DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

Individual Criminal Responsibility for Incitement to Commit International Crimes through Social Media and Online Platforms: Defining the Thresholds of Causation and Intent

Almoatuz A. Munsoor
Associate Professor of Public International Law, University of Jeddah, KSA
Email: aaadam@uj.edu.sa
ORCID: 0009-0004-5816-7142

ABSTRACT

The exponential growth of social media and online platforms has fundamentally transformed the landscape of international criminal law, particularly concerning incitement to commit international crimes. This research examines the legal frameworks governing individual criminal responsibility for incitement conducted through digital channels, focusing on the critical thresholds of causation and intent required for prosecution. Drawing upon jurisprudence from international criminal tribunals, domestic courts, and emerging legal scholarship, this study analyzes how traditional doctrines of incitement have adapted to address the unique characteristics of digital communication, including virality, anonymity, and transnational reach. The research identifies significant challenges in establishing direct causation between online speech and subsequent international crimes, while also exploring the evidentiary requirements for demonstrating specific intent in the digital context. Through comparative analysis of cases from the International Criminal Tribunal for Rwanda, the International Criminal Court, and national jurisdictions, this paper proposes refined criteria for determining when online content crosses the threshold from protected speech to criminal incitement. The findings reveal that courts increasingly recognize the amplifying power of social media while grappling with questions of intermediary liability, algorithmic amplification, and the temporal distance between incitement and criminal acts. This research contributes to ongoing debates about balancing freedom of expression with the imperative to prevent atrocity crimes in the digital age, offering practical recommendations for prosecutors, judges, and policymakers navigating this complex intersection of technology and international humanitarian law.

Keywords: incitement, international crimes, social media, criminal responsibility, causation, intent, genocide, crimes against humanity, freedom of expression, international criminal law.

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

1. Introduction

The advent of social media and digital communication platforms has revolutionized how information disseminates across borders, creating unprecedented opportunities for both connection and harm. Within the realm of international criminal law, these technological developments have introduced complex challenges regarding individual criminal responsibility for incitement to commit grave international crimes. The traditional legal frameworks developed in contexts of radio broadcasts and print media now confront questions about how tweets, Facebook posts, YouTube videos, and encrypted messaging apps can constitute criminal incitement under international law (Benesch, 2014). This transformation demands critical examination of the doctrinal elements that have historically defined incitement, particularly the thresholds of causation and intent that distinguish criminal speech from protected expression.

The prosecution of incitement as a distinct international crime gained prominence following the Rwandan genocide, where radio broadcasts and print media played instrumental roles in mobilizing mass violence against the Tutsi population. The International Criminal Tribunal for Rwanda (ICTR) established foundational jurisprudence on incitement to genocide, most notably in the landmark case of Prosecutor v. Nahimana et al., commonly known as the Media Case (ICTR, 2003). However, the factual circumstances underlying these precedents involved centralized media institutions operating within a confined geographical and temporal space, circumstances markedly different from the decentralized, instantaneous, and global nature of contemporary social media platforms (Gordon, 2014).

Today's digital ecosystem presents qualitatively different challenges for establishing individual criminal responsibility. Social media platforms enable individuals to reach millions of followers instantaneously, bypass traditional gatekeepers, and employ sophisticated techniques of persuasion and mobilization. The algorithmic curation of content can amplify inflammatory messages beyond the original speaker's intended audience, while anonymity features and encryption technologies complicate attribution and evidence gathering (Gagliardone et al., 2015). Furthermore, the transnational character of online communications raises jurisdictional questions and enforcement challenges that previous generations of international criminal lawyers rarely encountered.

Against this backdrop, this research investigates how international criminal law addresses individual responsibility for incitement conducted through social media and online platforms, with particular emphasis on defining the requisite thresholds of causation and intent. The central research questions guiding this inquiry include: First, how have international and domestic courts adapted traditional doctrines of incitement to address the unique characteristics of digital communication? Second, what causal connection must prosecutors establish between online incitement and subsequent international crimes? Third, how do courts assess the specific intent requirement in cases involving social media communications? Fourth, what role should intermediary platforms play in the analysis of criminal responsibility? Finally, how can legal

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

frameworks balance the prevention of atrocity crimes with the protection of freedom of expression in the digital public sphere?

The significance of this research extends beyond theoretical concerns to urgent practical implications. Recent years have witnessed numerous instances where social media platforms allegedly facilitated incitement to international crimes, from the persecution of Rohingya Muslims in Myanmar to ethnic violence in Ethiopia and the Central African Republic (Amnesty International, 2022). These cases underscore the necessity for clear legal standards that can guide prosecutorial strategy, judicial decision-making, and platform governance while respecting fundamental human rights.

This paper proceeds through several interconnected sections that build toward comprehensive analysis of individual criminal responsibility for online incitement. Following this introduction, the second section establishes the theoretical and legal foundations by examining the definition and elements of incitement under international criminal law, tracing its historical development and doctrinal evolution. The third section analyzes the specific challenges posed by social media platforms, including their technical characteristics, the role of algorithmic amplification, and questions of intermediary liability. The fourth section focuses on the causation threshold, exploring different theoretical approaches to establishing causal links between speech and subsequent crimes in the digital context. The fifth section examines the intent requirement, addressing how courts determine whether speakers possessed the specific intent to incite international crimes through online communications. The sixth section presents comparative case studies from international tribunals and domestic jurisdictions, identifying emerging patterns and persistent tensions in judicial reasoning. The seventh section addresses the critical balance between preventing incitement and protecting freedom of expression, considering both international human rights law and practical enforcement challenges. The final section synthesizes the findings and offers recommendations for refining legal standards applicable to individual criminal responsibility for online incitement.

2. Incitement Under International Criminal Law: Foundations and Elements

The prohibition of incitement to commit international crimes occupies a distinctive position within international criminal law, representing one of the rare instances where speech itself constitutes a prosecutable offense regardless of whether the incited crimes actually occur. This section examines the legal foundations of incitement, its constituent elements, and the evolution of relevant jurisprudence that informs contemporary applications to digital communications.

2.1 Historical Development and Legal Sources

The criminalization of incitement to genocide emerged from the ashes of the Holocaust, codified in Article III(c) of the 1948 Convention on the Prevention and Punishment of the Crime of Genocide (Genocide Convention). This provision explicitly prohibits "direct and public incitement to commit genocide" as a punishable act, even when no genocide ultimately occurs (United Nations, 1948). The drafters

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

recognized that preventing genocide required intervention before mass atrocities materialized, acknowledging that inflammatory speech could serve as a precursor to systematic violence (Schabas, 2009). This preventive rationale distinguishes incitement to genocide from other modes of liability that require actual commission of underlying crimes.

Beyond genocide, international criminal law addresses incitement through various liability theories applicable to crimes against humanity and war crimes. While these offenses lack specific provisions criminalizing incitement as a distinct offense, prosecutors can pursue incitement-related conduct through general principles of individual criminal responsibility, including ordering, instigating, or aiding and abetting (Ambos, 2013). The Rome Statute of the International Criminal Court (ICC) incorporates these modes of liability in Article 25, providing multiple pathways for holding individuals accountable for speech acts that contribute to international crimes (ICC, 1998).

International human rights law also addresses incitement, particularly through the International Covenant on Civil and Political Rights (ICCPR). Article 20(2) of the ICCPR mandates that states prohibit "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence" (United Nations, 1966). This provision creates tension between free speech protections under Article 19 and the duty to prevent dangerous advocacy, a tension that permeates contemporary debates about online incitement (Mendel, 2012).

2.2 Essential Elements of Incitement to Genocide

Incitement to genocide, the most developed area of incitement law, requires proof of several essential elements derived from the Genocide Convention and elaborated through tribunal jurisprudence. The ICTR's Media Case established authoritative guidance on these elements, which subsequent courts have refined and applied (ICTR, 2003).

First, the speech must constitute a direct call to commit genocide. The directness requirement distinguishes criminal incitement from abstract advocacy or general expressions of hatred. According to the ICTR Appeals Chamber in Nahimana, direct incitement "assumes a special meaning in the light of African, and particularly Rwandan, culture" where communication often employs allusion, metaphor, and coded language (ICTR, 2007, para. 692). Courts must therefore examine the cultural and linguistic context to determine whether reasonable persons in the specific circumstances would understand the communication as a call to commit genocide. This culturally-sensitive interpretation complicates analysis of social media content that circulates across diverse cultural contexts with varying interpretive frameworks.

