

Contested Realities: LGBTI+ Rights in the Age of Algorithmic Politics

A Consultation Paper from the Global LGBTI+ Rights Commission



About Kaleidoscope Trust

Kaleidoscope Trust is a UK-based international charity fighting for a future where LGBTI+ people everywhere can live free, safe, and equal lives. We are building a global movement to create this world for LGBTI+ people everywhere.

We work with political leaders in the UK and beyond to ensure global LGBTI+ issues are a priority. We ensure that activists have the resources, skills, and training they need to learn from one another in order to create an enabling environment

and change hearts and minds. We bring together grassroots organisations with those in power to create that change.

Kaleidoscope Trust is a founding member of The Commonwealth Equality Network (TCEN) and provides the operational and financial base for the network's Secretariat as its host. We have engaged in Commonwealth processes and with Commonwealth institutions, as a priority, for over a decade in support of our shared objectives with TCEN's other members.

About the Global LGBTI+ Rights Commission

The Global LGBTI+ Rights Commission is a time-limited, international initiative designed by Kaleidoscope Trust to confront the alarming global rollback of LGBTI+ rights we now face. This independent commission will bring together experts, activists, and community voices from around the world to analyse current threats, understand root causes, and chart a path forward by:

Bringing together diverse voices to ensure our findings are informed by lived

experience as well as thematic expert insight.

Influencing policy by producing working papers and a final report presenting a strategic global roadmap to inform the policy priorities of governments and international institutions. This work will seek to defend and advance global human rights.

Amplifying stories and voices from across the community who are not in a position to speak out freely and safely. Through

storytelling, testimony, and engagement, the Commission is not just about policy – it’s about supporting a global movement.

The Commission will produce working papers on key issues, which will be shared for public consultation through a green paper/white paper model. A green paper is an initial discussion document inviting feedback and debate, while a white paper

presents a more developed set of proposals based on that feedback. This process allows individuals and organisations to respond directly to the papers or submit their own evidence and insights – all of which will be acknowledged and credited. This combination of research, consultation, and storytelling will ensure the Commission’s work is inclusive, transparent, and grounded in real-world voices.



Learn more on our website.

CONTENT

11

Executive Summary

Purpose of this Consultation

How responses will be used

How to respond

List of consultation questions

Glossary

Introduction

25

I. SETTING THE SCENE

Section 1: Why Social Media and Online Spaces Matter
for LGBTI+ Rights

Introduction: Digital Life as a Lifeline

From Empowerment to Exposure: Risks in Digital Spaces

Social Media as a Tool for Activism and Community Building

Digital Activism and the Vulnerabilities of Networked Visibility

Intersectional Vulnerability and Differential Exposure

The Digital Environment and the Expansion of Human Rights

Conclusion

49

II. THE ANATOMY OF CYBERSPACE

Section 2: What Is Cyberspace?

The Mirage of the Digital

Gibson's Cyberspace: Expansion and Warning

The Collapse of the Digital-Physical Divide

Disinhibition and Anonymity

Data Colonialism

Cyberspace, Power, and Anti-Rights Mobilisation

Conclusion

64

Section 3: Online Spaces as Modern Civic Arenas

Digital Publics and the Transformation of Civic Space

Networked Movements and a New Cycle of Mobilisation

Disinformation, Attention Hacking, and Anti-Gender Mobilisation

Surveillance Capitalism and Behavioural Governance

Conclusion

85

Section 4: Algorithms - Amplification, Bias, and Free Speech

Platform Governance and the Architecture of Visibility

Engagement Optimisation and the Architecture of Amplification

Filter Bubbles and Fragmented Realities

The Subjectivity and Predictive Governance

Navigating the Architecture: Contested Uses of Algorithmic Systems

Conclusion

107

Section 5: AI-Generated Propaganda and Emotional Manipulation

Synthetic Propaganda Systems

Enabling Elements of AI-Generated Propaganda

The AI-Enhanced Misinformation Pipeline

The Dopamine Economy of Disinformation

Conclusion

120

III. TOWARDS A TYPOLOGY OF ANTI-RIGHTS DRIVERS IN ONLINE SPACES

Section 6: The Political Economy of Hate

Platform incentives and the political economy of hate

Alternative Platforms, Parallel Presences, and the Outrage Pipeline

The Transnational Anti-Gender Movement

Information Warfare and Realpolitik

From Online Hate to Offline Harm

Global Asymmetries and Unequal Consequences

Section 10: Shifting the Overton Window - Normalisation, Platform Abdication, and the Persistence of Harm

Naming Harm: Conceptual Instability in the Digital Age

Jurisdictional Lag: When Law Becomes Geographically Obsolete

Platform Governance: The Architecture of Amplification and Abdication

Harm as Persistence: The Failure of Removal

Epistemic Harm: The Collapse of Verifiable Reality

Institutional Strain: Delegated Enforcement and Unequal Burden

Conclusion

Section 11: Political Parties, Power, and Online Mobilisation

Digital Populism: Outrage as Political Infrastructure

Microtargeting and Algorithmic Manipulation

Platformed Politics: Movements as Networked Systems

Coordinated Disinformation and Civil Society Delegitimisation

Hybrid Repression: State and Platform Convergence

Conclusion

Section 12: The Unaccountability Paradox

Platform Evasion: Neutrality as a Governance Strategy

Governments Struggling to Catch Up

Legal Gaps: Human Rights Law in the Digital Environment

Civil Society Burden: The Displacement of Enforcement

The Need for Global Governance

Accountability Cannot Be Algorithmic

Conclusion

Conclusion: The Canary in the Coal Mine

Section 7: Far-Right Online Ecosystems and Anti-LGBTI+ Mobilisation

Transnational Hate Networks

Influencer-Driven Extremism

Hybrid Attack Pathways

The Intersectionality of Hate

Conclusion

Section 8: Hidden Infrastructures of Harm

Covert Organisation

Data Trafficking

Law-Enforcement Challenges

State Use and the Dual-Use Paradox

Conclusion

Section 9: Predatory Communities and Gendered Exploitation

Toxic Masculinity Online: The Manosphere as a Grievance Ecosystem

Digital Exploitation: Deepfakes, Image-Based Abuse, and the

Industrialisation of Harm

Competition and Control: Domination as Social Capital

Predatory Communities and Multi-Target Vulnerability

Conclusion

Executive Summary

Digital technologies have become indispensable to modern civic life. For many LGBTI+ people – particularly those living in hostile legal or social environments – online spaces provide access to community, identity, information, and activism that may be unavailable offline. Digital platforms have transformed how LGBTI+ communities organise, tell their stories, build solidarity across borders, and participate in public life.

Yet, these same technologies have also become powerful tools for surveillance, harassment, disinformation, and political repression. Governments, extremist movements, and coordinated anti-rights actors increasingly exploit digital platforms to identify, target, intimidate, and silence LGBTI+ people. Platform design, algorithmic amplification, and data-driven business models can unintentionally reinforce these harms by rewarding outrage, increasing visibility without adequate safeguards, and enabling the rapid spread of hostile narratives.

This is the fourth working paper of the Global LGBTI+ Rights Commission where we present the argument that these developments are not simply isolated challenges affecting one community. Rather, they reveal wider structural changes in how digital technologies shape democracy, human rights, and public discourse. The experiences of LGBTI+ communities demonstrate how online harms – including coordinated abuse, algorithmic discrimination, digital surveillance, and transnational disinformation – often emerge

first against marginalised groups before expanding to threaten democratic participation more broadly.

