

The Gab Hate Corpus

367.8MB

Public

0

...

Contributors: [Brendan Kennedy](#), [Mohammad Atari](#), [Aida Mostafazadeh Davani](#), [Leigh Yeh](#), [Ali Omrani](#), [Yehsong Kim](#), [Kris Coombs](#), [Gwenyth Portillo-Wightman](#), [Shreya Havalдар](#), [Elaine Gonzalez](#), [Joseph Hoover](#), [Aida Azatian](#), [Gabriel Cardenas](#), [Alyzeh Hussain](#), [Austin Lara](#), [Adam Omary](#), [Christina Park](#), [Xin Wang](#), [Clarisa Wijaya](#), [Yong Zhang](#), [Beth Meyerowitz](#), [Morteza Dehghani](#)

Date created: 2018-07-18 02:29 PM | Last Updated: 2020-04-05 08:03 PM

Identifier: DOI 10.17605/OSF.IO/EDUA3

Category:  Project

Description: The growing prominence of online hate speech is a threat to a safe and just society. This endangering phenomenon requires collaboration across the sciences in order to generate evidence-based knowledge of, and policies for, the dissemination of hatred in online spaces. To foster such collaborations, here we present the Gab Hate Corpus (GHC), consisting of 27,665 posts from the social network service [gab.ai](#), each annotated by a minimum of three trained annotators. Annotators were trained to label posts according to a coding typology derived from a synthesis of hate speech definitions across legal, computational, psychological, and sociological research.

We detail the development of the corpus, describe the resulting distributions of hate-based rhetoric, target group, and rhetorical framing labels, and establish baseline classification performance for each using standard natural language processing methods.

The GHC, which is the largest theoretically-justified, annotated corpus of hate speech to date, provides opportunities for training and evaluating hate speech classifiers and for scientific inquiries into the linguistic and network components of hate speech.

License: CC-BY Attribution 4.0 International 








Has supplemental materials for [The Gab Hate Corpus: A collection of 27k posts annotated for hate speech](#) on [PsyArXiv](#)

Files

Download as zip

Filter

i

Name	Modified
 The Gab Hate Corpus	
 OSF Storage (United States)	
 Code	
 Data	
 GabHateCorpus_annotatio...	2020-03-03 05:2...
 GabHateCorpus_full_predi...	2020-02-24 08:3...
 Documentation	

Citation

Tags


Hate Speech


Natural Language Processing


Text Analysis


Text Annotation


Recent Activity


 [Brendan Kennedy](#) created external identifier(s) doi:10.17605/OSF.IO/EDUA3 on [The Gab Hate Corpus](#) 2020-04-05 08:03 PM

 [Brendan Kennedy](#) reordered contributors for [The Gab Hate Corpus](#) 2020-04-05 08:03 PM

 [Brendan Kennedy](#) added Gabriel Cardenas as contributor(s) to [The Gab Hate Corpus](#) 2020-04-05 08:02 PM

 [Brendan Kennedy](#) removed Gabriel Cardenas as contributor(s) from [The Gab Hate Corpus](#) 2020-04-05 08:00 PM

 [Brendan Kennedy](#) reordered contributors for [The Gab Hate Corpus](#) 2020-04-05 07:56 PM

 [Brendan Kennedy](#) added Beth Meyerowitz as contributor(s) to [The Gab Hate Corpus](#) 2020-04-05 07:56 PM

<

1

2

3

4

...

10

>

This website relies on cookies to help provide a better user experience. By clicking Accept or continuing to use the site, you agree. For more information, see our [Privacy Policy](#) and information on [cookie use](#).

Accept