World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:19, No:06, 2025

# Automatic Moderation of Toxic Comments in the Face of Local Language Complexity in Senegal

**Authors :** Edouard Ngor Sarr, Abel Diatta, Serigne Mor Toure, Ousmane Sall, Lamine Faty

**Abstract :** Thanks to Web 2, we are witnessing a form of democratization of the spoken word, an exponential increase in the number of users on the web, but also, and above all, the accumulation of a daily flow of content that is becoming, at times, uncontrollable. Added to this is the rise of a violent social fabric characterised by hateful and racial comments, insults, and other content that contravenes social rules and the platforms' terms of use. Consequently, managing and regulating this mass of new content is proving increasingly difficult, requiring substantial human, technical, and technological resources. Without regulation and with the complicity of anonymity, this toxic content can pollute discussions and make these online spaces highly conducive to abuse, which very often has serious consequences for certain internet users, ranging from anxiety to suicide, depression, or withdrawal. The toxicity of a comment is defined as anything that is rude, disrespectful, or likely to cause someone to leave a discussion or to take violent action against a person or a community. Two levels of measures are needed to deal with this deleterious situation. The first measures are being taken by governments through draft laws with a dual objective: (i) to punish the perpetrators of these abuses and (ii) to make online platforms accountable for the mistakes made by their users. The second measure comes from the platforms themselves. By assessing the content left by users, they can set up filters to block and/or delete content or decide to suspend the user in question for good. However, the speed of discussions and the volume of data involved mean that platforms are unable to properly monitor the moderation of content produced by Internet users. That's why they use human moderators, either through recruitment or outsourcing. Moderating comments on the web means assessing and monitoring users' comments on online platforms in order to strike the right balance between protection against abuse and users' freedom of expression. It makes it possible to determine which publications and users are allowed to remain online and which are deleted or suspended, how authorised publications are displayed, and what actions accompany content deletions. In this study, we look at the problem of automatic moderation of toxic comments in the face of local African languages and, more specifically, on social network comments in Senegal. We review the state of the art, highlighting the different approaches, algorithms, and tools for moderating comments. We also study the issues and challenges of moderation in the face of web ecosystems with lesser-known languages, such as local languages.

**Keywords :** moderation, local languages, Senegal, toxic comments
**Conference Title :** ICDSAA 2025 : International Conference on Data Science and Advanced Analytics
**Conference Location :** Montreal, Canada
**Conference Dates :** June 12-13, 2025