

Annotation Codebook

Motivation: These annotations will be used in research that compares the ability of different machine learning models in identifying hateful dog whistles.

Guidelines:

The accompanying spreadsheet contains seven columns. The first ("Tweet ID") contains the tweet identification numbers. The second column contains a link to an HTML file with the full tweet content. You will see the tweet text and accompanying image. This image may be a thumbnail from a video or GIF. Information regarding the account which sent the tweet has been removed. Mentions (@Username) and hyperlinks have been replaced with the tokens `<user>` and `<url>` to help with readability for the classifier. You may use all the available information when making annotation decisions.

The five subsequent columns are where you record the annotations. The annotation structure shows the options for each of the columns. Please use the definitions and examples provided below to guide your decisions. If you select an option other than *Hateful* in the first column, columns four and five will be auto-crossed out. If you select *None* for the first column, columns two and three will also be auto-crossed out in addition to columns four and five. There is no need to fill out these cells. Also, note that to change selections, you must delete the contents of the cell and make a new selection. Selecting another option before deleting will result in multiple selections. Only in column five is that allowed.

Annotation structure

1. Primary [Select from drop down menu]
 - Hateful
 - Counter-speech
 - Reclaimed
 - None
2. Modality (If not None): [Select from drop down menu]
 - Text-Unimodal
 - Image- Unimodal
 - Multimodal
3. Target (If not None): [Enter all that apply]
4. Strength (If Hateful): [Select from drop down menu]
 - Animosity
 - Derogation
 - Extreme
5. Strategy (If Hateful): [Select all that apply from drop down menu]

- Explicit
- Pseudo-factual
- Normative
- Coded
- Creative

Definitions and Examples

Primary:

There are many forms of toxic online communication. *Hate* is a specific one. For this study, we define it as “language that is used to express hatred towards a targeted group or is intended to be derogatory, to humiliate, or to insult the members of the group” (Davidson, Warmesley, Macy, & Weber, 2017). Such comments should be marked as “Hateful”.

Not to be confused with hate is *Counter-speech*. *Counter-speech* is any direct response to hateful speech that undermines it.¹ Counter-speech expresses support to a group that has been targeted by hate, and such comments should be marked as “Counter-speech.”

Reclaimed captures uses of slurs self-referentially. These uses aim to reappropriate oppressive language to give them positive connotations during in-group use. If there is doubt over the identity of the tweet sender and whether the slur is used self-referentially, mark “Hateful”.

Communication falling outside of these categories should be labelled “None”.

In circumstances where the distinction between two or more categories is unclear, please input the option highest on the following hierarchy:

1. Counter-speech
2. Reclaimed
3. Hateful

Slur handling:

When deciding on tweets whose only indication of a Primary column label is a slur, consider the following guidance. If there is clear evidence of the sender’s identity and it matches the target of the

¹ <https://dangerousspeech.org/counterspeech/>



slur, mark “Reclaimed” (example 16). If their identity does not match the target group, mark “Hateful”. If the slur is used to draw attention to or undermine prejudice against a group, mark “Counter-speech” (example 11). In cases where the identity of the speaker is undetermined and use of the slur is not Counter-speech, use your best judgement given the negativity of the word’s connotation and the historical marginalization of the targeted group. Some examples are:



- For uses of the word *nigga*, please mark “None” (example 20)
- For uses of *cunt/twat*, please mark “Hateful” if the use is gender-based. If not, please mark “None” (example 21).
- For uses of words like *nigger/faggot*, please mark “Hateful” (examples 3 and 6)
- For examples of dog whistle slurs, please mark “Hateful”


Hateful

- Fill this out if the tweet meets the definition for *Hateful*
- Examples for this label are below...

#	Example	Explanation
1	 <p><user> So, you dumb twat, has anyone you’ve ever known done anything illegal? Is it your fault? <url></p>	While this tweet appears to be abuse targeted towards a user for their perceived stupidity, the tweet sender implies, through the use of the sexist slurs <i>bitch</i> and <i>twat</i> , that this lack of intelligence can be attributed to the gender of the target.

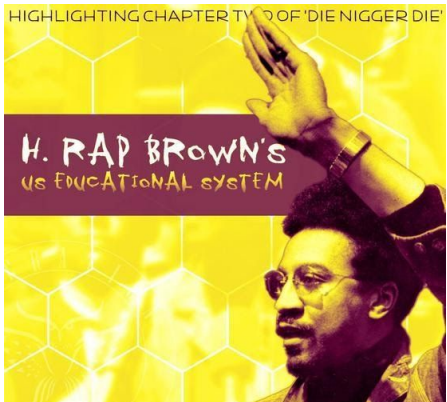
2	 <p><i>#ChristChurch don't you dare complain! #BanIslam <url></i></p>	<p>This tweet expresses hate towards Muslims and those who accept them by showing a lack of sympathy for the victims of the Christchurch mosque shooting. Rather than blame the far-right shooter, this tweet places the culpability on the victims suggesting they deserve to be killed.</p>
3	 <p><i>if you a coon i'll just address you as nigger. good riddance, coon ass nigger. <url></i></p>	<p>This tweet uses two racial slurs—<i>coon</i> (a derivative of racoon) and <i>nigger</i>—to disparage what appears to be a Black twitter user. These terms originated from an ongoing history of violence towards Black Americans and reinforce a racial hierarchy.</p>



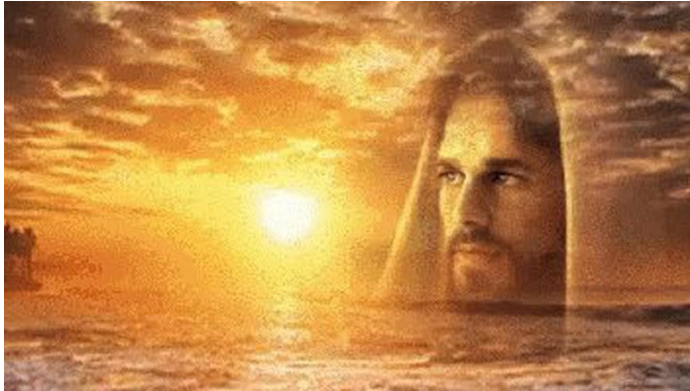
4	 <p><i>Nigga pounded that nut deep in his bitch tight asshole <url></i></p>	<p>There is an ongoing debate over whether certain forms of porn constitute the objectification of women or their sexual liberation. This tweet is objectifying by calling the woman a <i>bitch</i> and violently describing the sexual encounter (<i>pounded</i>).</p>
5	 <p><i>#larryelder An anti-Semite, two race card hustlers and a guilty white liberal walk into a bar... <url></i></p>	<p>On the surface, it might appear to be counter speech against Louis Farrakhan's (on the left) anti-Semitism. However, the intent of the tweet is to downplay the racism experienced by Black Americans rather than support Jews. The use of <i>race card</i> and <i>white guilt</i> in a common joke format (people walk into a bar) suggests a dismissive attitude toward the impact of the history of racism in the United States. This is both hateful and a dog whistle which is explained in more detail later.</p>



6	 <p><mention> Twitter being a faggot. <url></p>	<p>In this tweet, the sender refers to another user as a <i>faggot</i> to express their dissatisfaction with the exchange. Though the word is used to disparage another user for their perceived stupidity rather than sexual orientation, this tweet is hate rather than abuse. Using a gay slur in that manner strengthens the negative associations around homosexuality and contributes to its stigmatization.</p>
---	--	--

Counter-Speech

- Fill this out if the tweet meets the definition for *Counter-Speech*
- Examples for this label are below...

