Detecting Islamophobia | Islamophobia annotation guidelines

The annotation guidelines for labelling up Islamophobia in tweets is provided here.

[1 | Overview 2](#_Toc536717266)

[2 | Defining Islamophobia 2](#_Toc536717267)

[3 | Weak and Strong Islamophobia 3](#_Toc536717268)

[3.1 | Some notes on the examples 4](#_Toc536717269)

[4 | Determining negativity and generality 6](#_Toc536717270)

[5 | Annotation process 7](#_Toc536717271)

[6 | Data 8](#_Toc536717272)

[7 | Final advice 9](#_Toc536717273)

# 1 | Overview

The empirical focus of the present work is Islamophobic hate speech on social media. Building on previous academic work, Islamophobic hate speech is defined as:

“Any content which is produced or shared which expresses indiscriminate negativity against Islam or Muslims.”

We then distinguish between weak and strong manifestations of Islamophobia. Strong Islamophobia is defined as:

“Speech which explicitly expresses negativity against Muslims.”

Weak Islamophobia is defined as:

“Speech which implicitly expresses negativity against Muslims.”

# 2 | Identifying Islamophobia

Both Muslims and Islam are included within our definition as targets of Islamophobia. This is because anti-Islam negativity is often a proxy for negativity against Muslims. So, if you are trying to identify whether the ‘topic’ of Muslims appears in the dataset, bear these aspects in mind:

* Any reference to Muslims or Islam, or closely associated artefacts, events and practices (such as mosques, the Qu’ran, Mecca, the Hajj) means that we are potentially looking at Islamophobia.
* In particular, any explicit references to Muslims and Islam, or to Muslims *qua* group means that we are dealing with some form of ‘generality’. That is, just to use the label ‘Muslims’ means that someone is making some sort of general statement (however implicitly). In practice, this usually means any reference to Muslims as a group or to Islam in sweeping generalised terms. BUT! Remember that just because someone is being general, it does not mean that they are necessarily being Islamophobic – they might be expressing neutral or even positive sentiments about Muslims.

# 3 | Weak and Strong Islamophobia

A key innovation in the present work – and, sadly, a very difficult task for you – is to distinguish between weak and strong varieties of Islamophobic hate speech. Please note that these terms refer only to an analytical distinction, rather than to the morality or ‘impact’ of tweets – weak Islamophobic tweets may still cause victims considerable harm and should be treated as seriously as strong Islamophobia.

Strong Islamophobia can vary, and includes:

* Expressing explicitly negative *views*, such as describing Muslims as barbarians
* Calling for prejudicial *actions*, such as demanding that Muslims are forcibly banned from the UK
* Expressing negative *emotions* about Muslims, such as anger and distrust, which are often articulated through the use of profanities

Examples of strong Islamophobic tweets include:

1. “Muslim men groom and rape children”
2. “Muslim mothers want to practice FGM ni the UK!”
3. “Typical, another bloody Muslim just blew himself up. LOSER”
4. “Fuck alllll Muslims”
5. “Muslim invasion, they’re going to take over the UK”
6. “Top European Lawyer says that Muslims don’t obey rule of law and should not be allowed to remain in Europe whilst posing a threat”
7. “The Police target Muslims because they’re a problem, new #evidence”
8. “Huge rally atm against Loughborough Mosque – let’s take back our country”

In example 6 the speaker is supposedly reporting someone else’s claims (the ‘top lawyer’ that is referenced) – but nonetheless it is still the speaker who is engaging in Islamophobia as s/he is the one who has shared the content. Note also that in determining whether the tweet is Islamophobic, the ‘truth’ of the claims is not evaluated. Even if a claim is supported by supposed evidence, Islamophobia can still be expressed. In any given context, ‘truth’ is always contested, and there is no neutral objective position from which to judge the epistemology validity of any claim (B. J. Allen, 2017). Thus, whilst intuitively it seems like many Islamophobic tweets contain falsehoods, this is not the conceptual basis on which we decide whether or not they are Islamophobic.

