AD Label Studio Annotation Guidelines

Annette Rios

June 2025

1 Interface

Audio descriptions are presented as groups of three consecutive segments. This format allows annotators to identify errors that impact coherence and cohesion, which are often not apparent when evaluating individual segments in isolation. All three samples in a group have the same source, i.e. are produced by the same model(s).

There are two possible ways to do the annotations, both are fine, but the interface looks a bit different: You can directly clicking on any sample in the data manager, see Fig. 1 (1). In this annotation view, you will get a side panel with a list of all annotation tasks, see (1) in Fig. 2. You will have to manually click on the next task in the list to move forward (hitting *Submit* will not jump to the next task in the list).

Alternatively, you can click on *Label All Tasks* in the data manager, see (2) in Fig. 1. In this interface, you won't see the list of tasks, instead you jump to the next task once you've hit the *Submit* button. When you come back after a break, you can hit *Label All Tasks* again, and it will take you back to where you left off.

Figure 2 highlights the different elements in the annotation interface that can be used for navigation:

1. On the left side, all tasks are outlined, you can click on any of those to open them in the annotation interface, e.g. if you want to go back and look at a particular annotation. Otherwise, the interface automatically moves to the next task when you hit *Submit*.

2. You can manually jump to any frame in the clip by entering the frame number here.

3. Underneath the video, the frame spans for each AD are outlined. This is just a visual help, the buttons are not clickable.

4. For easier navigation in the video, use the *Regions* window on the left. There is a clickable button for each AD segment that will move the video to the first frame of that audio description. Note that there is a bug: when clicking on AD 1, the video jumps to the correct starting frame, but the

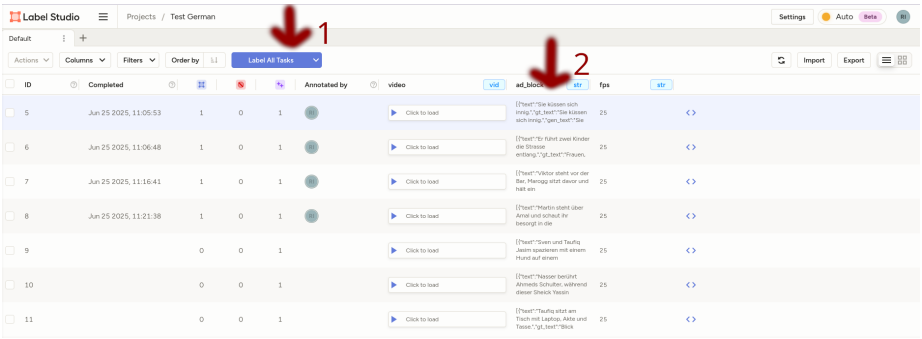

Figure 1: Data manager view. Select a sample here (2) to start annotating in order (or continue where you left off). Alternatively, click on "Label all tasks" (1) to annotate in a randomized order.

frame number and slider underneath the video will not update (for AD 2 and 3, everything works fine). It's a bit confusing, but the navigation works.

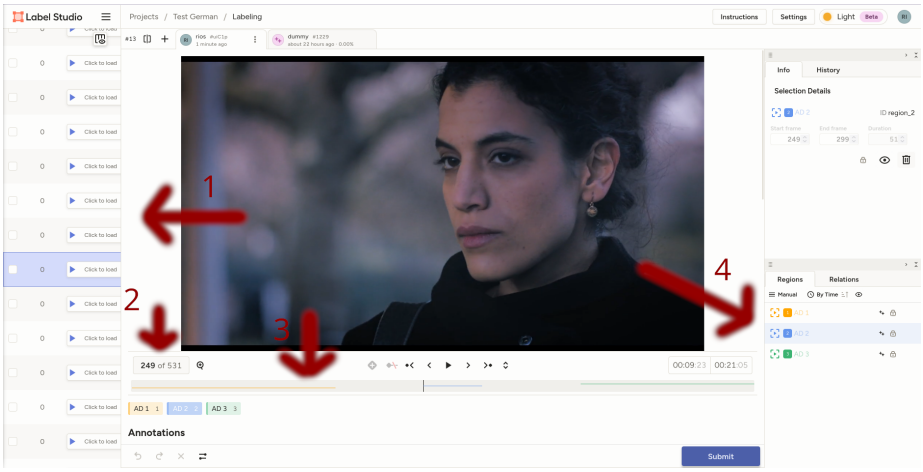

Figure 2: Navigation interface:
1. outline of tasks (groups of segments)
2. go to specific frame number
3. span for each AD
4. clickable buttons (will take you to the first frame of that AD)

2 Character List

To help with judging character recognition and use of correct names/pronouns, you can check this list of characters for German and Italian with headshots of the actors. The list is not complete, as the information has been compiled from IMDB and not all movies have the full set of characters. The main characters should be present however (this is the exact same list the models had to generate the audio descriptions).

