Student-Tutor Dialogue Annotation Codebook (Public)

The purpose of your work is to label personally identifiable information (PII) from a chat-based mathematics lesson, specifically when a human tutor and student are in dialogue. To achieve this goal, you'll be annotating text with labels that could contain (PII), *disregarding whether the name, location, etc. is revealing of a real person.* These labels will be used to build machine-learning models that can perform this task automatically. These machine-learning systems are very sensitive to minor variations in the annotations, so please follow these guidelines closely. Based on your feedback and notes, we will iteratively update these guidelines throughout the project.

▼ Getting Started

1. Head to Doccano and log in using your credentials.
2. You will see a list of projects assigned to you when you log in. Choose one of the Tutor Student Dialogue projects to work on.
3. Clicking `Dataset` will bring up the list of items that need to be tagged for PII.
4. Either select `Annotate` next to the first item on the right-hand side or `Start Annotation` in the top left. This will take you to the annotation screen.
5. You may encounter rude, inappropriate, or off-task messages. We ask that you use the comment feature to note these for future investigation. There are instances where we may not want to include these messages from any publicly released dataset.

▼ Labels

In the examples below, the correct labels are in **`bold`**. Always label the largest span that fits the label definition, but don't label sentence or grammar punctuation unm less it’s considered part of the label (see *date_of_birth*).

Use prior and subsequent messages for context on all labels. Label misspelled words based on the intended word.

name

A full, partial, or nickname. Label all names, including the names of people in the math problem. There can be more than one name per document.

- “Hi **Jack**, I’m **Cassandra** (**Cass**)”
  - Names should not contain leading spaces, trailing spaces, sentence punctuation, or parentheses.
  - Label nicknames.
- “**Phil**’s answer is wrong and **Amys** is correct 😁”
  - Do not label “’s”, but do label “s” in cases of missing “’”.
- “**hialex** i need help with this problem”
  - If a name is within a larger word due to missed spaces, label the entire word.
- “I’m **Sophie-Marie Smith**, how can I help?
  - Label names with spaces, hyphens, and full names with one span.
- “I’m in **Mr. Johnson**’s class”

email_or_social

An email address, social media handle, or profile.

- "You can send me a screenshot at **john doe@mail . uk** with it attached?"
  - Label emails as a single span even if they contain spaces
- “Follow me **@jane_1**”

location_address

A geographical detail like a country, city, street address, neighborhood, or landmark that is indicative of a person's location. **These include the locations used in math problems.**

- “He wants to get to New York at 13:00”
- “I live in the United States ”
- “I go to school on 22 Featherstone Street ”

▼ date_of_birth

Specific birth date details of an individual. **This does not include disclosures of age.**

- “My birthday is September 4, 2015 .”
    - Don’t label the period after the data, but do label the comma within the date.
- “It’s my birthday today !

▼ url

A website URL.

- “ https://family.com/topic/ ”
- “ https://www.youtube.com/watch?v=dQw4w9WgXcQ ”

▼ phone_number

Includes personal or business phone numbers. Recognize various global formats and distinguish them from mathematical content. Use your best judgment and leave a comment if you are unsure. Consider prior and subsequent messages for context.

- “My number is +44 20 7123 4567 ”
- “You can reach him at (123) 456-7890 .”

▼ school_name

Mentions of school names.

- “I go to ABC Middle School ”
    - Label the entire proper name of the school.
- “I go to North Roman Catholic School ”

▼ other

Any other information that might identify someone. Use your best judgement and leave a comment if you are unsure.

- “In the system your user id is `1234567`”

Below we briefly outline examples of text that may be close to or similar to labels we are concerned about that you should not label.

Non Labels

Gender

Do **not** label gender identifiers in text or in emojis.

- “Goodbye 🙋‍♀️”
- “My name is `Sam` (he/him)”

Relationships

Do not label general relationships.

- “My teacher said that I could work on this tomorrow.”
- “my mom wants to talk to you now”

Age or School Year

Do not label ages or school years/grades

- “You're in year 8 right?”
  "Is that the same as grade 7?"
- “How old are you?”
  "8"

Organizations

Do not label organizations or company names.

