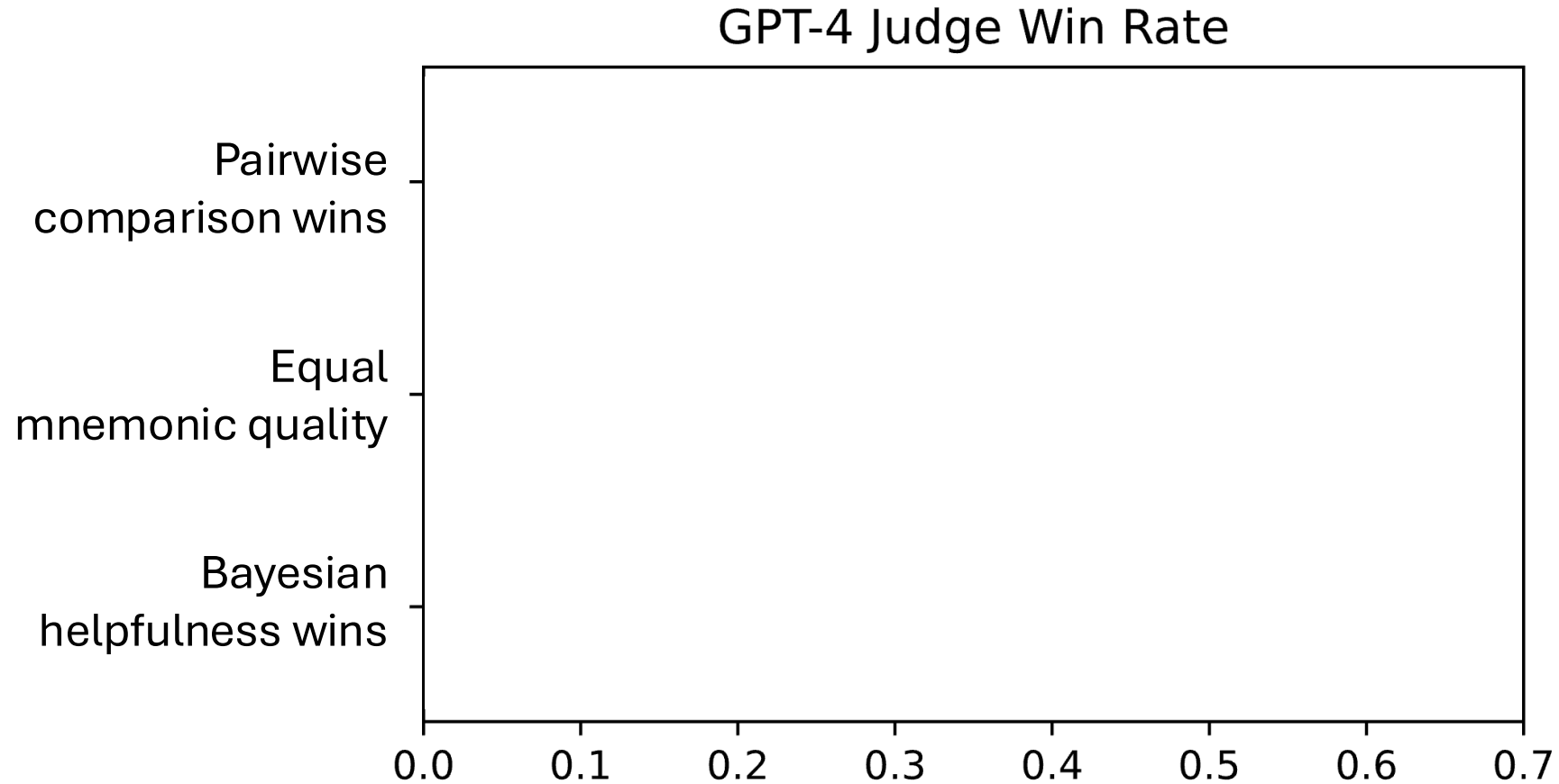
Chosen Response: **More Helpful** Mnemonic

