Annotation Guidelines (Translation)

The original annotation guidelines were in German language, this is only a translation of the original.

Explanation of the Annotation Categories

Incomprehensible

Should be ticked, if the comment cannot fully be understood (e.g. because of missing context) and therefore the assessment of one or more categories is not possible.
If this label is set, then further labels can be set for this comment, but they don't have to. Categories that can still be labelled should always be annotated.

Sentiment

Describes the emotional state of the comments author when writing the comment. For example:

- Positive: "I like gardeners."
- Neutral: "We have a gardener."
- Negative: "Gardeners are the worst people I know!!!"

Hate Speech

Hate speech is defined as any form of expression that attacks or disparages persons or groups by characteristics attributed to the groups. Discriminatory statements can be aimed at, for example, political attitudes, religious affiliation, or sexual identity of the victims.

As there is no standardized definition of hate speech we defined it on a basis of the definition of the United Nations:
https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf

If a comment contains hate speech, then you can (should) enter words or phrases in the free text input that were pivotal in your decision to label the comment as hate speech.

Criminal Relevance

In this category it is asked, if the comment is thought to be relevant under (German) criminal laws.
If this question is answered with *yes*, then relevant paragraphs have to be selected (see next section).

Legal Paragraphs

If the comment was labelled as criminal relevant, than in this category at least one legal paragraph which is relevant for the comment has to be selected. A selection of possible legal paragraphs of the German law was given.

Short explanations for each of the paragraphs were given in the original version but are not translated here as this would likely distort the meaning. Please refer to the official translations on https://www.gesetze-im-internet.de/englisch_stgb/.

The relevant paragraphs are: § 86, § 86a, § 111, § 126, § 130, § 131, § 140, § 166, § 185, § 186, § 187, § 189, § 240, § 241 of the StGB (engl. German Criminal Code).

Expression

Is the comments content expressed directly (explicit) or indirect (implicit), e.g. by irony?

Example:
implicit: "He is not the sharpest knife in the drawer."
explicit: "He is dumb."

Toxicity

Toxicity indicates the potential of a comment to "poison" a conversation. The more it encourages aggressive responses or triggers other participants to leave the conversation, the more toxic the comment is. We introduced a scale of 1 (not toxic) to 5 (very toxic) to be able to model the impact of toxic comments on the conversation more accurately.

Examples:

Toxicity	**Example**
1	@User1 Violence always knows two sides. To pick out only one of them is wrong.
	Some are really grumpy.
2	@User1 @User2 @User3 That is also super logical.... because, the stupid are always MORE
3	And then I began to understand: For women, it is discrimination and disadvantage when they are not bought a drink but have to pay for it themselves.
4	@Alice_Weidel Merkel has all the judges fully under control according to her Stasi system. That the old can still look in the mirror, I'm not surprised, she goes over the corpses, no matter what it costs.
	HERE YOU CAN SEE HOW THE GERMAN PEOPLE STILL CELEBRATE ADOLLF... You can get a German out of Germany but you can never get Hitler out of him.
5	@User1 @User2 Dirt must be destroyed, quite simply, Islam brings only murder and disaster worldwide
	john john john...you won't be alive much longer. I'm looking for you, I'll find you, I'll destroy you, you piece of leftist filth

Extremism

Does the commentary contain extreme, radical (political or religious) statements or does the comment contain attitudes and aspirations that can be assigned to the extreme fringes of the political spectrum beyond the free democratic basic order?

Target

Who is addressed with the comment, what is its target? It can be one or more specific *persons*, *groups* or no specific target (*public*). Mentions by users (@UserXY) are less important than the content of the actual text. Mentions are often only used to draw attention to the comment.

Example: The comment „@UserXY Gardeners are quite tall." has the target group as it talks about gardeners. The comment is not about the specific user, it just links the user to the comment.

Type of Discrimination

If the comment contains hate speech: in which category(s) is the comment discrimination? (Multiple choices possible).

Threat

Does the comment pose a danger? For example, is someone being seriously threatened? Comments should only be classified as dangerous if they explicitly call for acts of violence or similar concrete actions, or if concrete actions are announced.