

Modeling Future Conversation Turns to Teach LLMs to Ask Clarifying Questions

Michael J.Q. Zhang, W. Bradley Knox, and Eunsol Choi

Contact: michaelzhang@nyu.edu



NYU



TEXAS
The University of Texas at Austin

TL;DR

Our goal is to train LLMs to resolve ambiguity in user requests by asking clarifying questions.

We achieve this by modeling future outcomes from asking different clarifying questions to determine overall preference across users.

1. Background

Who is the highest paid football player?

As of 2024, the highest-paid football players are Cristiano Ronaldo and Karim Benzema...

GPT

The highest-paid football player in the NFL for 2024 is Joe Burrow...

Gemini

User Requests are often ambiguous:

- “**football**” can be or

LLM responses often...

- Do not distinguish different interpretations
- Rely on the user to identify the interpretation

3. SFT Training Data Generation (~5K Examples)

We generate clarifying interactions from multiple answers $Q \rightarrow \{A_1, A_2, \dots\}$

Answer sets are from

- **Models** by sampling multiple short answers
- **Human** annotators tasked with labeling all correct answers

How many medals did the US win in the 2016 olympics?

121

46

Are you asking about all medals or just gold?

All Medals

Gold Medals

121

46

3.5. DPO Training (~20K Examples)

We train a **User-Simulator model** to replicate each annotator’s response to a clarifying question and assume standard QA data with 1+ annotators per question

Final preference labels aggregated across annotators based on:

- **(1) Efficacy** — Which response satisfies the most users?
- **(2) Efficiency** — Which response leads to the least interaction turns?

2. Assigning Preferences from Interaction Outcomes

Expect

Expects

Who is the highest paid football player?

Response A

Cristiano Ronaldo...

Response B

Joe Burrow...

Response C

Are you asking about or ?

Response D

Are you asking about a specific time period?

I meant

Cristiano Ronaldo...

I meant

Joe Burrow...

Currently

Cristiano Ronaldo...

Standard Single-Turn Preference Annotation

	A	B	C	D
Expect	✓	✗	✗	✗
Expects	✓	✗	✗	✗
	✗	✓	✗	✗

Best Response: A

- Difficult to assess clarifying questions
- Incentivizes giving the majority’s choice answer

Ours Full Interaction Preference Annotation

	2 Turns		4 Turns	
	A	B	C	D
Expect	✓	✗	✓	✓
Expects	✓	✗	✓	✓
	✗	✓	✓	✗

Best Response: C

- Users assess clarifying questions after interacting
- Models can learn **when** to ask a clarifying question and **what** to ask

