This annotation task is designed to understand whether or not a set of words or phrases can reasonably be referred to as a personal identifier, which we define as *a word or phrase which, when read, can* ***be used to make a direct assumption about a social identity*** *(or set of social identities) held by an individual* ***without any additional context, and where it is likely that the user intends to present that inferred identity****.* Social identities are, in turn, defined as *words or phrases that are used to refer to a particular social group, category, or role*.

During the annotation task, you will see approximately [XXX] expressions extracted from Twitter bios. For each, you will be asked the following:

**Does this expression contain a personal identifier?**

* **Yes**
* **No, none**
* **Yes, but there is potentially more than one identity listed here (>1)** - For example, the phrase “actor/athlete” contains two words that are normally used to define two different identities, actor and athlete
* **Unclear** - Resort to this option when you are unsure of whether or not a phrase contains an identifier or not

**Details**

Personal identifiers fall into one of five categories:

1. **Social identities** - The most obvious case, a word or phrase that would make sense if it directly followed the statement “I am a”.
2. **Preferences**- Things that might follow ``I like ...'' or ``Fan of ...''. Examples include sports teams and television shows. Note this should reveal something that is not a universal statement e.g. “into everything” and should not need context (e.g. “and everything in between” … in between what?). Explicitly discounted here are references to other social media sites, because we would expect these (and these almost always are) followed by links; as such the term “facebook” is not an identity statement (as in “I like facebook”) but a description of a link.
3. **Personal Descriptors** -adjectives to describe oneself, e.g. “kind”, “old”. as in ``I am kind, smart, and old.'' Again, you should be able to have a reasonably specific idea of the identity here. One concrete example of a descriptor that is **not** a personal identifier is “experienced” … experienced in what? This needs additional context.
4. **Affiliations** - Indirect references to affiliations with social groups or movements. For example, hashtags like “#maga'' and “#blm”, and also @mentions of accounts of things like universities “@umich” and businesses @walmart
5. **Actions** - Many actions are **not** identifiers, because they are so general that almost anyone could do them - e.g. smiling, hugging are **not** identifiers. However, many actions convey specific identities - e.g. hooping -> basketball player, or running -> runner.

**Additional important notes**

* Any observed phrase in our labeling is a delimited expression extracted from a Twitter bio. Thus, things like “caring”, while perhaps not a personal identifier in other context, should be considered to be one here, because it can reasonably be inferred that this individual is a “caring person”. On the other hand, the phrase “be kind to others” would **not** be a personal identifier, because it is simply telling other people to be kind. Same for “for the kids,” because even though we know this is a person expressing it, we do not know which kids (additional context needed)
* We will assume that the users in this dataset are people. Thus, things like show and company names should be labeled as personal identifiers, because it can reasonably be inferred that an identity of this individual is as a fan of this show or as an employee of this company. Similarly, diseases listed can be used to reasonably infer that this person is affected by this particular disease