#	Example	Explanation
7	 <p><i>H. Rap Brown's US Educational System :: Highlighting Chapter Two Of 'Die Nigger Die'</i> <url></p>	<p>This tweet promotes a book discussing racial inequalities in the United States' education system. The book uses a racial slur (<i>nigger</i>) provocatively to draw attention to the book's contents which highlight a history of structural racism.</p>


8	 <p><i>Triggered Redneck In Texas Yells At Couple Protesting A Confederate Monument! <url></i></p>	<p>Even though this tweet uses disparaging language against Whites (<i>redneck</i>), the primary purpose is to expose racist ideas like supporting Confederate monuments.</p>
9	 <p><i>I drive Uber and listen to music saying nigga with white people in the car quite often. <url></i></p>	<p>The sender of this tweet is expressing discomfort with the belief that white people get a pass to say <i>nigga</i> when singing along to songs by Black artists. The idea of a pass speaks to the privilege of being able to freely enter and exit racialized spaces; an affordance that people of color do not have.</p>
10	 <p><i>This nigga be having y'all at war <url></i></p>	<p>This tweet expresses the hypocrisy of people who use religion to justify war and violence. This is counter-speech against religious fundamentalism which is unaccepting of other worldviews.</p>


11	 <p><i>Henry just spit in my face and called me a faggot, thanks for the shower <url></i></p>	<p>This tweet describes a homophobic encounter experienced by the sender. The nonchalance with which they responded to the attack undermines the power the homophobe tried to express over them. By embarrassing their gay identity, the sender is countering heteronormative discourse.</p>
12	 <p><i>I'm still shocked that el dude thought that changing nigger to negro would be better <url></i></p>	<p>This tweet calls out the use of racial slurs and <i>el dude's</i> attempt to sugarcoat racist ideas through the use of slightly less offensive language.</p>

Reclaimed

- Fill this out if the tweet meets the definition for Reclaimed
- Examples for this label are below...

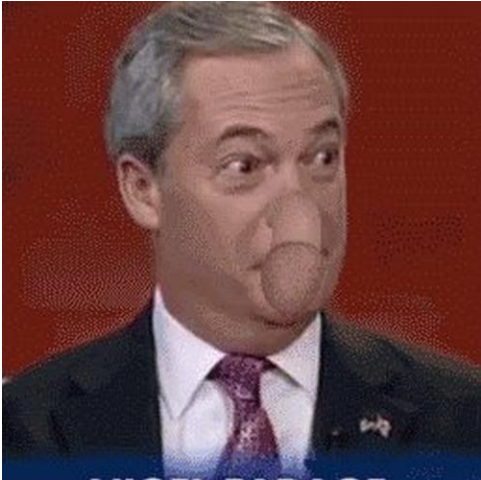
#	Example	Explanation
---	---------	-------------

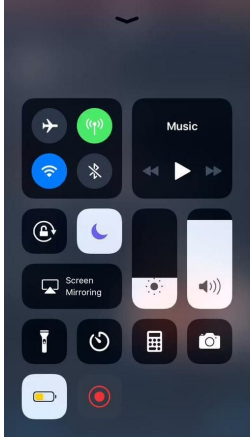



13	 <p>I LOVE MY FAGGOT FACE <url></p>	<p>The word <i>faggot</i> is being used here self-referentially to reclaim its meaning and express pride in the sender's sexuality.</p>
14	 <p>I think we call this redneck ingenuity <url></p>	<p>This is another example of a slur (<i>redneck</i>) being used self-referentially and its stereotypes (such as having lower education levels and engaging in manual labor) are being celebrated rather than derided.</p>



15	 <p>Fine ass bitch give a fuck bout a nigga 🤗💙 <url></p>	<p>The sender is reclaiming both <i>bitch</i> and <i>nigga</i> to express her independence and lack of reliance on a man.</p>
----	---	---

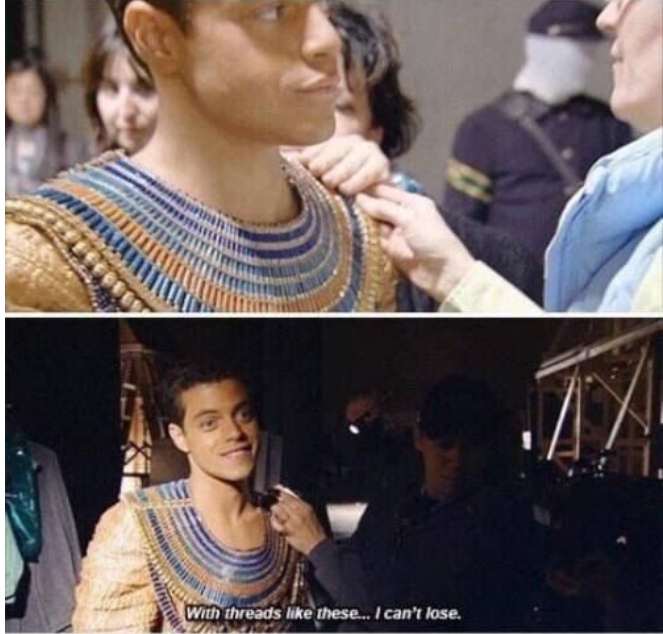
None

- Fill this out if the tweet meets the definition for *None*
- Examples for this label are below...

#	Example	Explanation
16	 <p><user> <user> Will this do? <user> is a fucking twat faced manfrog bell end <url></p>	<p>The insult <i>fucking twat faced manfrog</i> is directed at an individual (Nigel Farage) for their personal political views and not a shared group identity.</p>

17	 <p><user> I'm tired of your shit you giraffe neck ass nigga <url></p>	Following the slur flowchart, this use of <i>nigga</i> is non-hateful.
18	 <p><user> <user> You're seriously thick or just an overliteral retard, if you can't detect obvious humour. <url></p>	This tweet insults the intelligence of the recipient by calling them <i>thick</i> and a <i>retard</i> , but is not based on their group identity.
19	 <p>Trump called Sessions 'mentally retarded, dumb Southerner'... <url></p>	The tweet is reporting on Trump's actions which would be classified as abusive, but the tweet itself is not.
20	 <p>Do it for the Realist Nigga in the game right now 🤔🤔🤔 <link></p>	This use of the word <i>nigga</i> denotes a kinship with the subject rather than sense of superiority. It is often used familiarly within the Black community. For this reason, this tweet is not hateful. However, because there

		is no clear indication of the sender's identity, this is not reclaimed either.
21	 <p><user> cheeky wee cunt <url></p>	In certain Anglophone countries (like the United Kingdom and Australia), the words <i>cunt</i> and <i>twat</i> carry a positive connotation and are used between friends to call them a goof.
22	 <p>Kiara Cole - Cum In My Cunt 3 - 03/20/19 <url> March 20, 2019 at 11:43PM <url></p>	This falls in the category of porn as female sexual expression and depicts a woman consensually engaging in sexual acts with an explicit description.