4. Evaluations

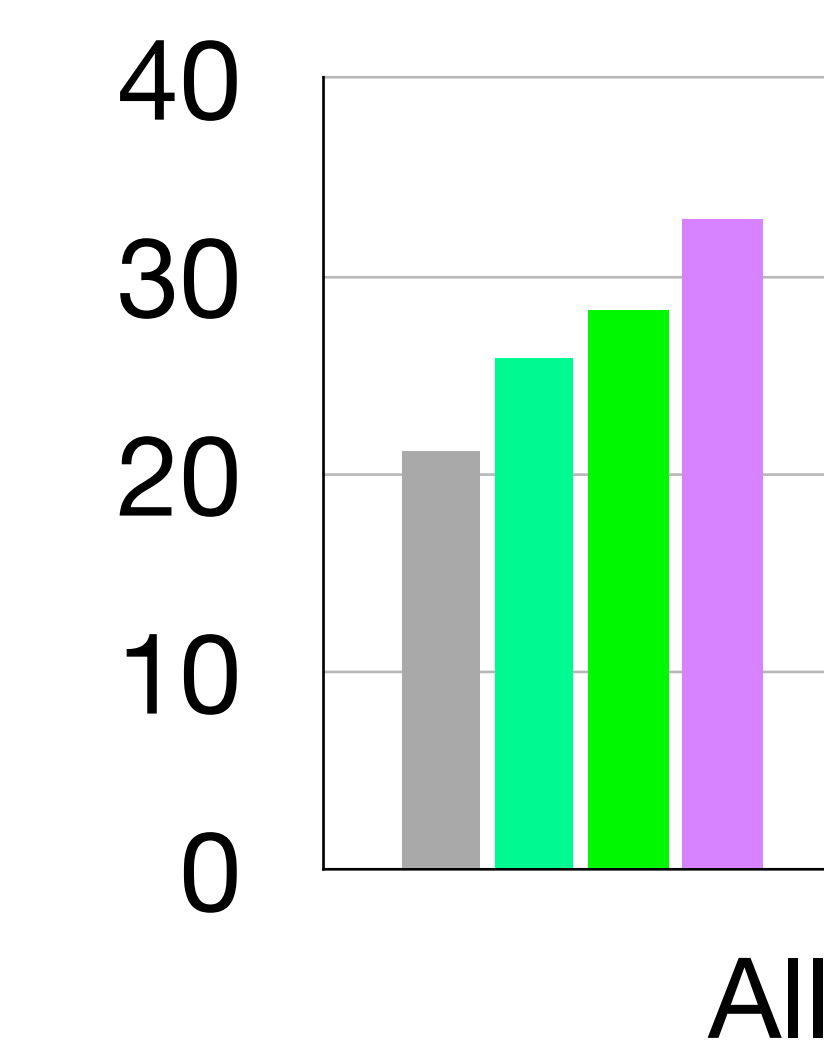
We evaluate on AmbigQA

We measure both QA performance and *interaction efficiency*:

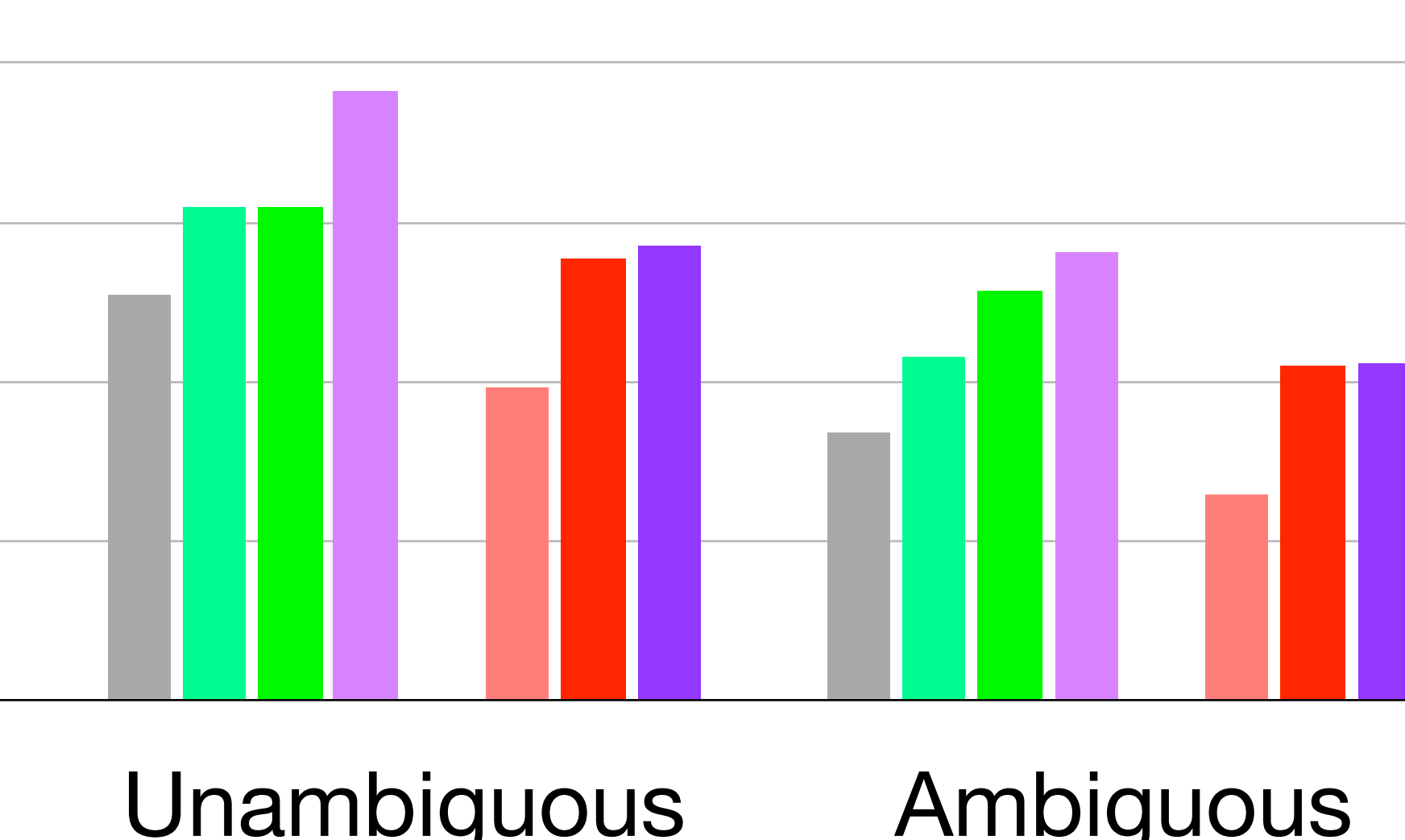
- Do models only ask clarifying questions when necessary?

