What Did You Just Call Me? Computers vs. Abusive Language

Zeerak Talat
University of Sheffield

Sapienza University of Rome
01/12/2017

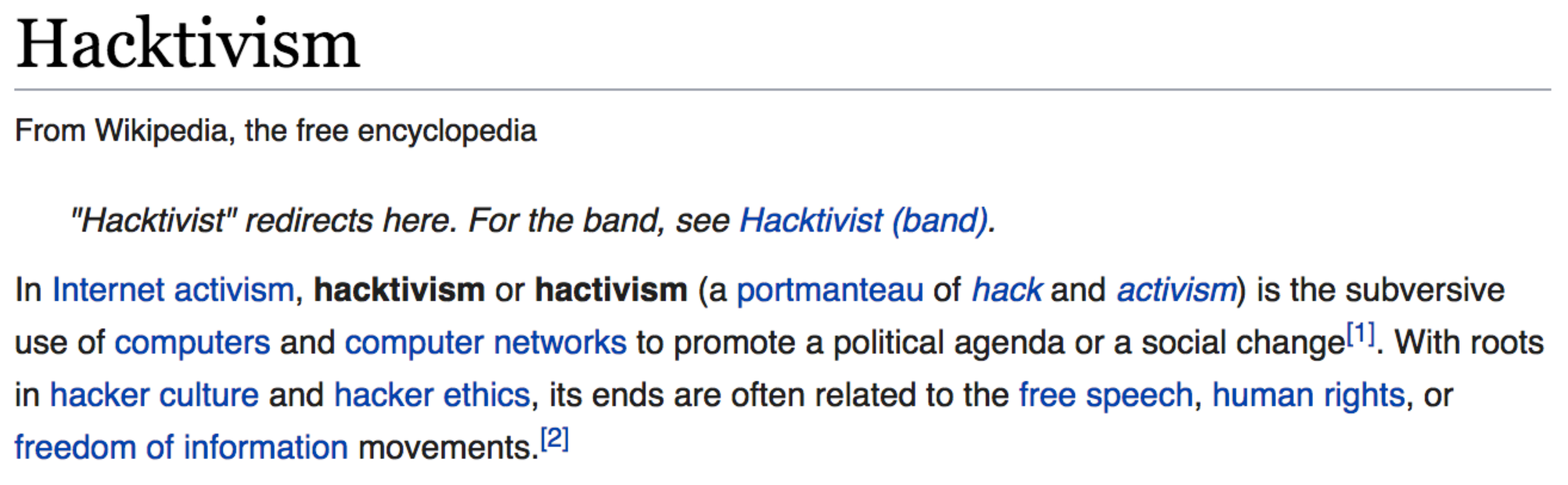
Why Should We Care?

Mental Health
Stigmatisation

Social Cohesion
Democratic inclusion
Integration
Political Depolarisation
Preventing Stigmatisation

Hate Crime

Research == Activism?



Hacktivism

From Wikipedia, the free encyclopedia

"Hacktivist" redirects here. For the band, see Hacktivist (band).

In Internet activism, **hacktivism** or **hactivism** (a portmanteau of *hack* and *activism*) is the subversive use of computers and computer networks to promote a political agenda or a social change[1]. With roots in hacker culture and hacker ethics, its ends are often related to the free speech, human rights, or freedom of information movements.[2]

Research == Activism?

Kimberly Crenshaw

How do we understand “hate speech”

Common Understanding

- Slurs
- Threats
- Violence (sometimes)
- Vandalism (sometimes)
- Stereotyping (sometimes)
- Disparaging speech (sometimes)