Input Prompt: **Vocab Term**

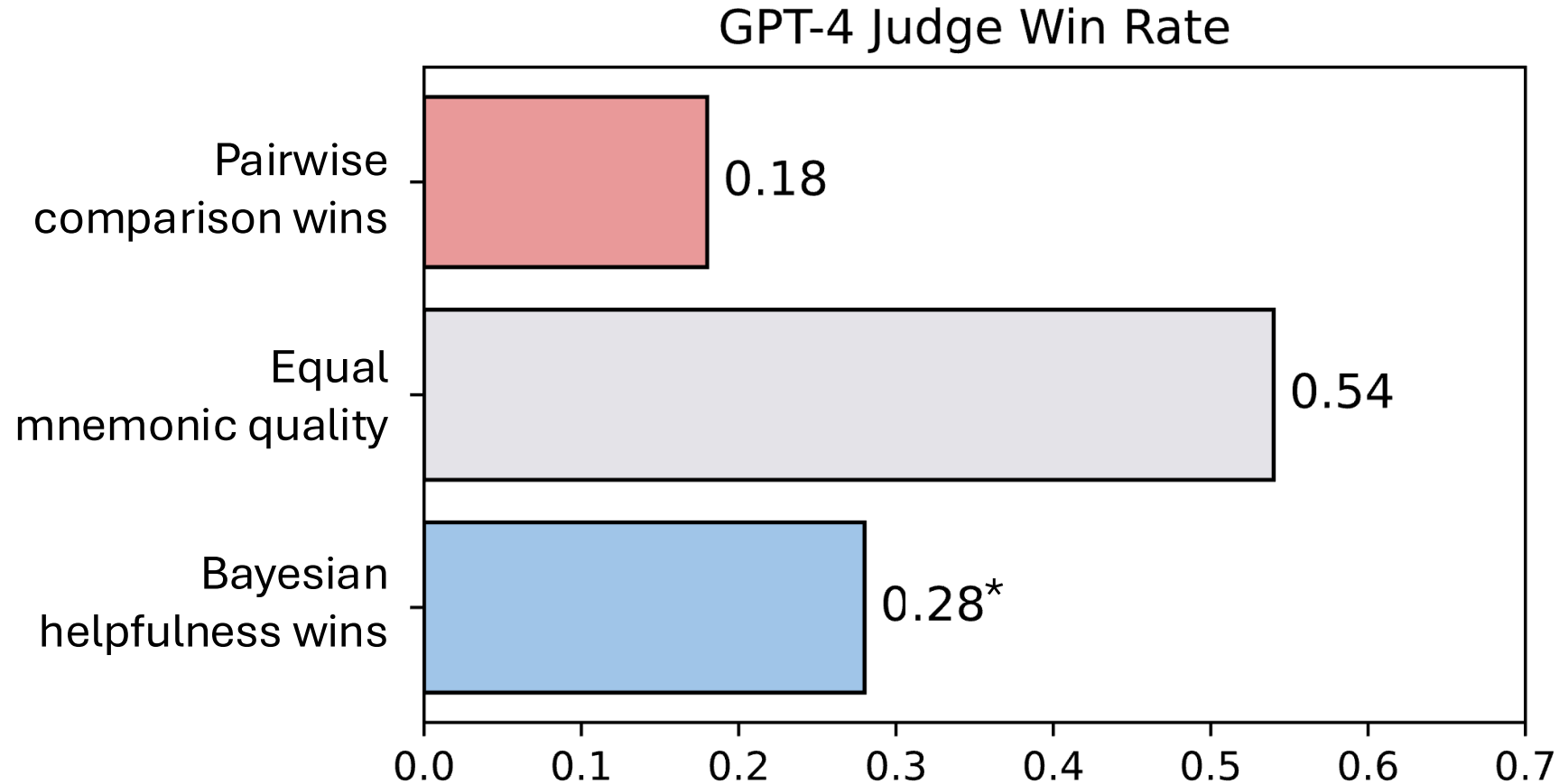
Our Initial SMART Model

[2] Direct Preference Optimization: Your Language Model is Secretly a Reward Model

Do we even need observed preferences?



Do we even need observed preferences?




➤ Training on observed preferences significantly improves mnemonic quality!

What do mnemonic experts think?

What do mnemonic experts think?

Two mnemonic researchers rate mnemonics from: SMART, GPT-4, and a professional writer

Benevolent
Benevolent sounds like "benefits", and a boss that gives their workers benefits is kind

