

REVIEWER'S REPORT

Manuscript No.: IJAR-57983

Title: Content Moderation & Free Speech: Legal limits of automated takedown systems on social platforms,

Recommendation:

Accept after minor revision

Rating	Excel.	Good	Fair	Poor
Originality			✓	
Techn. Quality		✓		
Clarity		✓		
Significance	✓			

Reviewer's ID: JPR-Abdul Haseeb Mir

Detailed Reviewer's Report

General Overview

The manuscript titled "*Content Moderation & Free Speech: Legal limits of automated takedown systems on social platforms*" presents a sophisticated, interdisciplinary analysis of a critical junction in contemporary digital governance. The primary objective of the article is to critically evaluate the legal, constitutional, and algorithmic boundaries of automated content moderation systems deployed by major social media platforms. To achieve this, the authors implement a robust comparative legal framework, examining how automated moderation mechanisms—such as machine-learning classifiers, hash-matching technologies, and artificial intelligence-driven sentiment analysis—interact with protected speech frameworks. The study primarily contrasts the First Amendment jurisprudence of the United States with the fundamental rights protections and regulatory mandates of the European Union, specifically focusing on the Digital Services Act (DSA).

The research topic holds exceptional relevance in the current socio-legal landscape, as the sheer velocity and volume of digital content have forced platforms to rely almost exclusively on automated filters to combat hate speech, misinformation, and copyright infringement. However, the systematic over-enforcement and inherent contextual blindness of these algorithms pose a direct threat to lawful expression, digital public squares, and democratic discourse. A major strength of this work lies in its

International Journal of Advanced Research

Publisher's Name: Jana Publication and Research LLP

www.journalijar.com

REVIEWER'S REPORT

balanced, multi-faceted analytical approach. The authors successfully avoid a purely legalistic or entirely technical focus, opting instead to blend algorithmic critique with deep constitutional analysis. The narrative flow is highly logical, guiding the reader seamlessly from the operational mechanics of natural language processing to statutory liability exemptions, and ultimately to human-in-the-loop regulatory recommendations.

In terms of overall contribution, the manuscript significantly enhances the existing literature by bridging the gap between algorithmic capability and public law. By analyzing the structural legal risks of privatized, automated censorship, the study fills an interpretive gap regarding platform accountability and due process in the digital age. The insights generated are well-grounded, highly insightful, and represent a premium scholarly effort that aligns perfectly with the standards of a top-tier journal specializing in law and technology.

Suggestions for Improvement

While the manuscript is exceptionally written and thoroughly researched, a few minor modifications and thematic expansions would further elevate its academic depth and policy-relevant impact prior to final publication.

- Clarification on Algorithmic Bias and Error Rates:** The section describing the technical limitations of machine-learning models notes that automated systems lack the capacity to interpret irony, satire, and political parody. To strengthen this argument, the authors should include a brief narrative discussion highlighting published empirical error rates or known algorithmic biases against marginalized dialects, which would further illustrate the real-world impact of these systems on vulnerable communities.
- Expanding the Discussion on Redress Mechanisms:** The manuscript correctly identifies the lack of procedural due process in immediate algorithmic takedowns. The analysis would be significantly enhanced by a more detailed exploration of effective user-redress models, specifically detailing how platforms can implement transparent, timely, and binding human-led appeal structures as mandated by emerging regulatory frameworks like the European Union's Digital Services Act.
- Contextualizing State-Action Doctrines:** In the portion of the text analyzing United States constitutional law, the discussion focuses heavily on how private platforms are generally exempt from First Amendment constraints. The authors should incorporate a concise, updated paragraph exploring recent legal debates or brief case updates surrounding the "state-action doctrine," particularly where government communication with platforms might be construed as coercive intervention or joint action.

REVIEWER'S REPORT

4. **Integration of a Global South Perspective:** The comparative legal analysis effectively isolates the United States and European Union regimes. To maximize the manuscript's global relevance, it is suggested that the authors add a short, narrative paragraph touching upon the broader global implications, such as how automated takedown systems operate under more restrictive cybersecurity or intermediary liability laws in jurisdictions outside the West.
5. **Standardization of Legal and Technical Terms:** The manuscript demonstrates a command of specialized terminology. To ensure complete accessibility for an interdisciplinary readership, the authors should verify that all technical machine-learning terms and distinct statutory clauses are clearly defined upon their first mention in the text, maintaining a uniform nomenclature across both the legal and technical chapters.
6. **Minor Citational and Formatting Refinements:** The review of legal precedents is thorough and comprehensive. A minor proofreading pass is recommended solely to guarantee that all legal citations, statute references, and footnotes strictly conform to standard bluebook or preferred international legal citation formatting rules, ensuring perfect typographic consistency throughout the final text.

Recommendation

Recommendation: Accept with Minor Revisions

Justification

The manuscript makes a meaningful, highly sophisticated, and timely contribution to the fields of internet law, technology policy, and constitutional jurisprudence. The methodology utilizing comparative legal analysis is sound, the algorithmic critiques are technically grounded, and the proposed policy solutions are legally realistic and well-reasoned. The identified areas for improvement are minor, focusing primarily on slight narrative expansions regarding algorithmic error rates, global contexts, and statutory adjustments. These minor modifications can be readily executed by the authors and will further polish an already outstanding piece of scholarship, making it fully suitable for immediate publication.