Second, the incitement must occur publicly. The public element ensures that the speech reaches a potentially significant audience capable of committing the incited acts. Traditional media broadcasts clearly satisfied this requirement, but digital communications present more nuanced questions about what constitutes public dissemination. Courts have generally interpreted "public" broadly to include any

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

communication accessible to the general public or a substantial group, even if initially directed to a limited audience (Benesch, 2008). Social media posts, even those shared within ostensibly private groups, may qualify as public when they reach audiences beyond the immediate circle of the speaker's intended recipients.

Third, prosecutors must establish that the accused possessed specific intent, or dolus specialis, to incite others to commit genocide. This mens rea requirement demands proof that the speaker consciously desired to provoke genocide, going beyond general knowledge that violence might result from inflammatory rhetoric. The specific intent requirement creates significant evidentiary challenges in social media cases where communications may appear ambiguous or susceptible to multiple interpretations (Mettraux, 2005). Courts examine various factors to infer intent, including the speaker's explicit statements, the context in which communications occurred, the speaker's position of authority or influence, and patterns of repeated inflammatory messaging.

Fourth, although not universally required, some jurisprudence suggests that the speech must create a clear and present danger of genocide actually occurring. This element, drawn partly from domestic free speech jurisprudence, considers whether the circumstances surrounding the incitement made violence an imminent and realistic possibility (Nybondas, 2014). The ICTR has not consistently applied this requirement, but considerations of temporal proximity, audience receptivity, and capability to commit genocide often factor into judicial assessments of whether particular speech crossed the threshold into criminal incitement.

2.3 Incitement Related to Crimes Against Humanity and War Crimes

Unlike genocide, crimes against humanity and war crimes lack explicit provisions criminalizing incitement as a separate offense. However, international criminal law addresses such conduct through alternative liability theories that capture the contribution of inflammatory speech to these crimes.

Instigation represents the most analogous mode of liability to incitement. Article 25(3)(b) of the Rome Statute provides for liability when an individual "orders, solicits or induces the commission" of crimes within the Court's jurisdiction (ICC, 1998). The International Criminal Tribunal for the former Yugoslavia (ICTY) defined instigation as prompting another to commit an offense, which can include public speeches that incite others to criminal conduct (ICTY, 1999). Unlike incitement to genocide, instigation generally requires proof that the instigation had a substantial effect on the commission of the crime, establishing a causal nexus between speech and subsequent criminal acts (Cassese, 2008).

The concept of persecution as a crime against humanity also addresses certain forms of hateful expression. Persecution involves the intentional and severe deprivation of fundamental rights on discriminatory grounds, and speech can constitute persecution when it creates an atmosphere of fear, hostility, and discrimination that fundamentally undermines human dignity (ICTY, 2001). The ICTY's Streicher jurisprudence, drawing upon the Nuremberg precedent regarding Julius Streicher's anti-Semitic

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

publications, recognizes that systematic hate propaganda can constitute persecution even absent direct incitement to violence (Klamberg, 2013). This mode of liability extends criminal responsibility to speech that may not directly call for violence but nonetheless contributes to a discriminatory attack on civilian populations.

Contribution to joint criminal enterprise or common purpose liability may also capture incitement-related conduct. When multiple individuals participate in a common criminal plan, those who make significant contributions through inflammatory speech may bear responsibility for crimes committed in furtherance of that plan (ICTY, 2003). This theory proves particularly relevant for cases involving coordinated disinformation campaigns or organized hate speech initiatives conducted through social media platforms.

2.4 Protected Speech and Legal Limitations

International criminal law's treatment of incitement necessarily confronts fundamental questions about the boundaries of protected expression. The Universal Declaration of Human Rights and the ICCPR enshrine freedom of expression as a fundamental right, subject only to narrowly tailored restrictions necessary to protect specified interests including public order, national security, and the rights of others (United Nations, 1948, 1966). The Rabat Plan of Action, adopted by the United Nations in 2012, provides guidance on distinguishing protected expression from prohibited incitement, emphasizing a six-part threshold test that examines context, speaker, intent, content, extent of dissemination, and likelihood of harm (United Nations, 2013).

Courts analyzing incitement claims must carefully balance these competing interests, ensuring that criminalization does not unduly chill legitimate political discourse, criticism of government policies, or advocacy for social change. The European Court of Human Rights has developed extensive jurisprudence distinguishing protected political speech from dangerous incitement, generally requiring clear evidence that speech created an imminent risk of violence or discrimination (European Court of Human Rights, 1976). This proportionality analysis becomes particularly complex in social media contexts where viral content can simultaneously advance legitimate debate and fuel dangerous mobilization.

Academic and political discourse about contentious historical events, including genocide and mass atrocities, receives heightened protection even when such discussions cause offense or distress to affected communities. International human rights law generally prohibits criminalization of genocide denial or historical revisionism absent additional elements demonstrating intent to incite hatred or violence (Buyse, 2014). However, several European jurisdictions have enacted laws criminalizing denial of specific historical genocides, creating ongoing tension between different conceptions of permissible speech restrictions.

3. Social Media Platforms and the Transformation of Incitement

The migration of public discourse to social media platforms has fundamentally altered the dynamics of incitement, introducing new mechanisms for dissemination,

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

amplification, and mobilization that challenge traditional legal frameworks. This section examines the distinctive characteristics of digital communication that affect analysis of individual criminal responsibility for incitement.

3.1 Technical Characteristics of Digital Communication

Social media platforms differ from traditional broadcast media in several legally significant respects. First, these platforms enable peer-to-peer communication without requiring access to centralized infrastructure, democratizing speech but also removing editorial gatekeeping that historically filtered inflammatory content (Sunstein, 2017). Any individual with internet access can potentially reach global audiences, fundamentally transforming who can engage in mass communication.

Second, social media facilitates unprecedented speed and scale of dissemination. Content can achieve viral spread within hours, reaching millions across multiple jurisdictions before authorities can respond (Starbird, 2019). This velocity challenges legal systems designed around slower information flows where authorities had opportunities to interdict dangerous speech before it reached critical mass. The instantaneous nature of digital communication compresses the temporal distance between incitement and potential criminal acts, intensifying questions about causation and foreseeability.

Third, algorithmic curation shapes content exposure through personalization and recommendation systems. Platform algorithms optimize for engagement, often amplifying emotionally charged content that generates strong reactions (Tufekci, 2018). This algorithmic amplification may spread inciteful content beyond what the original speaker intended or could have anticipated, raising questions about how to attribute responsibility for consequences of algorithmically-driven dissemination.

Fourth, social media enables targeted micro-messaging to specific demographic groups while maintaining plausible deniability for broader audiences. Speakers can employ dog whistles, coded language, and in-group references that communicate inciteful messages to receptive audiences while appearing innocuous to outsiders (Scheffler, 2020). This capacity for strategic ambiguity complicates evidentiary assessments of both the content of speech and the speaker's intent.

Fifth, platforms provide tools for coordination and organization that transform incitement from passive reception of messages into active mobilization. Features such as event creation, group messaging, and location-based services enable speakers to move beyond general calls for violence toward concrete planning and coordination (Zeitzoff, 2017). This organizational capacity blurs boundaries between incitement and more direct forms of participation in criminal enterprises.

3.2 Anonymity, Pseudonymity, and Attribution Challenges

Many social media platforms permit anonymous or pseudonymous accounts, creating significant obstacles for identifying individuals responsible for inciteful content. While attribution challenges existed in traditional media through unsigned pamphlets and anonymous radio broadcasts, digital technologies enable much more sophisticated

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

concealment of identity through virtual private networks, encrypted communications, and distributed networks (Shulman, 2018).

These attribution difficulties affect both the investigation and prosecution of incitement cases. Prosecutors must first identify the individuals behind anonymous accounts, often requiring cooperation from platform providers who may resist disclosure based on privacy concerns or jurisdictional objections. Even when platforms cooperate, technical limitations may prevent definitive attribution, particularly when sophisticated actors employ anti-forensic techniques designed to obscure their digital footprints (Kerr & Schneier, 2009).

The possibility of account hijacking, impersonation, and coordinated inauthentic behavior further complicates attribution. Prosecutors must establish not only that inciteful content appeared on a particular account but also that the accused individual actually controlled that account and authored the relevant communications (Jardine, 2018). This evidentiary burden becomes particularly challenging when defendants claim their accounts were hacked or that inciteful posts were created by others with access to their devices.