Weak Islamophobia

Weak Islamophobia is distinguished from strong based on whether the negativity is implicit or explicit. There are two main types of weak negativity. First, is emphasizing perceived differences between Muslims and other members of society, such as attributing to Muslims strange or unusual practices. Such content excludes and marginalizes Muslims in an insidious fashion; Muslims are not explicitly targeted and attacked but, rather, their incompatibility is highlighted. This can be seen as implicitly negative as perceived differences are not celebrated but problematized. Examples include:

1. “Muslims are just different!”
2. “Muslim food smells so weird”
3. “Wearing a Burkha doesn’t feel very #UK”

The second form of weak negativity is to take the tropes associated with strong negativity (such as claiming that Muslims are terrorists, barbarians or uneducated) and to ostensibly link them to only a small subset of Muslims (e.g. to just one individual terrorist or Muslims only living in one small geographical area, such as Rotherham) – and by doing so to implicitly forge a connection between the negative trope and *all* Muslims, By using the term ‘Muslims’ or ‘Islam’, even with caveats to heighten the specificity (such as ‘this Muslim terrorist’ or ‘Muslim Men in Rotherham’), an *implicit* connection is established with all Muslims. The key point here is that discourses about paedophiles, terrorists or FGM practitioners can often be articulated without the need to reference Muslim identity. Examples of this type of weak Islamophobia are provided below. In all of the cases, the speaker appears to be commenting on a specific case but still implicitly creates an association with the negative trope and all Muslims.

1. “Muslim terrorists attack London Bridge”
2. “Muslim radicals in the desert kill Christian hostage”
3. “Muslim pedos are sick”

# 4 | Annotation process

You will be provided with a csv file with your name in the file name. This file will contain the list of tweets, each tweets’ ID and some relevant metadata. Two columns will be of interest to you: ‘strength’ and ‘comments’. Strength is where you enter your annotation. Enter ‘0’ if there is no Islamophobia at all, ‘1’ if weak Islamophobia is expressed and ‘2’ if strong Islamophobia is expressed. Use the ‘comments’ section to explain your annotation, (if needed) flag any issues and to draw attention to any interesting features of the tweets.

# 5 | Data

You will be presented with 4,000 tweets. We have:

* Removed URLs from the tweets as these do not contain any semantic content and can make reading the remaining content in the tweets more difficult.
* Removed emojis from the tweets as these can be difficult to display and may result in annotators viewing different content. From our test studies we do not believe that this will make a substantive difference to your annotations – negativity against Muslims is Islamophobic, irrespective of whether it is preceded or followed by a smiley face.
* Not provided any links to photos and other forms of media as these will not be used to train our classifier. From our test studies we believe that this content is unlikely to make a considerable difference to your annotations.

We advise caution; if you think that viewing the additional media might render a tweet Islamophobic but the text content of the tweet – *by itself* – is not Islamophobic then do **not** label it as Islamophobic. Base your annotations solely on the content you are presented with.

# 6 | Final advice

* Be as literal as possible in applying the guidelines; do not over think it.
* Please take into account context! The tweets you will see are produced by far right Twitter accounts. Use your common sense to work out the nature of the tweets.
* In the UK many different groups may be victims of prejudice (likely targets include immigrants, refugees, people who are gay and people who are transgender0. Unless you think that they are being targeted as a proxy for Muslims (as with misdirected Islamophobia) do NOT include them in your annotation.
* Hashtags are equivalent to other forms of speech and should be analysed as such and taken literally – “#BanIslam” can be considered equivalent to writing “Islam should be banned”.
* If you are unsure whether a tweet is either not Islamophobic or weakly Islamophobic it is best to mark it as weak Islamophobic and flag the annotation in the ‘comments’ section. We will then revisit the annotation you have provided.
* Overall, we anticipate that most tweets will not be Islamophobic of any sort – so do not worry if you annotate many ‘0’ labels.

If you have any concerns or queries then please refer back to this guide and the examples provided. Do not hesitate to contact me.