Data Collection

Trace Selection Error

4.1 How was the data associated with each instance acquired? On what basis were the trace selection criteria chosen? The instances collected for this dataset are the tweets collected through the historic search API containing the keyphrase "call me sexist, but". The rationale behind this choice of query was that several Twitter users opine potentially sexist comments and signal so using the presence of this phrase, which arguably serves as a disclaimer for sexist opinions.

4.2 Was there any data that could not be adequately collected? For the parts of the dataset that stem from existing datasets used in the literature, only fractions of these datasets could be recovered, as many of their instances were deleted or removed from Twitter after having been posted and were thus unavailable for recollection.

4.3 Is any information missing from individual instances? Could there be a systematic bias? The keyphrase "call me sexist, but" was removed from all the tweets collected for the dataset to avoid its priming effect on annotators, who have been shown to be more likely to consider a tweet sexist if it included the keyphrase. However, since this was done for all of the tweets in the dataset, there should be no systematic bias arising from this decision.

4.4 Does the dataset include sensitive or confidential information? The Twitter dataset was pseudonymized by replacing any mentions (@username) with a placeholder (MENTION). For the adversarial examples written by the MTurkers, their IDs were also pseudonymized. Mentions of family names, identified via regular expressions and a NER model, were manually confirmed to actually be of family names and then shortened to only the initial letter using a regular expression (e.g., John Doe to John D.).
Even though the dataset might still contain tweets with sensitive or confidential contents, the described procedures prevent these contents from being easily associated with the corresponding individuals.

User Selection Error

The questions aiming at issues related to the "User Selection Error" do not apply to this dataset, as the instances do not represent individuals. Furthermore, the collection and sampling of the data happened on the trace (tweet) level, not the user (Twitter user) level.