Legal Understanding

- Disparaging speech
- Violence (sometimes)
- Threats (sometimes)
- Vandalism

Some more understandings

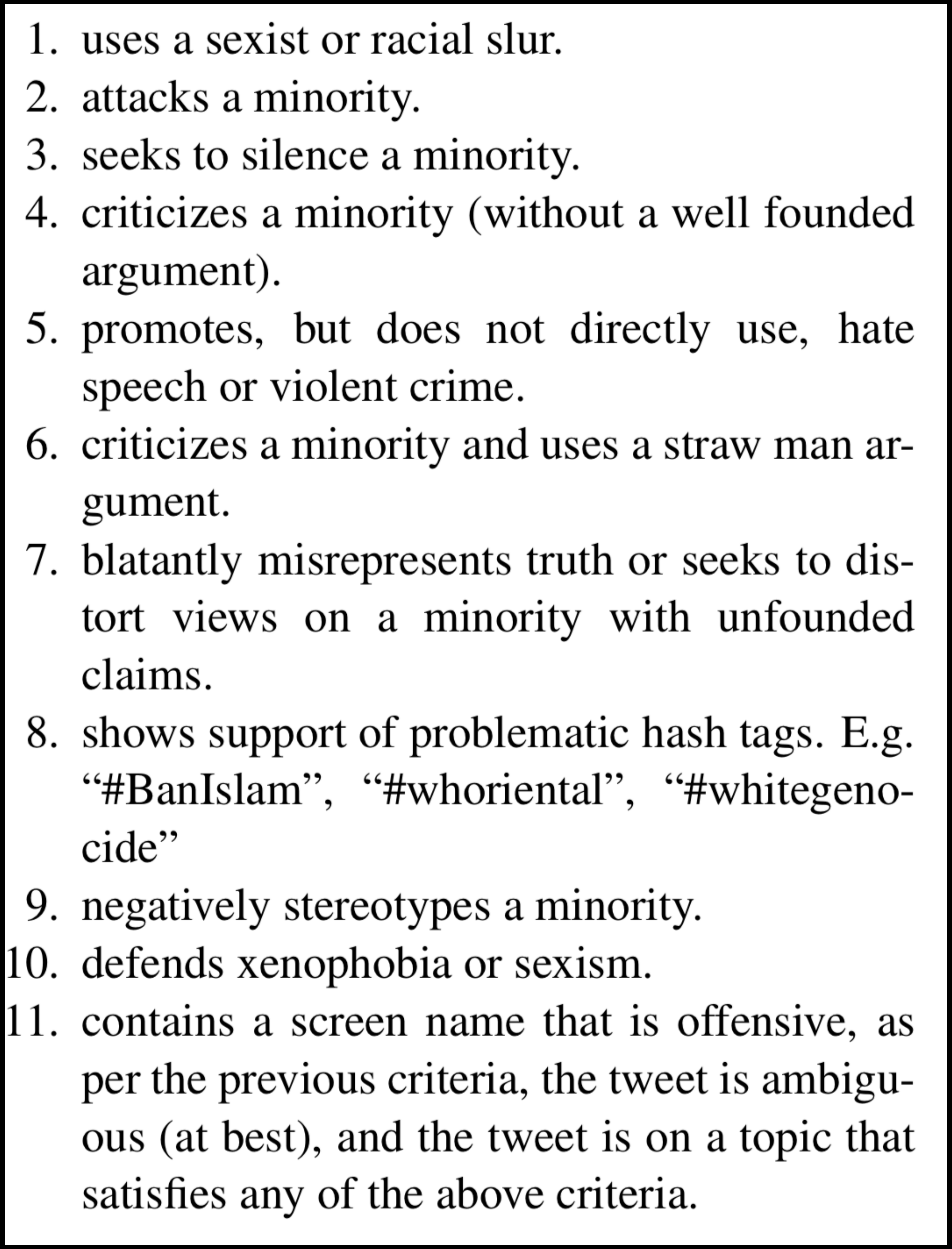
Academic Understanding

- Slurs
- Stereotyping
- Disparaging speech
- Ridicule
- Minimising
- Undermining
- Lack of representation
- Violence
- Threats
- Vandalism