23	 <p>highspirits-hoodvibes Ok but whys there a white guy playing an Egyptian</p> <p>angelrin89 Please tell me you're joking. That is Rami Malek, HE IS EGYPTIAN!</p> <p><i>When a SJW is crying about white washing but just proves they're...</i> <url></p>	<p>The aim of this tweet is less to undermine the aims of social justice advocates and more to highlight hypocrisies within some far-left activists. In this example, Rami Malek, an Egyptian-American, is falsely accused of cultural appropriation by a commenter for playing a Pharaoh in <i>Night at the Museum</i>. This is an example of people using social media to look like they care about social justice issues such as fair media representation while reproducing stereotypical representations; Malek did not fit into their generalization of what Egyptian people look like.</p>
----	--	---

Modality:

A *modality* is a channel of communication. In this dataset, two exist: text and image. Everything that falls within the image including text is part of the image modality whereas the text outside of the image is the text modality.

To make a decision in this column, ask yourself the following...

1. Was the text informative to my decision in the Primary column?
2. Was the image informative to my decision in the Primary column?

If you answered yes to both the above questions, mark *Multimodal*. If not, mark the modality for which you answered yes.

Text- Unimodal

- Fill this out if your decision in the Primary column was due to only the text content
- Examples for this label are below...

#	Example	Explanation
24	 <p data-bbox="253 850 587 886"><user> nigger fagot <url></p>	<p>The image is not hateful, but the text which contains slurs against black and gay people (nigger and faggot) is.</p>
25	 <p data-bbox="253 1453 912 1488"><user> Happy dance.... white trash be gone! <url></p>	<p>The image features a non-hateful still from a popular sitcom, but the text contains a slur against white people (white trash).</p>


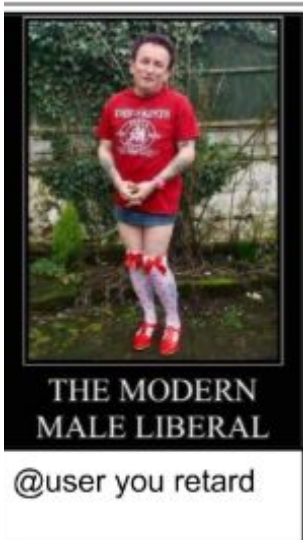
26	 <p><user> <user> You're seriously thick or just an overliteral retard, if you can't detect obvious humour. <url></p>	<p>The abuse is located in the text with the use of the word <i>retard</i>.</p>
----	--	---

Image- Unimodal

- Fill this out if your decision in the Primary column was due to only the image content
- Examples for this label are below...

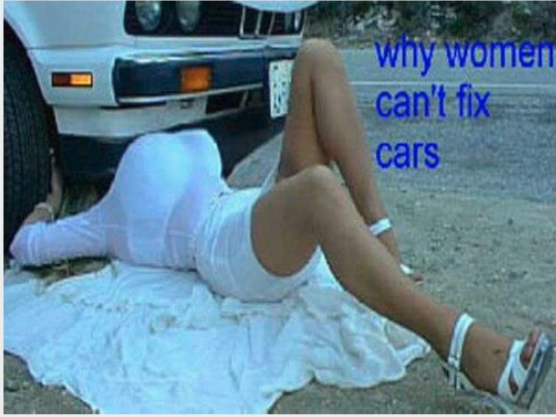
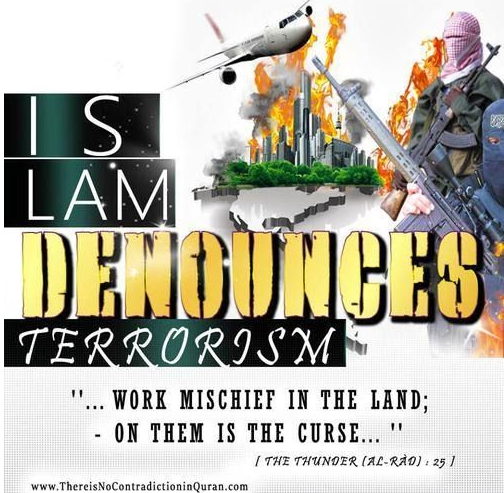
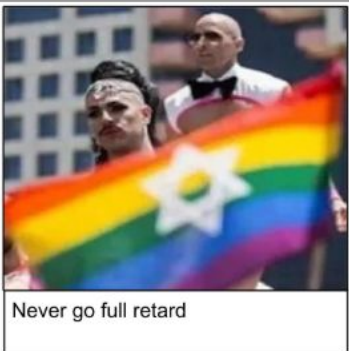
#	Example	Explanation
27		<p>Only the image contains information pertinent to the Primary column annotation of <i>Hateful</i>.</p>



28	 <p data-bbox="250 936 990 1005"><user> 🤔🤔 I guess I'm not black anymore y'all Ima pull up to the next trump meeting lookin like this nigga <url></p>	<p>The tweet is not hateful, but the image contains a confederate flag which is a white supremacist symbol disguised as a States' Rights symbol. Even though the image features a Black man supportive of its use, it does not undermine that fact that, overwhelmingly, Black Americans view it as a celebration of the history of slavery.</p>
----	--	--

Multimodal

- Fill this out if either...
 - Both the image and text content individually match your label in the Primary Column
 - Or...your label in the Primary column was due to jointly considering both the image and the text content. Independently, one or the other modality might not match the Primary column label, but together they do.
- Examples for this label are below...

