ChatGPT Role-play Dataset: Analysis of User Motives and Model Naturalness

Yufei Tao ⬨, Ameeta Agrawal ⬨, Judit Dombi‡, Tetyana Sydorenko ◦, Jung In Lee ◦

⬨ Department of Computer Science, Portland State University, Portland, USA
‡Department of English Linguistics, University of Pécs, Pécs, Hungary
◦ Department of Applied Linguistics, Portland State University, Portland, USA

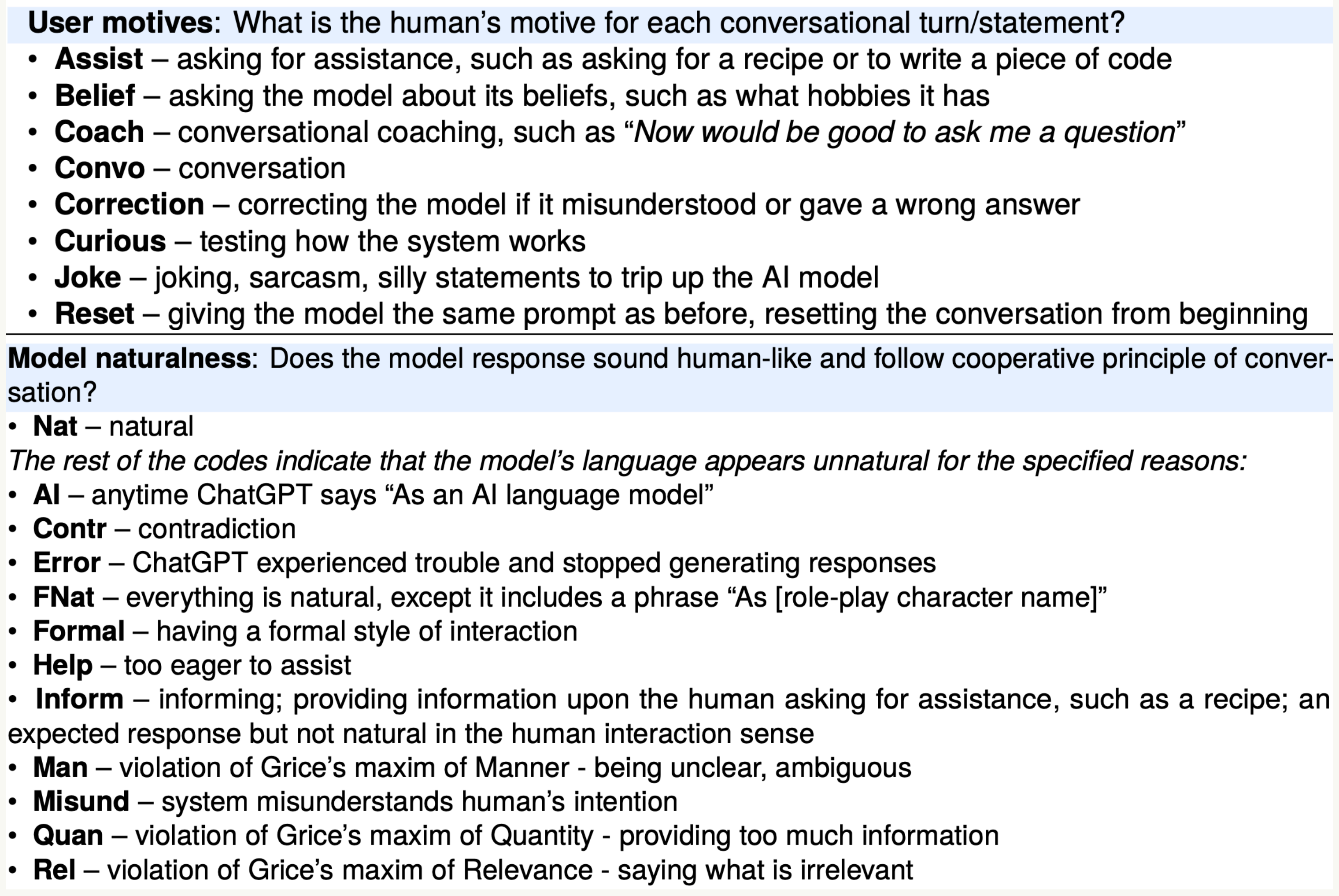
Link to our dataset:

DATASET OVERVIEW

- Three subsets: ***vanilla***, ***boss***, ***classmate***.
- Conversations when ChatGPT act "as is" and role-play.
- 57 participants, 85 unique conversations, 1742 utterances.
- Each utterance was manually annotated by three experts in linguistics specializing in pragmatics.

