


Chapter 8

Exploring Hate Speech Classification in Low-Resource Languages: A Comprehensive Review

Sargam Yadav

 <https://orcid.org/0000-0001-8115-6741>

Dundalk Institute of Technology, Ireland

Abhishek Kaushik

 <https://orcid.org/0000-0002-3329-1807>

Dundalk Institute of Technology, Ireland

Kevin McDaid

Dundalk Institute of Technology, Ireland

ABSTRACT

The widespread adoption of social media has profoundly impacted the lives of individuals, businesses, and governments, fostering greater connectivity. However, a byproduct of online anonymity is that individuals are more willing to display aggression, which can negatively impact groups and discourage their participation. This chapter surveys recent and pertinent literature that utilizes Natural Language Processing (NLP) techniques to automatically detect online hate speech, including studies that explore model explainability with Explainable Artificial Intelligence (XAI), evaluate model fairness, and measure unintended bias. Furthermore, the chapter reviews shared tasks and challenges that have contributed towards advancing research in hate speech detection. The findings suggest that despite the recent strides made in the development of Artificial Intelligence (AI) models for hate speech

DOI: 10.4018/979-8-3373-3063-1.ch008

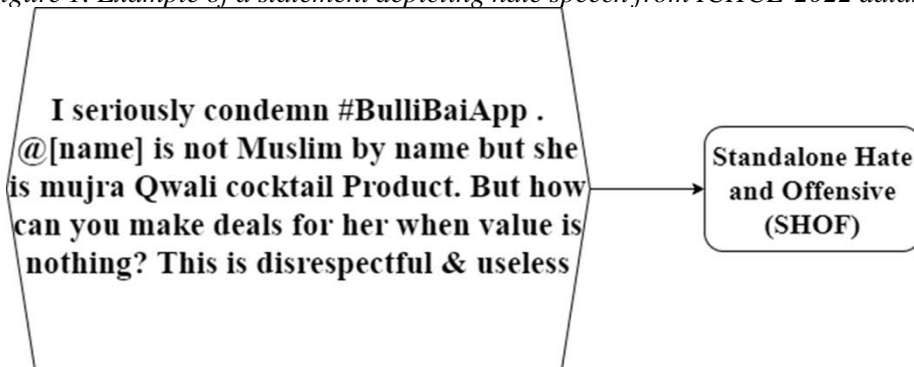
Copyright © 2026, IGI Global Scientific Publishing. Copying or distributing in print or electronic forms without written permission of IGI Global Scientific Publishing is prohibited. Use of this chapter to train generative artificial intelligence (AI) technologies is expressly prohibited. The publisher reserves all rights to license its use for generative AI training and machine learning model development.

detection, the problem requires further examination, creation of benchmark datasets, and examination of explainable methodologies, particularly in identification of misogynistic content.

INTRODUCTION

Hate speech, defined as discriminatory speech or language targeting a person or group based on identity-based characteristics such as race, religion, gender, and sexual orientation, poses a serious threat in the digital era. While social media provides several benefits such as visibility and networking, it has also enabled malicious individuals to spread hate (Djuric et al., 2015) and misinformation, and conduct trolling (Cheng et al., 2017) and cyber-bullying (Moreno et al., 2019). Social media serves as a hub for political discourse and can greatly impact social narratives. However, persistent harassment of targeted groups can severely impact their participation and limit their engagement. According to (Nockleby, 2000), hate speech is defined as ‘any communication that disparages a target group of people based on some characteristic such as race, color, ethnicity, gender, sexual orientation, nationality, religion, or other characteristics.’ Figure 1 displays an example of hate speech from the Identification of Conversational Hate-Speech in Code-Mixed Languages (ICHCL) 2022 dataset (Satapara et al., 2022), which has been categorized as ‘Standalone Hate’.

Figure 1. Example of a statement depicting hate speech from ICHCL-2022 dataset



(Satapara et al., 2022).

Cyberbullying, on the other hand, is a more general form of abuse which is not targeted towards a unique and non-controllable attribute of a group of people (*Cyberbullying and Hate Speech Online*, 2016) (Feinberg & Robey, 2009). Misogyny,

44 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/exploring-hate-speech-classification-in-low-resource-languages/393862

Related Content

The Ethical Dimensions of AI Development in the Future of Higher Education: Balancing Innovation with Responsibility

Megha Ojha, Amar kumar Mishra, Vinay Kandpal and Archana Singh (2025). *Ethical Dimensions of AI Development* (pp. 401-436).

www.irma-international.org/chapter/the-ethical-dimensions-of-ai-development-in-the-future-of-higher-education/359653

Value Co-Creation and Artificial Intelligence: Towards Responsible Innovation

Reda Lharti (2026). *The Ethical Landscape of AI: Global Issues and Solutions* (pp. 211-242).

www.irma-international.org/chapter/value-co-creation-and-artificial-intelligence/399867

Survey of Recent Cyber Security Attacks on Robotic Systems and Their Mitigation Approaches

Abdullahi Chowdhury, Gour Karmakar and Joarder Kamruzzaman (2019). *Cyber Law, Privacy, and Security: Concepts, Methodologies, Tools, and Applications* (pp. 1426-1441).

www.irma-international.org/chapter/survey-of-recent-cyber-security-attacks-on-robotic-systems-and-their-mitigation-approaches/228791

The Role of Multi-Modal Sentiment Analysis in Optimizing Leadership Communication

Ashish Khosla and Gaurav Gupta (2025). *Ethical Dimensions of AI Development* (pp. 369-400).

www.irma-international.org/chapter/the-role-of-multi-modal-sentiment-analysis-in-optimizing-leadership-communication/359652

Towards a Student Security Compliance Model (SSCM): Factors Predicting Student Compliance Intention to Information Security Policy

Felix Nti Koranteng (2020). *Modern Theories and Practices for Cyber Ethics and Security Compliance* (pp. 204-216).

www.irma-international.org/chapter/towards-a-student-security-compliance-model-sscm/253671