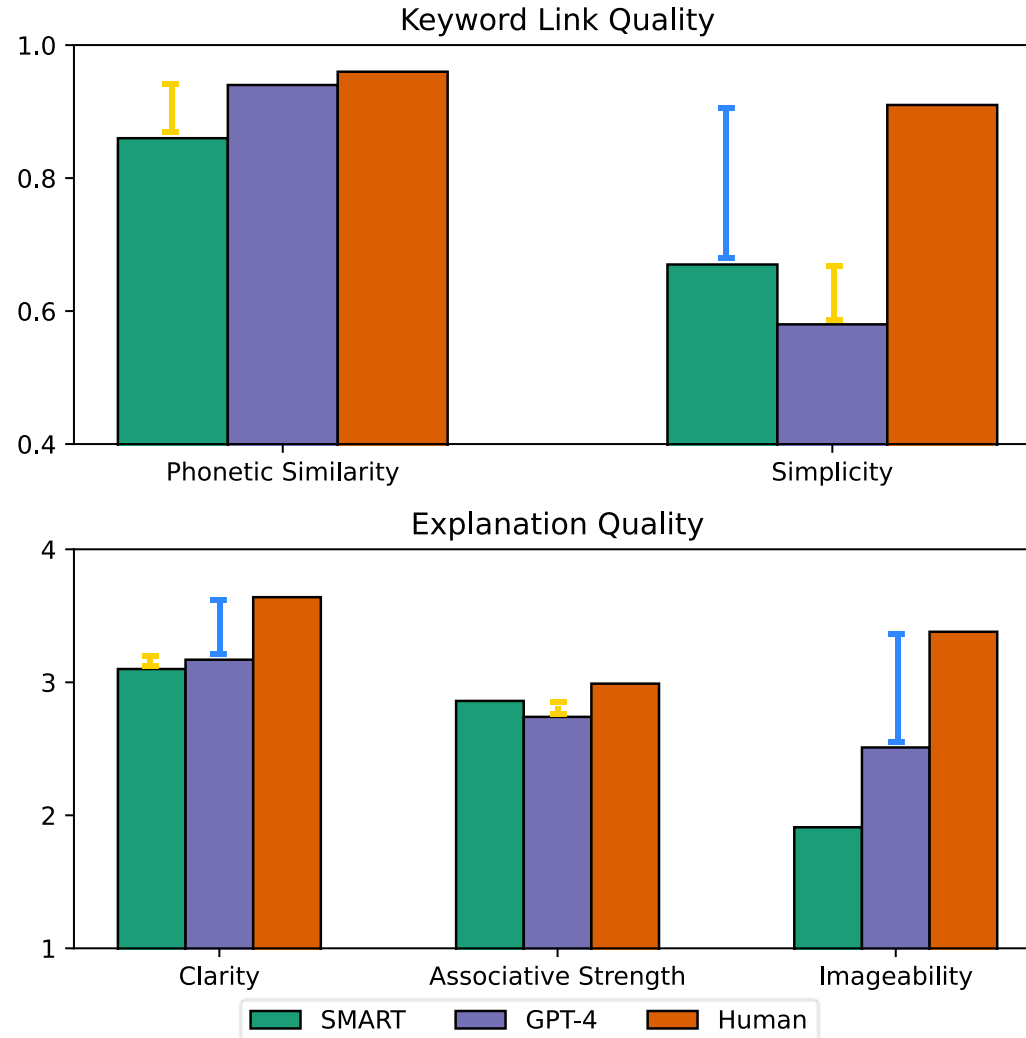
 **Keyword Link Quality**

Phonetic Similarity
Simplicity

 **Explanation Quality**

Clarity
Associative Strength
Imageability

What do mnemonic experts think?



1) **SMART** matches **GPT-4**!

SMART is a more efficient mnemonic generator!

2) Our **human writer** is much better than **SMART** and **GPT-4**!

Mnemonic generation is a challenging task!

Lionized

Lionized sounds like "lion-eyes," envisioning a lion being admired for its eyes. Lionized means to be admired or treated like a celebrity.

Haptic

Haptic sounds like "happy tic", which can be associated with a happy tickle, relating to touch. Haptic refers to the sense of touch.

Upbraid

Upbraided sounds like "up" + "braid". Imagine a teacher scolding a student for not braiding her hair properly. This scolding or reprimanding is upbraiding.

Conclusion: How did we build a more helpful QA system?

Overall:

- Students cannot predict which mnemonics best aid learning (expressed versus observed)
- Using observed preferences versus just expressed improve overall mnemonic quality
- Training smaller models on student feedback can match the helpfulness of larger systems



Typical preference training only gives models that appear helpful



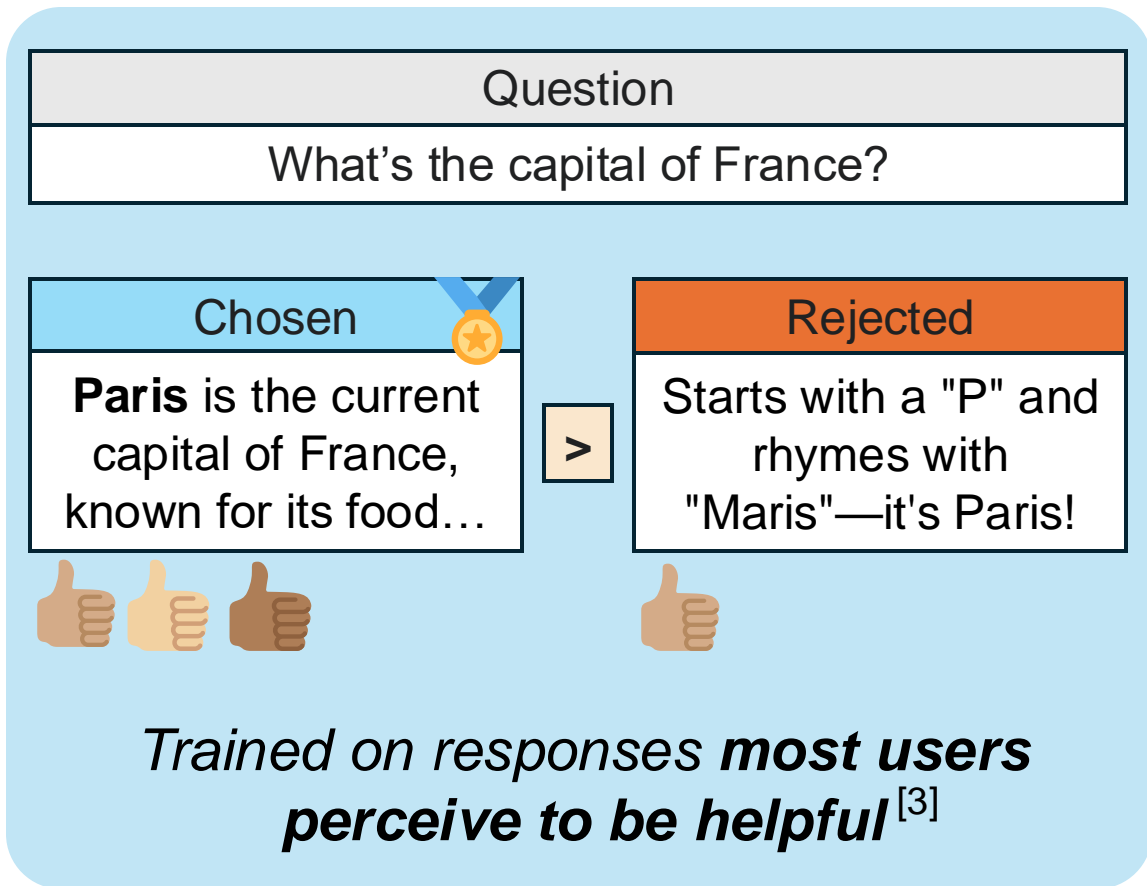
To build helpful systems, we must capture what truly helps users



