



Toolkit for analysing a case of hate speech

Intermediate guide

The main problem in analysing a case of hate speech is that there isn't a one-size-fits-all procedure for it, as the methodologies used by national institutions vary to a rather high extent.

The Rabat Plan of Action on the ban of advocacy of national, racial or religious hatred, recommends a distinction be made between criminally punishable hate speech, hate speech that calls for civil or administrative sanctions and hate speech that just “raises concerns in terms of tolerance, civility and respect for the rights of others”¹.

Moreover, the Rabat Plan of Action proposes a test that takes into account:

1. the Context of the speech
2. the Speaker
3. the Intent
4. the Content and the form of the speech
5. the Extent of the speech and
6. the Likelihood of the speech to produce immediate actions against its targets.

The methodology we are proposing starts from the criteria and recommendations expressed in the Rabat Plan of Action and aims at making them easier to use by adding more sub-criteria.

1. Context

Analysing the context means understanding the socio-cultural environment in which hate speech was used. This relates to how vulnerable the target of the hate speech is from a social, cultural, or political perspective. Hate speech can target any social group, but members of the majority population are less vulnerable than potentially marginalised groups who have been exposed to a long history of negative stereotyping, lack of access to services and weak political self-determination.

- a. **Determining whether the group targeted by the expression is a potentially vulnerable group.** This is a binary assessment (Yes / No) that can be done by looking into whether the group represents a minority from an ethnical / racial / religious /

¹ Rabat Plan of Action Article 20

sexual / gender / social status / other criteria. If group identities are overlapping (e.g. the target is a Roma business owner), the criterion should yield “Yes” if the person was attacked on grounds of being Roma and “No” if the person was attacked for being a large business owner.

- b. **Type of acts of violence / discrimination carried out in recent years against the group targeted by the expression.** The answer options we are proposing, in order of severity, are: “Verbal violence”, “Psychological violence”, “Generalised discrimination by fellow citizens”, “Institutionalised discrimination” “Property destruction”, “Generalised and institutionalised restrictions of human or civil rights”, “physical violence”, “murder motivated by hatred”. You should choose the answer based on the most severe situations in which members of the targeted group have found themselves in recent years.
- c. **Extent of negative stereotypes towards the group targeted by the expression.** We are proposing three possible answers: “Some extent”, “Moderate extent” and “High extent”. Some extent means that there are just a few people who hold negative stereotypes against the group, while “High extent” means that negative stereotypes are common in society.
- d. **Connection of the hate message with the negative stereotypes against the group targeted by the expression.** The answer options are: “No connection”, “allusions towards negative stereotypes”, “affirmation and/or consolidation of negative stereotypes”.
- e. **Political representation of the group targeted by the expression.** The answer options are: “Consolidated political representation”, “In-group political representation”, “Limited political representation”, “Lack of political representation”. Choose “In-group political representation” when the only well-known elected officials are exclusively members of a party that aims to represent the vulnerable group. Alternatively, choose “Consolidated political representation” when there are multiple well-known elected officials belonging to the targeted group who are members of different political parties.

2. Speaker

This analysis will help you understand the likeliness of the speakers’ hate message being positively received by the audience. It is important here to assess how much the speaker has abandoned their political/social/moral obligations when engaging in hate speech (engaging in hate speech is more serious for public servants, who must not discriminate among citizens, than it is for politicians, who are supposed to act as the voice of their constituency and which might hold rather radical views towards some social groups).

- a. **Status of the speaker.** The answer options, in order of severity, are: “Regular citizen”, “Political figure”, “Public figure or influencer”, “Educator”, “Public servant”. “Regular citizen” means that the hater’s status does not provide them with authority over the audience. “Political figure” refers to politicians or to similar figures (e.g. union leaders, NGO representatives). Choose “Public figure or influencer” when the speaker is a well-known figure (e.g. actors, vloggers, journalists, artists). “Educator” applies for teachers, university professors etc. “Public servant” refers to people who are supposed to serve all member of society without discriminating.
- b. **Capacity in which the speaker made the statement.** Most of the time, the status of the speaker is the same as the capacity in which they deliver the expression. However, sometimes the capacity differs from the status, such as when a politician’s

private conversation is leaked in the public space. The answer options are the same as for the status of the speaker: “Regular citizen”, “Political figure”, “Public figure or influencer”, “Educator”, “Public servant”.

