COMMUNICATION PLATFORM POWER

A PRACTITIONER'S GUIDE TO THE ARCHITECTURE OF AMPLIFICATION

A **MADE** BRIEF

BY ETHAN CHIU, HUGO CHUNG, KAJ LITCH, SARAH MARKEY & ISABELLA PANICO

2026

MADE
Mass Atrocities in the Digital Era
Yale MACMILLAN CENTER
Genocide Studies Program

Table of Contents

Introduction

We are deeply grateful to Nisheeeth Vishnoi, Julian Posada, and Artur Pericles Lima Monteiro for generously sharing their expertise and guidance throughout the development of this brief.

Understanding how mass violence unfolds in the digital era requires conceptual clarity about the role of platforms and the dynamics they produce. For the purpose of this module, communication platforms are digital infrastructures that mediate social interaction, information flow, and collective action through algorithmic systems. Unlike earlier internet infrastructure that passively hosted content, contemporary platforms actively curate, rank, and recommend what users see.

Communication platform dynamics refers to the systematic ways digital communication platforms shape visibility, virality, and social interaction through algorithmic curation, network effects, and two-sided market structures. These platforms exhibit a dual nature—designed simultaneously for capture (through lock-in mechanisms, network effects, and switching costs) and for shaping social outcomes through their amplification mechanisms.

This dual nature has important implications for mass atrocity risk. When platforms become mandatory infrastructure for accessing services, their amplification mechanisms can help facilitate, accelerate, and coordinate violence. This primer outlines the ways that communication platforms shape the speed, scale, and coordination capacity of violence through their technical architecture—shedding light on how platform governance fits into the scope of atrocity prevention in the digital era.

Architecture of Platform Power

Platforms as Infrastructure

Communication platforms are not merely optional services that users can freely adopt or abandon. As Poel and others have argued, platforms function as infrastructures—foundational systems that everyday people rely on to facilitate their daily tasks and interactions. However, they shape these experiences through systematic collection, algorithmic processing, and circulation of data. Understanding platforms as infrastructure rather than services reveals three critical pillars of their power: their role as essential infrastructure, their market-making capacity, and their concentration of power over information flows. While users can technically opt out of any given platform, the realistic possibility of doing so has been steadily eroded. Governments and institutions increasingly rely on communication platforms for essential public functions—as seen during the COVID-19 pandemic, when public health guidance and emergency updates were disseminated primarily through social media channels. When opting out means losing access to critical government communication, the choice becomes illusory.

The transition from optional service to mandatory infrastructure often occurs without notice. As Cory Doctorow has argued, platforms typically progress through a predictable arc of "enshittification": they first offer genuine value to attract users, then leverage network effects to lock those users in, and finally extract value from the captive user base for the benefit of advertisers and shareholders. WhatsApp began as a convenient messaging app, but became the primary channel for hospital appointment reminders, school communications, and local business transactions. WeChat offers an even more advanced case: originally a messaging app, it now encompasses payments, government services, transportation, and virtually every dimension of daily life in China.

When these platforms become the dominant—or only viable—means of accessing essential services, information, or communities, alternatives are systematically eliminated. Like water mains or electrical grids, there is no practical workaround: users face a stark choice between participating on the platform's terms or losing access to critical infrastructure. The mechanisms that produce this dependency—network effects, switching costs, accumulated data—are taken up in the following section; what matters here is that the dependency itself is real, and that it grants platforms a form of structural power that extends far beyond any individual transaction.

This infrastructural power creates acute mass atrocity risk. When a platform becomes mandatory infrastructure, those seeking to incite violence gain access to captive audiences who cannot easily leave. The platform mediates not just social connection but access to healthcare, education, economic opportunity, and civic participation. In contexts like Myanmar, where Facebook became synonymous with the internet itself, this infrastructural dependence enabled hate speech and coordinated violence to reach mass audiences with no viable alternatives for information or communication.

Architecture of Platform Power

Lock-in and Network Effects

Beyond attracting users, digital communication platforms bind them. Platforms gain significant power through self-reinforcing network effects that lock in users and increase switching costs. These mechanisms are not inherently harmful—they are precisely why platforms are so useful. But they become dangerous in high-risk contexts, where they trap people in information environments that reward escalation, distort perceptions, and enable coordinated harm.

Network effects operate through two distinct mechanisms. Direct network effects occur when a platform's value increases as more people from the same user group join. On social media, the platform's value lies in the dense web of relationships crossing it—friends, classmates, coworkers, and community members. Over time, participation becomes a baseline expectation for social belonging. Indirect network effects occur when value flows between different user groups. In two-sided marketplaces like ride-hailing, the platform shapes interaction terms, sets incentives, and becomes the default venue. Even dissatisfied users must remain because leaving means losing access to the other side of the market.