ANNOTATION

- **User motives**: Intent behind each human utterance
- **Model naturalness**: Evaluated against Grice's four maxims (Quantity, Quality, Relevance, Manner)



User motives: What is the human's motive for each conversational turn/statement?

- **Assist** – asking for assistance, such as asking for a recipe or to write a piece of code
- **Belief** – asking the model about its beliefs, such as what hobbies it has
- **Coach** – conversational coaching, such as "*Now would be good to ask me a question*"
- **Convo** – conversation
- **Correction** – correcting the model if it misunderstood or gave a wrong answer
- **Curious** – testing how the system works
- **Joke** – joking, sarcasm, silly statements to trip up the AI model
- **Reset** – giving the model the same prompt as before, resetting the conversation from beginning

Model naturalness: Does the model response sound human-like and follow cooperative principle of conversation?

- **Nat** – natural

The rest of the codes indicate that the model's language appears unnatural for the specified reasons:

- **AI** – anytime ChatGPT says "As an AI language model"
- **Contr** – contradiction
- **Error** – ChatGPT experienced trouble and stopped generating responses
- **FNat** – everything is natural, except it includes a phrase "As [role-play character name]"
- **Formal** – having a formal style of interaction
- **Help** – too eager to assist
- **Inform** – informing; providing information upon the human asking for assistance, such as a recipe; an expected response but not natural in the human interaction sense
- **Man** – violation of Grice's maxim of Manner - being unclear, ambiguous
- **Misund** – system misunderstands human's intention
- **Quan** – violation of Grice's maxim of Quantity - providing too much information
- **Rel** – violation of Grice's maxim of Relevance - saying what is irrelevant

CRD (ChatGPT Role-play Dataset)

Vanilla	... **H:** Can I be Frank with you? **C:** Of course, you can be frank with me! As an AI language model, I am here to listen and provide any help or support that I can. What's on your mind? ...
Boss	... **H:** So, I would like to arrange a meeting for us. **C:** Of course, that sounds like a good idea. Can you please let me know the purpose and agenda of the meeting, as well as the preferred date and time? ...
Classmate	... **H:** OMG I love it so much, it is so much better than the one we had before. What d you think **C:** I'm glad to hear that you're enjoying the class, Adam! As an exchange student from France, I'm finding the class to be quite different from what I'm used to back home. ...

Key Points:

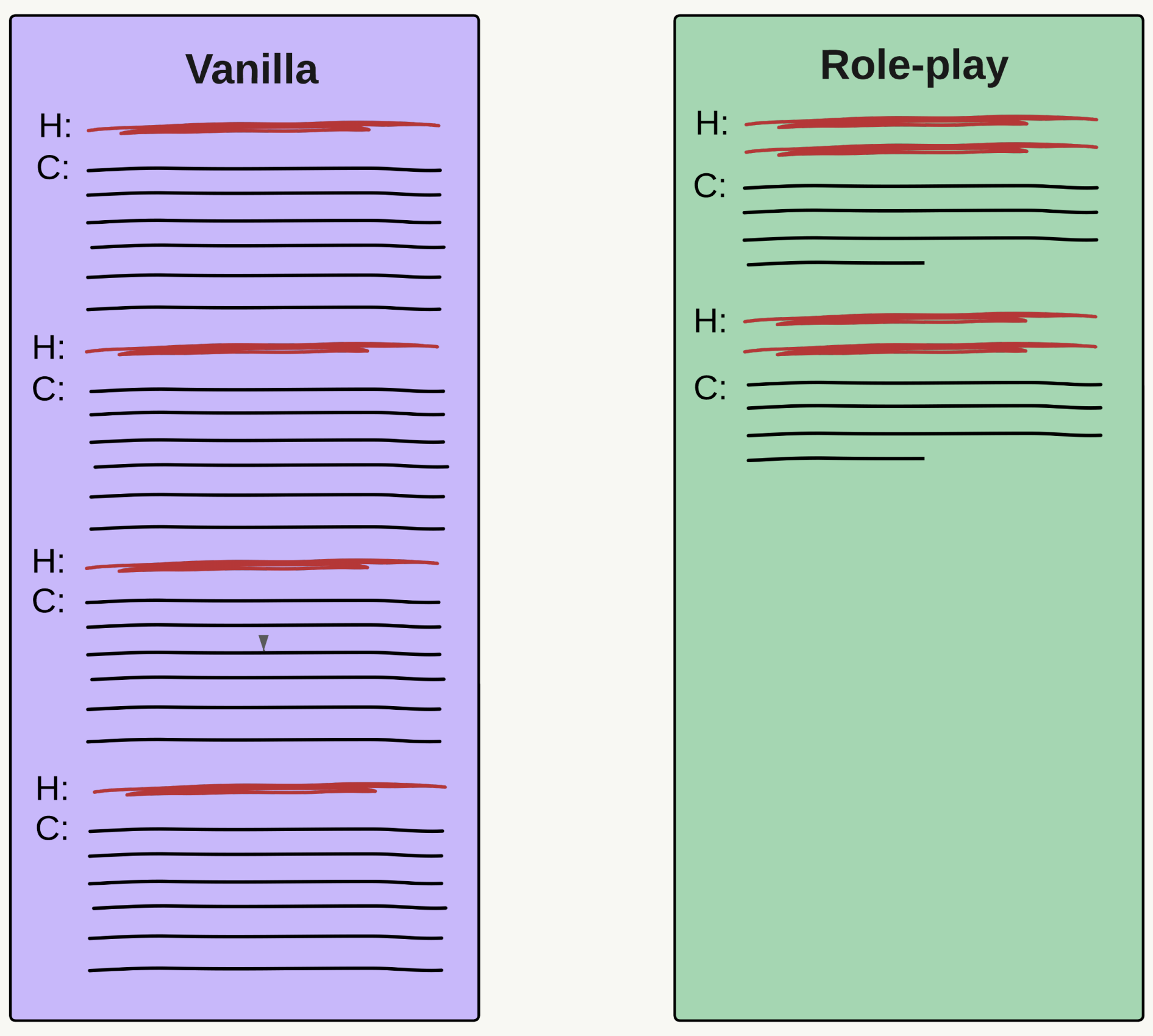
1. MODEL PERFORMANCE ANALYSIS:

