Annotation Guidelines for the Detection of Social Group Constructions

We are looking for mentions of social group constructions in Reddit comments. **Social group constructions are defined as two or more individuals who share common characteristics or beliefs. Commonalities a social group construction shares can include but are not limited to race, nationality, ethnicity, religion, gender, sexual orientation, socio-economic status, migration status, profession, family ties, organizational and institutional ties, common issue stances, or political views.**

Mark the text span referring to social group constructions starting with @@ and ending with ##. In the below examples, the ==yellow highlighted== part is always the correct annotation. More information about the annotation of social group constructions 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 an identifiable 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 construction (e.g. "==young white Americans=="), in this case, also annotate the adjectives, as they specify the social group constructions and provide more detailed information.
- Occasionally, mentions of social group constructions are preceded by the article "the" (e.g. "the ==democratic party=="). In such a case, annotate only the core mention of the social group construction (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.
- 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 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 group constructions, 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==##".
- When encountering nested cases such as "@@==US government officials working for Russia==##", both "US government officials" and "Russia" could be considered a social group construction. However, due to the sentence construction, it is clear that "working for Russia" is a description of the "US government officials". Therefore, in such cases, the bigger span is annotated, not separate instances.

- When a text includes symbols like *, ", or brackets, annotate the content within the symbols and exclude the symbols themselves (e.g. * German diplomats *). If a comma appears in the social group mention, annotate it only if it is part of the social group span (e.g., annotate "@@Black, Hispanic and Chinese workers##" but annotate @@Blacks## , @@Hispanics## and @@Chinese people## separately)
- 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 "North Korea called White House" where "White House" and "North Korea" refer to the members of the government in America and North Korea respectively, they are considered mentions of social group constructions.
- 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 group constructions if the context indicates the actions of the people working there. E.g.,@@ in the phrase "Fox News 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 construction because a news agency cannot misinform, just the people who work there.

More Example annotations

1- there are @@people on reddit that are n't from america## . we 're not all us .

2- this is really not a question anymore . the only question is whether anything will be done about it . the more interesting question is if it can be proven that trump was involved in criminal financial activity with companies tied to the @@russian government## .

3- very excited . he tore up the dnc plants at the cnn townhall like so much newspaper , doubt @@fox## will be able to do much better against him . edit : tune in he is tearing it uppp
absolutely false . clinton was impeached in @@the house## , but @@the senate## did not remove him from office . that came from @@the house## . your statement is 100 % incorrect .

4- LOOOOOL. Bruv, you trollin right? I'll bite the bullet, did @@uefa## pay @@Madrid## comeback against @@City## ?