Social Media Co's Understanding

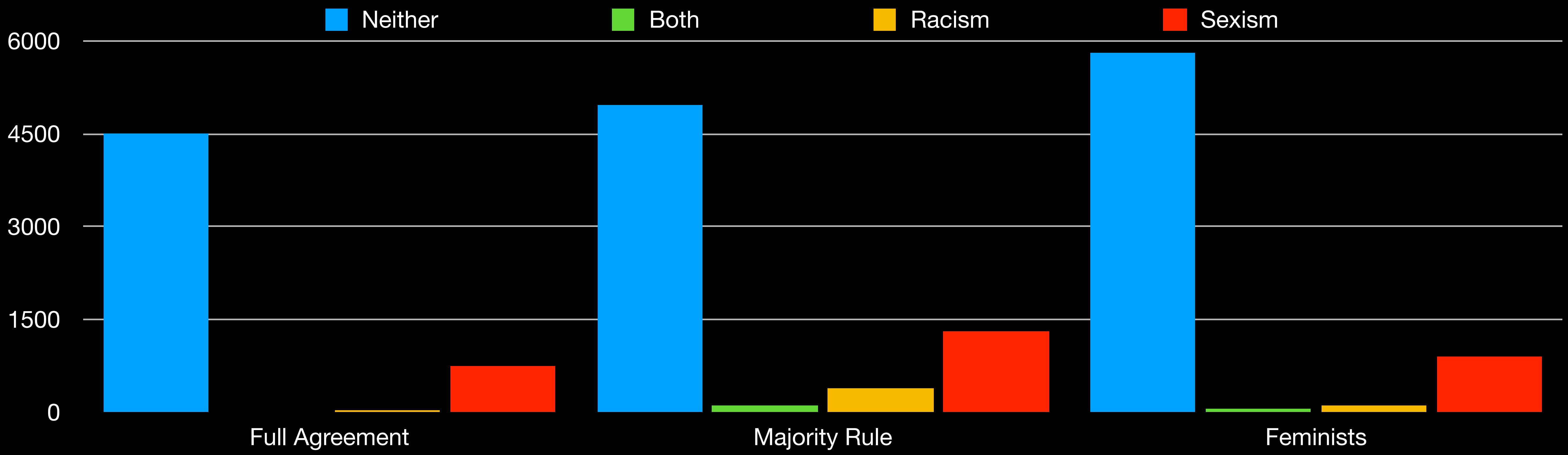
- Slurs (sometimes)
- Disparaging speech (sometimes)
- Threats (sometimes)
- Violence (sometimes)

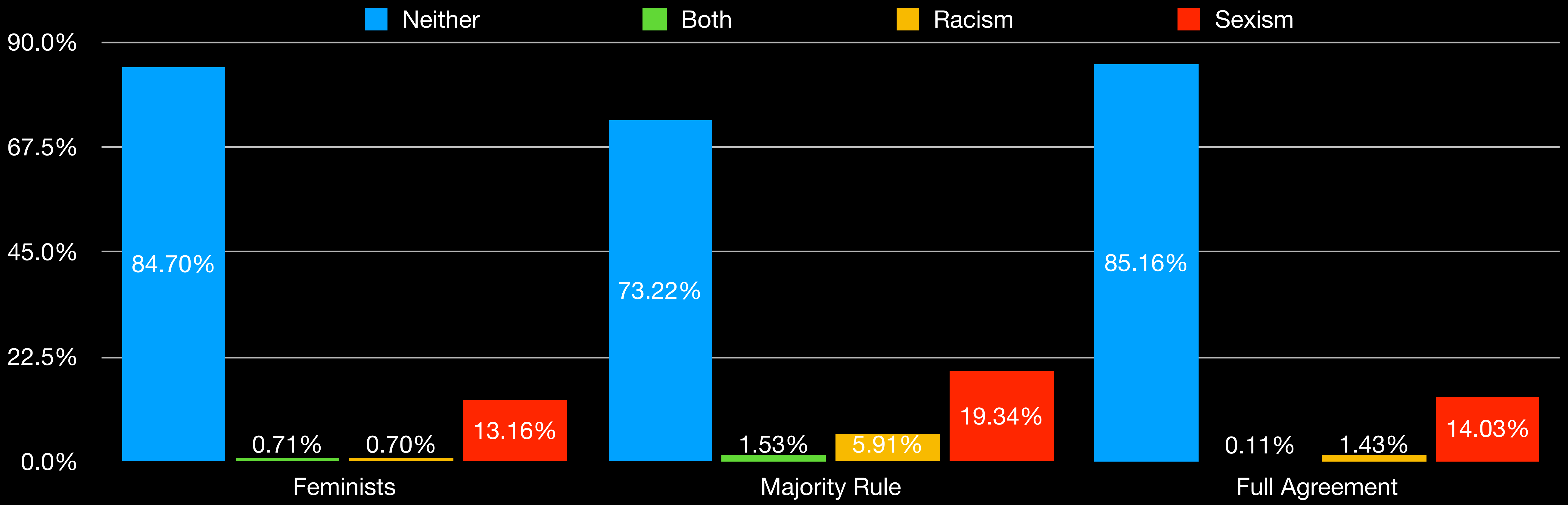
Annotation Guidelines



1. uses a sexist or racial slur.
2. attacks a minority.
3. seeks to silence a minority.
4. criticizes a minority (without a well founded argument).
5. promotes, but does not directly use, hate speech or violent crime.
6. criticizes a minority and uses a straw man argument.
7. blatantly misrepresents truth or seeks to distort views on a minority with unfounded claims.
8. shows support of problematic hash tags. E.g. “#BanIslam”, “#whoriental”, “#whitegenocide”
9. negatively stereotypes a minority.
10. defends xenophobia or sexism.
11. contains a screen name that is offensive, as per the previous criteria, the tweet is ambiguous (at best), and the tweet is on a topic that satisfies any of the above criteria.