These network effects harden into lock-in through accumulating switching costs. Users who consider leaving face the loss of their social graph, group memberships, message history, photo archives, content libraries, and reputation metrics built over years. They also face coordination costs (convincing others to switch), learning costs (mastering new interfaces), and status costs (losing established presence). The result is that even when platforms become sites of harmful content, users cannot easily exit without severe personal, social, or economic consequences—transforming infrastructure dependence into informational captivity. In high-tension contexts, this lock-in means vulnerable populations remain exposed to escalating cycles of dehumanization and incitement, unable to leave even as violence becomes imminent.

Architecture of Platform Power

Algorithmic Governance Systems

Algorithmic governance systems are the core mechanisms through which platforms exert control over visibility, interaction, and information flow. These systems determine not only what content circulates, but how, when, and to whom; shaping patterns of attention, opinion, and collective behavior at scale. These systems tend to prioritize extreme opinions. This can enable certain narratives to dominate the information environment while suppressing moderating or countervailing voices.

Two core mechanisms define today's algorithmic governance systems: recommendation engines and personalization systems. Recommendation engines play a particularly powerful role in directing user attention. On platforms such as YouTube, and through the "explore" pages of TikTok and Instagram, the majority of content consumption occurs through generated "suggested" content. Reviews of recommender systems indicate that these engines can narrow informational exposure and, in some cases, facilitate pathways toward more extreme or polarizing material by prioritizing engagement-maximizing content.

Personalization systems individualize information exposure based on behavioral data such as past engagement, social connections, and inferred preferences. This means users see more of what they have previously engaged with. While personalization is often framed as enhancing user experience, it reinforces polarization and information bias. Research on algorithmic radicalization shows that users who engage with extreme content are more likely to be shown increasingly similar material, intensifying grievance and identity-based hostility.

Underpinning both mechanisms is a classification process through which platforms categorize users, content, and interactions into algorithmically defined groups. Filter bubbles and echo chambers become the result of self-reinforcing information environments that harden social boundaries and intensify adversarial identities

Architecture of Platform Power

Algorithmic Governance Systems cont.

These enclosed information environments also cultivate a dynamic of interpersonal trust that amplifies atrocity risk. When users are locked into platforms and repeatedly exposed to the same voices within their algorithmically defined communities, they develop trust in frequent contacts—even when those contacts spread false or dehumanizing information. This trust operates as a form of social proof. This means that content shared by familiar sources is less likely to be questioned or challenged. In high-risk contexts disinformation circulated by trusted in-group members could spread with little resistance, effectively bypassing the scrutiny that might otherwise slow escalation.

Algorithmic governance is largely invisible. The algorithms that platforms use are their proprietary technology—and arguably the feature that most meaningfully differentiates platforms from one another, more so than interface design or visible product features. Users experience their curated feeds as organic or neutral. They remain largely unaware of the content they are not being shown and rarely question their spoon-fed experiences. Algorithmic governance systems are the silent influencers shaping the opinions and lives of nearly all platform users.

In high-tension contexts, these dynamics can have catastrophic consequences. Investigations into violence against the Rohingya in Myanmar found that anti-Rohingya and inflammatory content was repeatedly amplified on Facebook, contributing to a hostile information environment that helped legitimize and accelerate real-world violence. By shaping who is heard, what narratives dominate, and how rapidly information spreads, algorithmic governance systems can accelerate escalation, normalize hate speech, distort collective threat perception, and facilitate coordination toward violence—ultimately shaping the scale and speed of mass atrocities.

How Platforms Facilitate Violence

Algorithmic Amplification of Harmful Content

Algorithms do not invent values—they absorb and amplify the priorities encoded into them. Because outrage, fear, and anger reliably generate stronger engagement than neutral or conciliatory material, these systems systematically elevate emotionally charged content. This reflects human design choices embedded at every stage of the algorithmic pipeline—from what data is collected, to how relevance is defined, to which metrics are optimized.

In contexts marked by political polarization, ethnic tension, or social grievance, engagement optimization therefore disproportionately amplifies dehumanizing narratives, conspiracy theories, and inflammatory rhetoric. Algorithmic amplification reshapes the speed, scale, and structure of harm. Recommendation and personalization systems sort users into increasingly narrow information environments based on prior behavior, reinforcing echo chambers and filtering out countervailing perspectives. Content that once would have spread slowly or remained localized can now propagate exponentially, gaining legitimacy through repetition and visibility alone.

This compression of time accelerates the pathway from grievance to dehumanization to mobilization, reducing opportunities for social or institutional intervention. Crucially, these systems optimize for attention rather than wellbeing—users are repeatedly exposed to harmful content because it sustains engagement, even when distressing or destabilizing. Compared to a counterfactual world without platform-driven amplification—where coordination is slower, narratives less synchronized, and dehumanization diffuses unevenly—platform architectures dramatically lower the cost of mass persuasion and collective alignment.