The paper examines how digital environments have evolved from spaces of opportunity into contested political infrastructures. It explores the changing nature of cyberspace, the role of social media and artificial intelligence in shaping visibility and influence, the actors and networks driving anti-LGBTI+ mobilisation online, and the ways in which digital hostility increasingly translates into offline discrimination, violence, and democratic erosion. Throughout, the report argues that protecting LGBTI+ rights in the digital age requires moving beyond treating online harms as isolated incidents and instead recognising them as products of technological design, political incentives, and evolving systems of governance.

The report also highlights the unequal impact of these developments. Digital harms disproportionately affect those facing multiple forms of marginalisation, including transgender people, refugees, disabled people, racialised minorities, and human rights defenders, demonstrating that digital inequality often mirrors and reinforces existing social inequalities.

Ultimately, we contend that safeguarding LGBTI+ rights online is fundamental to protecting human rights more broadly. As digital technologies become increasingly central to political participation and civic life, ensuring that online spaces remain

safe, inclusive, and rights-respecting is no longer a niche concern but an essential component of democratic resilience.

The accompanying consultation to this report seeks evidence, perspectives, and lived experience to test and strengthen these findings. We particularly welcome

contributions from activists, researchers, civil society organisations, policymakers, and individuals with direct experience of digital harms. Responses will help refine the Commission's analysis and inform future recommendations aimed at strengthening protections for LGBTI+ people in an increasingly digital world.

Purpose of this Consultation

This consultation seeks views from activists, civil society organisations, academics, technology experts, policymakers, and other stakeholders on the findings presented in this paper. The report examines the evolving relationship between digital technologies and LGBTI+ rights, exploring how online spaces have become both vital sites of community, advocacy, and democratic participation, while also increasingly important arenas for surveillance, harassment, disinformation, and anti-rights mobilisation. This paper considers the legal, political, technological, and social factors shaping these developments, while the consultation aims to:

- test the findings of the report against the lived experiences of those most affected;
- identify additional examples and evidence not captured by the desk research;
- and gather perspectives on how future advocacy strategies can be strengthened at national, regional, and international levels.

We particularly welcome responses from:

- LGBTI+ rights activists and organisations;
- civil society groups working on human rights, democracy, and equality;
- academics and researchers in law, politics, and social sciences;
- international organisations, donor states, and development agencies;

- individuals with lived experience of LGBTI+ rights rollbacks and digital advocacy.

How responses will be used

The evidence and perspectives gathered through this consultation will inform the next iteration of the report and contribute to the development of future advocacy strategies by the Commission's Advisory Council. In the interest of transparency, a summary of responses will be published, setting out the key themes raised. Where permission is given, case studies and direct testimonies may be incorporated into future publications.

How to respond

Respondents are encouraged to answer the consultation questions set out under each report section on our website. You do not need to answer every question; partial responses are welcome. Submissions may be made individually or on behalf of an organisation.

List of consultation questions

Section 1 Questions: Why Social Media and Online Spaces Matter for LGBTI+ Rights

1. Have you or your community relied on platform anonymity/pseudonymity features that later proved more

fragile than expected (e.g., through geolocation, metadata, screenshots)? What happened, and what would have prevented it?

2. How does your organisation or community currently weigh the benefits of online visibility (support, advocacy reach, solidarity) against the risks (surveillance, hyper-visibility, targeting)? What would help you make this calculation more safely?

Section 2 Questions: What Is Cyber-space?

1. Has your organisation or community experienced harm connected to data collection you did not meaningfully consent to (e.g., inferred identity data, platform terms changed after the fact)? What would 'meaningful consent' look like in the digital environment your community depends on for survival?
2. How does your community currently balance the protective value of anonymity/pseudonymity against the risk that the same anonymity is used by hostile actors to harass or infiltrate your spaces? What governance changes (platform-level or legal) would improve this balance without eliminating protective anonymity?
3. Has strategic management of your online identity (secondary accounts, partial disclosure, coded language) changed your ability to build community, seek employment, or engage politically?
4. Would framing digital harm using concepts like 'data colonialism' or

'extractive infrastructure' help or hinder your advocacy with policymakers, funders, or the public in your context? What language has proven most persuasive for you so far?

Section 3 Questions: Online Spaces as Modern Civic Arenas

1. What strategies has your organisation used to maintain advocacy capacity during the 'droughts' between moments of high public attention, when algorithmic and media focus has moved elsewhere?
2. Have you observed coordinated anti-LGBTI+ campaigns (coordinated posting, meme networks, hashtag manipulation, bot amplification) in your country or platform environment? What form did they take, and how did mainstream media or institutions respond?
3. Have you encountered anti-LGBTI+ content that used humour, satire, or 'just joking' framing to spread narratives while evading moderation? How did this affect your ability to respond or report it?
4. If you operate in a context with restrictive digital laws (e.g., propaganda laws, extremist designations, VPN restrictions), how has this affected what you and your community feel able to search for, post, or organise online? What protective measures have proven effective?
5. Have you observed cases where anti-LGBTI+ narratives that originated or gained traction on social media were subsequently picked up and legitimised by mainstream news coverage?

What would help journalists avoid amplifying coordinated disinformation?

Section 4 Questions: Algorithms - Amplification, Bias, and Free Speech

1. Has your organisation or community experienced content being removed, downranked, or 'shadowbanned' for identity-affirming speech, while comparable hostile content targeting LGBTI+ people remained active? Can you provide specific examples (platform, content type, outcome)?
2. Have you experienced coordinated reporting/flagging campaigns used to suppress LGBTI+ content or accounts? What was the platform's response, and how effective - or ineffective - was it?
3. Are you aware of cases in your context where algorithmically amplified anti-LGBTI+ narratives influenced public debate, policy, or legislation?
4. What strategies (coded language, hashtag practices, format changes) has your community developed to maintain visibility within algorithmic constraints? What has this cost you in terms of time, reach, or psychological burden?
5. Where you or your organisation have experienced online harm that may meet a criminal threshold (harassment, incitement, coordinated abuse), what barriers did you face in reporting, evidence-gathering, or securing law enforcement action - particularly across jurisdictions?

Section 5 Questions: AI-Generated Propaganda and Emotional Manipulation

1. Has your organisation or community experienced fabricated statements, deepfake imagery, or synthetic 'evidence' targeting LGBTI+ individuals or advocacy work? If so, what form did it take, and how quickly - if at all - were you able to have it corrected or removed?
2. Have you or your organisation had authentic statements, footage, or testimony dismissed by opponents as 'AI-generated' or fabricated? How did this affect your credibility or advocacy work, and what response strategies - if any - proved effective?
3. Does your organisation currently have the technical capacity, training, or resources to identify synthetic media and respond to it?
4. What responsibility should platforms bear for the amplification of AI-generated disinformation targeting LGBTI+ communities, and what concrete platform-level interventions (labelling, detection tools, faster takedown processes) would you prioritise?
5. Should human rights and LGBTI+ advocacy organisations begin using generative AI defensively (e.g., synthetic content detection and rapid-response verification tools) even if they continue to avoid manipulative uses? What ethical boundaries should guide this?
6. In your jurisdiction, are there adequate legal mechanisms to address AI-generated defamation, harassment, or

impersonation targeting LGBTI+ individuals or organisations? What reforms are most urgently needed?

Section 6 Questions: The Political Economy of Hate

1. Have you seen evidence of specific platform design features (recommendation algorithms, monetisation/verification systems, moderation policy changes) measurably increase anti-LGBTI+ content visibility or engagement?
2. In your experience, do alternative/low-moderation platforms (e.g., Telegram, Rumble, Truth Social) function as radicalisation or coordination spaces that later surface on mainstream platforms? What would effective intervention look like?
3. What indicators – in your experience – most reliably signal that online anti-LGBTI+ hostility is escalating toward offline harm (protests, threats, violence)?
4. Given that identical online narratives produce vastly different real-world consequences depending on local legal and political context, what support (legal, technological, advocacy) is most urgently needed in high-risk jurisdictions?

