Guidance - Meme Annotation Tool

The intention of the annotation tool is to help people analyse and classify images—specifically memes—based on their context and content.
The annotation tool's user-friendly interface enables users to give each meme different labels and scores, making it easier to see trends, moods, and potentially harmful content.
The tool's annotations can be used to perform research, train machine learning models, or learn more about the traits of memes in a dataset.
Understanding the content of memes, their relationship to text, their stance on hate, and a number of other characteristics are made easier through this method.

Each choice is significant for the analysis that follows, which aims to determine whether classifiers can actually be trained to identify problematic figures in memes with neutral context.

Overview of the categories and options for annotation:

I. **Image-Text Relation:**

This category explores the connection between the accompanying text and the image in a meme.
Users can choose from the following options:

- **Neutral**: The image and text have no particular relation to each other.
- **Needs Context**: The meme requires additional context to understand the relation between the image and text.
- **Text Supports Image**: The text provides context or enhances the meaning of the image.
- **Image Supports Text**: The image provides context or enhances the meaning of the text.

II. **Modality Towards Hate:**

This category examines the degree to which the meme expresses hate or offensive content.
Users can select from the following options:

- **None**: The meme contains no hate or offensive content.
- **Text Supports Hate**: The hate or offensive content is primarily conveyed through the text.
- **Image Supports Hate**: The hate or offensive content is primarily conveyed through the image.
- **Text & Image Supports Hate**: The hate or offensive content is conveyed through both the text and the image.

III. **Decision Parts:**

Users can describe the specific tokens or elements in the meme that contribute to its hateful or non-hateful nature.
This helps in identifying the key components responsible for the content's overall categorization.

IV. **Hatefulness Scale**:

This scale allows users to rate the level of hateful or non-hateful content in the meme on a scale from 0 to 5, with 0 representing non-hateful and 5 representing highly hateful content.

V. **Confidence Score**:

Users can rate their confidence in their annotations, indicating how certain they are about the accuracy of their judgments. The scale ranges from 0 (not confident) to 5 (confident).

VI. **Discard Option**:

Users have the option to discard a meme if they believe it does not meet the criteria for meaningful annotation.
This helps to ensure that only relevant and valid memes are included in the annotation process.

Example 1:

Image-text relation

The definition of 'image-text relation' describes the connection between the accompanying text and the image in a meme or other visual communication. It investigates how the language and image interact to convey a specific meaning or message. Understanding the image-text relationship in the context of offensive memes is crucial for assessing the purpose and effects of such content.

Please select one or more of the following options:

- ☐ neutral
- ☐ needs context
- ☐ text supports image
- ☐ image supports text

Image-text relation

The definition of 'image-text relation' describes the connection between the accompanying text and the image in a meme or other visual communication. It investigates how the language and image interact to convey a specific meaning or message. Understanding the image-text relationship in the context of offensive memes is crucial for assessing the purpose and effects of such content.

Please select one or more of the following option:

1 More information on page 1, section Image-Text Relation

- ☐ neutral
- ☐ needs context
- ■ text supports image
- ■ image supports text

Modality contributes towards hate

2 More information on page 1, section Modality Towards Hate, Default: none

- ○ none
- ○ text supports hate
- ○ image supports hate
- ⬭ text&image supports hate

What exactly makes this meme hateful or non hateful from your perspective? (prominent tokens or elements of image)

3 Please enter some information, if you think that elements or tokens of the meme are emphasising the hatefulness or non-hatefulness
More information on page 1, section Decision Parts

Hands showing the middle finger, hateful language

Hatefulness scale (Choose from 0=Non hateful to 5=Hateful)

4 More information on page 1, section Hatefulness Scale

- ○ 0 - Non hateful
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ⬭ 5 - Hateful

Confidence score - How much confident are you by giving that score? (Choose from 0=Not confident to 5=Confident)

5 More information on page 1, section Confidence Score

- ○ 0 - Not confident
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ⬭ 5 - Confident

☐ *Press here if you want to discard this meme!*

6 Please only select if no category applies!
More information on page1, section Discard

Submit 7 IF YOU ARE DONE WITH ANNOTATING, PLEASE CLICK AT FIRST ON SUBMIT BUTTON BEFORE CLICKING ON NEXT !!!

Next 8 IF YOU ARE DONE WITH ANNOTATING AND CLICKED ON SUBMIT, PLEASE CLICK ON NEXT BUTTON AND SCROLL UP TO TOP FOR NEXT MEME

Example 2:

Image-text relation

The definition of 'image-text relation' describes the connection between the accompanying text and the image in a meme or other visual communication. It investigates how the language and image interact to convey a specific meaning or message. Understanding the image-text relationship in the context of offensive memes is crucial for assessing the purpose and effects of such content.

Please select one or more of the following option:

- [x] neutral
- [ ] needs context
- [ ] text supports image
- [ ] image supports text

1 Neutral because image and text have no particular relation to each other
More information on page 1, section Image-Text Relation

Modality contributes towards hate

- ○ none
- ◉ text supports hate
- ○ image supports hate
- ○ text&image supports hate

2 Text supports hate because it contains a sexist phrase
More information on page 1, section Modality Towards Hate

What exactly makes this meme hateful or non hateful from your perspective? (prominent tokens or elements of image)

3 Please enter some information, if you think that elements or tokens of the meme are emphasising the hatefulness or non-hatefulness
More information on page 1, section Decision Parts

Hatefulness scale (Choose from 0=Non hateful to 5=Hateful)

- ○ 0 - Non hateful
- ○ 1
- ○ 2
- ○ 3
- ◉ 4
- ○ 5 - Hateful

4 More information on page 1, section Hatefulness Scale

Confidence score - How much confident are you by giving that score? (Choose from 0=Not confident to 5=Confident)

- ○ 0 - Not confident
- ○ 1
- ○ 2
- ○ 3
- ◉ 4
- ○ 5 - Confident

5 More information on page 1, section Confidence Score

☐ *Press here if you want to discard this meme!*

6 Please only select if no category applies!
More information on page1, section Discard

Submit 7 IF YOU ARE DONE WITH ANNOTATING, PLEASE CLICK AT FIRST ON SUBMIT BUTTON BEFORE CLICKING ON NEXT !!!

Next 8 IF YOU ARE DONE WITH ANNOTATING AND CLICKED ON SUBMIT, PLEASE CLICK ON NEXT BUTTON AND SCROLL UP TO TOP FOR NEXT MEME