#	Example	Explanation
---	---------	-------------




29	 <p>RT @CarSalesBossMan: I'm not sexist but http://t.co/tkYiFGes16</p>	<p>The image text is what makes the text mode hateful because it perpetuates the idea that women cannot do manual labor. The image is also hateful because it furthers this belief with a caricature of a woman's bust preventing her from doing the job.</p>
30	 <p>Islam Denounces Terrorism. ISLAM is the Religion of the PEACE and SECURITY 試合終了 Dan C...</p>	<p>The image contains a stereotypical image of the Muslim terrorist and the text denounces Islam.</p>
31		<p>In isolation, the image would be labeled None and the text <i>Interpersonal Abusive</i>. However, when combined, it becomes clear that the sender is being homophobic and anti-semitic.</p>

32	 <p><user> Shock the monkey. Florida #democrat #racecard strategery for governor. #AndrewGillum #DontMonkeyThisUp <url></p>	<p>Even though a monkey is not inherently hateful, it becomes racist when it is associated with Black people. In this case, the monkey symbolizes Andrew Gillum, a black Democratic candidate for Florida's governorship. If the image is used to support or clarify the hate within the text, then the label should be multimodal.</p>
33	 <p>Fine ass bitch give a fuck bout a nigga 🥰💙 <url></p>	<p>The image of the sender who is a Black woman supports the reappropriation of the slurs found in the text. The image alone does not suggest Reclaimed but, together with the text, it does.</p>



Target:




Target specifies the group towards which the hate is conveyed. This may be expressed towards the group as a whole or against an individual based on their group membership. This is a free form entry field. Please write as many targets as are applicable. If there are multiple, separate them with commas. If *Interpersonal Abuse* was filled in the Primary column, write *individual*. For instances of counter-speech, write the group against which the initial prejudice was targeted and the counter-speech aims to support.

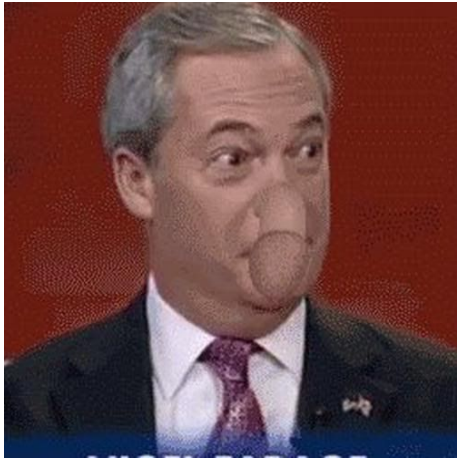
#	Example	Target	Explanation
---	---------	--------	-------------

34		Black people	The tweet uses a derogatory racial slur (nigger) and makes light of the history of slavery by joking about cotton picking.
35		Black people	The tweet stereotypes basketball as being a black sport while using a derogatory racial slur (nigger).
36		White people	This tweet uses the derogatory term for Whites "hillbilly" and depicts the negative stereotype of them being gun-toting, uneducated people. Other common slurs include white trash, cracker, and redneck.

37	 <p>THE MODERN MALE LIBERAL</p> <p>@user you retard</p>	Liberal men	This tweet mocks the masculinity of men who consider themselves politically liberal.
38	 <p>@user shhh, shut the fuck up already cunt ht</p>	Women	This tweet perpetuates notions of women as submissive, objectifies them by dismissing what they have to say, and uses a gender-based slur (cunt).
39	 <p>If this doesn't say 'faggot', I don't know what does <url></p>	Gay people	The tweet insinuates that the two men in the image share romantic feelings and describes it in a derogatory way (faggot).

40		Muslims	This tweet expresses Islamophobic sentiment by using the anti-Muslim slur “raghead” and depicting violence toward an Islamic symbol, the star and crescent.
41	 <p data-bbox="240 1203 894 1304"><user> We should SUE the MIGRANTS For BREAKING OUR LAWS !! 👉 That !!! Turn around !!! #BuildTheWall <url></p>	Immigrants, Latinx people	The tweet expresses a militaristic opposition to immigration particularly to that by Central American immigrants on the Southern border of the United States.

<p>42</p>	<p>This is Your Democrat Party</p> <p>What Media will Never Tell You. Why Would You Vote Democrat?</p>  <p><user> <user> <user> Using the race card you are weak and lazy. <url></p>	<p>Black people, Muslims, Jews, Latinx people, Immigrants</p>	<p>The tweet equates the views of prominent Black leaders with those of the extremist Louis Farrakhan, promotes the conspiracy theory the George Soros (who is Jewish) is encouraging mass immigration to the United States and Europe to undermine the white populations, and depicts progressive activist Linda Sarsour as a radical fundamentalist Muslim.</p>
<p>43</p>	 <p><user> <user> <user> Dems going full retard.. <url></p>	<p>Democrats</p>	<p>Rather than express an ideological disagreement, this tweet expresses hate towards all Democrats by calling them <i>retards</i>.</p>
<p>44</p>	 <p>Henry just spit in my face and called me a faggot, thanks for the shower <url></p>	<p>Gay people</p>	<p>This tweet counters homophobia and takes pride in the sender's sexual orientation. The target should be <i>gay people</i> as this is the group the message supports.</p>

45	 <p><user> <user> Will this do? <user> is a fucking twat faced manfrog bell end <url></p>	Individual	Since this tweet was marked as <i>Abuse</i> the identified target should be <i>Individual</i> .
----	--	------------	---

Strength:

The strength of the hate expressed by the tweet is measured ordinally on the scale *Animosity*, *Derogation*, and *Dehumanization*. Select the highest level of strength possessed by a tweet.

Animosity entails targeting one or more of the following towards a group: downplaying (whether mocking or ridiculing) their maltreatment or experiences of marginalization, accusations of special treatment, subtle implications of a threat, expressing stereotype with negative connotations, caveating use of derogation towards some members of group.

Derogation entails targeting one or more of the following towards a group: extremely negative representations, portrayals as evil, intensely negative feelings or emotions, expression of the negative impacts on society, suggestion that they are not welcome, ascribing evil or conspiratorial intentions, praising their negative condition, and/or threats of social exclusion.

Extreme hate entails targeting one or more of the following towards a group: descriptions in animalistic or subhuman terms, threats of violence or a desire for them to experience pain and suffering, advocacy of exclusion or expulsion from society.

Animosity



- Fill this out if the tweet meets the definition for *Animosity*

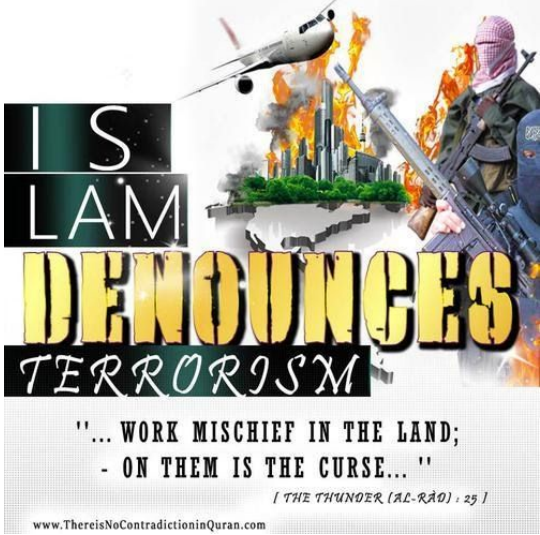
- Examples for this label are below...