We may need to steer LLMs toward multiple facets of helpfulness

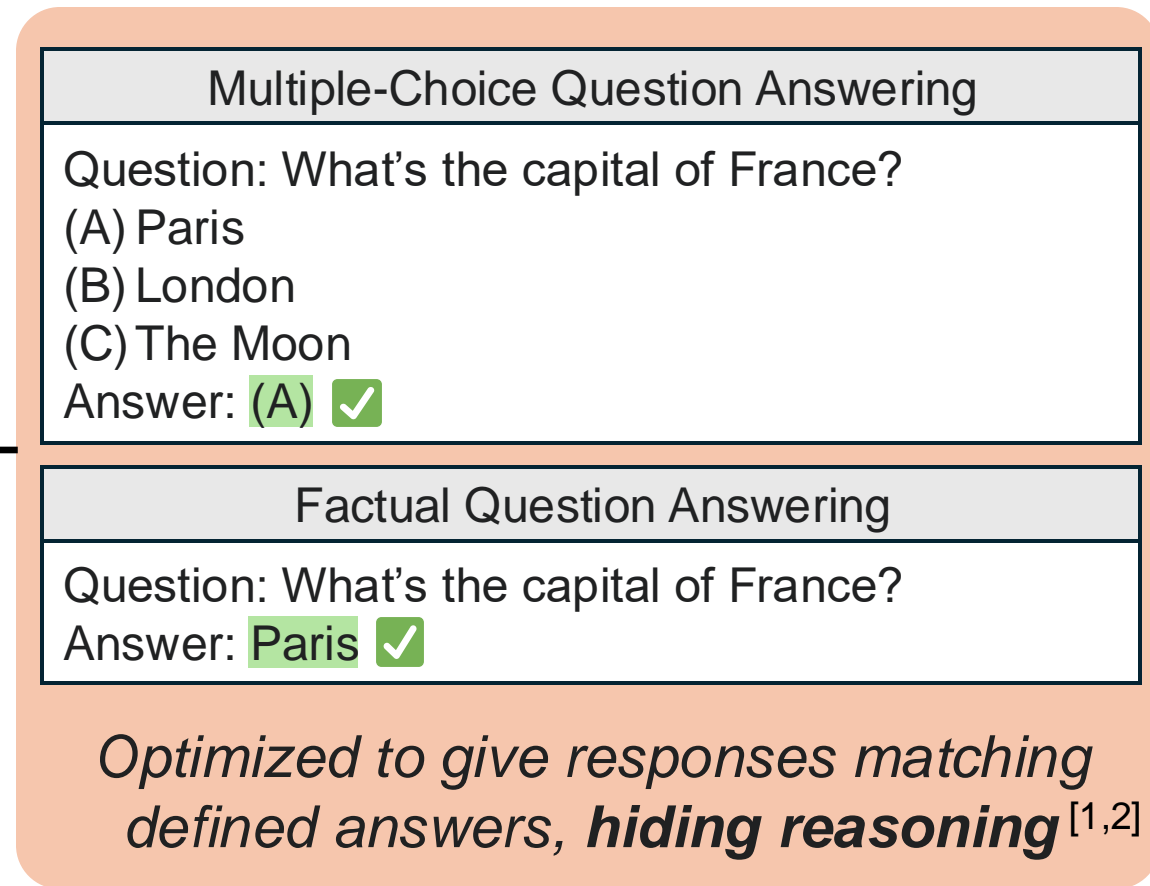
Recap: How can we teach LLMs to help users in QA?

Preference Training



[3] A SMART Mnemonic Sounds like "Glue Tonic" (EMNLP 2024)