3.3 Transnational Reach and Jurisdictional Complexity

Social media platforms operate across borders, enabling speakers in one jurisdiction to incite crimes in distant locations while remaining physically removed from the sites of violence. This transnational character raises complex questions about jurisdiction, enforcement, and the applicable substantive law governing incitement (Ryngaert, 2015).

Traditional principles of territorial jurisdiction prove inadequate when the speaker, platform, audience, and site of incited crimes span multiple countries. The effects doctrine permits states to exercise jurisdiction over extraterritorial conduct that produces harmful effects within their territory, potentially enabling prosecution of foreign nationals who incite domestic violence through online platforms (Colangelo, 2014). However, this approach risks creating conflicts of jurisdiction and may enable authoritarian regimes to assert criminal jurisdiction over protected political speech occurring in liberal democracies.

Questions also arise regarding which state's law governs incitement cases with international dimensions. While international criminal tribunals apply international law directly, domestic prosecutions may invoke either domestic incitement statutes or direct application of international criminal law principles (Sadat, 2012). The choice of law can significantly affect the applicable mens rea standards, evidentiary requirements, and available defenses, creating potential forum shopping incentives and inconsistent outcomes across jurisdictions.

3.4 Platform Architecture and Intermediary Liability

The role of social media platforms themselves presents novel questions about intermediary liability that intersect with individual criminal responsibility. Platforms provide the infrastructure that enables incitement to reach vast audiences, yet they

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

typically claim protection under safe harbor provisions that shield intermediaries from liability for user-generated content (Citron & Wittes, 2017).

Most jurisdictions distinguish between active publishers who exercise editorial control and passive conduits who merely transmit third-party content. Social media platforms occupy an ambiguous position between these categories, as they curate and algorithmically amplify content without exercising traditional editorial judgment (Gillespie, 2018). This ambiguity affects questions about whether platforms owe duties to moderate inciteful content and whether their failure to remove such content can constitute complicity in subsequent crimes.

Recent developments suggest increasing recognition of platform responsibilities regarding incitement. The United Nations Guiding Principles on Business and Human Rights establish that private actors, including technology companies, bear responsibility to respect human rights and remediate adverse impacts associated with their operations (United Nations, 2011). Several jurisdictions have enacted legislation requiring platforms to expeditiously remove illegal content, including incitement, upon notice (Kaye, 2019). However, these regulatory approaches raise concerns about privatized censorship, over-removal of protected speech, and the capacity of platforms to accurately assess complex questions of criminal incitement.

The intersection of platform liability and individual criminal responsibility becomes particularly complex when analyzing whether platform decisions to allow, promote, or remove content affect the criminal culpability of individual speakers. If a platform's algorithms amplify inciteful content to audiences far beyond what the speaker could have reached organically, should this algorithmic intervention affect assessments of the speaker's mens rea or the causation analysis? Conversely, if platforms implement moderation policies that limit the reach of inflammatory content, does this mitigation affect the speaker's criminal liability?

4. The Causation Threshold in Digital Incitement Cases

Establishing the requisite causal connection between incitement and subsequent international crimes presents one of the most significant challenges in prosecuting online speech. This section examines different approaches to causation in incitement cases and their application to social media contexts.

4.1 Theoretical Approaches to Causation in Incitement

International criminal jurisprudence has developed several approaches to the causation requirement in incitement cases, reflecting ongoing tension between strict causal standards and the preventive objectives of incitement law. These approaches differ in the degree of causal connection required and in their treatment of cases where incitement does not produce immediate criminal acts.

The substantial contribution standard requires prosecutors to demonstrate that the incitement made a substantial contribution to the commission of the underlying crime. This approach, articulated in various ICTY decisions regarding instigation, demands proof that "there is a sufficient link between the instigation and the physical

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

perpetration of the crime" such that without the instigation, the crime would not have occurred in the same manner (ICTY, 2004, para. 442). Applied to social media incitement, this standard would require showing that the accused's online communications substantially contributed to decisions by others to commit international crimes.

The but-for causation test asks whether the crime would have occurred absent the incitement. This approach sets a high evidentiary bar, as prosecutors must prove a counterfactual proposition about what would have happened in hypothetical circumstances where the incitement did not occur (Jain, 2014). In complex social media environments with multiple sources of inflammatory content, establishing but-for causation becomes extremely difficult, as courts must determine whether other speakers or factors would have independently produced the same criminal outcome.

The risk creation approach focuses on whether the incitement created a substantial risk that crimes would be committed, regardless of whether such crimes actually materialized. This standard, consistent with the preventive purpose of criminalizing incitement to genocide, allows conviction when speech creates dangerous conditions even if no genocide occurs (Gordon, 2013). For social media cases, this approach would examine whether online communications generated sufficient risk of international crimes based on factors such as audience reach, receptivity to mobilization, and capability to commit violence.

Some scholars advocate for a "substantial likelihood" standard that balances preventive objectives with rule of law concerns about criminalizing speech that never produces harmful results. This approach would require prosecutors to prove that the incitement created a substantial likelihood of criminal acts occurring, considering both the nature of the speech and the circumstances surrounding its dissemination (Waldron, 2012). Social media platforms' capacity for viral spread and rapid mobilization might make substantial likelihood easier to establish than under traditional media, though proving likelihood remains inherently speculative.

4.2 Temporal Distance and Intermediate Causation

Social media incitement often involves significant temporal distance between inflammatory speech and subsequent criminal acts, complicating causal analysis. Unlike incitement that immediately precedes violence, online content may circulate for extended periods, gradually building animosity and preparing audiences for eventual mobilization (Straus, 2007). This temporal gap raises questions about how long causal chains can extend before the connection between incitement and crime becomes too attenuated to support criminal liability.

Intermediate causation presents particular challenges when multiple factors contribute to criminal decisions. Perpetrators who commit international crimes after exposure to online incitement also respond to numerous other influences, including local political conditions, peer pressure, economic grievances, and their own predispositions toward violence (Harff, 2003). Isolating the causal contribution of specific social media posts from this complex web of causation requires sophisticated analysis that courts may

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

struggle to conduct rigorously.

The presence of intervening actors further complicates causation. When inciteful content passes through multiple intermediaries who amplify, modify, or contextualize the original message, determining whose speech caused subsequent crimes becomes conceptually challenging. Social media's viral nature means that content often reaches ultimate perpetrators through chains of sharing and reposting that significantly alter the message's context and meaning (Marwick & Lewis, 2017). Courts must decide whether to attribute causal responsibility to the original speaker, the intermediaries who amplified the message, or both.

4.3 Probabilistic Causation in Mass Incitement Scenarios

Mass incitement through social media presents distinct causation challenges because inflammatory content reaches vast audiences, only some of whom may commit crimes in response. When speakers broadcast inciteful messages to millions of followers, establishing which specific criminal acts resulted from the incitement and which would have occurred independently proves methodologically difficult (Yanagizawa-Drott, 2014).

Some scholars propose probabilistic approaches to causation that would hold speakers responsible when their incitement increases the likelihood of crimes occurring, even if specific causal pathways cannot be traced. This approach draws upon statistical reasoning to infer that inflammatory content contributed to an overall increase in violence, without requiring proof of causation regarding particular criminal acts (Padilla, 2020). For example, if researchers could demonstrate that areas with higher exposure to inciteful social media content experienced significantly elevated rates of atrocity crimes, this correlation might support probabilistic causal inferences.

However, probabilistic causation raises significant rule of law concerns. Criminal liability traditionally requires proof beyond reasonable doubt that the accused caused or contributed to specific criminal acts, not merely that the accused's conduct correlated with increased crime rates (Ashworth & Horder, 2013). Convicting individuals based on statistical associations rather than particularized proof of causation arguably violates fundamental principles of individual culpability and fair notice.

4.4 Evidentiary Challenges in Proving Digital Causation

Proving causation in social media incitement cases requires prosecutors to overcome substantial evidentiary obstacles. Unlike traditional media where audiences could be identified and interviewed about whether they heard inflammatory broadcasts and how those broadcasts influenced their conduct, social media audiences are often anonymous, dispersed, and difficult to locate (Greenawalt, 1989).

Digital evidence trails provide some advantages for reconstruction of causal pathways. Prosecutors can potentially demonstrate that perpetrators accessed specific inciteful content through social media platforms' data logs, showing temporal correlation between content exposure and subsequent criminal activity. Platform

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

analytics can reveal the reach and engagement of inflammatory posts, establishing that content reached audiences capable of committing crimes (Faris et al., 2016). However, platform cooperation with investigators varies widely based on jurisdiction, corporate policies, and technical capabilities.

