

Why Do We Laugh? Annotation and Taxonomy Generation for Laughable Contexts in Spontaneous Text Conversation

Koji Inoue, Mikey Elmers, Divesh Lala, Tatsuya Kawahara
Kyoto University, Japan



Summary

- We introduce a semi-automated method using an LLM to annotate and generate a ten-category taxonomy of laughable contexts in Japanese text conversations, aiming to enhance conversational AI's ability to understand and generate laughter.
- While an LLM (GPT-4o) could recognize laughable contexts with an F1 score of 43.14%, significantly above chance, challenges remain in capturing nuanced humor, particularly for categories like "Nostalgia and Fondness" and "Positive Energy."

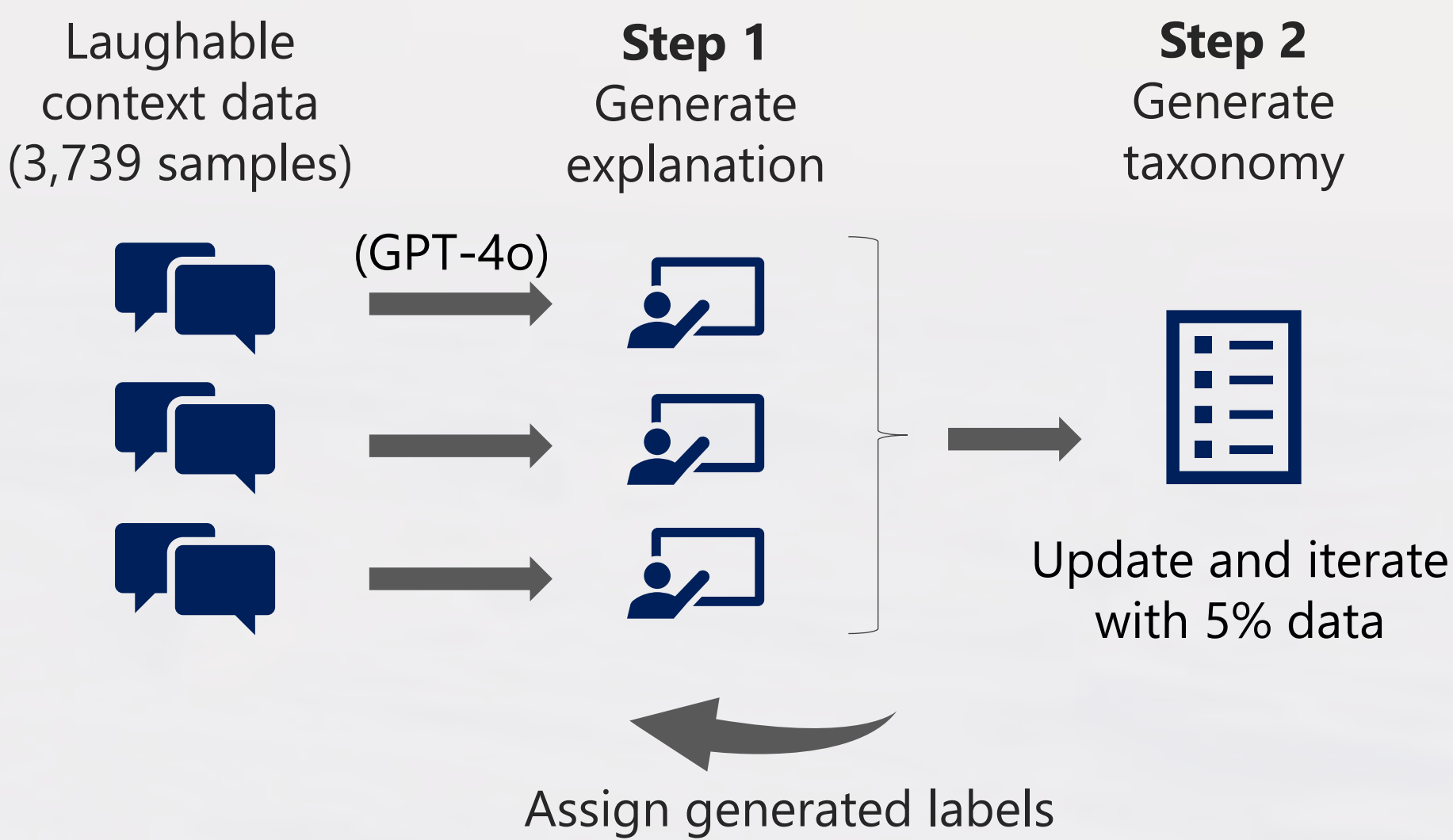
Annotation of Laughable Context

- We annotated laughable contexts in the RealPersonaChat [Yamashita 23].
- 900 dialogues were annotated, with each dialogue having 30 utterances.
- 5 annotators made a binary decision (laughable / not) for each utterance.
- Annotation agreement varied, highlighting the subjectivity of laughter, with 14.8% (3,739 contexts) labeled as laughable by majority vote.