Combining Perplexity and Response Length

- **Findings:** ChatGPT shows significantly higher perplexity in vanilla settings compared to role-play scenarios (boss and classmate), indicating challenges in more structured interactions. Human responses typically have higher perplexity in vanilla due to their unpredictable nature.
- **Impact:** Longer utterances in the role-play settings correlate with decreased perplexity, suggesting that contextual engagement improves model performance.

Naturalness and Verbosity

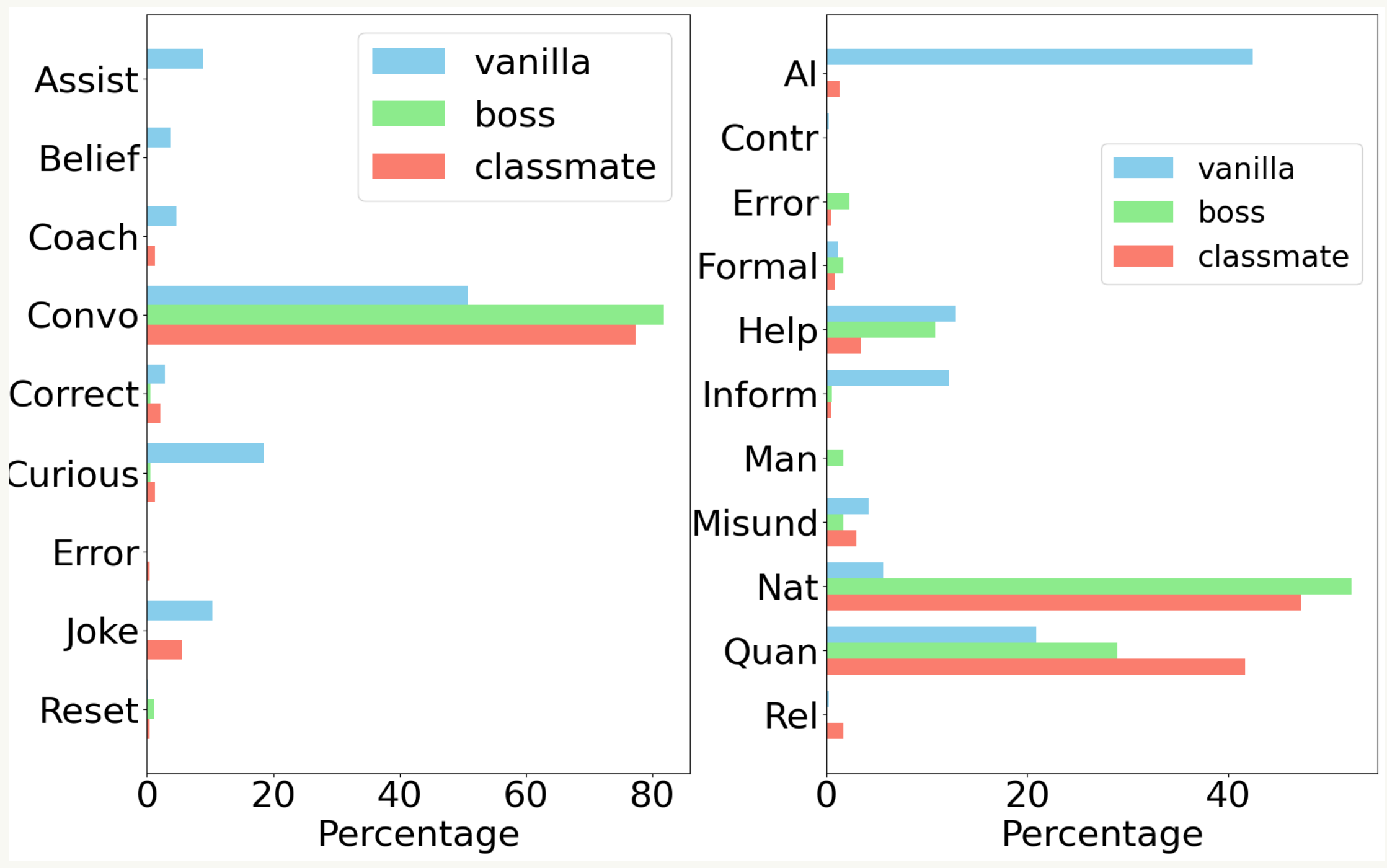
- **Observations:** Naturalness improve in role-play settings, with 52% in boss and 47% in classmate settings being rated as natural, compared to only 5.6% in vanilla.
- **Challenge:** Despite being more natural, responses in role-play settings tend to be overly verbose: boss (28%) and classmate (41%), highlighting a trade-off between fluency and brevity.