Perpetrator testimony regarding motivations presents another evidentiary approach but introduces reliability concerns. Individuals accused of international crimes may have incentives to attribute their actions to incitement by others rather than accepting personal responsibility (Osiel, 2009). Conversely, some perpetrators may minimize the influence of online incitement to protect speakers with whom they share ideological commitments. Courts must carefully evaluate such testimony in light of potential biases and corroborate self-serving claims about causation through independent evidence.

Expert testimony about the effects of inflammatory speech on audience behavior can help establish causal connections, though such testimony often remains contested. Social scientists, psychologists, and media experts can offer opinions about how inflammatory content influences cognition, emotion, and behavior, potentially bridging evidentiary gaps (Braddock & Horgan, 2016). However, the methodological challenges of studying incitement's causal effects limit the certainty with which experts can opine about whether specific speech caused particular criminal acts.

5. The Intent Requirement in Online Incitement

Proving the specific intent element of incitement presents unique challenges in social media contexts where communications may be ambiguous, sarcastic, or susceptible to multiple interpretations. This section examines how courts assess intent in digital incitement cases and the evidentiary factors that inform these determinations.

5.1 Specific Intent versus General Intent

Incitement to genocide requires proof of specific intent, or dolus specialis, meaning that the accused must have consciously desired to bring about genocide through their communications. This mens rea standard exceeds the general intent required for many other international crimes, reflecting the gravity of attempting to destroy protected groups and the need to distinguish criminal incitement from inflammatory but protected speech (Greenawalt, 1999).

The specific intent requirement creates high evidentiary burdens for prosecutors, particularly when social media communications employ coded language, humor, or ambiguity that obscures the speaker's true purpose. Speakers may intentionally maintain plausible deniability by framing inflammatory content as satire, political commentary, or hypothetical speculation while simultaneously conveying inciteful messages to receptive audiences (Massaro & Norton, 2016). Courts must penetrate these rhetorical strategies to determine whether speakers actually intended to provoke genocide or merely engaged in offensive but protected expression.

For crimes against humanity and war crimes prosecuted through instigation or other liability theories, the intent requirements vary depending on the specific mode of

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

liability alleged. Instigation generally requires proof that the accused intended to provoke the commission of crimes, though some authorities suggest that reckless indifference may suffice (Ambos, 2013). This lower mens rea threshold potentially encompasses social media speakers who broadcast inflammatory content knowing it might provoke violence but without specifically desiring such outcomes.

5.2 Inferring Intent from Digital Communications

In the absence of explicit confessions of intent, courts must infer speakers' mental states from available evidence, including the content and context of their communications. Social media platforms generate vast digital records that can inform intent analysis, though interpreting these records requires careful attention to linguistic and cultural context (Benesch et al., 2020).

The content of inflammatory posts provides primary evidence of intent. Direct calls for violence against protected groups, use of dehumanizing language, and explicit endorsement of genocide or mass atrocities strongly suggest specific intent to incite such crimes (Leader Maynard, 2014). However, many social media communications employ indirection and coded language that requires interpretation. Courts examine the overall message conveyed by communications, considering how reasonable members of the target audience would understand the content rather than applying overly literal readings that might miss implied meanings.

Patterns of repeated messaging support inferences of intent. Individuals who consistently post inflammatory content targeting particular groups over extended periods demonstrate systematic commitment to inciteful messaging that exceeds isolated inflammatory outbursts (Straus, 2007). Social media platforms' permanent records enable prosecutors to document these patterns comprehensively, presenting courts with longitudinal evidence of the accused's sustained focus on provoking violence.

The speaker's position and influence affect intent analysis. Individuals in positions of authority or with substantial social media followings who broadcast inflammatory content demonstrate greater culpability than ordinary citizens making similar statements with minimal reach (Nahimana et al., 2007). Platform analytics documenting follower counts, engagement rates, and viral spread help establish that speakers understood their capacity to influence audiences and knowingly exploited that influence for inciteful purposes.

5.3 Circumstantial Evidence and Contextual Analysis

Beyond the content of specific communications, courts examine broader contextual factors to infer intent. The circumstances surrounding social media posts can reveal whether speakers aimed to incite international crimes or merely engaged in heated rhetoric within broader political debates (Mchangama & Alkiviadou, 2021).

Temporal correlation between inflammatory posts and outbreaks of violence provides circumstantial evidence of intent. When speakers escalate inflammatory messaging immediately before episodes of mass violence, this timing suggests purposeful

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

mobilization rather than coincidental commentary (Yanagizawa-Drott, 2014). However, correlation does not establish causation or intent without additional evidence that the speaker anticipated and desired the violent outcomes.

The speaker's knowledge of on-the-ground conditions informs assessments of intent. Individuals who post inflammatory content while aware that audiences possess capability and readiness to commit violence demonstrate greater culpability than those addressing purely hypothetical scenarios (Gordon, 2017). Social media communications sometimes reference current events, local tensions, or preparatory activities that establish the speaker's awareness of volatile conditions, supporting inferences that inflammatory posts aimed to ignite imminent violence.

Responses to violence that follows inflammatory posts can reveal retrospective evidence of intent. Speakers who celebrate, encourage, or call for intensification of violence after it begins demonstrate approval of criminal outcomes that tends to corroborate initial intent to provoke such acts (Benesch, 2014). Conversely, speakers who express shock, condemnation, or regret about violence may undermine prosecutorial claims that they specifically intended to incite crimes, though such post-hoc disavowals require skeptical evaluation.

5.4 Distinguishing Intent from Recklessness and Negligence

The boundary between specific intent and lesser mental states proves critical for determining criminal liability. Social media speakers who broadcast inflammatory content without specifically desiring to provoke international crimes but with awareness that such crimes might result occupy a grey area between criminal incitement and irresponsible but lawful speech (Elagab, 2020).

Some speakers may act with reckless indifference to whether their inflammatory posts provoke violence, focusing primarily on gaining followers, engagement, or political influence rather than specifically seeking to incite crimes. These speakers foresee violence as a possible consequence of their communications but lack the purposeful desire required for specific intent (Saul, 2019). Whether recklessness suffices for liability depends on the specific legal framework applied, with incitement to genocide requiring specific intent while some forms of persecution or instigation may accept reckless mens rea.

Negligent or merely offensive speech falls outside criminal incitement's scope entirely. Speakers who fail to appreciate that their communications might provoke violence, even when such failure reflects poor judgment or insensitivity, lack the mens rea required for criminal responsibility (Novic, 2016). The challenge lies in distinguishing between genuinely negligent speakers unaware of their speech's dangerous potential and sophisticated actors who feign ignorance while deliberately inciting violence through strategic ambiguity.

6. Comparative Case Studies: Emerging Patterns in Jurisprudence

Examination of how international tribunals and domestic courts have addressed incitement through media and digital platforms reveals emerging patterns, persistent

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

tensions, and areas requiring further doctrinal development. This section analyzes representative cases that illustrate the application of incitement law to various forms of mass communication.

6.1 The ICTR Media Case: Foundation for Modern Incitement Law

Prosecutor v. Nahimana, Barayagwiza, and Ngeze remains the most comprehensive international judgment on media incitement to genocide. The case involved three defendants who used radio broadcasts and print media to incite genocide against Tutsi populations in Rwanda during 1994. Ferdinand Nahimana directed Radio Télévision Libre des Mille Collines (RTLM), Jean-Bosco Barayagwiza served as a founder and policy maker for the radio station, and Hassan Ngeze edited the newspaper Kangura (ICTR, 2003).

The Trial Chamber found that RTLM broadcasts explicitly called for extermination of Tutsi civilians, using dehumanizing language that characterized Tutsis as "inyenzi" (cockroaches) and "snakes" deserving elimination. The radio station provided real-time coordination for genocidal violence, directing killers to specific locations where Tutsi populations had sought refuge. The court held that this conduct satisfied the directness requirement despite occasional use of euphemisms and coded language, applying contextual interpretation that considered how Rwandan audiences would understand the broadcasts (ICTR, 2003).

Regarding intent, the Trial Chamber inferred specific intent from the systematic nature of inflammatory broadcasts, the defendants' knowledge of ongoing genocide, and their continued incitement despite witnessing its deadly effects. The court rejected defense arguments that broadcasts merely reflected legitimate journalism or political commentary, finding that the content transcended protected speech by explicitly calling for genocide (ICTR, 2003, para. 970).