#	Example	Explanation
46	 <p><user> Promises made. Promises kept. #MAGA #AmericaFirst #QAnon #BuildTheWall <url></p>	The tweet expresses the perceived negative impact on society of immigrants but caveats it by only saying <i>illegal aliens</i> , the suggestion being that legal immigrants might be acceptable.
47	 <p><user> <user> Si sick of the race card!!! <mention></p>	The tweet accuses people of color of special treatment by playing the <i>race card</i> to gain advantages.
48	 <p><user> NIGGER BASKETBALL ASSOCIATION. MILES. <url></p>	This tweet expresses the stereotype that basketball is a black sport and does so in a way that uses disparaging language.

Derogation


- Fill this out if the tweet meets the definition for *Derogation*
- Examples for this label are below...

#	Example	Explanation
49	 <p data-bbox="261 877 1040 940"><user> We should SUE the MIGRANTS For BREAKING OUR LAWS !! 👉 That !!! Turn around !!! #BuildTheWall <url></p>	<p>The tweet explicitly expresses that immigrants are not welcome. This stops short of being dehumanizing because, while military defense of the border is implied, direct violence is not advocated.</p>
50	 <p data-bbox="285 1423 670 1497">@user shhh, shut the fuck up already cunt ht</p>	<p>This tweet contains an extremely negative depiction of women as subservient as well as expresses strongly negative feelings towards them.</p>

51	 <p>Islam Denounces Terrorism. ISLAM is the Religion of the PEACE and SECURITY 試合終了 Dan C...</p>	<p>The tweet depicts muslims as evil and ascribes to them the intention of committing acts of terror.</p>
----	---	---

Extreme

- Fill this out if the tweet meets the definition for *Extreme* hate
- Examples for this label are below...

#	Example	Explanation
52	 <p>Shock the monkey. Florida #democrat #racecard strategy for governor. #AndrewGillum #DontMonkeyThisUp</p>	<p>The tweet dehumanizes the black former Florida gubernatorial candidate, Andrew Gillum, by describing him as a monkey.</p>

53		<p>This tweet expresses hate towards Muslims and those who accept them by showing a lack of sympathy for the victims of the Christchurch mosque shooting. Rather than blame the far-right shooter, this tweet places the culpability on the victims suggesting they deserve to be killed.</p>
----	---	---

Strategy:

Purveyors of hate employ many strategies to communicate their message including both explicit and implicit ones. Dog whistling falls under implicit hate by using *polysemy*, the capacity of words and phrases to possess multiple meanings, to elicit different responses from different groups. The term originates from the physical object of a dog whistle which is designed to emit a high-pitch sound that dogs hear while humans cannot. Like the object, hateful dog-whistles are used to communicate a hateful idea which is difficult for the out-group to recognize but is understood by the in-group. In this dynamic, the in-group is composed of those who have sufficient knowledge of the socio-cultural context which can be used to decipher the hateful message. Dog-whistles provide the communicator with plausible deniability over their intentions; their utterance can be interpreted as innocuous without the knowledge possessed by the in-group. Dog-whistling is achieved through a combination of the following strategies: *pseudo-factual*, *coded*, and *creative*. For the annotations in this column, please select all the strategies present in the tweet.

Explicit hate encompasses manifestations of hate where there is little ambiguity over the sender's hateful beliefs/intentions and that require little cultural knowledge to reach this conclusion.

Pseudo-factual hate is stereotypical and prejudicial thinking that is supported by facts that are presented as impartial.

Normative hate attempts to pass off prejudicial viewpoints as good faith arguments in discussions of popular democratic or egalitarian ideals.

Coded hate conflates the meanings of neutral terms with hateful ones or uses negative terms to proxy for hateful grievances.

Creative hate reaffirms and exploits stereotypes of group identities in the name of artistic license through expressions like humor, music, and art.

Explicit

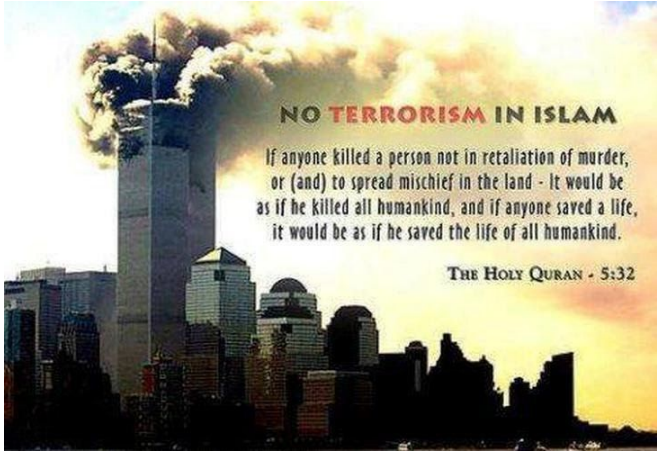
- Fill this out if the tweet meets the definition for *Explicit*
- Examples for this label are below...



#	Example	Explanation
54		The Rainbow flag and Star of David are widely known symbols of gay pride and the Jewish culture respectively. Therefore, the sender's intention to defame those groups is clear.
55	 <p><user> Happy dance.... white trash be gone! <url></p>	The sender does not hide their hatred for white people by celebrating (<i>happy dance</i>) when they (referred to as <i>white trash</i>) are <i>gone</i> .


56	 <p>the....nigger agenda... <link></p>	<p>The tweet uses the slur <i>nigger</i> non-self-referentially to describe black people and ascribes to them a nefarious <i>agenda</i>.</p>
----	---	--

Pseudo-factual

- Select this if the tweet meets the definition for *Pseudo-factual*
- Examples for this label are below...

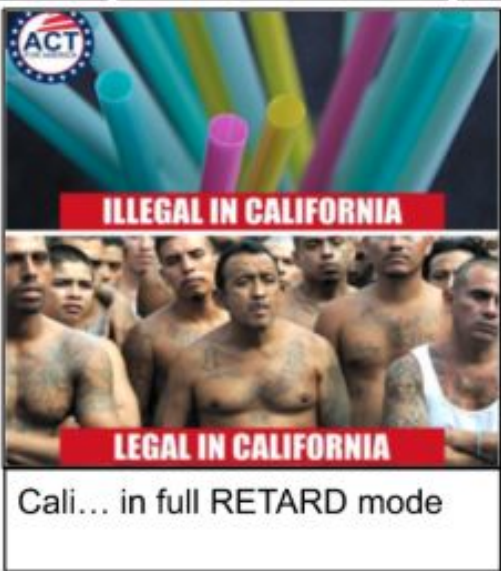
#	Example	Explanation
57	 <p><user> No terrorism in Islam. <url></p>	<p>This tweet presents an “objective”[*] fact by quoting the Quran. This quote is superimposed onto an image of the Twin Towers burning. Rather than making a distinction between fundamentalist forms, it implies terroristic thinking is central to Islam.</p> <p>[*]Objective is in quotes because this quoted translation of the Quran has not been fact-checked and is likely to have been cherry-picked</p>




