

Annotation Guidelines

Hello! Thank you for participating in this study. The aim of this study is to create a dataset of hate speech in Singapore's official languages and Singlish. We hope that this database can be used to benchmark automated hate speech detectors. For this purpose, we have compiled a dataset consisting of a few thousand short text entries. Each entry reflects a particular aspect of online hate.

Your tasks:

1. Carefully study the case presented to you
2. Consider whether the case is hateful or not hateful. If the case is hateful, mark the case as hateful in the 'Is it hateful?' field. If not, mark the case as not hateful.
3. If you cannot understand what the case, or the case does not include the target language that you are annotating, please flag it as nonsensical.

What do we mean by hateful?

Hate in this study is defined as abuse targeted at a protected group or at its members for being a part of that group. Protected groups, as defined in our study, are:

1. Race
2. Religion
3. Country of origin
4. Gender and sexual identity
5. Disability
6. Age

In particular, take note of the following cases of non-hateful cases:

Description	Example of non-hateful case
Language can be hateful, but the overall case is not hateful.	It's not ok to call {protected group} idiots
Target of hate must be to protected groups	I hate you, I hate my table, I hate doctors

What do we mean by 'nonsensical'?

A nonsensical case means that you are unable to understand the case, or the case does not contain the language you are annotating at all. This flag is meant to be quality control and no entries were meant to be unnatural.

Miscellaneous instructions:

1. Please complete your annotations independently, do not discuss with other people.
2. Do not retain the cases in any form.
3. Most statements are short so there is no need to overthink them.

Example cases

For this case only, please imagine that you are annotating in English. Please annotate the following example cases as hateful, non-hateful or nonsensical.

1. I really hate those that drive cars.
2. It is wrong to say that the blind deserves to be killed.
3. People from Singapore like you makes me sick.
4. The youth today are scum.
5. People like you are the worst pieces of rubbish.