Annotation


Neither
Both
Racism
Sexism
6000
4500
3000
1500
0
Full Agreement
Majority Rule
Feminists


Neither
Both
Racism
Sexism
90.0%
67.5%
45.0%
22.5%
0.0%
84.70%
0.71%
0.70%
13.16%
73.22%
1.53%
5.91%
19.34%
85.16%
0.11%
1.43%
14.03%
Feminists
Majority Rule
Full Agreement

Annotation

	Majority v. Feminists	Full Agreement v. Feminists	All CF Annotators
Cohen's kappa	0.34	0.70	0.57
Krippendorf's alpha	0.32	0.70	

Classification

F1-Scores

	Amateur (Majority Vote)	Expert
Character n-*gram*	86.41	91.24
Token n-*gram*	86.37	91.55
3-Skip-grams	86.27	91.53
Binary Gender	76.64	77.77
GenderProbability	86.37	81.30
Brown Clusters	84.50	87.74
AHST	71.71	55.40

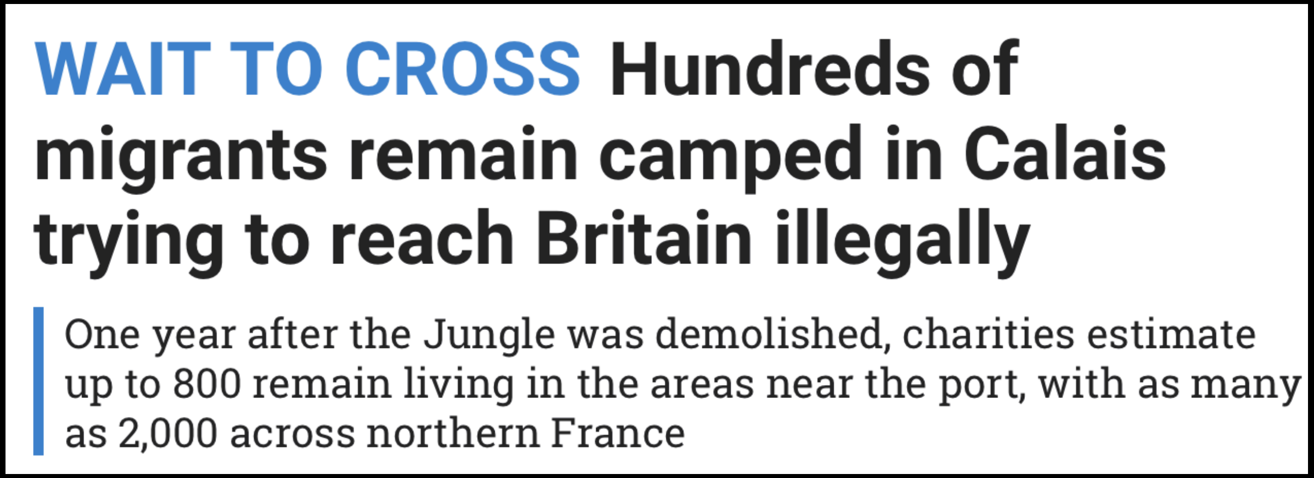
Typology

	Explicit	Implicit
Directed	“@User shut yo beaner ass up sp*c and hop your f*ggot ass back across the border little n*gga”	“(((@User))) and what is your job? Writing cuck articles and slurping Google balls? #Dumbgoogles”
Generalized	“So an 11 year old n*gger girl killed herself over my tweets? ^ ^ thats another n*gger off the streets!!”	“Totally fed up with the way this country has turned into a haven for terrorists. Send them all back home.”