Correctness Evaluation

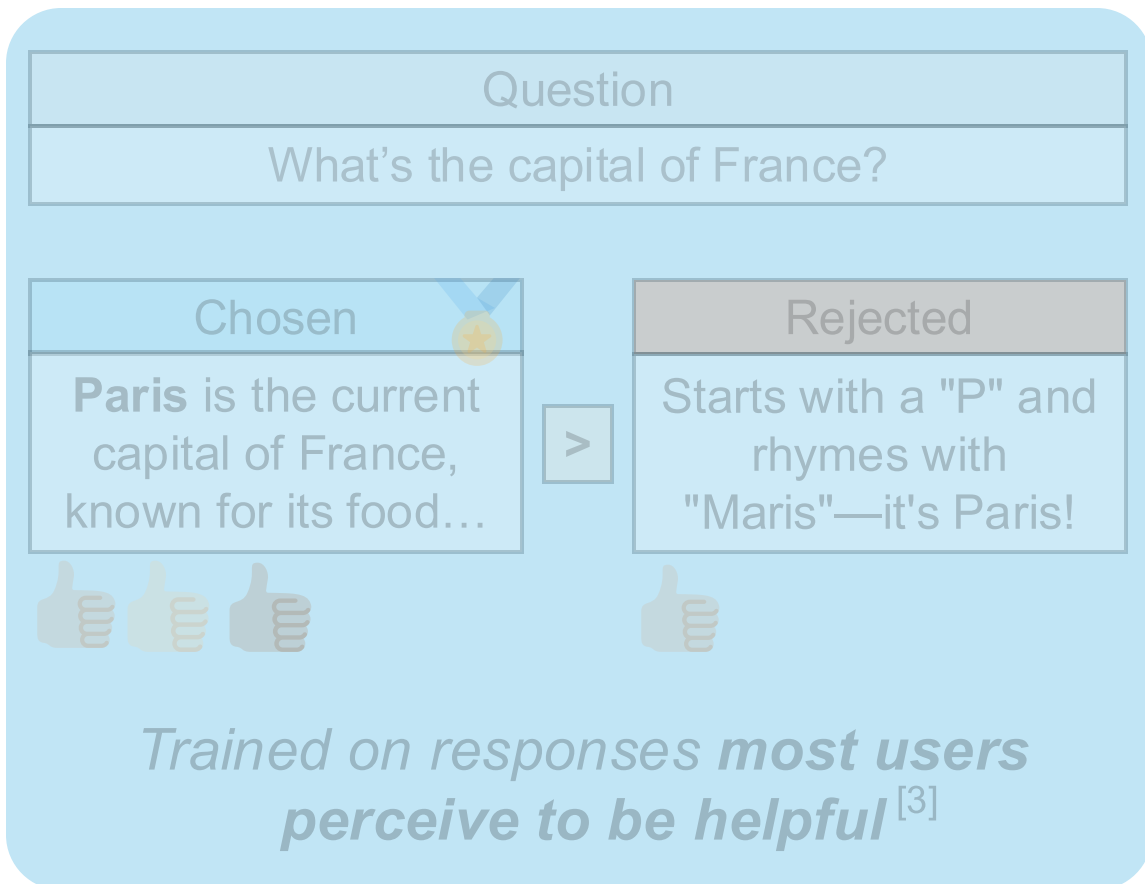


[1] It's Not Easy Being Wrong (ACL 2024, Findings)

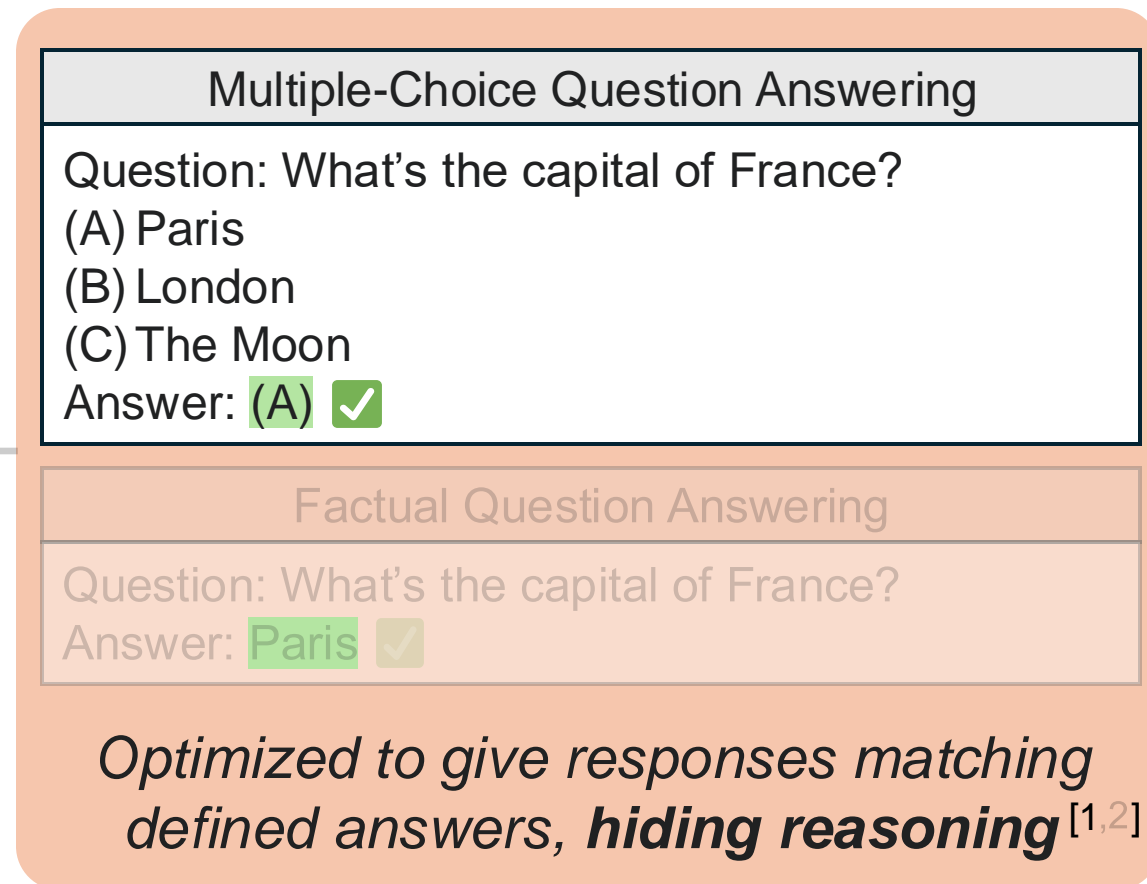
[2] Reverse Question Answering (NAACL 2025)