F1 over All Answers

We evaluate over **All** questions

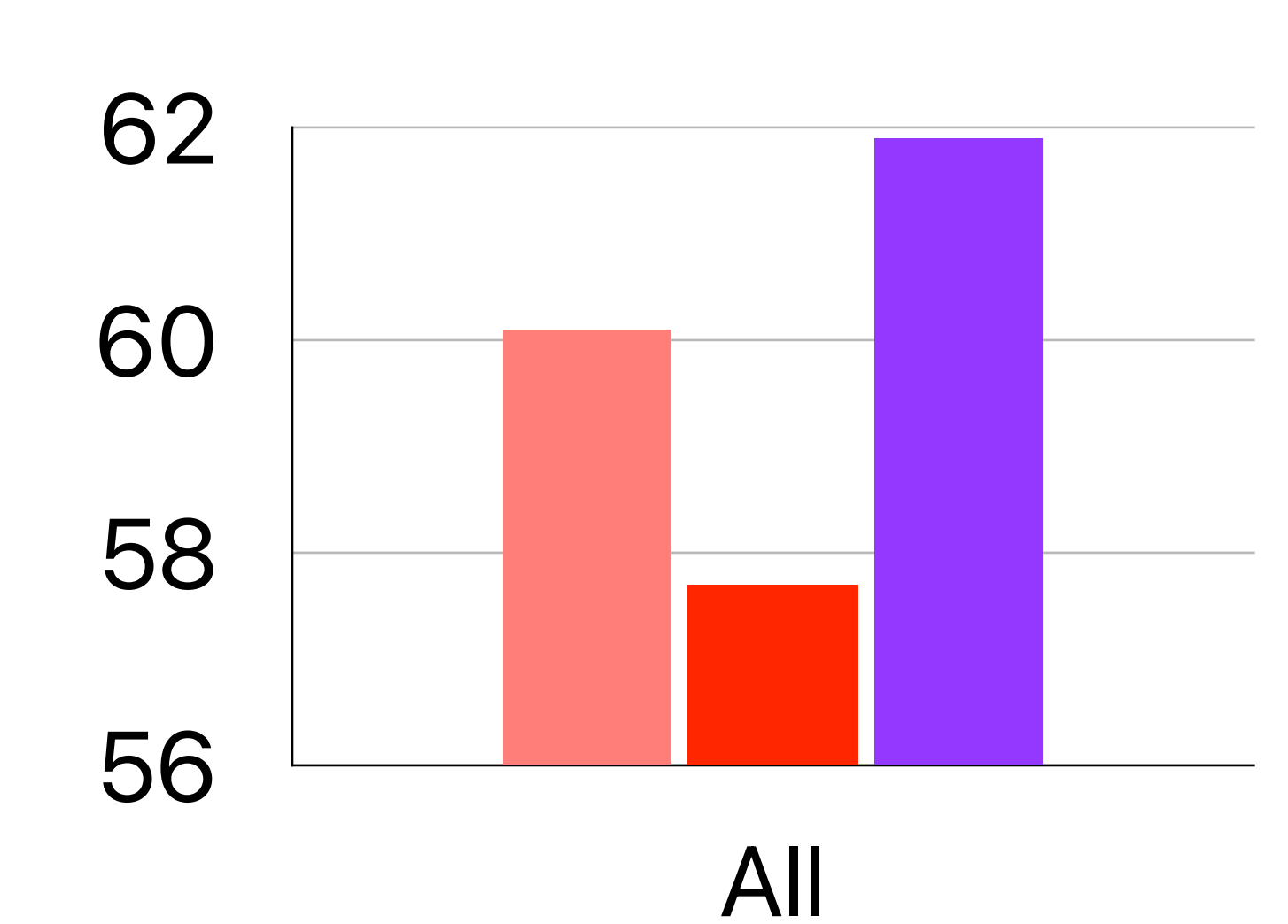


Gains are over both **Ambiguous** and **Unambiguous** questions



Direct-Answer Accuracy

Directly evaluates clarify vs. direct answer decisions



Compared Systems:
Model (Avg Turns)

Always answers →
Always clarifies {

- SFT Only (1)
- SFT Only (2)
- Single-Turn Prefs (2)
- Our — Full Interaction Prefs (2)

Predicts to clarify or answer {

- Ours w/ ProCoT (1.56)
- Ours w/ PPDPP (1.65)
- Ours — Full (1.54)

CoT Prompt to predict whether to clarify
Auxiliary LLM trained to predict whether to clarify