58	<p>Every year in America, 4,745 people are killed by terrorist bombings. Oh, wait, I mean they're killed by illegal aliens who were driving drunk.</p> <p><user> Promises made. Promises kept. #MAGA #AmericaFirst #QAnon #BuildTheWall <url></p>	<p>This tweet uses an “objective”* fact of the number of deaths cause by intoxicated undocumented immigrants to support the idea that the restriction of immigration, likely from Latin American counties give the reference to the wall, is needed.</p> <p>*Objective is in quotes because the numbers have not been verified and may stretch the truth</p>
59	 <p><user> Happy Halloween from Blackface hillary & Hillbilly willy <url></p>	<p>This is a fake picture of Hilary and Bill Clinton which circulated the web in the run up to the 2016 election. It attempts to paint her as duplicitous by publicly advocating racial justice and privately wearing Black face. As explained above, her gender largely made her a target of these conspiracies.</p>
60	 <p>#BuildTheWall #NoDACA deals or amnesty! #AmericaFirst 🇺🇸 #ICE <url></p>	<p>On the surface, this may seem to be counter-speech against the Obama Administration’s large deportation efforts. However, it is using Obama’s stance on the issue as cover for Trump’s stricter policies.</p>

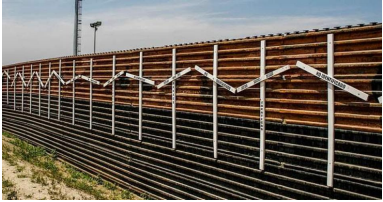
61	 <p>US news · 3 hours ago</p> <p>University looks into video of student being escorted out of biology class</p> <p>13K people Tweeting about this</p> <p>Professor: "there are 2 genders" SJW: FFFFFFFFUUUUUUUUUUUUUU- <url></p>	<p>Conversations around "biological truths" such as there being two discrete genders are often dog whistles for transphobic views rather than authentic scientific discussion. By asserting there are two biological genders, it blames those transitioning for bringing their psychological and physical harm on themselves.</p>
----	---	---


Normative

- Select this if the tweet meets the definition for *Normative*
- Examples for this label are below...

#	Example	Explanation
62		<p>This tweet references the debate over whether undocumented immigrants in the United States should be granted clemency. It suggests that California's sanctuary city policy will increase crime and laments the fact that it was able to ban plastic straws for environmental reasons but is unwilling to restrict immigration. In staking this position, it depicts an offensive Hispanic <i>cholo</i> stereotype by showing a group of men that appear to be inmates or gang members.</p>


63	 <p>California is retarded and Criminal traitors to America <url></p>	<p>Similar to the one above which views undocumented immigration as a larger issue than environmental pollution largely due to racialized views of crime.</p>
64	 <p>Sick of companies making leftwing SJW-politics. No more <user> for me <link></p>	<p>This tweet criticizes Nike for its social justice branding. Many on the right disagree with Nike's selection of Colin Kaepernick as one of its spokespeople. Kaepernick is known for protesting police brutality against the Black community by kneeling during the national anthem. Arguments against his protest on the grounds it was disrespectful to the country are racially tinged.</p>
65	 <p><user> <user> Si sick of the race card!!!! <mention></p>	<p>The race card refers to the belief that racial minorities are not marginalized. Rather, their minority status affords them special privileges access to which White people do not have. This downplays the existence of current forms of structural racism.</p>

66	<p><i>No Crime No Chaos No Caravans Yes #BuildTheWall Yes #ICE Yes Law & Order Yes #StopTheInvasion <url></i></p>	<p>There is a long history in the United States of issues of law and order and security being used to hide underlying racial animosity. Disseminators of this dog whistle claim their policies target criminals which, by their definitions, are disproportionately people of color.</p>
67	 <p><i>Migrants Not Stopping — Vow To Storm U.S. Southern Border <url> #AmericaFirst #BuildTheWall #MAGA <url></i></p>	<p>This is a more subtle version of prior concerns over immigration on the United States' Southern border. The language used promotes an adversarial in/out group dynamic between citizens and "storming" migrants. The expressions America First and Build the Wall place inordinate focus on a certain subset of immigration from Central American countries. While these are presented as the non-hateful prioritization of American citizens, there is a heavily racialized component to the alienation of immigrants. The #MAGA (Make American Great Again) further implies this as the subtext is that America was better when the country was whiter and White Americans were more powerful relative to racial minorities. Another variety of MAGA is KAG (Keep America Great) which is used in the runup to the 2020 re-election campaign.</p>

68	 <p>NOOOOO!!! Give him his soul back you muslim commie twat!!! <url></p>	<p>Barack Obama was frequently questioned over his citizenship and being a secret Muslim. Using his middle name “Hussein” and suggesting he was Muslim was a dog whistle to suggest he was less American and, therefore, unfit for office. These are both Islamophobic dog whistles, but also racial. As the first Black president, he was also the first to face repeated allegations of not being a citizen. These allegations imply American identity is tied to a White identity.</p>
----	---	---

Coded

- Select this if the tweet meets the definition for *Coded*
- Examples for this label are below...

#	Example	Explanation
69	 <p><user> Good.... #BuildTheWall <url></p>	<p>“Build the wall” references the desire to keep out undocumented immigrants primarily of Latinx ethnicity. The wall has become a symbol of nativist sentiment.</p>

70



🔥 YASSS It's time for a great show ♥.Classy
Redneck. ♥:#Chillin #Sleep <url>



The confederate flag is argued by its proponents to symbolize the democratic value of states' rights. However, it is the symbol of a secessionist movement rooted in the idea that black Americans were the rightful property of Southern slave owners.


71



<user> maybe not beeing retarded helps <url>

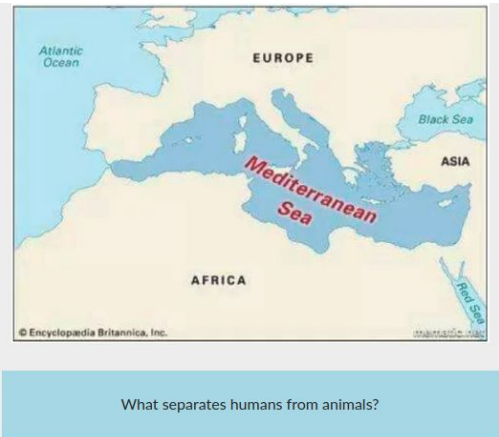
Featured in the image is Pepe the Frog who was successfully turned into a [hate symbol](#) by alt-right activists.