- c. **Influence of the speaker on the group targeted by the expression.** The answer options, in order of severity are: “Little to no influence”, “Limited influence”, “Moderate influence”, “High influence”. In order to assess this, you should look into how much damage the actions of the speaker acting in accordance to their status can cause to the targeted group targeted. A regular citizen engaging in hate speech will have less influence than a public servant whose daily work involves protecting the human rights of people against whom (s)he is speaking.

3. Assumed Intent

Assessing the intent of the speaker is important for determining the action that needs to be taken against the speaker or to compensate for the expression. Intent can be extremely hard to determine, but the speaker’s past actions, the way they have selected the audience of the message and the way they reacted after the speech are elements that can be rather easily determined.

- a. **Past actions of the speaker with regards to the group targeted by the expression.** Looking into the past actions of the speaker towards the group targeted by the expression can reveal whether the speaker holds negative feelings towards the group. If the speaker has fought for the rights of members of that group, then it is unlikely that there were bad intentions behind the negative expression. The answer options, in order of severity: “Positive actions”, “Mixed actions / no actions”, “Negative actions”. When choosing the answer option, we recommend, when possible, to consider more recent actions of the speaker.
- b. **Reaction of the speaker after promoting the hate message.** The way speakers react after distributing hateful narratives can provide clues as to the speaker’s actual intentions (e.g. showing true remorse or continuing incitement). The answer options, in order of severity are: “Apologies offered”, “No reaction”, “Continued incitement”.
- c. **Intended audience of the hate message.** Different audiences tend to react differently to the specific messages they are being presented. The answer options, in order of severity, are: “audience not likely to have negative feelings towards the targets of the expression”, “audience likely to have negative feelings towards the targets of the expression”, “audience having strong negative feelings towards the targets of the expression”.

4. Content and form

Analysing this part is a bit more complicated so your experience will play a crucial role in determining the content and form of the hate speech.

- a. **Degree to which the expression is provocative or aggressiveness of the message.** The answer options, in order of severity: “Low degree of violence”, “Moderate degree of violence”, “High degree of violence”. You should thus pay

attention to whether the speech contained loaded words known to evoke negative reactions in the audience towards the targeted group.

- b. **Form taken by the expression.** Some forms of expression benefit from a higher degree of protection than others (e.g. artistic expressions, religious expressions, academic discourse, etc.). However, there are cases where hate speech is disguised as academic discourse by citing obscure theories generally considered by the academic community as untrustworthy or backed by references to religious texts disguised as religious expressions.
- c. **How direct was the message?** The expression that is being analyzed can be openly hateful, or it can try to just suggest the hateful message by using metaphors or other figures of styles. Openly hateful messages containing calls for action tend to be more easily understood as such by the audiences. Therefore, they also tend to be more severe than the hidden ones. The answer options we are proposing are “Direct” and “Indirect”.
- d. **Legal status of hate message.** Some countries have clear provisions on what types of hate speech are punishable under criminal law. To see a collection of national law provisions on this topic, you can access the country-specific information provided by the members of the International Network against Cyber Hate at this [link](#).

5. Extent/reach of the speech act

When analysing a case of hate speech, it is also important to consider the extent/reach of the hate speech act. You will thus have to analyse the medium in which the speech has been shared and the frequency of the dissemination.

- a. **Nature of the expression.** This means identifying if the message was expressed in a public or a private context. Expressions shared in private contexts do not aim to produce harmful results and are protected by the right to privacy.
- b. **Means of dissemination.** The medium in which the speech was shared is very important as it might help you understand the speaker’s potential to reach either large audiences or their intended audience. The answer options are “Likely inefficient at reaching the intended audience”, “Likely moderately efficient at reaching the intended audience”, “Likely efficient at reaching the intended audience”. Social media and new media can be considered more efficient than traditional media when it comes to younger audiences.
- c. **Frequency of the dissemination of the hate message.** Here you should look into how many times the speaker has repeated the hate message. The answer options are: “Single time dissemination”, “Moderate frequency of dissemination”, “High frequency of dissemination”.

6. Likelihood of generating violent / discriminatory events

This criterion is the hardest to determine and relies heavily on your experience. The aim here is to establish a clear cause and effect relationship between the expression and the potential of the audience to act against the targets of the expression.

- a. **Effects produced by the hate message.** The answer options are: “No effects produced”, “Audience engaged in verbal violent conduct” (a significant number of its audience must engage in hate speech using similar ideas from the original expression), “Audience engaged in violent / discriminatory actions” (sufficient for one member of the audience to engage in such actions, but a clear connection between the expression and the actions must be established).