Users need *adaptable reasoning*, but correctness doesn't capture this

Preference Training



Correctness Evaluation



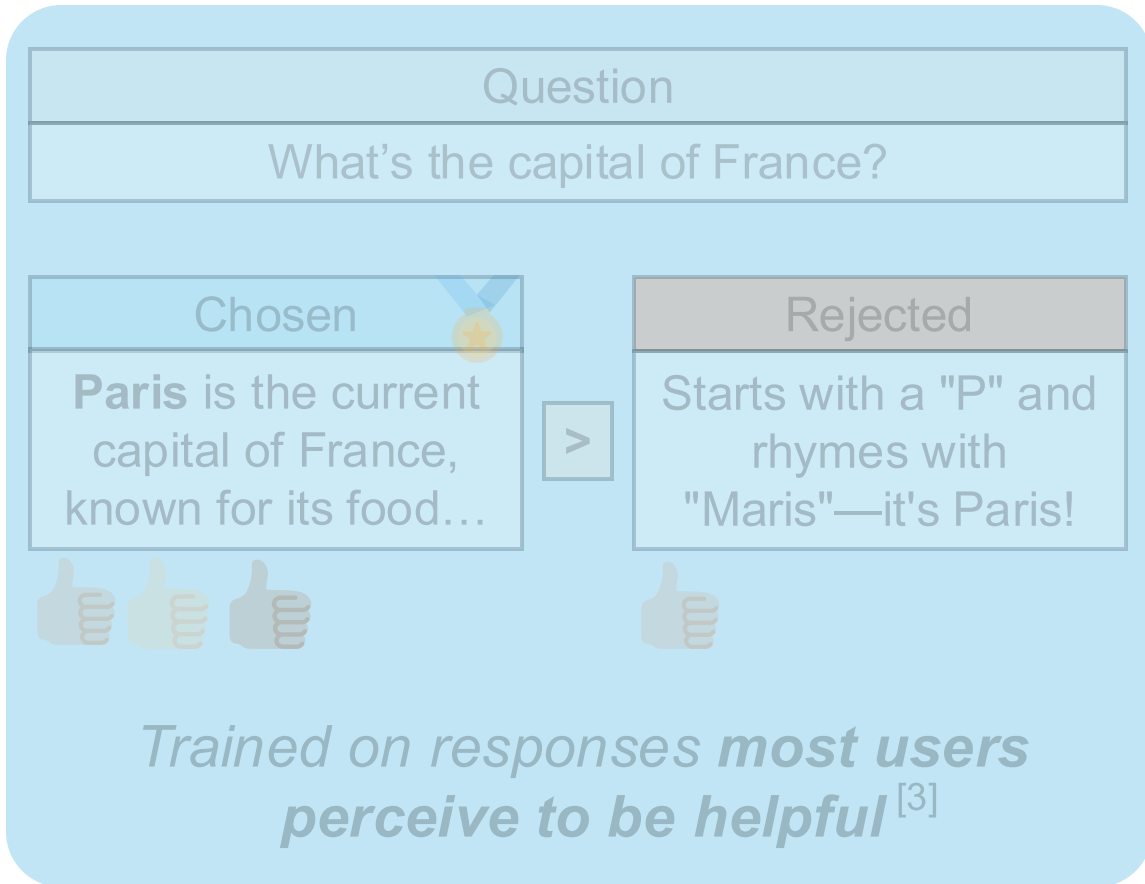
[3] A SMART Mnemonic Sounds like "Glue Tonic" (EMNLP 2024)

[1] It's Not Easy Being Wrong (ACL 2024, Findings)