The Appeals Chamber modified certain aspects of the Trial Chamber's analysis while affirming the core findings. Importantly, the Appeals Chamber clarified that the directness requirement must be assessed in light of cultural and linguistic context, acknowledging that communications employing allusion or implication may constitute direct incitement when audiences clearly understand them as calls for genocide (ICTR, 2007). This culturally-sensitive interpretive approach has significant implications for social media cases involving communications across diverse linguistic and cultural contexts.

The Media Case established several principles relevant to contemporary digital incitement. First, the scale and systematicity of inflammatory communications supports inferences of specific intent and demonstrates the direct nature of incitement. Second, perpetrators cannot shield themselves from liability by using euphemisms or coded language when contextual analysis reveals clearly inciteful meanings. Third, speakers in positions of media authority bear enhanced responsibility for the content they disseminate. Fourth, continuing to broadcast inflammatory content after violence begins provides strong evidence of intent to incite further crimes.

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

6.2 ICTR Jurisprudence on Individual Speech Acts

Beyond the Media Case, the ICTR addressed incitement through various cases involving individual speakers who used public forums to encourage genocide. In Prosecutor v. Akayesu, the first international conviction for incitement to genocide, the Trial Chamber found that a local mayor's public speeches calling for extermination of Tutsis constituted direct and public incitement (ICTR, 1998). Akayesu addressed public gatherings and made statements that community members understood as authorization and encouragement for killing Tutsi neighbors.

The Akayesu judgment emphasized that even single speech acts can constitute incitement when circumstances indicate that the communication directly called for genocide and reached audiences capable of committing such crimes (ICTR, 1998). This principle applies to social media contexts where individual posts can reach massive audiences instantaneously, potentially triggering liability even absent sustained campaigns of inflammatory messaging.

Prosecutor v. Ruggiu involved a Belgian radio broadcaster who worked for RTLM and made numerous broadcasts inciting violence against Tutsis. The Trial Chamber found that Ruggiu's broadcasts, which explicitly called listeners to "wipe them out" and "exterminate them," constituted direct incitement to genocide (ICTR, 2000). The court rejected defenses based on claims that broadcasts reflected RTLM's editorial policies rather than Ruggiu's personal views, holding individuals responsible for incitement they personally broadcast regardless of institutional context.

These cases establish that individual criminal responsibility attaches to persons who make inciteful communications, even when operating within larger institutional frameworks. For social media contexts, this principle means that individual users who post inflammatory content cannot escape liability by arguing they merely followed online communities' norms or amplified content created by others.

6.3 ICC Preliminary Examinations and Investigations

While the ICC has not yet prosecuted completed cases focused primarily on social media incitement, various preliminary examinations and investigations have considered the role of online platforms in facilitating international crimes. The Office of the Prosecutor's examination of the situation in Kenya following the 2007-2008 post-election violence considered allegations that text messages and emails had been used to coordinate ethnic violence (ICC OTP, 2010). Though the Kenya cases ultimately focused on other forms of liability, the preliminary examination recognized that digital communications could constitute relevant evidence of criminal planning and coordination.

More recently, the ICC's investigation into alleged crimes against the Rohingya in Myanmar has examined Facebook's role in disseminating hate speech and incitement to violence against Rohingya Muslims. Investigative reports documented systematic use of Facebook to spread inflammatory content characterizing Rohingya as illegal immigrants and national security threats (Amnesty International, 2022). While the investigation focuses primarily on senior military and civilian officials, the

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

examination of social media's role illustrates growing recognition that digital platforms facilitate incitement to international crimes.

The ICC's 2021 Policy on Children emphasizes that online exploitation and abuse of children, including through social media platforms, falls within the Court's jurisdiction when connected to situations under investigation (ICC, 2016). This policy acknowledges that digital technologies create new modalities for committing international crimes that require prosecutorial attention, though the policy does not specifically address incitement through social media.

6.4 Domestic Prosecutions of Social Media Incitement

Several domestic jurisdictions have prosecuted individuals for using social media to incite violence or hatred, though most cases do not involve international crimes specifically. These domestic prosecutions nevertheless provide valuable precedents for understanding how courts assess online communications under incitement frameworks.

In the United Kingdom, various prosecutions under the Public Order Act 1986 have addressed social media posts inciting racial or religious hatred. R v. Sheppard and Whittle involved defendants who operated websites publishing racist material designed to stir up racial hatred (2010). The court found that website publications constituted public dissemination even though accessed primarily by like-minded extremists, establishing that online content need not reach mainstream audiences to qualify as public incitement.

German courts have prosecuted numerous cases involving online hate speech and incitement under provisions of the Criminal Code prohibiting incitement to hatred and public approval of crimes. These cases illustrate how civil law jurisdictions navigate tensions between free expression and prevention of dangerous speech, generally permitting prosecution when online content creates concrete dangers of violence or severely undermines human dignity (Herz & Molnar, 2012).

In Rwanda itself, domestic courts have prosecuted several individuals for incitement to genocide involving social media or online forums. These cases demonstrate continued application of incitement law in digital contexts, though concerns about due process and political instrumentalization of prosecutions complicate assessment of these domestic proceedings (Waldorf, 2009).

6.5 Cases from Other Regions

Beyond Europe and Africa, courts in other regions have addressed digital incitement in various contexts. In India, numerous prosecutions have targeted social media users accused of posting inflammatory content designed to provoke communal violence. While many of these prosecutions raise free speech concerns and allegations of politically motivated enforcement, they illustrate widespread recognition that social media can facilitate dangerous incitement (Nayak, 2020).

In the United States, First Amendment protections create high barriers to prosecuting incitement, generally requiring proof that speech is directed to inciting imminent

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

lawless action and likely to produce such action (Brandenburg v. Ohio, 1969). This demanding standard has limited prosecutions of online incitement, though cases involving material support for terrorism have addressed some forms of online radicalization and mobilization (Cole, 2003). The U.S. approach illustrates one end of the spectrum regarding permissible restrictions on inflammatory speech, contrasting with jurisdictions that permit earlier intervention.

7. Balancing Prevention and Expression: The Human Rights Framework

Any legal regime governing incitement must navigate the fundamental tension between preventing atrocity crimes and protecting freedom of expression. This section examines how international human rights law seeks to balance these competing interests and the implications for social media regulation.

7.1 Freedom of Expression under International Law

Freedom of expression receives robust protection under international human rights instruments, including Article 19 of the Universal Declaration of Human Rights and Article 19 of the ICCPR. These provisions establish that everyone has the right to hold opinions without interference and to seek, receive, and impart information through any media regardless of frontiers (United Nations, 1948, 1966). The Inter-American Convention on Human Rights and the European Convention on Human Rights provide similar protections, establishing freedom of expression as a cornerstone of democratic societies (Organization of American States, 1969; Council of Europe, 1950).

International human rights law recognizes that freedom of expression serves multiple essential functions, including enabling democratic self-governance, facilitating the search for truth, promoting individual autonomy and dignity, and checking governmental abuses of power (Barendt, 2005). Protection extends not only to popular or inoffensive speech but also to expression that offends, shocks, or disturbs, as pluralistic societies must tolerate diverse viewpoints even when controversial (European Court of Human Rights, 1976).

However, freedom of expression is not absolute. Article 19(3) of the ICCPR permits restrictions on expression when necessary to protect the rights or reputations of others, national security, public order, or public health or morals (United Nations, 1966). Restrictions must meet tests of legality, legitimate purpose, and proportionality, ensuring that limitations on expression represent the least restrictive means of achieving compelling governmental interests (Mendel, 2012).

Article 20(2) of the ICCPR specifically requires states to prohibit "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence" (United Nations, 1966). This provision creates an affirmative duty to restrict certain forms of expression, but the scope of this obligation has generated considerable debate. The United Nations Human Rights Committee has clarified that Article 20(2) does not mandate criminalization of all hateful speech but

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

rather requires prohibition of severe incitement meeting high thresholds (UN Human Rights Committee, 2011).

7.2 The Rabat Plan of Action and Threshold Tests

Recognizing widespread confusion about where to draw lines between protected expression and prohibited incitement, the United Nations Office of the High Commissioner for Human Rights convened expert workshops that produced the Rabat Plan of Action in 2012. This guidance provides a six-part threshold test for identifying expression that states may legitimately prohibit as incitement to hatred (United Nations, 2013).

The first factor examines the social and political context in which expression occurs. Speech that might constitute dangerous incitement in volatile circumstances may represent protected political debate in stable contexts. Courts must assess whether tensions, discrimination, or violence against targeted groups create conditions where inflammatory speech could trigger harm (Article 19, 2012). For social media cases, this requires examining the environment both where speakers post content and where audiences receive messages, which may differ substantially given platforms' global reach.

