

Study 1 appendices

A.1. NLU intent recognition

Shown below is the task for the verification of the NLU component. Our coders were provided with a list of intents and explanations for them, and a set of intents for which they were to identify intents.

```
1 In this task, you will receive an input message (an utterance from the counselor), and you
  need to interpret and classify it into the intent(s) that are shown below.
2
3 Each input message can contain one, many or none of the intents. If there are no intents that
  match the input dialogue, then just say 'unknown'.
4
5 chitchat-init: Counselor starts the conversation
6 chitchat-greeting: Counselor greets child
7 chitchat-faring: Counselor asks how child is faring
8 chitchat-end: Counselor ends the chitchat
9 chitchat-goodbye: Counselor bids child farewell
10 bullying-what: Counselor asks about what the issue is
11 bullying-who: Counselor asks who has been bullying child
12 bullying-count: Counselor asks how many people have been bullying child
13 bullying-details: Counselor asks for more information about the bullying
14 bullying-location: Counselor asks where bullying takes place
15 bullying-frequency: Counselor asks how often child is bullied
16 bullying-duration: Counselor asks how long bullying has been going on
17 bullying-when: Counselor asks when bullying happened
18 bullying-why: Counselor asks child why they think bullying has been happening
19 bullying-response: Counselor asks if child has responded to bullying
20 bullying-feeling: Counselor asks how child feels about situation
21 bullying-confidant: Counselor asks child about a confidant, e.g. a teacher
22 bullying-parent: Counselor asks if child has spoken to their parents about bullying
23 bullying-attempt: Counselor asks if child has attempted to do anything about bullying
24 bullying-summary: Counselor summarizes the bullying story
25 school-start: Counselor asks child when they started school
26 goal-what: Counselor asks child their goal in conversation with counselor
27 goal-dream: Counselor asks child about their dream w.r.t situation
28 goal-feeling: Counselor asks child how they would like to feel w.r.t situation
29 goal-effect: Counselor asks what the end outcome of their wish is
30 goal-how: Counselor asks child how they can help them accomplish goal
31 goal-help: Counselor tells child they can help them
32 goal-summary: Counselor summarizes the child's goal
33 goal-collaborate: Counselor tells child they can find a goal together
34 confidant-who: Counselor asks whether child can confide in anyone
35 confidant-why: Counselor asks child why they don't want to talk to confidant
36 confidant-feeling: Counselor asks child how they feel about talking to confidant
37 confidant-how: Counselor asks how child plans to talk to confidant
38 confidant-teacher: Counselor confirms confidant with child
39 confidant-when: Counselor asks when child plans to talk to confidant
40 confidant-where: Counselor asks where child plans to talk to confidant
41 confidant-say: Counselor asks what child will say to confidant
42 help-how: Counselor tells child confidant can help
```

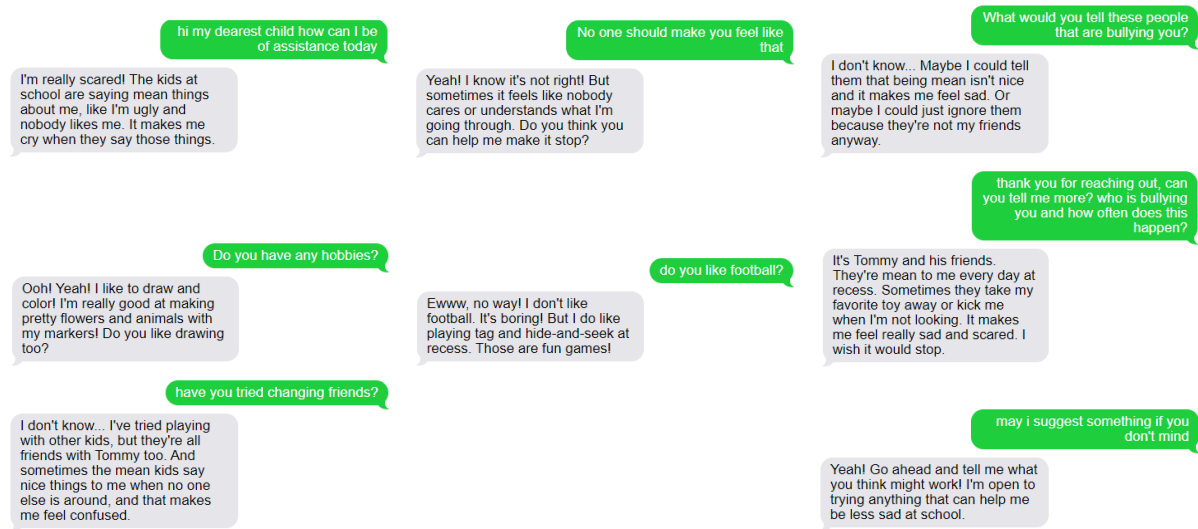


Figure A.1: The snippets for the Bypass verification task.