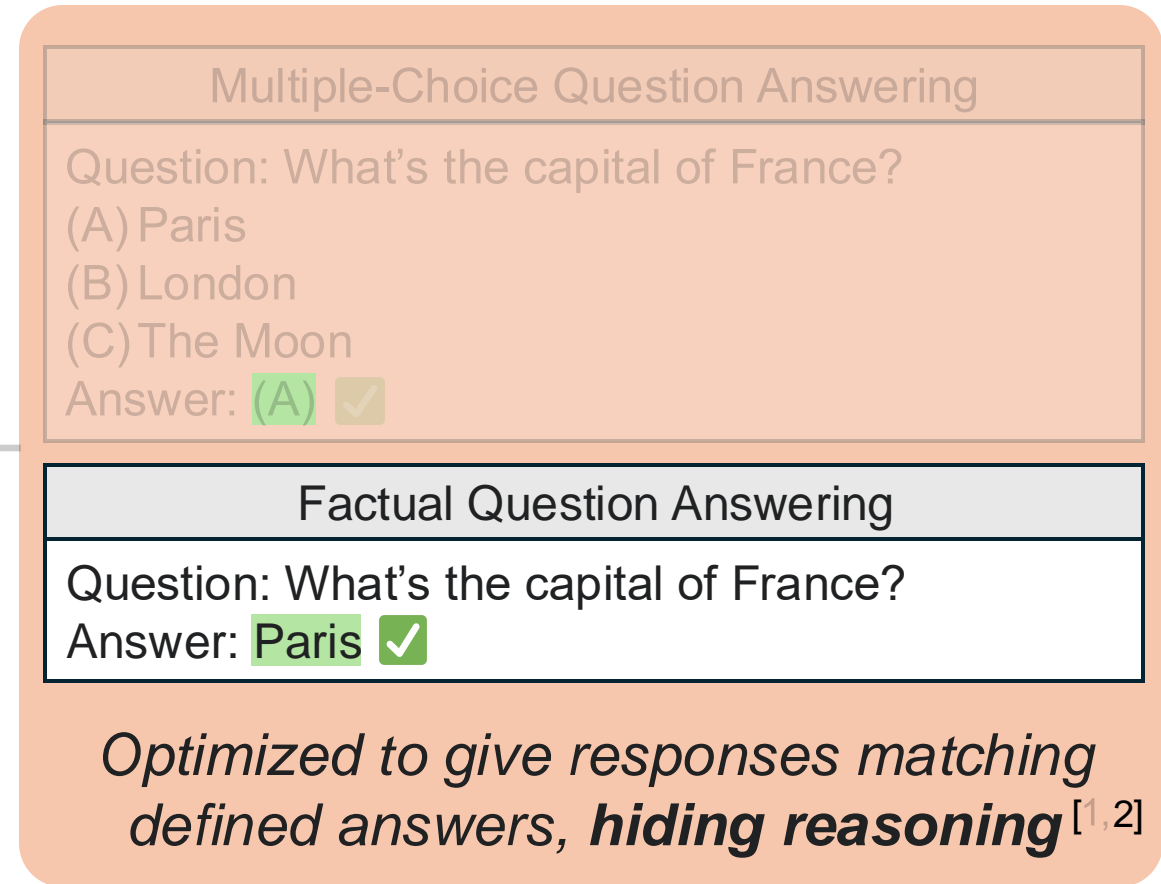
[2] Reverse Question Answering (NAACL 2025)

LLMs may also generate reasoning that solely *appears helpful*...

Preference Training



Correctness Evaluation



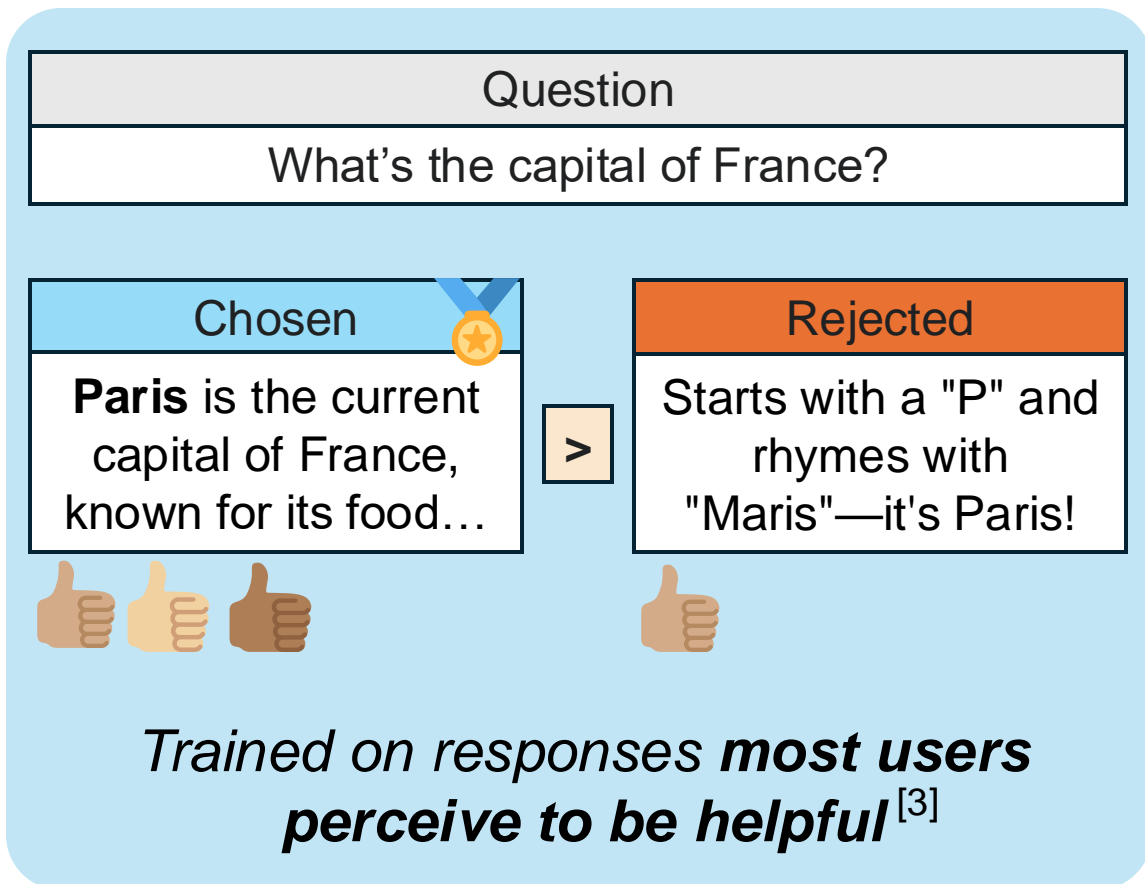
[3] A SMART Mnemonic Sounds like "Glue Tonic" (EMNLP 2024)