Section 7 Questions: Far-Right Online Ecosystems and Anti-LGBTI+ Mobilisation

1. Have you observed cooperation or coordination between domestic and

international anti-rights actors? If so, how has this affected your work or the communities you support?

2. Which online narratives or campaigns have been most effective in fuelling hostility towards LGBTI+ people in your context?
3. Within the LGBTI+ community, which groups are most affected by these digital campaigns, and how do factors such as race, gender identity, disability, migration status, or other characteristics shape those experiences?
4. What strategies have been most effective in disrupting, countering, or building resilience against coordinated anti-LGBTI+ mobilisation online?
5. What actions should governments, technology companies, international organisations, and donors prioritise to better protect LGBTI+ activists and communities from coordinated digital anti-rights campaigns?

Section 8 Questions: Hidden Infrastructures of Harm

1. Have you or your organisation experienced coordinated online harassment, surveillance, doxxing, or other forms of abuse that appeared to be organised through private or encrypted digital channels?
2. Are hidden digital infrastructures – such as encrypted messaging groups, anonymous forums, or illicit data sharing – playing an increasing role in anti-LGBTI+ mobilisation in your country or region?

3. What practical digital security challenges do LGBTI+ organisations face when balancing the need for privacy, anonymity, and secure communications with the risks of misuse by hostile actors? Have data breaches, doxxing, impersonation, or the misuse of personal information affected LGBTI+ activists, organisations, or communities where you work?
4. What support, tools, or resources would most strengthen the digital security and resilience of LGBTI+ activists and organisations in your context?
5. What responsibilities should governments, technology companies, and encrypted messaging providers have in addressing coordinated digital harms while protecting privacy, freedom of expression, and secure communications?

Section 9 Questions: Predatory Communities and Gendered Exploitation

1. To what extent have online misogyny, anti-gender movements, or ‘manosphere’ communities contributed to hostility towards LGBTI+ people in your context?
2. Have emerging technologies – including generative AI and deepfakes – changed the nature or scale of digital abuse faced by your organisation or the communities you support?
3. What barriers do survivors encounter when reporting or seeking redress for technology-facilitated gender-based violence, including image-based abuse or synthetic sexual content?

4. What practical measures should governments, technology companies, and international organisations take to better prevent and respond to these forms of abuse while protecting freedom of expression and privacy?
5. Are there examples of effective laws, platform policies, survivor support services, or community-led initiatives that have successfully addressed technology-facilitated gender-based violence affecting LGBTI+ people?

Section 10 Questions: Shifting the Overton Window - Normalisation, Platform Abdication, and the Persistence of Harm

6. Which communities within the LGBTI+ movement are most affected by the transition from online hostility to offline harm, and what factors contribute to their increased vulnerability?
7. Have digital campaigns influenced public opinion, political debate, or legislative developments affecting LGBTI+ rights in your country or region? If so, how?
8. What measures have been most effective in protecting LGBTI+ activists and organisations from the real-world impacts of coordinated digital hostility?
9. What additional actions should governments, technology companies, international organisations, and donors take to strengthen accountability and prevent digital harms from escalating into offline violence or discrimination?

Section 11 Questions: Political Parties, Power, and Online Mobilisation

1. Which policy or legal reforms should governments prioritise to better protect the digital rights of LGBTI+ people?
2. What additional responsibilities should technology companies have in preventing and responding to digital harms affecting LGBTI+ communities?
3. How can civil society organisations, activists, and community networks be better supported to strengthen digital resilience and respond to emerging online threats?
4. Are there examples of effective laws, regulatory frameworks, platform policies, or community-led initiatives?

5. What safeguards are needed to ensure that efforts to regulate digital platforms do not undermine privacy, secure communications, freedom of expression, or the work of human rights defenders?

Section 12 Questions: The Unaccountability Paradox

1. In this report, which findings most closely reflect your experience or the experiences of the communities you work with?
2. What additional evidence, case studies, or examples should be included before the report is finalised?
3. Looking ahead, what emerging technologies, digital trends, or policy challenges should our community prioritise?

Glossary

Biphobia: Biphobia is the aversion, fear, or discrimination against bisexual individuals, which can manifest through negative attitudes, stereotypes, and prejudices towards those who are attracted to both men and women.

Bisexual: A term used to describe a romantic and/or sexual orientation towards more than one gender. Bisexual people may describe themselves using one or more of a wide variety of terms, including, but not limited to, bisexual, pansexual, queer, and many others.

Civil society: Refers to the space for collective action around shared interests, purposes, and values, generally distinct from government and commercial for-profit actors. This can include charities, nongovernmental organisations, community groups, women's organisations, faith-based organisations, trade unions, social movements, coalitions, and advocacy groups.

Cisgender: Someone whose sense of gender identity is the same as the sex assigned at their birth.

Gay: Refers to a man who has a romantic and/ or sexual orientation towards men. Also, a generic term for lesbian and gay sexuality – some women define themselves as gay rather than lesbian. Some non-binary people may also identify with this term.

Gender expression: How a person chooses to outwardly express their gender within the context of societal expectations of gender. A person who does not conform to societal expectations of gender may not, however, identify as transgender.

Gender identity: A person's innate sense of their own gender, whether male, female or something else, which may or may not correspond to their sex assigned at birth.

Homophobia: The aversion, fear, or discrimination against homosexual individuals, which can manifest through negative attitudes, stereotypes, and prejudices towards those who are attracted to people of the same sex. This can result in social exclusion, harassment, and violence against lesbian, gay, and other non-heterosexual people.

Homosexual: A person attracted to the same sex.

ICCPR: International Covenant on Civil and Political Rights.

Intersex: A term used to describe a person who may have the biological attributes of both sexes or whose biological attributes do not fit with societal assumptions about what constitutes male or female. Intersex people may identify as male, female or non-binary.

Intersectionality: A theoretical framework for understanding how aspects of one's social and political identities (gender, race, class, sexuality, ability, etc.) might combine to create unique modes of discrimination.

Intersexphobia: The aversion, fear, or discrimination against intersex individuals, who are born with physical sex characteristics that do not fit typical definitions of male or female.

Lesbian: Refers to a woman who has a romantic and/or sexual orientation towards women. Some non-binary people may also identify with this term.

LGBTI+: Lesbian, Gay, Bisexual, Trans, and Intersex.

NGO: Non-governmental organisation.

Sex characteristics: A person's physical characteristics relating to sex, including genitalia and other reproductive anatomy, chromosomes and hormones, as well as secondary physical characteristics emerging from puberty.

Sexual orientation: A person's sexual identity in relation to the gender to which they are attracted; the fact of being het-

erosexual, homosexual, or bisexual; a person's sexual attraction to other people, or lack thereof.

Trans/Transgender: An umbrella term to describe people whose gender is not the same as, or does not conform to, the sex they were assigned at birth. Transgender may also include people who belong to a third gender, or else conceptualise transgender people as a third gender.

Transphobia: The aversion, fear, or discrimination against transgender individuals, who have a gender identity or expression that differs from the sex they were assigned at birth. This can manifest through negative attitudes, stereotypes, social exclusion, harassment, and violence against transgender people.

UDHR: The Universal Declaration of Human Rights is an international document adopted by the United Nations General Assembly that enshrines the rights and freedoms of all human beings.