The second factor considers the position or status of the speaker. Individuals in positions of authority, including government officials, religious leaders, or influential media personalities, bear greater responsibility for their communications than ordinary citizens. Social media introduces complexity to this factor because influence derives from follower counts and algorithmic amplification rather than traditional institutional authority (Benesch et al., 2020). Courts must assess whether speakers understood their capacity to shape audience behavior through online platforms.

The third factor addresses the speaker's intent to incite hatred, discrimination, or violence. This element aligns with criminal law mens rea requirements, demanding proof that speakers consciously aimed to provoke harmful outcomes. As discussed previously, intent may be inferred from content, context, and patterns of behavior, but prosecutors must demonstrate purposeful incitement rather than reckless disregard or negligence (United Nations, 2013).

The fourth factor examines the content and form of expression. Speech explicitly and directly calling for violence against targeted groups more clearly crosses the threshold into prohibited incitement than abstract expressions of hatred or criticism. The Rabat Plan emphasizes that context matters for assessing content, as coded language or dog whistles may constitute incitement when audiences understand implied meanings (United Nations, 2013). Social media's propensity for viral spread may intensify the harmful character of content by exposing larger audiences to inflammatory messaging.

The fifth factor considers the extent of dissemination. Expression reaching vast audiences through mainstream media or viral social media posts poses greater risks than communications within small private groups. However, the Rabat Plan cautions against assuming that limited initial audiences prevent speech from qualifying as

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

incitement, as content may subsequently spread to broader publics (United Nations, 2013). This consideration proves particularly relevant for social media content that can achieve unexpected viral reach.

The sixth factor assesses the likelihood of harm occurring, including imminence of potential violence or discrimination. This element requires examination of whether circumstances made harm probable rather than merely possible, considering audience receptivity and capability to commit violence (United Nations, 2013). The temporal compression enabled by social media may increase imminence, as online mobilization can produce rapid escalation from inflammatory speech to collective violence.

7.3 Proportionality Analysis and Least Restrictive Means

Even when expression satisfies the Rabat threshold test, states must ensure that restrictions represent proportionate responses that employ the least restrictive means available. Proportionality analysis requires careful balancing of the severity of threatened harm against the degree of interference with expression rights (Cohen-Eliya & Porat, 2013).

Criminal prosecution represents the most severe form of speech restriction and should be reserved for the most serious forms of incitement creating genuine risks of grave harm. Lesser restrictions, including civil liability, content removal, counter-speech initiatives, or community-based interventions, may adequately address many instances of inflammatory expression without resorting to criminalization (Gagliardone et al., 2015).

The availability of counter-speech as an alternative to prohibition affects proportionality assessments. When marketplace of ideas mechanisms can effectively refute inflammatory messaging and prevent mobilization to violence, criminalization may prove disproportionate (Post, 2009). However, counter-speech strategies face limitations in situations involving systematic propaganda campaigns, where targeted groups lack equal access to communication channels, or when violence occurs so rapidly that refutation cannot prevent harm.

7.4 Platform Governance and Content Moderation

The private governance of speech through social media platform content moderation policies introduces additional dimensions to the expression-incitement balance. Platforms operate under terms of service that typically prohibit hate speech and incitement while claiming to support free expression values (Klonick, 2017). However, these corporate speech policies do not necessarily align with international human rights standards, creating potential for both under-enforcement that permits dangerous incitement and over-enforcement that suppresses protected expression.

The Santa Clara Principles and other initiatives advocate for procedural safeguards in content moderation, including transparency regarding moderation rules, notice to users about content removal, and opportunities for meaningful appeal (Santa Clara Principles, 2018). These procedural protections help ensure that platform content moderation respects expression rights while addressing harmful speech, though

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

implementation remains inconsistent across platforms and jurisdictions.

Recent regulatory initiatives, including the European Union's Digital Services Act, impose obligations on large platforms to assess and mitigate systemic risks associated with their services, including risks related to civic discourse and fundamental rights (European Union, 2022). These regulatory frameworks increasingly recognize that platform design choices affect the prevalence and impact of incitement, requiring companies to consider how algorithmic systems may amplify dangerous content.

8. Challenges and Recommendations for Legal Development

The application of incitement law to social media contexts reveals numerous challenges requiring doctrinal refinement, institutional adaptation, and technological innovation. This section identifies key areas needing development and proposes recommendations for strengthening legal responses to online incitement.

8.1 Clarifying Causation Standards

Current inconsistencies in causation requirements across different modes of liability and jurisdictions create uncertainty and complicate prosecutions. International criminal law would benefit from clearer articulation of the causal nexus required for social media incitement cases, particularly regarding temporal distance, intermediate causation, and probabilistic approaches.

Recommendation: International criminal tribunals and domestic courts should develop specialized guidance addressing causation in mass incitement scenarios, acknowledging that strict but-for causation may prove unworkable while ensuring that liability attaches only when inflammatory speech makes substantial contributions to criminal outcomes. Courts might adopt a "material increase in risk" standard that requires proof that incitement significantly elevated the probability of crimes occurring, combined with requirements that prosecutors demonstrate reach and impact of inflammatory content through platform data and expert analysis.

8.2 Developing Evidentiary Frameworks for Digital Proof

The unique characteristics of digital evidence require specialized forensic capabilities and evidentiary standards. Many jurisdictions lack frameworks for authenticating social media content, establishing chains of custody for digital evidence, or assessing algorithmic amplification effects.

Recommendation: International organizations should develop model laws and best practices for digital evidence in incitement cases, addressing authentication procedures, standards for platform data disclosure, protection of privacy interests, and admissibility of expert testimony regarding online influence mechanisms. Prosecutors need access to specialized units with technical expertise in digital forensics, social media analysis, and computational methods for assessing content reach and engagement.

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

8.3 Addressing Algorithmic Amplification

The role of platform algorithms in amplifying inciteful content requires careful legal analysis that current frameworks inadequately address. Questions persist about whether algorithmic promotion affects assessments of speaker intent, whether platforms bear responsibility for amplification, and how foreseeability standards should account for viral spread.

Recommendation: Legal scholarship and jurisprudence should develop principles for attributing responsibility when algorithmic systems amplify incitement beyond speakers' organic reach. One approach might distinguish between speakers who deliberately exploit algorithmic systems to maximize inflammatory content's reach and those whose content achieves unexpected viral spread. Platforms themselves might face obligations to design systems that avoid amplifying incitement to international crimes, with failure to implement reasonable safeguards potentially triggering regulatory consequences distinct from criminal liability.

8.4 Enhancing International Cooperation

The transnational character of social media incitement demands enhanced cooperation mechanisms among states, international organizations, and private platforms. Current mutual legal assistance frameworks often prove too slow for addressing rapidly spreading inflammatory content, while platform policies regarding government requests vary significantly across jurisdictions.

Recommendation: States should negotiate multilateral agreements establishing streamlined procedures for gathering digital evidence in incitement investigations, respecting both human rights standards and legitimate platform concerns about unwarranted government interference. International organizations might establish specialized bodies to facilitate cooperation among prosecutors investigating transnational incitement, similar to Europol's coordination of cross-border law enforcement. Platforms should develop standardized processes for responding to lawful evidence requests related to international crimes, with transparent reporting about government requests and decisions.

8.5 Balancing Prevention and Expression

Ongoing tensions between preventing incitement and protecting expression require continuous refinement of legal standards that respect both imperatives. Current approaches sometimes favor either prevention at the expense of free speech or absolute speech protection that tolerates dangerous incitement.

Recommendation: Jurisdictions should adopt the Rabat Plan of Action's six-part threshold test as a baseline for distinguishing protected expression from prohibited incitement, while allowing adaptation to local contexts. Legal frameworks should emphasize procedural safeguards including judicial oversight of restrictions, rights to appeal, and sunset provisions for emergency measures limiting expression. Civil society organizations should receive support for implementing counter-speech initiatives and resilience-building programs that address inflammatory narratives

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

without resorting to criminalization.

8.6 Investing in Research and Monitoring

Limited empirical research on social media's role in facilitating international crimes hampers evidence-based policy development. Better understanding of how online incitement influences behavior, which platform features amplify inflammatory content, and which interventions effectively prevent mobilization to violence would inform more effective legal responses.

Recommendation: Governments, international organizations, and research institutions should fund rigorous empirical studies examining the relationship between online incitement and offline violence, including field research, computational analysis of platform data, and evaluation of intervention strategies. Early warning systems should integrate social media monitoring to identify emerging incitement campaigns before they produce mass violence, though such monitoring must respect privacy rights and avoid chilling protected expression.