[1] It's Not Easy Being Wrong (ACL 2024, Findings)

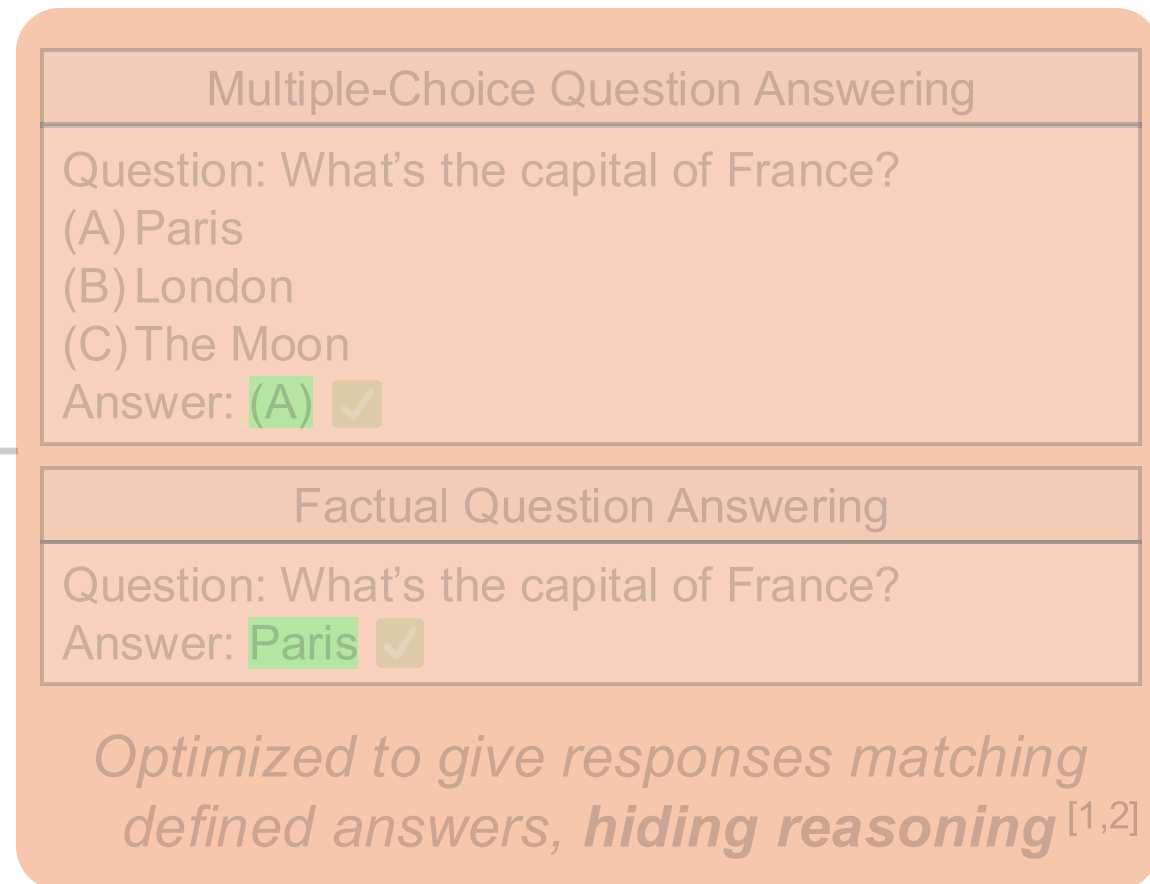
[2] Reverse Question Answering (NAACL 2025)

...requiring us to measure *what actually helps users* downstream

Preference Training



Correctness Evaluation



[3] A SMART Mnemonic Sounds like "Glue Tonic" (EMNLP 2024)

[1] It's Not Easy Being Wrong (ACL 2024, Findings)

[2] Reverse Question Answering (NAACL 2025)

So what's next?

Comparing human and model helpfulness

What is the nationality of Don Quixote's author?

Irish



1. Find who wrote Don Quixote
2. Look where they were born

Spanish

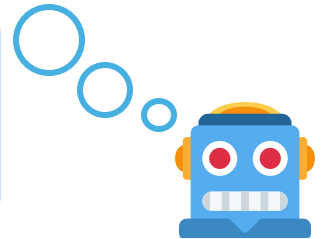


Improving Model Personalization



What should I get for dinner tonight?

This user likes short responses, Thai cuisine, prefers eating alone, dancing, drinking... which are relevant?



Balancing artist and consumer preferences

✓ Whose Boat Does it Float?... ✗

✗ Improving Personalization... ✓



Artist Preferences



Consumer Preferences



Evaluating "true helpfulness" efficiently

Benevolent sounds like "benefits". A boss giving benefits is considered kind



Benevolent sounds like "Ben Franklin", known to be a very kind individual

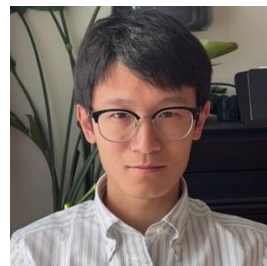


Thank you :)

My amazing advisors



UNIVERSITY OF
MARYLAND



THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC



EMPIRE STATE
UNIVERSITY



Ai2



cohere