2. USER ENGAGEMENT AND CONVERSATION LENGTH:

Influence of User Motives on Conversation Dynamics

- **Conversation length:** Vanilla conversations are nearly twice as long as role-play conversations, with user motives significantly influencing conversational length. In role-play settings, the motive of engaging in a conversation leads to longer conversations.
- **Correlation:** There is a strong correlation between the number of questions asked by users and the overall length of the conversations across all datasets. This indicates that engaging users with questions is an effective strategy to extend conversations.

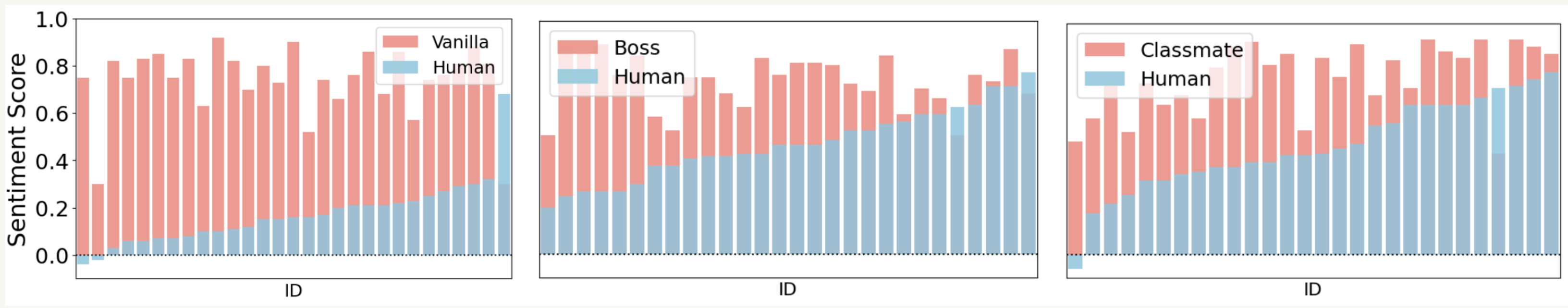


User Motives(left) and Model Naturalness(right)

3. SENTIMENT AND SATISFACTION:

Overall Sentiment in Role-Play Settings

- **Sentiment Analysis:** Users exhibit more positive sentiments in role-play settings compared to vanilla, where responses often feel inexpressive. The boss and classmate settings show a trend toward positive sentiments, suggesting that fitting the conversational role increases satisfaction.
- **Key Takeaway:** Adhering to expected conversational roles significantly enhances participant's experience, as seen by the alignment of model responses with user expectations in role-play scenarios.


Sentiment Score
Vanilla
Human
ID
Boss
Human
ID
Classmate
Human
ID