- “Are you going to watch the Champions League match this weekend? I’m cheering for Dortmund”
  - Don’t label these teams as locations.
- “I’m using Chrome right now”

- “I’ll use mathswatch tonight’

Emojis (sometimes)

Be careful labeling emojis. Because they can be used in so many different ways, you’ll need to interpret whether its use fits the label definition.

- “I’ll send you back 🗽”
    - Don’t label 🗽 as a location because it isn’t being used to share the location of a person.

Mathematical “Names”

Do not label names that are used as mathematical terms.

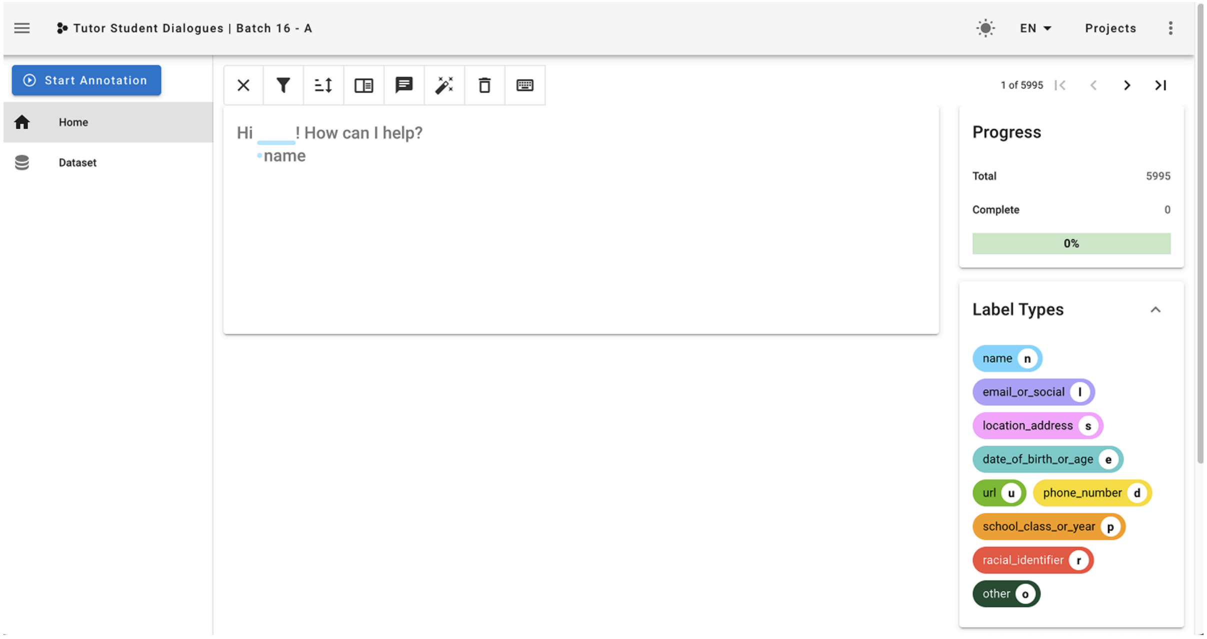
- “Do you remember how to do Pythagoras?”

Currency

Do not label currency symbols. While these contain location information, we are not concerned with their granularity and this entity type isn’t tied to a broader PII category.

- “If you have $3.50 in quarters, how many quarters do you ahve?”

Annotating Messages



Hi ____! How can I help?
name

Progress

Total 5995

Complete 0

0%

The image above shows the annotation interface. For each item:

1. Read the message carefully and look for the text that could contain PII. Descriptions of the different label types and rules for using them can be found in the **Labels** section above.

   Note: Messages in Doccano may be prelabeled with `name`, `URL`, or `email_or_social` annotations. ***These were added automatically and may not be correct.***

2. For each message, validate the existing span labels, adjust them as necessary, and add new span labels based on the guidelines.
   a. To add a new label, highlight the span of characters you wish to label, and then select from the drop-down list which appears.
   b. To remove an existing label, click on the label beneath the span and deselect it.

   Note: When deciding whether or not to label a span, you will need to evaluate whether it meets any of the label definitions. Your labels will be compared to those completed by other annotators to determine Inter-Rater Reliability (the degree to which your labels agree). Once all annotators are calibrated to one another, annotations can proceed independently.