Coordination Infrastructure

Since the advent of the internet, technology has been employed by violent actors to recruit and coordinate—a ubiquitous and consistent reality. Digital platforms have accelerated and scaled this capacity through three interconnected mechanisms.

Private groups and encrypted channels create spaces where violent actors organize away from public view. Platforms like Telegram and WhatsApp enable large-scale coordination through features designed for privacy—group sizes reaching hundreds of thousands, ephemeral messaging, and minimal content moderation. These spaces function as incubators where extreme content is developed, tested, and refined before being amplified across public platforms. Real-time coordination enables synchronized action across geography. Platforms allow coordinated attacks, simultaneous protests that turn violent, and rapid mobilization that would have required extensive in-person organizing in earlier eras. Cross-platform amplification creates multiple pathways for mobilization as content originates on one platform and spreads to others. This "out-linking" strategy ensures that when one account or channel is taken down, the network's connective tissue remains intact.

Dehumanization at Scale

Since the advent of the internet, technology has been employed by violent actors to recruit and coordinate—a ubiquitous and consistent reality. Digital platforms have accelerated and scaled this capacity through three interconnected mechanisms.

Private groups and encrypted channels create spaces where violent actors organize away from public view. Platforms like Telegram and WhatsApp enable large-scale coordination through features designed for privacy—group sizes reaching hundreds of thousands, ephemeral messaging, and minimal content moderation. These spaces function as incubators where extreme content is developed, tested, and refined before being amplified across public platforms. Real-time coordination enables synchronized action across geography. Platforms allow coordinated attacks, simultaneous protests that turn violent, and rapid mobilization that would have required extensive in-person organizing in earlier eras. Cross-platform amplification creates multiple pathways for mobilization as content originates on one platform and spreads to others. This "out-linking" strategy ensures that when one account or channel is taken down, the network's connective tissue remains intact.

Recognizing Platform-Facilitated Violence

The warning signs of platform-facilitated violence are not categorically different from the early indicators that atrocity prevention practitioners have long monitored—dehumanizing language, coordinated incitement, and the targeting of specific groups. What platforms change is how these signs manifest, how quickly they escalate, and how they interact with one another. Practitioners must therefore apply familiar frameworks with attuned attention to the specific ways platform dynamics accelerate and amplify each warning sign.

Dehumanizing content

- Sudden spikes in dehumanizing language or imagery, particularly zoomorphic metaphors comparing groups to animals, insects, or disease, and rhetoric of toxification
- Hate speech achieving rapid virality
- Recommendation systems steering users toward increasingly extreme material, indicating the active role of platform architecture in normalizing violence-enabling narratives

Coordination signals

- The emergence of closed coordination groups and private channels
- Calls to action spreading rapidly across platforms
- Organized brigading or harassment campaigns targeting specific communities
- Cross-platform amplification, where content originates in private channels and spreads to public ones, suggesting a network with both organizational depth and broad reach

Targeting and identification

- Rumors, dehumanizing memes, and location-sharing converging around a specific group
- Direct calls for violence against identified individuals or communities

Practitioners should treat the co-occurrence of multiple warning signs as especially significant: platform dynamics allow these indicators to reinforce one another at a speed that leaves little time for intervention.

Why This Control Matters

Platforms fundamentally lower the marginal cost of coordinating violence. What used to require extensive in-person organizing can now happen at scale, instantly, and globally. Algorithmic amplification accelerates the pathway from rhetoric to action—dehumanization that might take years through traditional media can happen in weeks or months on platforms.

Most importantly the architecture is exportable: the same platform dynamics that facilitated violence in Myanmar operate in Ethiopia (coordinated hate campaigns during the Tigray conflict), India (WhatsApp lynch mobs), and Sri Lanka (Facebook-enabled anti-Muslim violence). The model travels globally, making platform governance critical for atrocity prevention.

Recommendations

The dynamics outlined make clear that platform-facilitated violence is a structural governance issue. Because risk emerges from platform architecture, effective prevention must operate at multiple levels simultaneously. The recommendations below translate these priorities into concrete actions for practitioners seeking to reduce atrocity risk in digitally mediated environments.

01 Monitoring Platform Architecture and Information Ecosystems

Effective early warning requires systematic attention to both platform-level changes and content-level trends—and translating these signals into action requires both internal platform governance and external oversight, whether from civil society monitors, regulatory bodies, or state actors. Practitioners should establish monitoring systems that track:

Platform structural changes: Monitor algorithm updates and feature rollouts, especially changes to ranking, recommendation, or moderation systems. Document when platforms remove safety features, downgrade fact-checking, or reduce content moderation resources in at-risk regions. Track when platforms transition from optional services to mandatory infrastructure—for instance, when WhatsApp becomes the primary channel for government services or Facebook becomes the dominant news source in news deserts.