Introduction

For many LGBTI+ people, the internet is simultaneously a site of visibility, community, and survival as well as a terrain of surveillance, harassment, and organised hostility. Digital spaces that were once considered to be spaces of connection and self-expression have increasingly become environments in which identity can be weaponised, amplified through algorithmic systems, and targeted at scale. The consequences extend far beyond the screen; online abuse shapes political discourse, legitimises discrimination, undermines democratic participation, and – in many cases – translates directly into material and physical harm.

This paper is structured in four parts in order to explore how digital infrastructures have become central to both the protection and the erosion of LGBTI+ rights.

Furthermore, the report asks what these dynamics reveal about the changing relationship between technology, power, and democratic society.

The first section establishes the conceptual and political foundations of the analysis. It examines why digital spaces hold particular significance for LGBTI+ communities, not merely as supplementary channels of interaction but as essential social, political, and economic infrastructure. It further considers what it means for human rights protections to extend into digitally mediated forms of life.

The second section analyses the structure and dynamics of cyberspace itself. It explores the ways in which online identities are constructed, commodified, circulated, and – at times – exploited online; how

algorithmic systems shape visibility, exclusion, and participation processes; and how artificial intelligence is transforming the architecture of influence and information governance in a manner which impacts all aspects of democratic participation and human rights-focused advocacy.

The third section develops a typology of the forces and actors driving anti-LGBTI+ harm in digital environments. Rather than treating harassment, disinformation, and organised hate as isolated phenomena, it maps the broader ecosystem through which such harms are produced and sustained, including far-right networks, shadow infrastructures, predatory online communities, and the mechanisms through which intolerance becomes normalised.

The fourth section examines the relationship between online hostility and offline

consequences. It investigates how digitally mediated forms of hate and exclusion translate into political mobilisation, physical violence, and democratic erosion while also considering the challenges of accountability in an environment where responsibility is frequently fragmented and obscured.

Throughout, this paper advances a central argument: that LGBTI+ communities function as a critical indicator of wider democratic vulnerability in digital society. The forces disproportionately directed against them – including surveillance, disinformation, algorithmic suppression, transnational hate networks, and forms of state-platform convergence – are not exceptional or isolated phenomena. Rather, they reveal broader structural conditions that increasingly threaten democratic life as a whole.



I. SETTING THE SCENE

Before examining the architecture of digital platforms, the actors who exploit them, or the offline consequences of on-line harm, it is necessary to establish why digital space matters to LGBTI+ rights in the first place.

For many LGBTI+ people, particularly those living in socially conservative, legally restrictive, or otherwise hostile environments, online platforms are not a peripheral or optional part of social life. They are frequently the first – and sometimes the only – accessible space in which identity can be explored, community found, and information sought without immediate exposure to family, employers, or state au-

thorities. Digital access in these contexts is not a convenience; it is foundational infrastructure for survival, self-recognition, and connection.

At the same time, the very features that make digital spaces valuable (visibility, connectivity, searchability, and speed) are the same features that generate risk. Anonymity that protects can also embolden those who wish to do harm. Visibility that builds solidarity can also expose individuals to surveillance, harassment, or violence. Empowerment and exposure are not sequential stages of digital life but two effects of the same underlying mechanisms operating simultaneously.

Section 1: Why Social Media and Online Spaces Matter for LGBTI+ Rights

Introduction: Digital Life as a Lifeline

Over the past two decades, digital technologies have fundamentally reshaped how individuals connect, organise, and participate in society.^{1 2} For many LGBTI+ people around the world, these transformations have been particularly profound.

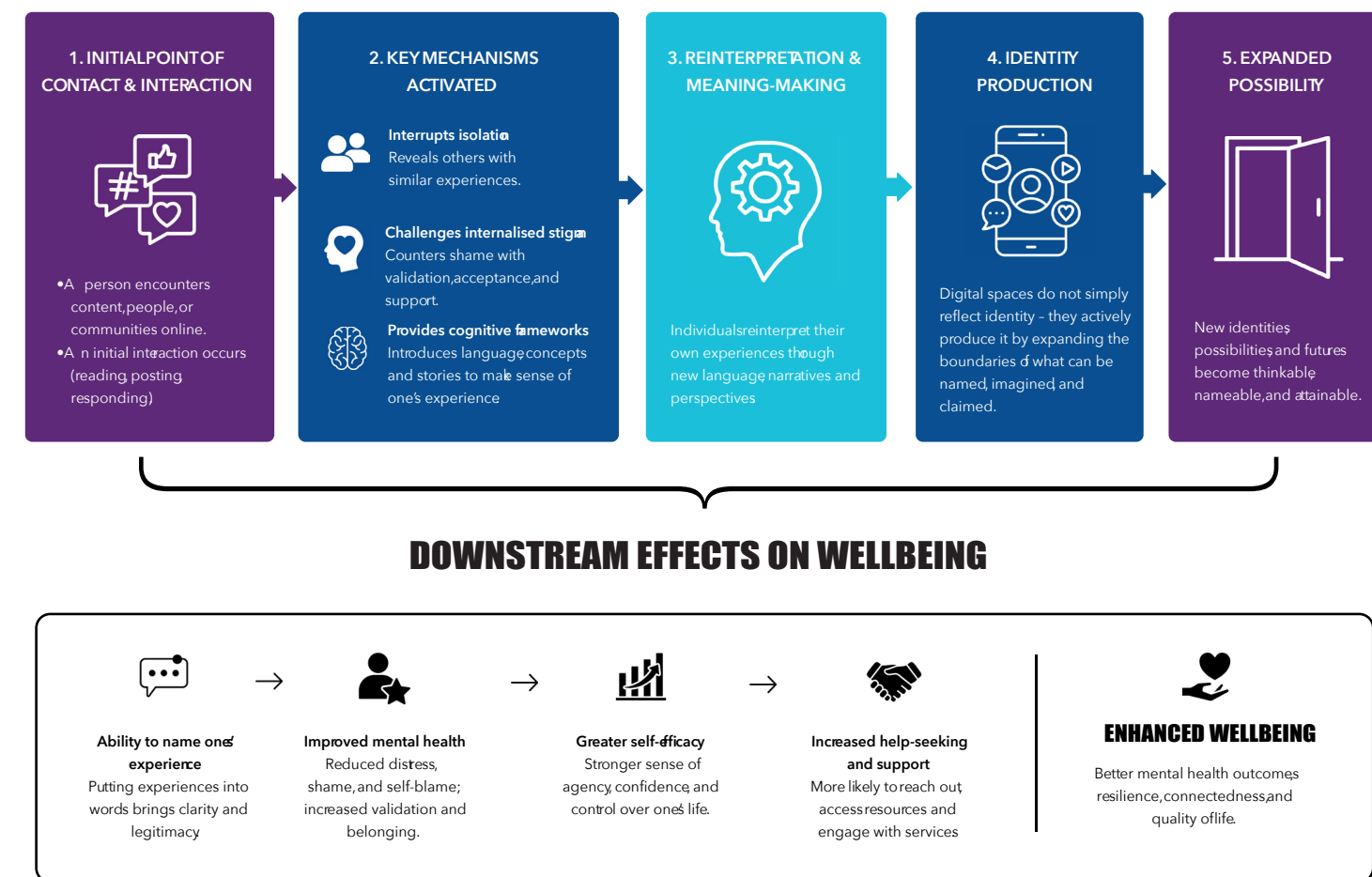
Online platforms now function not merely as tools of communication but as essential infrastructures of community, identity formation, and political activism.³

Especially for individuals living in socially conservative or legally restrictive environments, the internet may provide the first opportunity to encounter others who

share similar experiences.⁴ In contexts where public expressions of queer identity remain heavily stigmatised or criminalised, physical LGBTI+ spaces – such as community centres, advocacy organisations, or social venues – may be limited or entirely absent. Additionally, access to

terminology, personal testimonies, and shared narratives plays a critical role in identity recognition and self-acceptance, particularly where formal education systems exclude or erase LGBTI+ experiences. In such contexts, digital access is not supplementary but foundational;

How Digital Spaces Produce Identity and Improve Wellbeing



This is a reinforcing cycle: positive interactions and identity affirmation online can lead to improved wellbeing which further enables connection, participation, and growth.