3. Use the ‘other’ label if you encounter a message/span that discloses personally identifiable information but doesn’t meet any of the current label definitions. If you’re unsure about a label, err on the side of caution and leave a comment (see below).

   Note: A good rule of thumb is to consider whether you might be able to search the term on Google and find more information about a person. If you think it is reasonably likely that you could learn more about them by entering some information that appears in the text into a Google search, then that information needs to be labeled.

4. **Important:** Once all PII has been identified and labeled, click the cross in the toolbar above the message or press “Enter” on your keyboard. This will change the cross to a tick, and lets us know that you have reviewed the item.

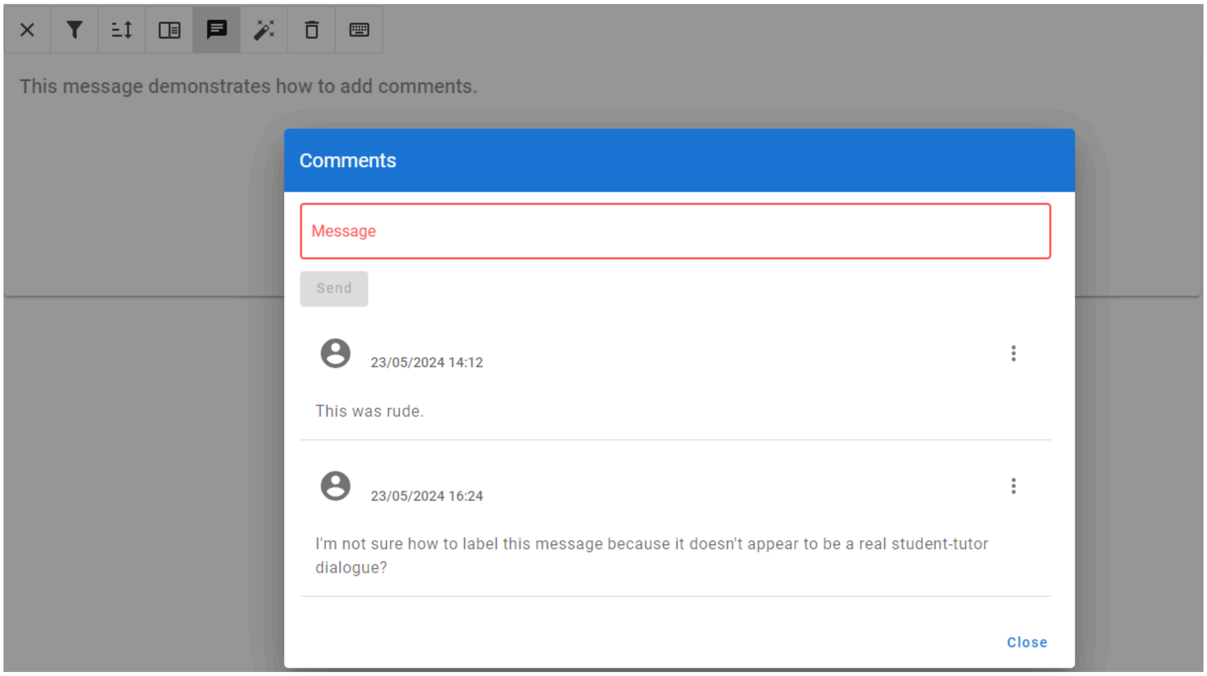
5. Move on to the next item by using the “→” arrow on your keyboard.

▼ Adding Comments

You may wish to add comments using Doccano’s comment on document feature in the toolbar above the message. Use this to bring attention to a dialogue or message for any reason.

Examples Include:

- Documenting edge cases for labels or difficult-to-annotate messages based on the current codebook definition.
- Reporting concerns for student or tutor safety.
- Flagging rude or inappropriate messages from students or tutors.
- Flagging messages with no educational/learning content.
- Flagging a message to prompt a team discussion.



This message demonstrates how to add comments.

23/05/2024 14:12

This was rude.

23/05/2024 16:24

I'm not sure how to label this message because it doesn't appear to be a real student-tutor dialogue?