72	 <p><user> delete your account you cunt you cum sucker <url></p>	<p>“Trash” is a coded reference to Donald Trump’s harsh stance on immigration. It refers primarily to Latinx people who Trump conflates with “criminals, drug dealers, rapists” and “bad hombres”. This has been the source of his desire to build a wall to drastically reduce immigration on the Mexican border. He has also expressed a desire to limit people from “shithole” African and Caribbean countries (while simultaneously welcoming more Norwegians) and has issued bans against majority-Muslim countries.</p>
73	 <p><i>((StanLee))) was nothing hut a sjw lib cuck good riddance. WPower #MondayMotivation #stanlee <url></i></p>	<p>While the addition of white power makes this tweet overtly hateful, it includes several symbols which are also used more covertly as dog whistles. The three parentheses around Stan Lee is a symbol used by neo-Nazis to out Jews. Further, the okay symbol depicted by the clouds is emerging as a new hate symbol which avoids detection because of its more common use.</p>


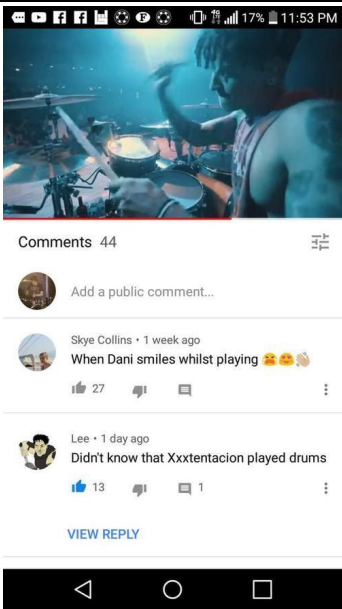
74	 <p><i>That feeling when someone uses a slogan that makes zero sense in place of an argument. #NPC #SJW <url></i></p>	<p>The man in the image is Jordan Peterson who is a conservative Psychologist who rose to prominence for his criticism of Canadian legislation that added gender expression to the Human Rights Act which spurred protest from trans-advocates. The tweet text is referring to these protesters as SJWs (social justice warriors) and NPCs (non-player characters). SJW is often used pejoratively to dismiss the underlying issues they are trying to address by suggesting the activism is about attention rather than action. NPC originates from the computer programmed characters that appear in video games but is used to denote someone who does not think for themselves. The use of it here implies that trans rights are illogical.</p>
----	--	---

Creative

- Select this if the tweet meets the definition for *Creative*
- Examples for this label are below...

#	Example	Explanation
75		<p>The tweet, under the pretense of a joke, compares people from the MENA region to animals.</p>

76	<div data-bbox="289 197 349 262"></div> <div data-bbox="362 205 496 252"> Imam of Peace ✓ @Imamofpeace </div> <div data-bbox="289 275 641 365"> <p>Islamists raid Christian minority areas after Asia Bibi is given her freedom and not executed.</p> </div> <div data-bbox="289 386 665 758"></div> <div data-bbox="289 772 412 793"> <p>Tweet your reply</p> </div> <div data-bbox="250 842 911 911"> <p><user> Yep, can't get enough of that Islamic peace and love! #banislam #banmuslims <url></p> </div>	<p>The tweet uses sarcasm to suggest that Muslims are violent. It is used to parody the belief that Islam means peace in Arabic (the meaning is closer to submission but has a similar root to the word for peace- <i>salam</i>)</p>
77	<div data-bbox="276 919 531 1373"></div> <div data-bbox="250 1409 792 1440"> <p>succ a nigga dick or sumthin' <mention> <url></p> </div>	<p>Even if the tweet is repeating song lyrics from Three3 Six6 Mafia's <i>Slob on My Knob</i>, the song's original lyrics encourage an attitude of objectifying women.</p>

78	<div data-bbox="274 193 652 697">  </div> <p data-bbox="248 730 873 800"><i>My boyfriend said I drive like an Asian today. LOL I'm always hitting the curbs <url></i></p>	<p>The sender jokes about their bad driving by comparing it to that of Asians which reinforces the stereotype.</p>
79	<div data-bbox="274 829 613 1434">  </div> <p data-bbox="248 1467 907 1537"><i>Imagine Smokey The Bear being an SJW. "Only YOU can prevent Rape Culture." <url></i></p>	<p>Smokey the Bear was a fire safety mascot used by the United States Forest Service. The text combines Smokey the Bear's tagline "only you can prevent forest fires" with the concerns over the sexual assaults faced by women. By drawing this comparison, it trivializes the experience of women and downplays the importance of consent.</p>



Happy Halloween..... LOL #BuildTheWall
 #LockThemUp #DrainTheDeepState #WalkAway &
 #KAG <url>

The image depicts Hillary Clinton and Huma Abedin (a staffer on the Clinton campaign) as witches which plays on the stereotypes of women being *shrill* and older ones being *hags*. Hillary Clinton was disproportionately the target of numerous right-wing conspiracy theories around her alleged connection to the *Deep State* and other nefarious attempts at holding power. Her gender played a large role in this as society treats aspirational women with distrust especially in United States' national politics which is one of the most gender imbalanced government bodies².

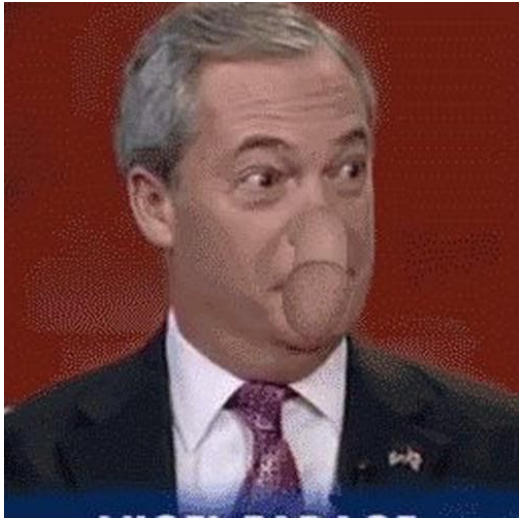
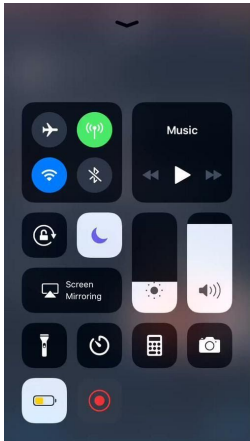
² https://data.worldbank.org/indicator/SG.GEN.PARL.ZS?most_recent_value_desc=true


IGNORE BELOW THE LINE- examples from previous versions

Forms of toxic communication which target entities based on their individual characteristics rather than group membership constitute *abuse*. *Interpersonal abuse* is also derogatory, humiliating, and insulting but is not group-based. Please mark these as “Interpersonal Abusive.”

Interpersonal Abusive

- Fill this out if the tweet meets the definition for Interpersonal *Abusive*
- Examples for this label are below...

#	Example	Explanation
14	 <user> <user> Will this do? <user> is a fucking twat faced manfrog bell end <url>	The insult <i>fucking twat faced manfrog</i> is directed at an individual (Nigel Farage) for their personal political views and not a shared group identity.
15	 <user> I'm tired of your shit you giraffe neck ass nigga <url>	The use of nigga appears to be self-referential and not hateful. Instead, the insult of “giraffe neck ass” makes the tweet abusive.


