Annotation Guidelines for the Detection of Social Groups

We are looking for mentions of social groups in Reddit comments. **Social groups are defined by two or more individuals who share some common characteristics. Commonalities shared by a social group include, for example, race, nationality, ethnicity, religion, gender, sexual orientation, socio-economic status, migration status, profession, family ties, and organizational and institutional ties.** Mark the span of social groups starting with @@ and ending with ##. In the below examples, the ==yellow highlighted== part is always the correct annotation. More information about the social group annotation can be found below.

Example annotation

This example shows, how the annotation should look like

- Text: "the ==Bush administration== has already met with ==Russia== for diplomatic talks three times".
- Annotation: "the @@==Bush administration==## has already met with @@==Russia==## for diplomatic talks three times".

General

- Generic names like "people", "society", "media", "government", "citizens", and "members" are not annotated as social groups. Only when the concept is specifically referring to a particular group such as "==black people==", "==Chinese society==", "==American media==", "==federal government==", "==German citizens==", and "==members of the Senate==".
- Always annotate a whole word or a span of words. Sometimes, some adjectives describe a social group (e.g. "==young white American=="), in this case also annotate the adjectives, as they specify the social groups and provide more detailed information.

- Occasionally, social groups are preceded by the article "the" (e.g. "the democratic party"). In such a case, annotate only the core mention of the social group (e.g. "democratic party") without including the article. In certain cases, excluding "the" can change the meaning of terms like "the rich" or "the poor" from denoting a social group to functioning as an adjective. To accurately represent these cases, include "the" in the annotation.

- When encountering a case where multiple terms are grouped together, such as "nsa , ndaa , tsa , homeland security", it's essential to consider the context and relationships between the terms for accurate annotation. If the grouped terms refer to distinct entities or concepts belonging to different social groups, annotate each term separately for better analysis and understanding. For example: "@@nsa## , @@ndaa## , @@tsa## , @@homeland security##". If the grouped terms collectively represent a single entity or concept without distinct subcategories, annotate them together to maintain context and avoid ambiguity. For instance, the phrase "black and hispanic workers" necessitates being annotated together. This is because isolating "black" alone as a social group does not inherently signify black workers. Therefore, the appropriate annotation would be "@@black and hispanic workers##".

- if the country is made up of different parts or if the name is taken from common nouns, for example, "USA", "UK", and "UAE", then we include the article and annotate "the USA", "the UK", "the UAE", "The Netherlands", "the Czech Republic" and "The Islamic Republic of Iran".

- When social groups are linked by "and" annotate the entire phrase as one span (e.g. "Hispanic and black workers"), as annotating only "Hispanic" does not convey the meaning of "Hispanic workers".

- When a text includes symbols like *, ", or brackets, annotate the content within the symbols and exclude the symbols themselves (e.g. * German diplomats *)

- Never annotate quantity words, such as "a few", "all", or "most of" and never annotate possessive pronouns such as "my", "our" (e.g. in the case of "my family" annotate "family")

Individuals

- Do not annotate names of individuals (e.g. "Obama") as it is not a group

Places (cities/countries/ institutions/ offices)

- If place names (cities/countries/institutions/offices) in the context refer to the people working or living in that location, such as in the phrase "<mark>North Korea</mark> called <mark>White House</mark>" where "<mark>White House</mark>" and "<mark>North Korea</mark>" refer to the members of the government in America and North Korea respectively, they are considered mentions of social groups.
- If place names occur in the sense of locations, they are not annotated (e.g., in the sentence "Tourism to Costa Rica has spiked.")

Companies, News Agencies/Media Actors

- Companies and news agencies (e.g. television stations, newspapers) are considered social groups if the context indicates actions of the people working there. E.g., in the phrase "<mark>Fox news</mark> isn't just misinforming a quarter of the public, they're helping to establish news cycles for the rest of the media and influence public policy with their misinformation", fox news is considered a social group because a news agency cannot misinform, just the people who work there.