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

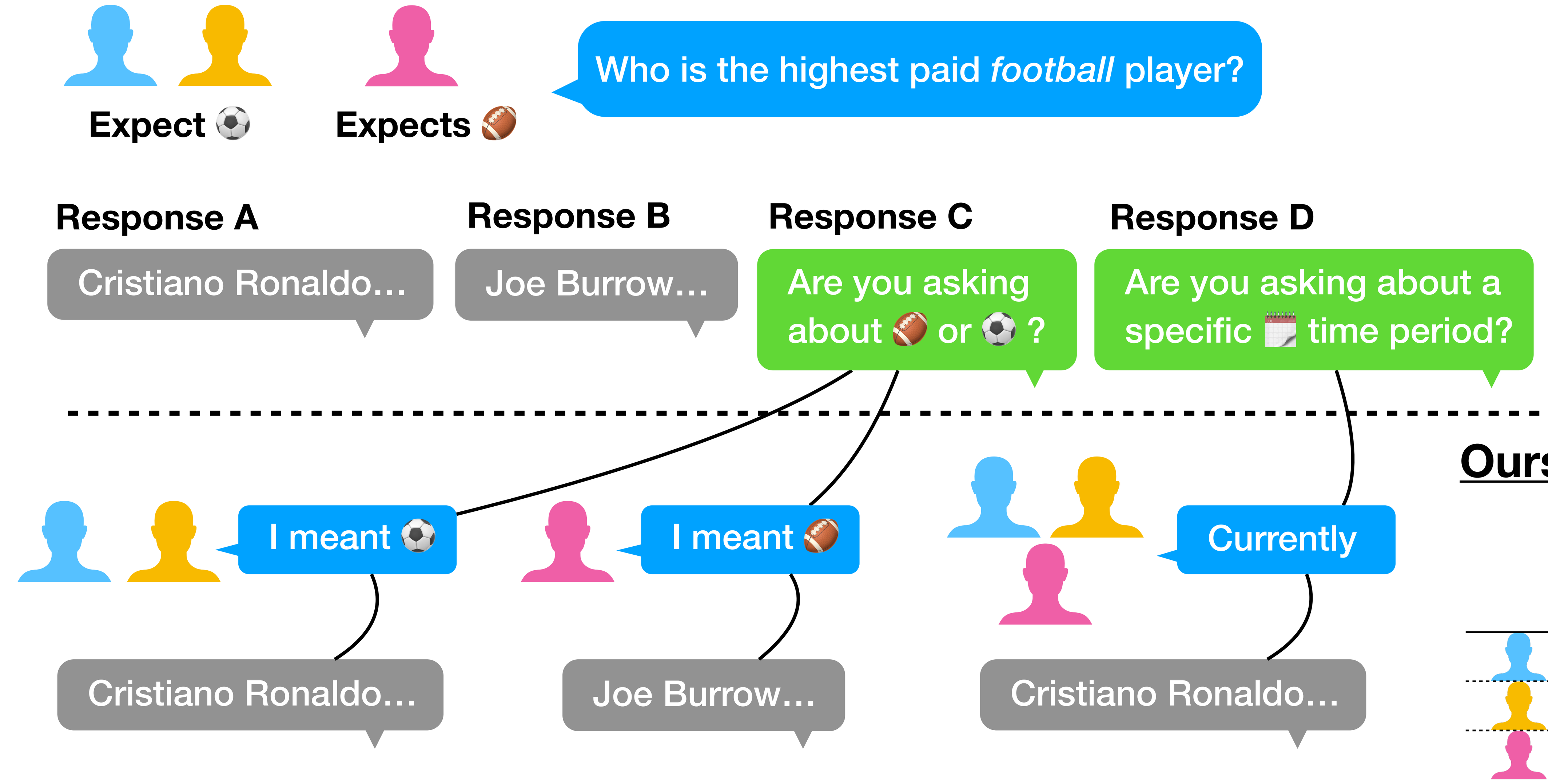
User Requests are often ambiguous:

- ***"football"*** *can be* 🏈 or ⚽

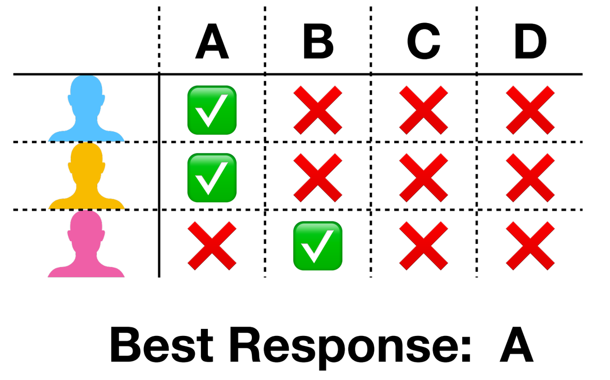
LLM responses often…

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

2. Assigning Preferences from Interaction Outcomes

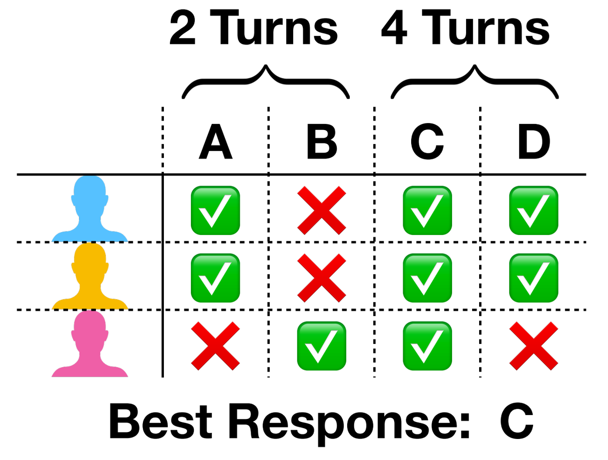


Standard Single-Turn Preference Annotation



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

Ours Full Interaction Preference Annotation



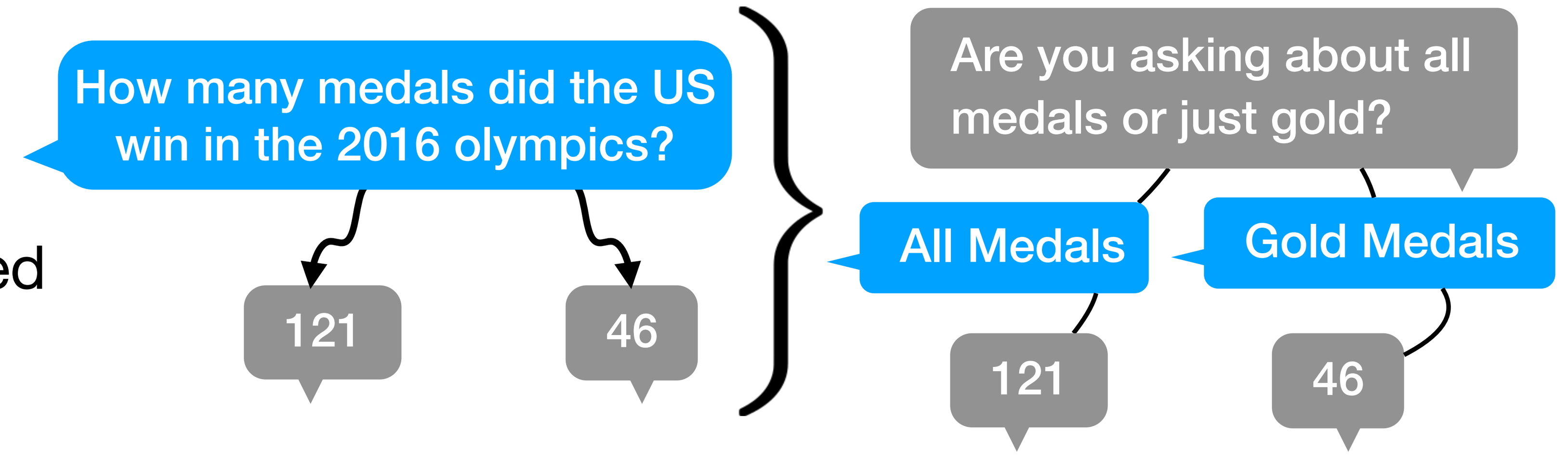
- Users asses clarifying questions after interacting
- Models can learn ***when*** to ask a clarifying question and ***what*** to ask

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

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

Answer sets are from

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



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?