1 Manuel Castells, *Networks of Outrage and Hope: Social Movements in the Internet Age*, 2nd ed. (Cambridge: Polity Press, 2015).

2 danah boyd, *It's Complicated: The Social Lives of Networked Teens* (New Haven, CT: Yale University Press, 2014), <https://www.danah.org/books/ItsComplicated.pdf>

3 Carolina Are, Catherine Talbot, and Pam Briggs, "Social Media Affordances of LGBTQIA+ Expression and Community Formation," *Convergence* 31, no. 4 (2025): 1401-22, <https://doi.org/10.1177/13548565241296628>

4 Ibid.

emerging research shows that first encounters with queer identity, language, and community are increasingly occurring online, particularly among young people in restrictive settings.⁵

This initial point of contact and interaction is critical; it can interrupt isolation, challenge internalised stigma, and provide cognitive frameworks through which individuals reinterpret their own experiences. In this sense, digital spaces do not simply reflect identity; they actively produce it by expanding the boundaries of what can be named, imagined, and claimed.⁶ This has downstream effects on wellbeing, as the ability to name one's experience is closely tied to mental health, self-efficacy, and the capacity to seek support.

Scholars have described these environments as forms of "networked refuge," replicating many of the functions histori-

cally served by physical community institutions as well as enabling peer support, knowledge exchange, and collective identity formation across geographic boundaries.⁷

However, unlike physical spaces, these networks can be entered anonymously, navigated selectively, and exited quickly, offering forms of safety that are otherwise unavailable in hostile environments. This conditional visibility allows individuals to calibrate risk in real time – a crucial feature in contexts where exposure can result in social exclusion, violence, or legal repercussions.⁸ Without meaningful safeguards, anonymity and selective visibility can produce only a perception of safety rather than safety itself, particularly within platforms that continuously collect, store, and infer information about user behaviour.⁹

Case Study: Perceive Anonymity and Location-Based Exposure on Grindr¹⁰

Location-based dating applications such as Grindr are often experienced by users as relatively private or anonymous spaces, particularly in contexts where offline LGBTI+ visibility carries significant social or legal risk. Features such as pseudonymous profiles, selective self-disclosure, and the absence of publicly searchable identities can create a perception of safety and control.

However, digital security researchers and human rights organisations have repeatedly demonstrated that these protections can be fragile or incomplete. In several countries, activists and hostile actors have used geolocation features to identify users' approximate locations through a process known as triangulation, in which the relative distance displayed between profiles is used to narrow down a user's physical position. Even where exact coordinates are not publicly visible, repeated location measurements from different points can make identification possible.

Digital spaces also enable forms of identity exploration that may not be possible offline. Features such as pseudonymity, anonymity, and selective self-disclosure allow individuals to control how and when aspects of their identity are revealed.¹¹ This creates opportunities for gradual, self-directed exploration of sexual orientation and gender identity. Such control is particularly significant in environments characterised by uncertainty or threat,

where premature or involuntary disclosure can carry serious social or material consequences.¹²

Scholars describe this process as "identity rehearsal": a staged form of self-expression in which individuals test and refine aspects of their identity within relatively lower-risk environments.¹³ Individuals may come out first within online communities before disclosing their identities in offline contexts. This sequencing allows

5 Celia B. Fisher, Xiangyu Tao, Michael D. Ford, and Madeline Ford, "Social Media: A Double-Edged Sword for LGBTQ+ Youth," *Computers in Human Behavior* 156 (2024): 108194, <https://doi.org/10.1016/j.chb.2024.108194>

6 Vincenzo Auriemma, Daniele Battista, and Serena Quarta, "Digital Embodiment as a Tool for Constructing the Self in Politics," *Societies* 13, no. 12 (2023): 261, <https://doi.org/10.3390/soc1312026>

7 Fisher et al., "Social Media."

8 Human Rights Watch, *All This Terror Because of a Photo: Digital Targeting and Its Offline Consequences for LGBT People in the Middle East and North Africa* (New York: Human Rights Watch, 2023), <https://www.hrw.org/report/2023/02/21/all-terror-because-photo/digital-targeting-and-its-offline-consequences-lgbt>

9 Tahrir Institute for Middle East Policy, "Digital Persecution: Surveillance in Queer Digital Spaces," August 25, 2023, <https://timep.org/2023/08/25/digital-persecution-surveillance-in-queer-digital-spaces/>

10 Human Rights Watch, *All This Terror Because of a Photo*.

11 boyd, *It's Complicated*.

12 Oz, Mustafa, Akan Yanik, and Mikail Batu. "Under the Shadow of Culture and Politics: Understanding LGBTQ Social Media Activists' Perceptions, Concerns, and Strategies." *Social Media + Society* 9, no. 3 (2023). <https://doi.org/10.1177/20563051231196554>

13 Ibid.

individuals to build confidence, develop language, and anticipate responses, effectively rehearsing both identity and interaction.¹⁴ In doing so, digital spaces function as preparatory environments for navigating offline life.

In environments where premature disclosure could result in family rejection, social exclusion, or violence, this process becomes not merely beneficial but necessary.¹⁵ Digital platforms thus provide a degree of agency that may be otherwise unavailable.

Case Study: Online Queer Support Networks in Uganda¹⁶

For many young LGBTI+ people in Uganda, online spaces have become one of the few available environments in which sexual orientation or gender identity can be explored with relative safety. Following the introduction and later expansion of anti-LGBTI+ legislation – including the 2023 Anti-Homosexuality Act – activists and community organisations reported increased fear surrounding public visibility, offline organising, and disclosure within family or educational settings.

In this context, encrypted messaging applications, closed social media groups, and pseudonymous online accounts became critical mechanisms for connection and support. Community members used digital platforms to access information about sexuality and gender identity, seek mental health support, share security advice, and communicate with peers without immediately exposing themselves within hostile offline environments.

For many individuals, these online interactions represented a first opportunity to encounter affirming language and narratives about queer identity. Activists noted that individuals frequently disclosed their identities online before doing so in offline contexts, allowing them to test forms of self-expression, build confidence, and assess potential risks gradually rather than through immediate public disclosure.

14 Tufekci, Zeynep. "Social Media and Human Rights: A Timeline." EachOther, 2023. <https://eachother.org.uk/social-media-and-human-rights-a-timeline/>.

15 Fisher et al., "Social Media."

16 Human Rights Watch, All This Terror Because of a Photo.

Representation within these spaces further shapes identity formation. Online creators, activists, and influencers often function as informal role models, offering diverse representations of queer life that challenge dominant narratives of devi-

ance or immorality. These representations expand the horizon of possibility.¹⁷ Seeing other LGBTI+ individuals living openly can reshape perceptions of what a future might look like, transforming abstract identity into imaginable life pathways.

Case Study: Queer Digital Creators and Visibility in India¹⁸

In India, digital platforms played a significant role in expanding public visibility for LGBTI+ communities both before and after the 2018 decriminalisation of same-sex relations under Section 377 of the Indian Penal Code. While mainstream media representations of queer life have historically been limited, stereotypical, or stigmatising, social media platforms and video-sharing sites enabled a growing number of LGBTI+ creators to speak publicly about their experiences.