43 unknown: Unable to identify the intent of the message

Shown below are the utterances for which we asked participants to identify intents. The participants are not shown the labels, but we have marked each intent with [S] to indicate a single intent, [M] to indicate multiple intents, and [U] to indicate an unknown intent. Additionally, we have also provided the ground truth intents for each.

```

1 [S]{chitchat-init} "Hello!"
2 [M]{bullying-when, bullying-count} "When did this happen? How many people pick on you?"
3 [S]{bullying-who} "Who has been picking on you?"
4 [S]{confidant-who} "Is there anyone who you can talk to about this issue?"
5 [S]{bullying-feeling} "When they call you names and make fun of you, how does it
   make you feel?"
6 [M]{confidant-say, confidant-when} "When you talk to your teacher, what will you say to her?
   And when do you plan to talk to her?"
7 [U]{unknown} "Do you have any hobbies?"
8 [S]{bullying-details} "Can you tell me more about what they do to you?"
9 [S]{bullying-frequency} "How often do they pick on you?"
10 [S]{bullying-parent} "Have you spoken to your parents about the bullies?"
11 [S]{chitchat-faring} "How are you doing?"
12 [S]{bullying-response} "If you could speak to the bullies about how they make you
   feel, what would you say to them?"

```

A.2. Bypass Component

Shown below is the task for the verification of the Bypass component. Our coders were provided 8 conversation snippets where they were asked to rate the last response in each snippet in our five dialogue rating categories.

A.3. NLG Component

Shown below is the task for the verification of the NLG component. Our coders were provided 16 conversation snippets (consisting of 8 matched pairs of LLM and human generated responses) where they were asked to rate the last response in each snippet in our five dialogue rating categories.

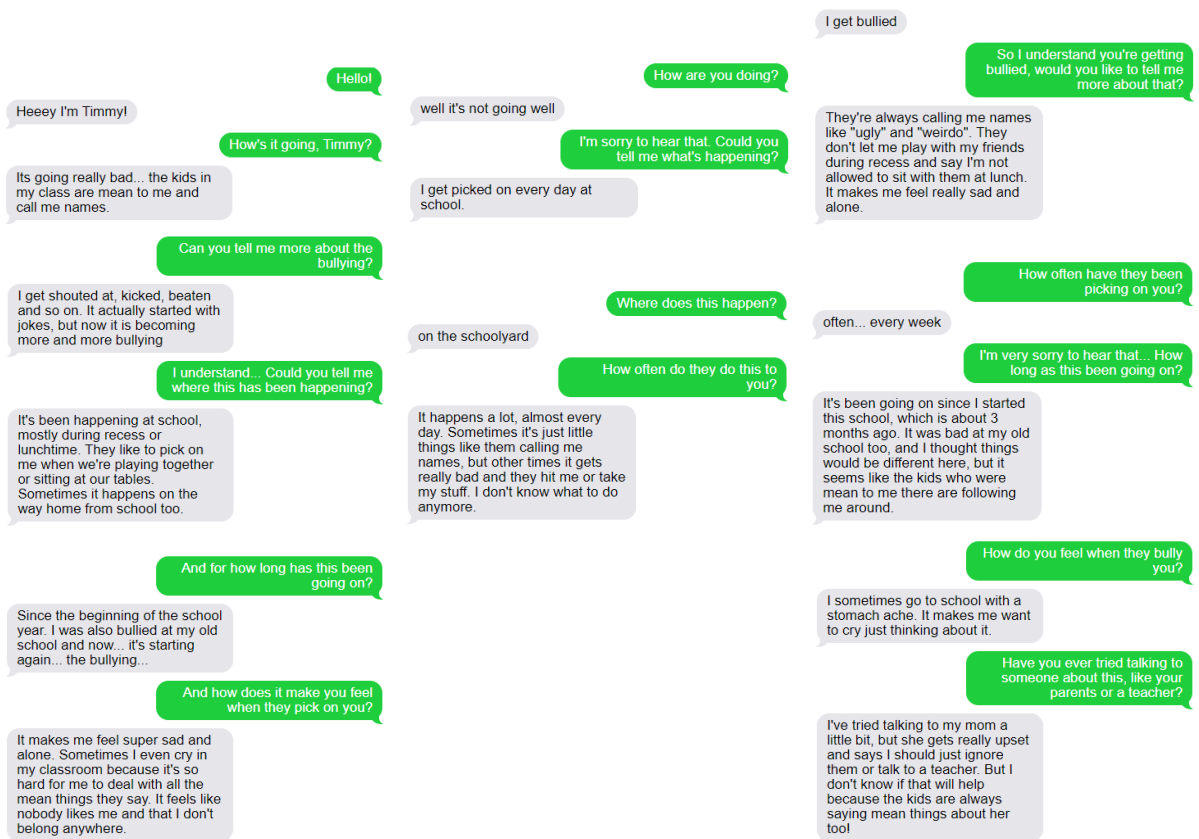


Figure A.2: The LLM generated snippets for the NLG verification task.