Some suggested features

	Explicit	Implicit
Directed	Mentions, bad-words dictionary, POS, NER	bad-words dictionary, Semi-supervised methods for finding euphemisms, mentions, word-embeddings
Generalized	Lexical features, named, demographies, bad-words dictionary	Semi-supervised methods for building euphemisms, lexical features, named demographies, word-embeddings

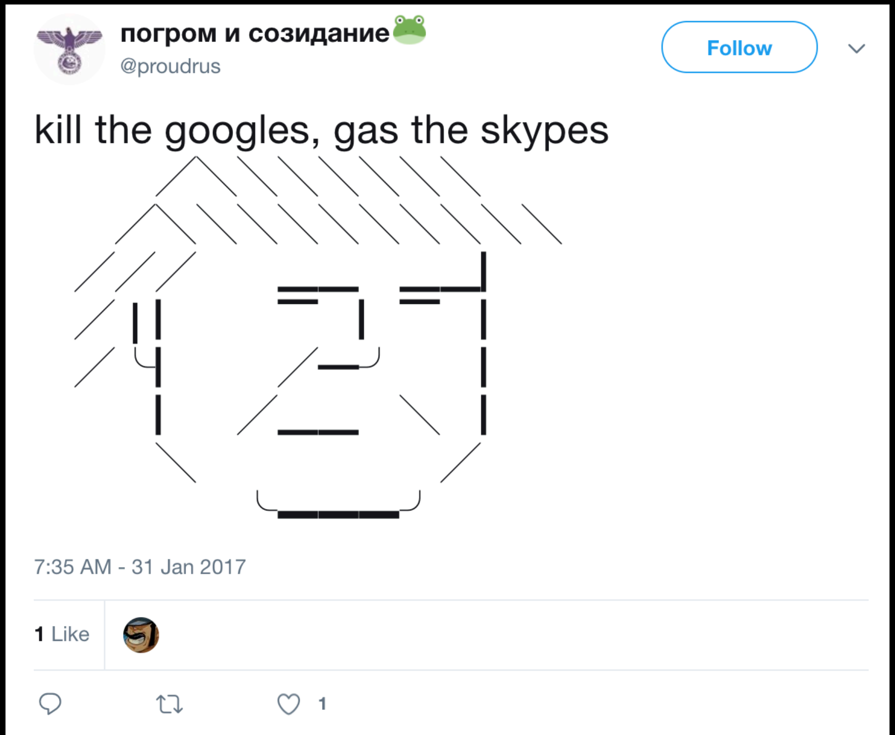
Current Work



WAIT TO CROSS Hundreds of migrants remain camped in Calais trying to reach Britain illegally

> One year after the Jungle was demolished, charities estimate up to 800 remain living in the areas near the port, with as many as 2,000 across northern France