8.7 Addressing Platform Accountability

While individual criminal responsibility properly focuses on speakers who incite crimes, the broader challenge of preventing online incitement requires attention to platform governance and corporate responsibility. Current approaches inadequately address how platform design choices, business models, and moderation practices affect the prevalence of incitement.

Recommendation: Regulatory frameworks should impose transparency obligations on platforms regarding content moderation decisions, algorithmic systems, and actions taken to address incitement to international crimes. Platforms should implement human rights impact assessments before launching services in conflict-affected regions and develop specialized responses to situations where their services facilitate atrocity crimes. International organizations might develop certification systems recognizing platforms that implement robust safeguards against incitement while respecting expression rights.

9. Conclusion

The migration of incitement to social media and online platforms has fundamentally challenged traditional legal frameworks for establishing individual criminal responsibility. While foundational principles developed through cases like the ICTR's Media Case remain relevant, courts and prosecutors must adapt these doctrines to address the unique characteristics of digital communication, including unprecedented speed and scale of dissemination, algorithmic amplification, anonymity, and transnational reach. This research has examined how international criminal law confronts these challenges, identifying both progress and persistent gaps in addressing online incitement to international crimes.

The analysis reveals that proving causation between social media incitement and subsequent international crimes presents formidable obstacles. The temporal distance

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

separating inflammatory posts from eventual violence, the multitude of factors influencing criminal decisions, and the difficulty of isolating specific speech acts' causal contributions all complicate prosecutorial efforts. Courts have begun developing approaches that account for social media's mass reach while maintaining individual culpability principles, but significant inconsistencies persist across jurisdictions regarding the requisite causal nexus. The material increase in risk standard proposed in this research offers a potential middle ground that serves incitement law's preventive purpose while respecting rule of law concerns about criminalizing speech based solely on speculative harms.

Establishing specific intent in social media cases similarly requires sophisticated evidentiary analysis that considers not only the content of inflammatory posts but also broader patterns of messaging, speakers' positions of influence, and contextual factors indicating purposeful mobilization. The permanent digital records created by social media platforms provide prosecutors with unprecedented evidence of systematic incitement campaigns, yet the interpretive challenges posed by coded language, cultural context, and strategic ambiguity demand careful judicial analysis. The Rabat Plan of Action's six-part threshold test offers valuable guidance for distinguishing criminal incitement from protected expression, though application to rapidly evolving digital contexts requires continued refinement.

The examination of comparative jurisprudence demonstrates growing judicial recognition of social media's capacity to facilitate incitement to international crimes. From the ICTR's foundational decisions on media incitement through emerging cases addressing digital communications, courts have shown willingness to adapt traditional doctrines while maintaining fundamental protections for freedom of expression. However, the relative paucity of completed prosecutions focused specifically on social media incitement means that many critical legal questions remain unsettled, creating uncertainty for both prosecutors and speakers engaging in online advocacy.

The human rights framework governing expression rights and permissible restrictions provides essential boundaries for incitement law's application to social media. Any legal regime must carefully balance the imperative to prevent atrocity crimes against the fundamental importance of protecting robust public discourse. The procedural safeguards recommended by international human rights bodies, including requirements of legality, legitimate purpose, proportionality, and least restrictive means, serve as vital constraints on state power to criminalize speech. Platform governance and content moderation introduce additional complexity, as private companies increasingly exercise quasi-regulatory authority over online expression without the accountability mechanisms applicable to state actors.

Looking forward, several developments will shape the evolution of legal responses to online incitement. Technological advances in artificial intelligence and machine learning may enable both more sophisticated incitement campaigns and enhanced detection capabilities, requiring ongoing adaptation of legal frameworks. The growing recognition of platforms' role in amplifying dangerous content will likely produce regulatory initiatives that impose obligations on companies to design systems that

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

resist exploitation for incitement purposes. International cooperation mechanisms will need enhancement to address the transnational character of digital incitement effectively while respecting sovereignty concerns and human rights standards.

The recommendations proposed in this research emphasize the need for clear causation standards adapted to mass incitement scenarios, specialized evidentiary frameworks for digital proof, legal principles addressing algorithmic amplification, enhanced international cooperation, balanced approaches that protect expression while preventing atrocities, investment in empirical research, and appropriate platform accountability mechanisms. Implementation of these recommendations requires collaboration among diverse stakeholders, including international tribunals, domestic courts, legislators, technology companies, civil society organizations, and academic researchers.

Ultimately, the challenge of addressing individual criminal responsibility for incitement through social media and online platforms reflects broader questions about how legal systems adapt to technological change while preserving fundamental values. The specific doctrinal issues examined in this research regarding causation and intent thresholds represent only part of this larger challenge, which encompasses questions of jurisdiction, evidence, procedure, and the proper balance between security and liberty in democratic societies. As social media continues evolving and new communication technologies emerge, international criminal law must maintain its capacity for principled adaptation, learning from experience while remaining grounded in core commitments to individual culpability, due process, and human rights.

The stakes could hardly be higher. Effective legal responses to online incitement can contribute to preventing atrocity crimes by establishing accountability for those who exploit digital platforms to mobilize mass violence. Conversely, overly broad or poorly calibrated approaches risk chilling legitimate political discourse and enabling authoritarian suppression of dissent. Navigating this treacherous terrain requires careful attention to both the technical characteristics of social media platforms and the fundamental principles that should govern any system of criminal justice. This research has sought to contribute to that ongoing project by clarifying the legal standards applicable to causation and intent in online incitement cases while identifying areas requiring continued development. The international community's capacity to prevent future atrocities while protecting freedom of expression depends significantly on how successfully legal systems rise to these challenges in the years ahead.

References

1. Ambos, K. (2013). *Treatise on international criminal law: Volume I: Foundations and general part*. Oxford University Press.
2. Amnesty International. (2022). *Myanmar: The social atrocity: Meta and the right to remedy for the Rohingya*. Amnesty International.
3. Article 19. (2012). *Prohibiting incitement to discrimination, hostility or violence*.

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

Article 19.
4. Ashworth, A., & Horder, J. (2013). *Principles of criminal law* (7th ed.). Oxford University Press.
5. Barendt, E. (2005). *Freedom of speech* (2nd ed.). Oxford University Press.
6. Benesch, S. (2008). Vile crime or inalienable right: Defining incitement to genocide. *Virginia Journal of International Law*, *48*(3), 485-528.
7. Benesch, S. (2014). Countering dangerous speech: New ideas for genocide prevention. United States Holocaust Memorial Museum.
8. Benesch, S., Buerger, C., Hlavka, J., & Octaviano, E. (2020). Dangerous speech: A practical guide. Dangerous Speech Project.
9. Braddock, K., & Horgan, J. (2016). Towards a guide for constructing and disseminating counternarratives to reduce support for terrorism. *Studies in Conflict & Terrorism*, *39*(5), 381-404.
10. Brandenburg v. Ohio, 395 U.S. 444 (1969).
11. Buyse, A. (2014). Dangerous expressions: The ECHR, violence and free speech. *International and Comparative Law Quarterly*, *63*(2), 491-503.
12. Cassese, A. (2008). *International criminal law* (2nd ed.). Oxford University Press.
13. Citron, D. K., & Wittes, B. (2017). The Internet will not break: Denying bad Samaritans Section 230 immunity. *Fordham Law Review*, *86*(2), 401-423.
14. Cohen-Eliya, M., & Porat, I. (2013). *Proportionality and constitutional culture*. Cambridge University Press.
15. Colangelo, A. J. (2014). What is extraterritorial jurisdiction? *Cornell Law Review*, *99*(6), 1303-1352.
16. Cole, D. (2003). The new McCarthyism: Repeating history in the war on terrorism. *Harvard Civil Rights-Civil Liberties Law Review*, *38*(1), 1-30.
17. Council of Europe. (1950). *European Convention on Human Rights*. Council of Europe.
18. Elagab, O. (2020). The criminalization of hate speech and incitement to hatred in international human rights law. *Netherlands Quarterly of Human Rights*, *38*(4), 258-276.
19. European Court of Human Rights. (1976). *Handyside v. United Kingdom*, Application No. 5493/72.
20. European Union. (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act).
21. Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., & Benkler, Y. (2016). Partisanship, propaganda, and disinformation: Online media and the 2016 U.S. presidential election. Berkman Klein Center for Internet & Society Research Paper.
22. Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). *Countering online hate speech*. UNESCO.
23. Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
24. Gordon, G. S. (2013). Music and genocide: Harmonizing coherence, freedom and