Queer creators on platforms such as YouTube and Instagram began producing content discussing coming out, relationships, mental health, family dynamics, gender identity, and everyday queer life. Rather than presenting LGBTI+ identity solely through the framework of crisis or criminality, these creators documented ordinary forms of existence, friendship, humour, intimacy, and professional success.

For many young viewers – particularly those living outside major urban centres or within socially conservative households – this visibility provided one of the first accessible representations of openly queer life within a familiar cultural and linguistic context. Activists and researchers noted that exposure to such content often helped individuals develop language for understanding their own identities and imagine futures that extended beyond isolation, secrecy, or social exclusion.

This case illustrates how digital representation can reshape the perceived boundaries of possibility. Online visibility does not merely communicate information; it can alter what individuals believe is survivable, attainable, and socially imaginable within their own societies.

17 Are, Talbot, and Briggs, "Social Media Affordances".

18 Ibid.

Importantly, these dynamics extend beyond LGBTI+ communities themselves. Exposure to personal narratives online can foster empathy among broader audiences, positioning digital spaces as sites not only of intra-community support but of wider social learning.¹⁹ In this way, the individual-level impacts of digital engagement are not contained but ripple outward, contributing to gradual shifts in social attitudes that – over time – can alter the very conditions that produce isolation.

Digital spaces also function as archives of cultural survival. Through social media, blogs, and online publications, LGBTI+ individuals document lived experiences, preserve community histories, and produce cultural expression. These digital traces accumulate into forms of collective memory, sustaining communities across time as well as space. They also counter state and societal attempts at erasure by creating distributed, difficult-to-destroy repositories of knowledge, ensuring that even where physical spaces are dismantled, cultural continuity can persist.²⁰

From Empowerment to Exposure: Risks in Digital Spaces

For many individuals, digital spaces are not simply extensions of offline communi-

ties but the foundation upon which those communities are built. This does not make digital technology an uncomplicated good. It means that digital infrastructures have become unavoidable and a necessary part of modern life; they are now among the places where care happens, where harm happens, and where the boundary between the two is constantly negotiated.

Therefore, the opportunities created by digital platforms for LGBTI+ individuals cannot be separated from the risks embedded within them. Online platforms make connection possible through the same mechanisms that make exposure possible. Their importance lies not only in what they enable but in what they substitute; every substitution carries a cost. Digital participation, therefore, often requires individuals to navigate infrastructures that facilitate care and community while simultaneously generating new forms of visibility and vulnerability.²¹ These systems are largely privately owned, data-driven environments that were not designed primarily for care, safety, or rights protection.²²

Even at the earliest stages of identity exploration, this trade-off is already present. The search for language, identity, and representation leaves traces. The

act of finding others can be observed. Someone’s online profile, messages, likes, search terms, photographs, or locations are all signals that make connection possible, but they can also become evidence that may be used for leverage, surveillance, or targeting.²³ Without such access, identity formation can be delayed, fragmented, or shaped exclusively by hostile narratives; with access, identity formation may become more possible but also more surveillable.

Scholars have described these environments as forms of “networked refuge,” replicating many of the functions historically served by physical community institutions by enabling peer support, knowledge exchange, and collective identity formation across geographic boundaries. Unlike physical spaces, however, these networks can be entered anonymously, navigated selectively, and exited quickly, offering forms of safety that are otherwise unavailable in hostile environments. This conditional visibility allows individuals to calibrate risk in real time, a crucial feature in contexts where exposure can result in social exclusion, violence, or legal repercussions. Yet, conditional visibility is still visibility.²⁴ It may be safer than forced exposure, but it is not the same as privacy. The refuge is real, but it is built inside

infrastructures that record, rank, recommend, and monetise behaviour. Safety within digital environments is, therefore, often provisional rather than guaranteed, dependent on constantly shifting relationships between visibility, anonymity, platform governance, and external political conditions.

Digital spaces also enable forms of identity exploration that may not be possible offline. Features such as pseudonymity, anonymity, and selective self-disclosure allow individuals to control how and when aspects of their identity are revealed. This creates opportunities for gradual, self-directed exploration of sexual orientation and gender identity. Such control is particularly significant in environments characterised by uncertainty or threat, where premature or involuntary disclosure can carry serious social or material consequences.

Individuals may come out first within online communities before disclosing their identities in offline contexts. This sequencing allows individuals to build confidence, develop language, and anticipate responses, effectively rehearsing both identity and interaction. In doing so, digital spaces function as preparatory environments for navigating offline life.

19 Castells, *Networks of Outrage and Hope*

20 Ibid.

21 Organisation for Economic Co-operation and Development (OECD), *Rights in the Digital Age* (Paris: OECD Publishing, 2024), https://www.oecd.org/en/publications/rights-in-the-digital-age_deb707a8-en.html

22 Ibid.

23 Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs, 2019.

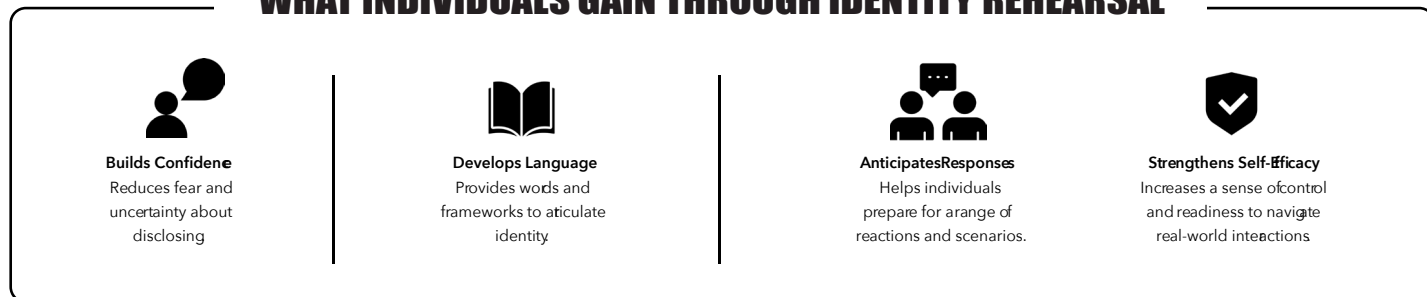
24 OECD, *Rights in the Digital Age*.

Identity Rehearsal: Digital Spaces as Preparatory Environments

A staged process in which individuals first explore and refine their identity online before disclosing in offline contexts.



WHAT INDIVIDUALS GAIN THROUGH IDENTITY REHEARSAL



Digital spaces function as preparatory environments, enabling individuals to rehearse both identity and interaction - bridging from online exploration to offline living with greater safety and confidence

At the same time, access to affirming communities is strongly associated with improved mental health outcomes.²⁵ Social connection functions as a protective factor, reducing psychological distress

and strengthening resilience. Crucially, this is not simply about the presence of others but about being recognised in ways that counteract experiences of erasure or rejection. This recognition oper-

ates as a psychological counterweight to stigma, disrupting cycles of shame and self-surveillance that often emerge in hostile environments.

Yet, this also introduces a second vulnerability; to be seen by those who affirm may also mean becoming visible to those who condemn.²⁶ The emotional value of online recognition, therefore, cannot be separated from the risks created by searchable, shareable, and persistent digital traces. Agency within these environments is always partial. Users may choose a pseudonym, but they do not control platform metadata practices, the screenshots taken by others, recommendation systems that infer interest, or the state and private actors who may seek access to their devices.²⁷

The same outward movement that enables solidarity can also carry hostile attention back toward the individual. A story that humanises one audience can be clipped, mocked, misrepresented, or circulated within hostile networks by another.²⁸ Digital visibility, therefore, does not move in one moral direction; it opens routes for solidarity and attack at the same time.