Platform dominance and gaps: Identify which platforms dominate in different languages and regions, as platform dynamics operate differently across linguistic and cultural contexts. Map language-specific gaps in content moderation capacity, noting where platforms lack sufficient moderators, automated detection systems, or cultural expertise. Document cross-platform amplification pathways to understand how content moves between platforms and which platform combinations create the greatest risk.

Content and coordination patterns: Track trends in dehumanizing content including volume, virality, and the emergence of new memes or metaphors. Monitor for zoomorphic language (comparing groups to animals, insects, or disease), rhetoric of toxification, and explicit calls for violence. Document coordination patterns including private group formation, call-and-response patterns between platforms, and organized brigading or harassment campaigns. Sudden changes in these patterns often precede offline violence.

Recommendations

02 Advocacy for Transparency and Structural Reform

Platform governance requires structural changes to how platforms operate, not merely reactive content moderation. Advocacy efforts should focus on:

Researcher access and transparency: Push for independent researcher access to platform data, including information on algorithmic ranking, recommendation systems, and content virality patterns. Demand algorithmic impact assessments before platforms deploy new features, particularly in high-risk contexts. Support initiatives like the Digital Services Act's researcher access provisions that mandate transparency while protecting user privacy.

Independent audits and accountability: Advocate for independent audits of ranking and recommendation systems, with particular attention to how algorithms handle content in at-risk regions and minority languages. Support public reporting requirements that make platforms disclose content moderation decisions, appeal processes, and the resources allocated to different regions and languages. Push for algorithmic accountability mechanisms that enable affected communities to challenge harmful platform design choices.

Interoperability and user agency: Advocate for interoperability mandates and data portability that allow users to control their social graph and content. Reducing switching costs weakens platform lock-in and enables users to leave platforms that become hostile information environments. Support standards that enable cross-platform communication while maintaining user safety and privacy protections.

Human oversight in high-stakes contexts: Push for human-in-the-loop systems for high-stakes decisions including account suspension, content removal in sensitive contexts, and changes to algorithmic systems in at-risk regions. Automated systems alone cannot account for cultural context, evolving hate speech tactics, or the nuanced ways dehumanization operates across different communities.

Meaningful content moderation: Demand content moderation systems with clear rules, transparent appeals processes, and public reporting on enforcement. Ensure platforms adequately resource content moderation in all languages and regions, not just dominant markets. Advocate for the inclusion of civil society, affected communities, and human rights experts in platform policy decisions, particularly when platforms operate in conflict-affected regions.

Recommendations

03 Building Community Resilience

Technology-focused interventions alone cannot prevent platform-facilitated violence, as platforms continue to evolve and as new technologies emerge, practitioners must remain vigilant to how design choices shape the pathways to mass atrocity.

Digital and media literacy: Help communities understand how algorithmic curation works, including how recommendation engines prioritize engagement over accuracy and how personalization creates filter bubbles. Build capacity to recognize manipulation tactics including dehumanizing memes, coordinated amplification, and the use of humor to normalize hatred. Focus particularly on vulnerable populations and communities at risk of being targeted by hate campaigns.

Counter-speech and counter-narratives: Create and amplify counter-speech that challenges dehumanizing narratives while avoiding simply boosting engagement with harmful content. Support community-led efforts to develop counter-narratives that resonate culturally and linguistically. Recognize that counter-speech operates within the same algorithmic constraints as harmful content—effective counter-narratives must be designed for virality and emotional resonance while maintaining accuracy and human dignity.

Local civil society capacity: Support local civil society organizations in digital safety, platform monitoring, and rapid response capabilities. Ensure these organizations have resources to document platform-facilitated violence, advocate for platform accountability, and support affected communities. Build networks that connect local monitors with platform trust and safety teams, enabling rapid escalation when warning signs emerge.

Trusted messenger networks: Develop networks of credible local voices who can effectively counter misinformation and dehumanizing narratives within their communities. These messengers must understand both the local context and platform dynamics to craft effective interventions. Support their safety and security, as those who challenge violent narratives often face coordinated harassment and threats.

About MADE

The Mass Atrocities in the Digital Era (MADE) initiative is a first-of-its-kind program within Yale University's Genocide Studies Program, created to examine how digital technologies are transforming the landscape of mass violence and protection. MADE studies how information systems shape the commission, prevention, and accountability of atrocities, while cultivating a new generation of scholars and practitioners equipped to navigate these challenges. The initiative's 2025 programming tackles these issues through four core modules — digital authoritarianism and extremism, platform dynamics and violence, militarization of technology, and digital infrastructures and identity — each producing practitioner-oriented primers that translate complex research into accessible, action-focused tools for early detection, prevention, and accountability.

MADE

Mass Atrocities in the Digital Era

Yale MACMILLAN CENTER

Genocide Studies Program