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

nonviolence in incitement law. *Santa Clara Law Review*, *53*(3), 607-678.
25. Gordon, G. S. (2014). Hate speech and persecution: A contextual approach. *Vanderbilt Journal of Transnational Law*, *46*(2), 303-372.
26. Gordon, G. S. (2017). *Atrocity speech law: Foundation, fragmentation, fruition*. Oxford University Press.
27. Greenawalt, K. (1989). *Speech, crime, and the uses of language*. Oxford University Press.
28. Greenawalt, K. (1999). Clear and present danger and criminal speech. In L. Bollinger & G. Stone (Eds.), *Eternally vigilant: Free speech in the modern era* (pp. 97-121). University of Chicago Press.
29. Harff, B. (2003). No lessons learned from the Holocaust? Assessing risks of genocide and political mass murder since 1955. *American Political Science Review*, *97*(1), 57-73.
30. Herz, M., & Molnar, P. (Eds.). (2012). *The content and context of hate speech: Rethinking regulation and responses*. Cambridge University Press.
31. International Criminal Court. (1998). *Rome Statute of the International Criminal Court*. UN Doc. A/CONF.183/9.
32. International Criminal Court. (2016). *Policy on children*. Office of the Prosecutor.
33. International Criminal Court, Office of the Prosecutor. (2010). *Request for authorization of an investigation pursuant to Article 15*. ICC-01/09.
34. International Criminal Tribunal for Rwanda. (1998). *Prosecutor v. Akayesu*, Case No. ICTR-96-4-T.
35. International Criminal Tribunal for Rwanda. (2000). *Prosecutor v. Ruggiu*, Case No. ICTR-97-32-I.
36. International Criminal Tribunal for Rwanda. (2003). *Prosecutor v. Nahimana, Barayagwiza and Ngeze*, Case No. ICTR-99-52-T (Media Case).
37. International Criminal Tribunal for Rwanda. (2007). *Prosecutor v. Nahimana, Barayagwiza and Ngeze*, Case No. ICTR-99-52-A (Appeals Chamber).
38. International Criminal Tribunal for the former Yugoslavia. (1999). *Prosecutor v. Tadić*, Case No. IT-94-1-A.
39. International Criminal Tribunal for the former Yugoslavia. (2001). *Prosecutor v. Kordić and Čerkez*, Case No. IT-95-14/2-T.
40. International Criminal Tribunal for the former Yugoslavia. (2003). *Prosecutor v. Krnojelac*, Case No. IT-97-25-A.
41. International Criminal Tribunal for the former Yugoslavia. (2004). *Prosecutor v. Blaškić*, Case No. IT-95-14-A.
42. Jain, N. (2014). The control theory of accomplice liability. *University of Pennsylvania Law Review*, *162*(2), 453-521.
43. Jardine, E. (2018). *The dark web dilemma: Tor, anonymity and online policing*. Global Commission on Internet Governance Paper Series No. 21.
44. Kaye, D. (2019). *Speech police: The global struggle to govern the internet*. Columbia Global Reports.
45. Kerr, O., & Schneier, B. (2009). Encryption workarounds. *Georgetown Law*

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

Journal, *106*(4), 989-1018.
46. Klamberg, M. (Ed.). (2013). *Commentary on the Law of the International Criminal Court*. Torkel Opsahl Academic EPublisher.
47. Klonick, K. (2017). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, *131*(6), 1598-1670.
48. Leader Maynard, J. (2014). Rethinking the role of ideology in mass atrocities. *Terrorism and Political Violence*, *26*(5), 821-841.
49. Marwick, A., & Lewis, R. (2017). *Media manipulation and disinformation online*. Data & Society Research Institute.
50. Massaro, T. M., & Norton, H. (2016). Siri-ously? Free speech rights and artificial intelligence. *Northwestern University Law Review*, *110*(5), 1169-1194.
51. Mchangama, J., & Alkiviadou, N. (2021). Hate speech and the European Court of Human Rights: Whatever happened to the right to offend, shock or disturb? *Human Rights Law Review*, *21*(4), 1008-1042.
52. Mendel, T. (2012). Does international law provide for consistent rules on hate speech? In M. Herz & P. Molnar (Eds.), *The content and context of hate speech* (pp. 417-429). Cambridge University Press.
53. Mettraux, G. (2005). *International crimes and the ad hoc tribunals*. Oxford University Press.
54. Nayak, P. (2020). *Online hate speech: Challenges and counter measures in India*. Observer Research Foundation.
55. Novic, E. (2016). *The concept of cultural genocide: An international law perspective*. Oxford University Press.
56. Nybondas, M. (2014). Testing the clear and present danger test in international criminal law. *Baltic Yearbook of International Law*, *13*(1), 99-122.
57. Organization of American States. (1969). *American Convention on Human Rights*. OAS Treaty Series No. 36.
58. Osiel, M. J. (2009). *Making sense of mass atrocity*. Cambridge University Press.
59. Padilla, D. (2020). Causation in international criminal law. In K. J. Heller, F. Mégret, S. M. H. Nouwen, J. D. Ohlin, & D. Robinson (Eds.), *The Oxford handbook of international criminal law* (pp. 857-880). Oxford University Press.
60. Post, R. C. (2009). Hate speech. In I. Hare & J. Weinstein (Eds.), *Extreme speech and democracy* (pp. 123-138). Oxford University Press.
61. R v. Sheppard and Whittle, [2010] EWCA Crim 65.
62. Ryngaert, C. (2015). *Jurisdiction in international law* (2nd ed.). Oxford University Press.
63. Sadat, L. N. (2012). *Forging a convention for crimes against humanity*. Cambridge University Press.
64. Santa Clara Principles. (2018). *Santa Clara Principles on Transparency and Accountability in Content Moderation*. https://santaclaraprinciples.org
65. Saul, B. (2019). The International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights. In D. Shelton (Ed.), *The Oxford handbook of international human rights law* (pp. 217-242).

DOI: https://doi.org/10.33193/JEAHS.49.2026.740
e-ISSN: 2709-071X
p-ISSN: 2709-0701
مجلة العلوم التربوية والإنسانية
Journal of Educational and Human Sciences
www.jeahs.com
Volume (49) January 2026
العدد (49) يناير 2026

Oxford University Press.
66. Schabas, W. A. (2009). *Genocide in international law: The crime of crimes* (2nd ed.). Cambridge University Press.
67. Scheffler, T. (2020). Talking about terrorism: Anti-Muslim prejudice in social media discourse. *Ethnicities*, *20*(4), 720-741.
68. Shulman, S. W. (2018). *The case against the internet: Transparency, information asymmetry, and the modern political economy*. Routledge.
69. Starbird, K. (2019). Disinformation's spread: Bots, trolls and all of us. *Nature*, *571*(7766), 449.
70. Straus, S. (2007). What is the relationship between hate radio and violence? Rethinking Rwanda's "Radio Machete." *Politics & Society*, *35*(4), 609-637.
71. Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.
72. Tufekci, Z. (2018). YouTube, the great radicalizer. *New York Times*, March 10, 2018.
73. United Nations. (1948). *Universal Declaration of Human Rights*. UN Doc. A/RES/217(III).
74. United Nations. (1948). *Convention on the Prevention and Punishment of the Crime of Genocide*. UN Treaty Series, Vol. 78, p. 277.
75. United Nations. (1966). *International Covenant on Civil and Political Rights*. UN Treaty Series, Vol. 999, p. 171.
76. United Nations. (2011). *Guiding Principles on Business and Human Rights*. UN Doc. A/HRC/17/31.
77. United Nations. (2013). *Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence*. UN Doc. A/HRC/22/17/Add.4.
78. United Nations Human Rights Committee. (2011). *General Comment No. 34: Article 19 (Freedoms of opinion and expression)*. UN Doc. CCPR/C/GC/34.
79. Waldorf, L. (2009). Revisiting Hotel Rwanda: Genocide ideology, reconciliation, and rescuers. *Journal of Genocide Research*, *11*(1), 101-125.
80. Waldron, J. (2012). *The harm in hate speech*. Harvard University Press.
81. Yanagizawa-Drott, D. (2014). Propaganda and conflict: Evidence from the Rwandan genocide. *Quarterly Journal of Economics*, *129*(4), 1947-1994.
82. Zeitzoff, T. (2017). How social media is changing conflict. *Journal of Conflict Resolution*, *61*(9), 1970-1991.