Utterance	Laughable?	Ratio (Yes)	#sample
I think that's a wonderful attitude. I always end up talking about myself, so I should follow your example.	NO	1.0 (5/5)	163 (0.64%)
Is that so? But does your husband listen to your stories?	NO	0.8 (4/5)	845 (3.34%)
Yes, yes, he listens to me. I wonder if I'm putting too much on him?	NO	0.6 (3/5)	2,731 (10.80%)
I don't think so! He's so kind. My husband doesn't seem to listen to me. Huh, that's strange.	YES 🤔	0.4 (2/5)	8,143 (32.20%)
		0.2 (1/5)	11,928 (47.17%)
		0.0 (0/5)	1,479 (5.85%)

Generating Taxonomy of Laughable Reason

- To understand why contexts were labeled laughable, we analyzed the 3,739 samples marked as laughable by majority vote.
- Step 1:** GPT-4o was utilized to generate explanations for these human judgments due to the high cost of manual annotation.
- Step 2:** A taxonomy of laughable reasons was then generated iteratively: GPT-4o generated initial labels from subsets of reason data, which were manually validated, and then the LLM refined the taxonomy with new subsets until all data were processed.



Generated reason (from the top-right example)

If we were to speculate on the reasons a third party might judge that Person A laughed at Person B's final remark in this conversation, the following points can be considered:

Element of Humor: Person B's comment, "My husband doesn't seem to listen to me. Huh, that's strange.," contains a touch of self-deprecating humor. This lighthearted tone, making fun of their own situation, can be amusing to the listener. (...)

Relaxed Atmosphere: The overall tone of the conversation seems light and relaxed, and Person B's comment might have been perceived as a playful joke in line with this mood.

A combination of these factors may have led the third party to interpret that Person A laughed in response to Person B's remark.

- This process resulted in ten distinct taxonomy labels, including prominent categories like "(1) Empathy and Affinity" and "(2) Humor and Surprise".
- The LLM then assigned these taxonomy labels to each reason sample, allowing for multiple labels per sample, with the results showing a broad distribution across categories.
- The generated taxonomy's explanatory power was further substantiated by reviewing related studies in conversational analysis (see Table 3 in paper)

Generated taxonomy labels

	Label name	Explanation	#sample
(1)	Empathy and Affinity	Situations where a sense of closeness and laughter is generated by sharing common experiences or emotions in a conversation. This includes empathy for shared hobbies or everyday relatable situations.	3013 (80.6%)
(2)	Humor and Surprise	Cases where humor or an element of surprise in the statement triggers laughter. This includes unexpected twists, wordplay, and exaggeration.	3233 (86.5%)
(3)	Relaxed Atmosphere	Situations where the conversation progresses in a calm, relaxed atmosphere, naturally leading to laughter. Lighthearted exchanges and conversations with jokes fall into this category.	2955 (79.0%)
(4)	Self-Disclosure and Friendliness	Situations where sharing personal stories or past mistakes creates a sense of approachability and triggers laughter. Self-disclosure that makes the other person feel at ease is also included.	475 (12.7%)
(5)	Cultural Background and Shared Understanding	Laughter based on specific cultural backgrounds or shared understandings. This includes jokes related to a particular region or culture or remarks based on common superstitions or folklore.	176 (4.7%)
(6)	Nostalgia and Fondness	Situations where past memories or nostalgic topics trigger laughter. This includes shared past experiences and the enjoyment of recalling familiar events.	204 (5.5%)
(7)	Self-Deprecating Humor	Situations where turning one's flaws or mistakes into humor makes the other person feel closer and triggers laughter. Slightly putting oneself down can also give the other person a sense of ease.	404 (10.8%)
(8)	Defying Expectations	Situations where intentionally defying the flow of conversation or the other person's expectations creates an element of surprise and triggers laughter. This includes unexpected responses or developments.	323 (8.6%)
(9)	Positive Energy	Situations where positive emotions or energy in the conversation bring a smile to the other person. Enjoyable topics and positive comments fall into this category.	338 (9.0%)
(10)	Exaggeration	Situations where exaggerating things gives a comical impression and triggers laughter. Exaggerated expressions can be amusing to the listener.	478 (12.8%)

LLM's Performance on Laughable Context Recognition

- We evaluated GPT-4o's ability to recognize laughable contexts using a zero-shot setting with a chain-of-thought (CoT) reasoning approach.
- GPT-4o achieved an F1 score of 43.14%, significantly above the chance level.

- 👍 40-50% Accuracy Primary labels from (1) to (3)
- 🔴 Higher accuracy (5) Cultural Background and Shared Understanding, (7) Self-Deprecating Humor (8) Defying Expectations
- 🔴 Lower accuracy (6) Nostalgia and Fondness (9) Positive Energy

	Label	Laughable (Correct 🟡)	Non-laughable (Incorrect 🔴)
(1)	Empathy and Affinity	1226 (40.7%)	1787 (59.3%)
(2)	Humor and Surprise	1571 (48.6%)	1662 (51.4%)
(3)	Relaxed Atmosphere	1257 (42.5%)	1698 (47.5%)
(4)	Self-Disclosure and Friendliness	232 (48.8%)	243 (51.2%)
(5)	Cultural Background & Shared ...	102 (57.9%)	74 (42.1%)
(6)	Nostalgia and Fondness	62 (30.4%)	142 (69.6%)
(7)	Self-Deprecating Humor	255 (63.1%)	149 (36.9%)
(8)	Defying Expectations	227 (70.3%)	96 (29.7%)
(9)	Positive Energy	50 (14.8%)	288 (85.2%)
(10)	Exaggeration	239 (50.0%)	239 (50.0%)