16	 <p><user> <user> You're seriously thick or just an overliteral retard, if you can't detect obvious humour. <url></p>	<p>This tweet insults the intelligence of the recipient by calling them <i>thick</i> and a <i>retard</i>.</p>
----	--	---

Secondary:

This category determines the level of directedness of the hate if *Hateful* was marked in the Primary column. Mark *Directed (Individual)* if the hate is expressed towards an individual on the basis of their group identity. Mark *Undirected (Group)* if the hate is targeted to the group as a whole.

Directed (Individual)


- Fill this out if the tweet meets the definition for *Directed (Individual)*
- Examples for this label are below...

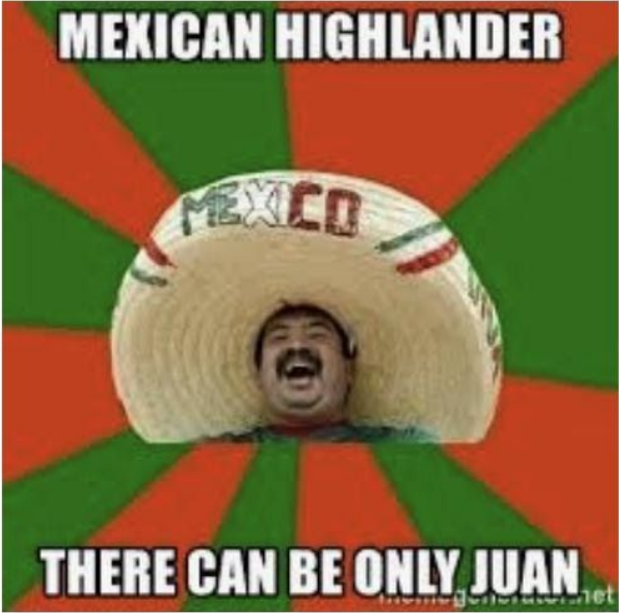
#	Example	Explanation
47	 <p><user> Fuck that nigger Obama <url></p>	<p>The hatred is directed towards Barack Obama based on his racial identity.</p>

48	 <p><user> So, you dumb twat, has anyone you've ever known done anything illegal? Is it your fault? <url></p>	<p>This tweet targets a specific twitter user based on her gender.</p>
----	--	--

Undirected (Group)

- Fill this out if the tweet meets the definition for *Undirected (Group)*
- Examples for this label are below...

#	Example	Explanation
49	 <p>Hey... fucck raghead</p>	<p>This tweet targets hate towards muslims in general rather than any specific one.</p>

50	 <p>Fucking beaners! I came up with right now! #BuildTheWall <url></p>	<p>This tweet directs hate towards all Mexicans by referring to them with the slur <i>beaners</i> and by caricaturing them in a sombrero and with the name Juan.</p>
----	---	--

Modality:

A *modality* is a channel of communication. In this dataset, two exist: text and image. Please consider all text, even that appearing within the image, as part of the text mode. All strictly visual cues fall into the image mode.

Text- Unimodal

- Fill this out if your decision in the Primary column was due to only the text content
- Examples for this label are below...

#	Example	Explanation
---	---------	-------------

25	 <p data-bbox="253 695 586 726"><user> nigger fagot <url></p>	<p>The image is not hateful, but the text which contains slurs against black and gay people (nigger and faggot) is.</p>
26	 <p data-bbox="253 1299 911 1331"><user> Happy dance.... white trash be gone! <url></p>	<p>The image features a non-hateful still from a popular sitcom, but the text contains a slur against white people (white trash).</p>
27	 <p data-bbox="253 1812 922 1843"><user> <user> in the real world.... #BUILDTHEWALL <url></p>	<p>This is considered uni-modal text and not multi-modal because it is the text within the image that makes it hateful.</p>



28	 <p><user> <user> You're seriously thick or just an overliteral retard, if you can't detect obvious humour. <url></p>	<p>The abuse is located in the text with the use of the word <i>retard</i>.</p>
----	--	---


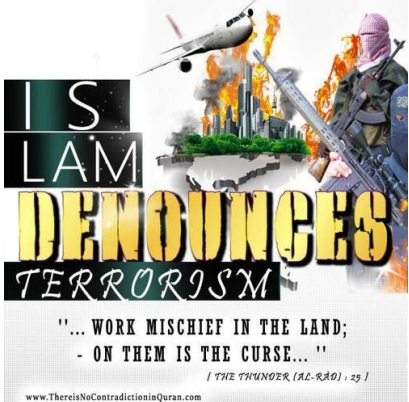
Image- Unimodal

- Fill this out if your decision in the Primary column was due to only the image content
- Examples for this label are below...

#	Example	Explanation
29	 <p><user> 🤔🤔 I guess I'm not black anymore y'all Ima pull up to the next trump meeting lookin like this nigga <url></p>	<p>The tweet is not hateful, but the image contains a confederate flag which is a white supremacist symbol disguised as a States' Rights symbol. Even though the image features a Black man supportive of its use, it does not undermine that fact that, overwhelmingly, Black Americans view it as a celebration of the history of slavery.</p>

Bimodal




- Fill this out if both the image and text content individually match your label in the Primary Column
- Examples for this label are below...

#	Example	Explanation
30	 <p>RT @CarSalesBossMan: I'm not sexist but http://t.co/tkYIFGEs16</p>	The image text is what makes the text mode hateful because it perpetuates the idea that women cannot do manual labor. The image is also hateful because it furthers this belief with a caricature of a woman's bust preventing her from doing the job.
31	 <p>Islam Denounces Terrorism. ISLAM is the Religion of the PEACE and SECURITY 試合終了 Dan C...</p>	The image contains a stereotypical image of the Muslim terrorist and the text denounces Islam.

Multimodal

- Fill this out if your decision in the Primary column was due to jointly considering both the image and the text content. Independently, one or the other mode might not match the Primary column label, but together they do.
- Examples for this label are below...

#	Example	Explanation
---	---------	-------------

32	 <p>Never go full retard</p>	<p>In isolation, the image would be labeled None and the text <i>Interpersonal Abusive</i>. However, when combined, it becomes clear that the sender is being homophobic and anti-semitic.</p>
33	 <p><user> Shock the monkey. Florida #democrat #racecard strategery for governor. #AndrewGillum #DontMonkeyThisUp <url></p>	<p>Even though a monkey is not inherently hateful, it becomes racist when it is associated with Black people. In this case, the monkey symbolizes Andrew Gillum, a black Democratic candidate for Florida's governorship. If the image is used to support or clarify the hate within the text, then the label should be multimodal.</p>
34	 <p>Fine ass bitch give a fuck bout a nigga 🥰💙 <url></p>	<p>The image of the sender who is a Black woman supports the reappropriation of the slurs found in the text. The image alone does not suggest Reclaimed but, together with the text, it does.</p>