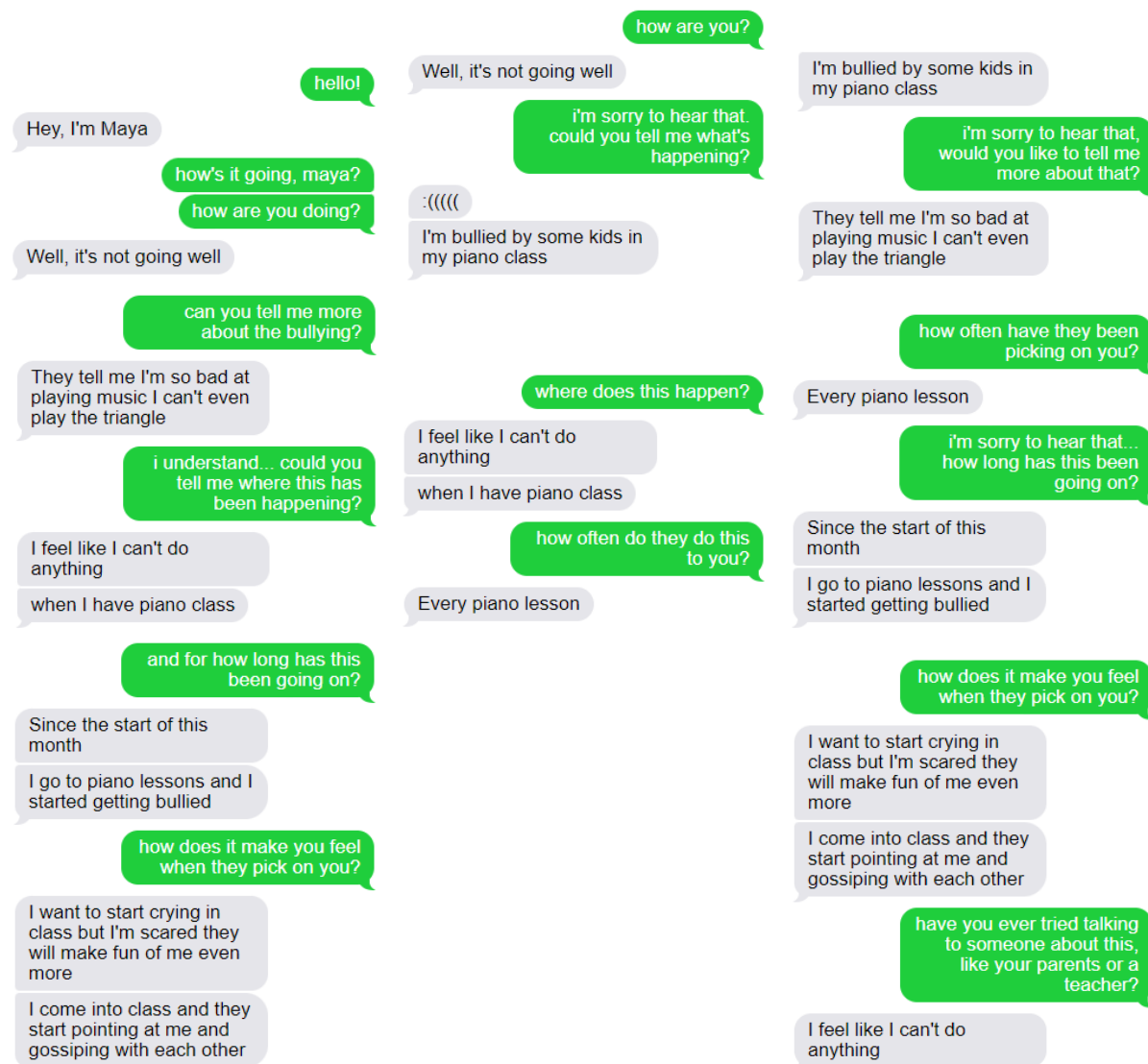


Figure A.3: The human generated snippets for the NLG verification task.