3 Annotation

The error tags are divided into these categories:

1. content
2. grammar
3. coherence
4. characters

3.1 Content errors

irrelevant: The audio description contains information that is not relevant for the viewer, e.g. a description of a background object that is not needed for the story telling.

missing: The audio description is missing crucial information that is needed to follow the plot, or, the audio description is too vague to be useful, e.g. *she holds an object*.

redundant: The audio description has redundant information, e.g. *the phone is ringing*, but the ringing is obvious from the audio.

subjective and/or patronizing: The audio description is not neutral, e.g. *the kitchen is a mess* instead of *dirty dishes are stacked on the counter*.

wrong action: The audio description has an action that is wrong, e.g. *a man walks down the street* but in the video, the character is running.

wrong object: The audio description has an object that was not correctly identified, e.g. *he holds a book* but in the video, the character is holding a phone.

missing text: Relevant information from text in the clip is missing in the audio description, e.g., a character is reading a note, the note is shown on screen, but the AD does not mention the text content. Another example is text overlay, e.g., credits or information such as "5 years earlier" to introduce a flashback scene.

other inaccuracy: The audio description has wrong information of another type that is not covered by the given categories (not an action, an object, or character that is wrong).

3.2 Grammar

ungrammatical/not fluent: The audio description contains grammatical errors or awkward phrasing, e.g. *Gregi betrachtet den Kuchen, Wanda ihn.*

too long/complex: The description is too long and/or complex, e.g., with nested sentences.

wrong tense: The audio description is not in present tense, e.g. *Lisa sat down* instead of *Lisa sits down.*

English wording/phrasing (Denglisch/Itanglese): The audio description contains phrases or words that are literal translations from English, e.g. *Sie betrachtet sich im Spiegel und justiert ihr Haar*, or *Lei si rompe giù in lacrime* ("*she breaks down in tears*").

other AD standards violation: The description violates AD standards or guidelines in a way that is not covered by the other tags.

3.3 Coherence

contextual gap There is an information gap to the preceding description that leads to confusion, in other words, there is a logical jump that leads to confusion. Example:

AD 1: *Paul and Lisa are talking to each other in front of the car.*

AD 2: *Paul closes the door and leans back.*

Paul getting in the car was not mentioned.

name repeated The same name is repeated from the previous segment, instead of using a pronoun:

AD 1 *Paul sits down at the table.*

AD 2 *Paul starts eating.*

content repeated Content is repeated from the previous segment:

AD 1 *Paul is talking on the phone, walking down the street.*

AD 2 *Paul talks on the phone.*

makes no sense/unrelated content: The description makes no sense with respect to the clip and/or has completely unrelated content.

other incoherence: The audio description is not properly "connected", but none of the above categories match.

3.4 Characters

wrong character The audio description contains a character name that does not appear in the clip.

wrong pronoun The audio description uses the wrong pronoun (e.g. *she/her* instead *he/him*).

redundant (first and last name) The audio description contains the correct character name, but uses first and last name in context where the last name is not needed and sounds unnatural.

missing name The audio description refers to an already introduced character with a description, or uses a pronoun when the name is needed to avoid confusion with another character.

bad description The audio description describes a character by physical attributes, but it's either wrong (*a blonde man*, but the person has dark hair) or inappropriate or too vague (*a person in a shirt*) to be useful.

misattributed action The characters are recognized correctly, but the action has confused subject/object. For example, the scene is *A looks at B*, but the AD is *B looks at A*. Or, the scene is *A and B are walking through a park, B is smoking* but the AD is *A and B are walking through a park, A is smoking.*

3.5 Quality rating

Not all errors affect the overall quality to the same extent, e.g. an audio description with redundant information might be better than a description with wrong content. Also, the AD might have wrong, but also correct content, so not completely bad. The rating is an overall measure of how useful a segment is.

3.6 Comments and other errors

Selecting this “tag” will open a text field for comments, explanations, additional observations etc.