The opportunities created by digital platforms are, therefore, inseparable from the harms they can produce. This is not a tem-

porary flaw that better design alone will resolve, but a structural feature of digital life itself. The same traceability that allows people to find community allows hostile actors to reconstruct networks. The same connectivity that allows solidarity to cross borders allows campaigns of abuse to do the same. Visibility, traceability, and connectivity operate simultaneously as mechanisms of empowerment and mechanisms of vulnerability.²⁹

This is why risk cannot be treated as a caveat to an otherwise positive story. It is the other half of the story.

Digital technologies do not first empower and later expose. They empower through exposure. They create community through discoverability. They enable mobilisation through circulation. They offer safety through distance, anonymity, and speed. Yet, each of those protective features can also be reversed: Anonymity can protect queer users or embolden abusers; speed can mobilise rescue or accelerate harassment; distance can create refuge or make accountability more difficult. The central question, therefore, is not whether digital technologies are good or bad for LGBTI+ rights but who benefits from these trade-offs, who absorbs the risks, and what forms of governance exist to limit the resulting harms.

25 Are, Talbot, and Briggs, "Social Media Affordances".

26 Tahrir Institute for Middle East Policy, "Digital Persecution."

27 Ibid.

28 Ibid.

29 Zuboff, The Age of Surveillance Capitalism.

In many contexts, governments have expanded their capacity to monitor online activity.³⁰ Social media platforms, messaging applications, and dating services are increasingly used to identify individuals considered politically subversive or morally deviant.³¹ This reflects a broader shift toward digitalised governance, where identity is not only expressed online but also inferred, tracked, and classified through data traces – often without transparency or due process.

For LGBTI+ individuals living under criminalised regimes, digital communication can, therefore, become a site of vulnerability. Interactions perceived as private may be monitored, intercepted, or used as evidence in legal proceedings. The boundary between private and public collapses in these settings, meaning that everyday forms of communication, such as messages, profile views, or digital ‘likes,’ can acquire legal or punitive significance.

Case Study: Digital Entrapment in Egypt³²

In Egypt, security forces have repeatedly used social media and dating applications to entrap individuals suspected of being gay. Officers create fake profiles and initiate conversations with LGBTI+ users before persuading them to meet in person. When victims arrive at the arranged location, they are arrested.

Human rights organisations have documented hundreds of arrests based on perceived sexual orientation or gender identity in recent years, many of which involved digital entrapment. Authorities frequently use photographs, chat histories, and dating applications found on individuals’ phones as evidence in prosecutions under morality laws.

In some reported cases, security forces have fabricated digital evidence by installing dating applications or creating fake conversations on confiscated devices.

Non-state actors also play a significant role in digital violence. Harassment campaigns, doxxing, forced outing, and blackmail are frequently directed at LGBTI+ individuals. These actions often produce severe offline consequences, including

loss of employment, family violence, and forced displacement. Unlike state-led surveillance, these harms are often decentralised and socially embedded, making them harder to regulate and more unpredictable in their escalation.

The cumulative effect is a chilling environment in which individuals may withdraw from online participation or self-censor in order to reduce risk. Visibility, which enables empowerment, also becomes a liability. This does not cancel the value of visibility; it complicates it. For many LGBTI+ people, invisibility is not safety but erasure, while visibility is not liberation but exposure.

Case Study: Forced Outing Campaign in Morocco³³

During COVID-19 lockdowns in 2020, a Moroccan social media influencer encouraged followers to identify nearby users on the dating application Grindr and publicly expose them online.

The campaign spread rapidly across social media platforms. Screenshots of users’ profiles were circulated widely, leading to the forced outing of numerous individuals. Victims reported being expelled from their homes, losing employment, and experiencing violence from family members.

The incident demonstrated how digital platforms can be weaponised by private actors to target vulnerable communities and how quickly online harassment campaigns can escalate into serious offline harm.

Digital life for LGBTI+ people living in dangerous environments, therefore, forces them to make an impossible calculation: to remain visible enough to access support, community, and recognition while remaining invisible enough to avoid becoming a target.

30 Human Rights Watch, All This Terror Because of a Photo.

31 Zuboff, The Age of Surveillance Capitalism.

32 Human Rights Watch, All This Terror Because of a Photo.

33 Human Rights Watch, All This Terror Because of a Photo.

Social Media as a Tool for Activism and Community Building

Historically, LGBTI+ communities were often misrepresented or excluded from mainstream media narratives. Digital platforms have partially disrupted this dynamic, enabling individuals and organisations to communicate directly with audiences.³⁴ Activists can now share lived experiences, document human rights abuses, and mobilise support without mediation. This capacity to bypass institutional filters allows marginalised communities to shape their own narratives and contest misrepresentation.³⁵ This disintermediation changes not only who speaks but also whose knowledge is recognised as legitimate, allowing lived experience to function as a form of evidence rather than anecdote.

Digital activism also expands the reach of advocacy. By targeting individuals who may not hold fixed positions – the so-called ‘movable middle’ – online campaigns can humanise abstract political debates and challenge entrenched stereotypes.³⁶ Through personal storytelling, visual media, and direct engagement, digital platforms can shift public sentiment incrementally, creating openings

for policy change that might not emerge through institutional advocacy alone.³⁷

Digital activism also facilitates the formation of transnational networks. Organisations and individuals can exchange strategies, coordinate campaigns, and build solidarity across borders. These networks enable the circulation of knowledge and resources among communities facing similar challenges in different contexts.³⁸

The speed and scale of digital communication have further transformed mobilisation. Hashtag campaigns, viral content, and online petitions allow advocacy messages to reach global audiences rapidly.³⁹ These dynamics have contributed to the emergence of networked social movements, characterised by decentralised coordination and distributed participation. This horizontal exchange can be particularly valuable in restrictive environments where local civil society infrastructures are weak or under threat, allowing activists to draw on external expertise, funding pathways, and advocacy models.⁴⁰

Additionally, in contexts where physical protest is restricted, digital platforms enable alternative forms of mobilisation. ‘Digital Pride’ initiatives – including online

events, livestream discussions, and virtual exhibitions – allow individuals to partic-

ipate in expressions of identity and solidarity while mitigating physical risk.⁴¹

Case Study: Global Attention to the Chechnya Crackdown⁴²

In 2017, reports emerged that authorities in Chechnya had detained and tortured men suspected of being gay. Local activists and journalists documented testimonies from survivors, but information initially struggled to reach international audiences due to state repression and limited press freedom.

Social media played a critical role in amplifying the story. Reports circulated rapidly across online networks, drawing global attention to the situation. Activists used digital platforms to share survivor testimonies, organise advocacy campaigns, and mobilise diplomatic pressure.

International organisations also used online communication channels to coordinate evacuation efforts for individuals at risk. The rapid dissemination of information through social media helped transform a local human rights crisis into an issue of global concern.

Digital Activism and the Vulnerabilities of Networked Visibility

Beyond individual identity formation and community building, digital platforms have transformed the landscape of collective action. Social media has altered how information circulates, reducing reliance on traditional gatekeepers such as mainstream media and state institutions. This shift is particularly significant for marginalised groups, including LGBTI+ communi-

ties, whose access to public discourse has historically been constrained by exclusion, misrepresentation, or censorship.