Futures

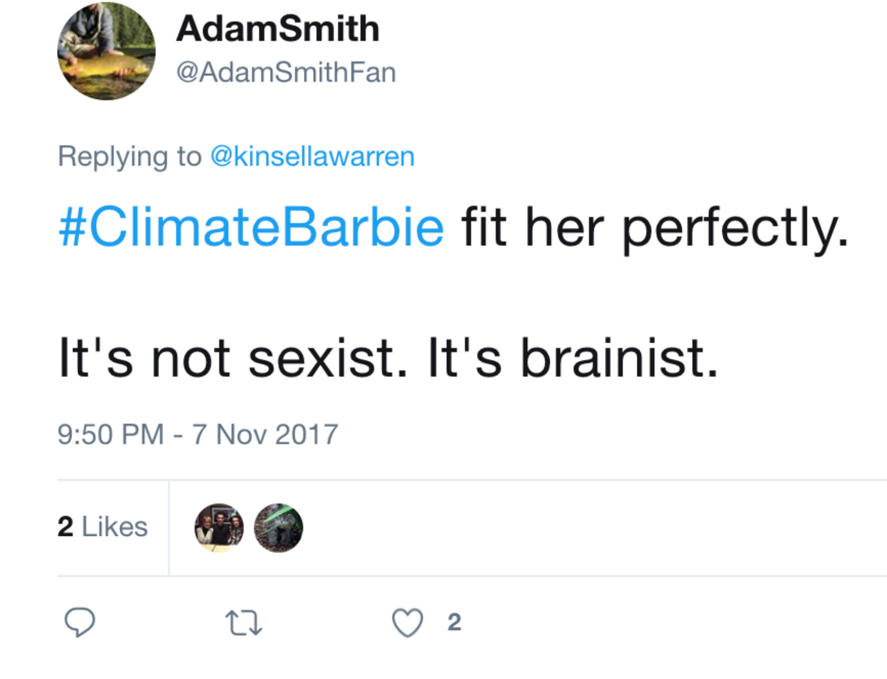


погром и созидание🐸
@proudrus

kill the googles, gas the skypes

7:35 AM - 31 Jan 2017

1 Like



AdamSmith
@AdamSmithFan

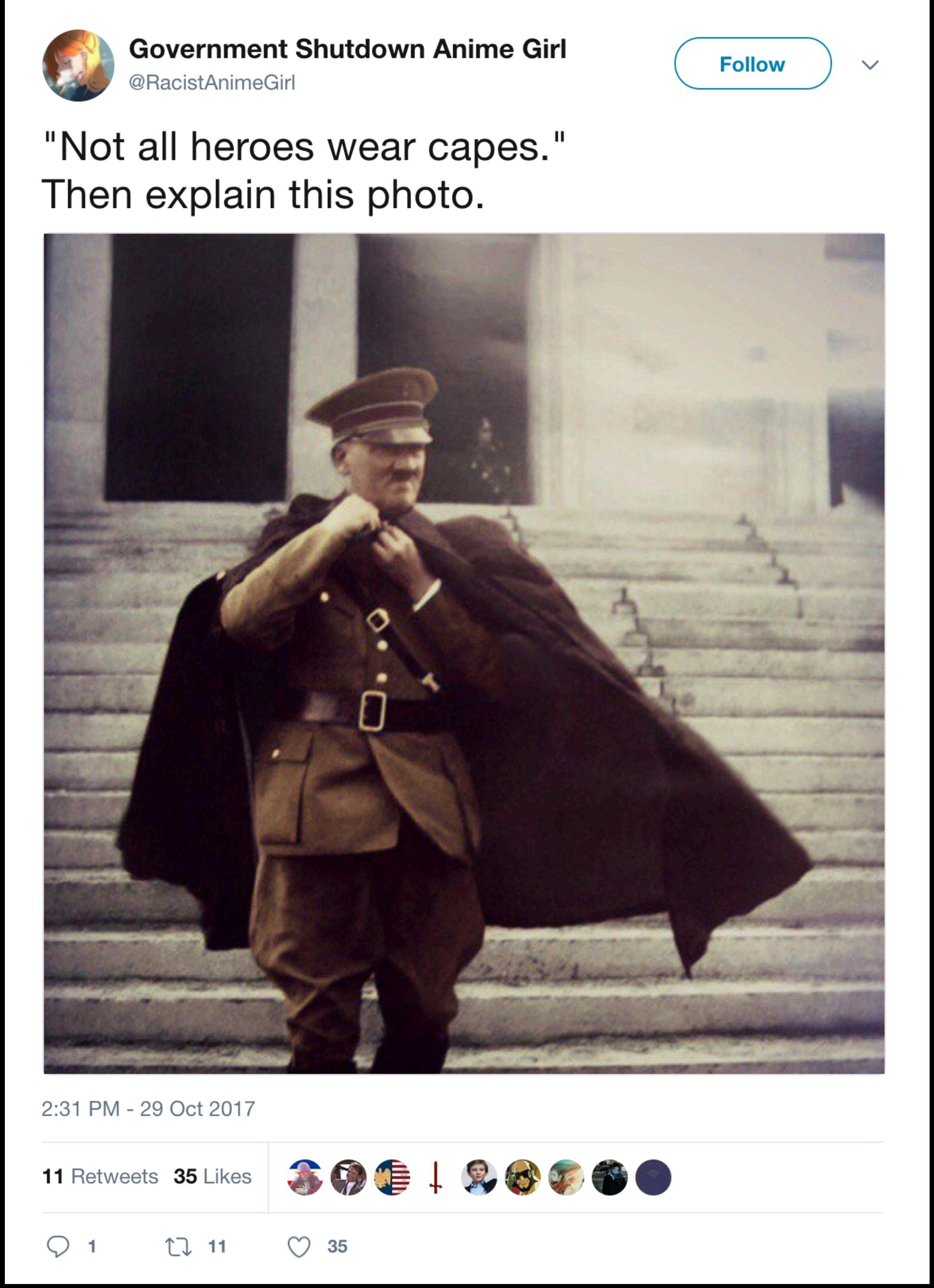
Replying to @kinsellawarren

#ClimateBarbie fit her perfectly.

It's not sexist. It's brainist.

9:50 PM - 7 Nov 2017

2 Likes



Government Shutdown Anime Girl
@RacistAnimeGirl

"Not all heroes wear capes."
Then explain this photo.

2:31 PM - 29 Oct 2017

11 Retweets 35 Likes

Questions?