4. Evaluations

We evaluate on AmbigQA

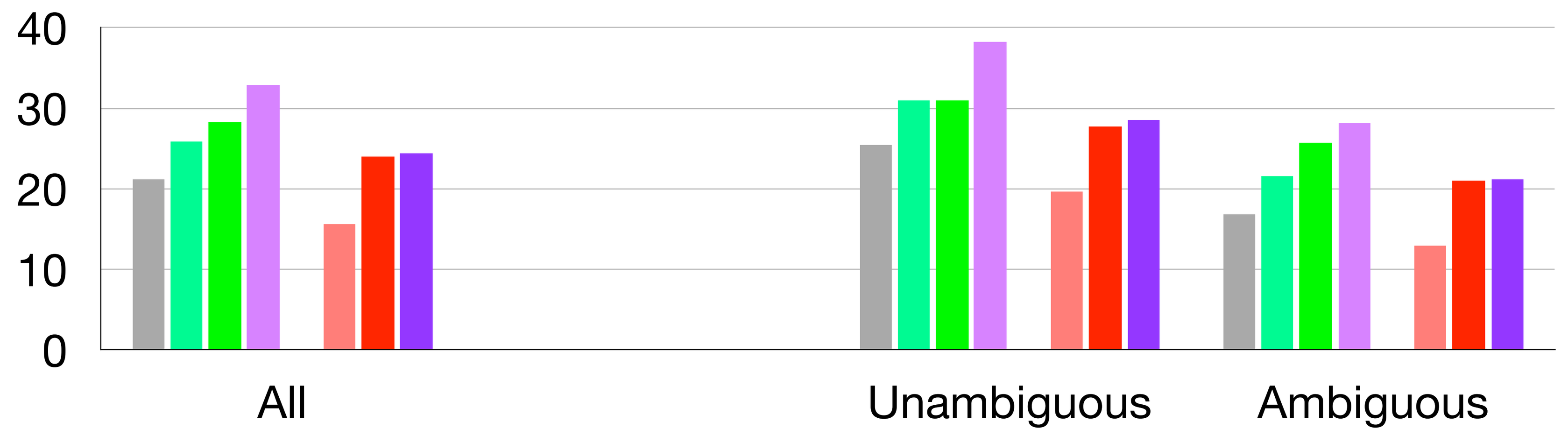
We measure both QA performance and *interaction efficiency:*

- Do models only ask clarifying questions when necessary?

F1 over All Answers

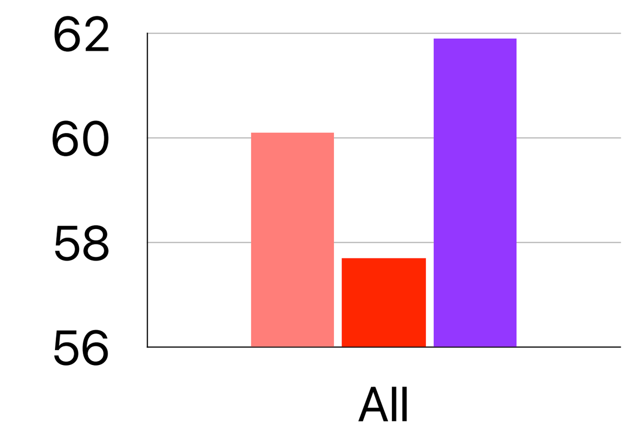
We evaluate over **All** questions

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 → SFT Only (1)

Always clarifies: SFT Only (2); Single-Turn Prefs (2); Our — Full Interaction Prefs (2)

Predicts to clarify or answer: Ours w/ ProCoT (1.56) ← CoT Prompt to predict whether to clarify; Ours w/ PPDPP (1.65) ← Auxiliary LLM trained to predict whether to clarify; Ours — Full (1.54)