Toolkit for analysing a case of hate speech

Beginners guide

The main problem in analysing a case of hate speech is that there isn't a one-size-fits-all procedure for it. 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 to make them easier to use by adding more sub-criteria.

1. Context

When analysing the context, you will need to understand the socio-cultural environment in which the hate speech was used. This relates to how vulnerable the target of the hate speech is from a social, cultural or political perspective.

a. **Determining whether the group targeted by the expression is a potentially vulnerable group.** This is a Yes/No evaluation that you can make by looking into whether the group represents a minority from an ethnical / racial / religious / sexual / gender / social status / other criteria. If group identities are overlapping (e.g. the target is a Roma business owner), you should choose "Yes" if the person was attacked on grounds of being Roma and "No" if the person was attacked for being a large business owner.

---

[1] Rabat Plan of Action Article 20

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", "Institutionalized discrimination" "Property destruction", "Generalised and institutionalized 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.

2. Speaker

Better understanding the status of the speaker is important as it might help you determine whether the speaker's hate message will be positively received by their audience. You might also want to consider how much the speaker has abandoned their political, social or moral obligations when engaging in hate speech (e.g. are they a public servant or just a regular citizen?).

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.

3. Assumed Intent

When analysing a case of hate speech, you should also consider assessing the intent of the speaker who is spreading hateful messages. This will help you determine the action you should take to compensate for their expression. Intent can be extremely hard to determine, but the speaker's past actions and their reaction 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 targeted group targeted 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 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 sharing hateful messages can also provide clues about their actual

intentions (e.g. showing true remorse or continuing incitement). The answer options, in order of severity are: "Apologies offered", "No reaction", "Continued incitement".

4. Content and form

Analysing this part is a bit more complicated so 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.

5. Extent of the speech act

When analysing a case of hate speech, it is also important to consider the extent 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.

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).