Historically, LGBTI+ communities were often excluded from mainstream media narratives or represented through hostile and pathologising frames. Digital platforms have partially disrupted this dynamic by enabling individuals and organisations to communicate directly with audiences. Activists can share lived experiences, document abuses, mobilise support, and

34 Ibid.

35 Kate Klonick, "Beyond the Public Square: Imagining Digital Democracy," Yale Law Journal 131, no. 8 (2022), <https://yalelawjournal.org/essay/beyond-the-public-square-imagining-digital-democracy>

36 Are, Talbot, and Briggs, "Social Media Affordances."

37 Ibid.

38 Castells, Networks of Outrage and Hope

39 Ibid.

40 Ibid.

41 Are, Talbot, and Briggs, "Social Media Affordances".

42 Human Rights Watch, All This Terror Because of a Photo.

contest misinformation without relying entirely on institutional intermediaries. This disintermediation changes not only who speaks but also whose knowledge is recognised as legitimate, allowing lived experience to function as a form of political evidence rather than anecdote.

Digital activism also expands the reach of advocacy. By targeting individuals who may not hold fixed positions, online campaigns can humanise abstract political debates and challenge entrenched stereotypes through personal storytelling, visual media, and direct engagement. Yet, reach is never neutral. The same post that reaches a potential ally can also reach a hostile influencer; the same hashtag that builds solidarity can be hijacked; the same viral testimony that generates pressure can expose the speaker to harassment. Scale is, therefore, both a political opportunity and a risk multiplier.

Digital platforms also facilitate the formation of transnational advocacy networks. Organisations and activists can exchange strategies, coordinate campaigns, circulate resources, and build solidarity across borders. These dynamics have contributed to the emergence of networked social movements characterised by decentralised coordination and distributed participation. In restrictive environments, such horizontal exchange can provide access to expertise, visibility, and support structures that may not exist locally.

However, the speed and scale of digital communication also introduce new forms of vulnerability. Information can travel faster than verification, protection

planning, or consent. Images of victims, names of activists, and details of local networks may circulate globally before those most affected have had the opportunity to assess what exposure will mean for their safety. Visibility that generates international attention can simultaneously increase local risk, particularly in contexts where governments, hostile organisations, or private actors monitor online activity closely.

Additionally, in contexts where physical protest is restricted, digital platforms enable alternative forms of mobilisation. ‘Digital Pride’ initiatives, including livestreamed events, online discussions, and virtual exhibitions, allow individuals to participate in collective expressions of identity and solidarity while reducing some forms of physical danger. Yet, mitigation is not elimination. Moving Pride online may reduce the risk of being attacked in a street, but it can increase the risk of being identified through an account, a comment, a livestream, a payment trail, or a device. Digital activism, therefore, does not remove danger from public life; it relocates and reshapes it.

These tensions reveal broader limitations in how digital space is often conceptualised within democratic discourse. Online platforms are frequently described through the language of the ‘public square’ and imagined as open civic arenas in which individuals gather, exchange ideas, and participate in public debate. For LGBTI+ communities, this analogy is especially appealing because digital platforms can provide access to forms of visibility and participation that may be inaccessible offline.⁴³

43 Klonick, "Beyond the Public Square."

The appeal of the public square framework lies partly in its democratic promise. The metaphor suggests openness, participation, accessibility, and the free exchange of ideas across social boundaries. For advocates of digital freedom, framing social media platforms as public squares helped legitimise the internet as a space for civic participation and political expression, particularly during periods in which online communication appeared to expand democratic engagement and reduce dependence on traditional institutional gatekeepers. The framework also aligned closely with libertarian models of internet governance that prioritised minimal intervention, openness, and the decentralisation of authority.⁴⁴

Importantly, however, the public square metaphor also performs political and institutional work. By framing platforms primarily as neutral hosts of public discourse rather than as active architects of information environments, responsibility for harm is often displaced away from platform design and governance structures and onto the users themselves.⁴⁵ Harassment, disinformation, coordinated abuse, and algorithmic amplification can, therefore, be interpreted as unfortunate byproducts of open participation rather than as outcomes shaped by commercial systems

44 John Perry Barlow, "A Declaration of the Independence of Cyberspace," Electronic Frontier Foundation, February 8, 1996, <https://www EFF.org/cyberspace-independence>

45 Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (New Haven, CT: Yale University Press, 2018).

46 Ibid.

47 Martin J. Riedl, "Rethinking Rights in Social Media Governance: Human Rights, Ideology and Inequality," *European Law Open* 3, no. 1 (2024), <https://doi.org/10.1017/elo.2024.3>.

48 Ibid.

optimised for engagement, visibility, and behavioural capture.

This framing has also enabled digital platforms to scale rapidly while avoiding many of the obligations historically associated with institutions that structure public communication. Traditional broadcasters, publishers, and public venues have generally operated within clearer systems of editorial accountability, legal oversight, or civic responsibility.⁴⁶ Social media companies, by contrast, were often able to position themselves simultaneously as neutral intermediaries when confronted with harmful content and as active curators when designing systems that maximise user engagement and advertising revenue.⁴⁷ This ambiguity allowed platforms to expand globally before implementing meaningful safeguards against harassment, targeted abuse, disinformation, or coordinated manipulation, particularly for marginalised communities disproportionately exposed to such harms.⁴⁸

For these reasons, scholars have increasingly proposed alternative frameworks for understanding digital platforms. Some argue that platforms function less like public squares and more like privately governed infrastructures or digital utilities upon which contemporary social,

political, and economic life increasingly depends.⁴⁹ Others describe them as surveillance-based ecosystems organised around data extraction and behavioural prediction, where participation itself becomes a source of commercial value.⁵⁰ Still others conceptualise platforms as algorithmically governed attention markets in which visibility is unevenly distributed according to engagement incentives rather than democratic principles.⁵¹

Each of these frameworks shifts attention away from the idealised image of open civic exchange and toward questions of power, ownership, governance, and infrastructure. For LGBTI+ communities in particular, these distinctions matter because they foreground how vulnerability is not produced solely through individual acts of prejudice but also through the design and political economy of the systems within which visibility occurs.

At the same time, the very features that made the public square metaphor politically attractive also limited its ability to fully account for how digital platforms actually operate. As social media became increasingly central to public life, it became harder to sustain the idea that these systems functioned merely as passive arenas for expression rather than as in-

frastructures that actively shape visibility, participation, and power.⁵²

Yet, the public square analogy obscures the extent to which digital environments are privately owned, commercially governed, and algorithmically structured systems. Unlike traditional civic spaces, social media platforms do not merely host discourse; they sort, amplify, suppress, monetise, and algorithmically distribute it. Visibility within these systems is, therefore, neither neutral nor evenly distributed but shaped by opaque governance structures, advertising incentives, moderation practices, and engagement-driven recommendation systems.⁵³

For LGBTI+ movements, this distinction is critical. Harassment, disinformation, forced outing, and coordinated abuse are not simply unfortunate behaviours occurring within otherwise neutral infrastructures. They are often intensified by the architecture of the platforms themselves. The same infrastructures that enable rights exercise also generate the conditions under which those rights can be undermined, making protection dependent not only on legal frameworks but also on platform design, governance, and political will.

Intersectional Vulnerability and Differential Exposure

The harms generated by anti-LGBTI+ digital ecosystems are not distributed evenly. Instead, exposure to harassment, surveillance, disinformation, and violence is structured through intersectional hierarchies of visibility and vulnerability. LGBTI+ individuals who are also racialised minorities, migrants, refugees, disabled persons, sex workers, or economically marginalised frequently face compounded forms of digital and offline harm.⁵⁴

Transgender individuals remain among the most heavily targeted groups within online hate ecosystems. Research from the Institute for Strategic Dialogue (ISD) demonstrates significant increases in anti-trans narratives across mainstream and fringe digital platforms over recent years, particularly surrounding healthcare access, education, and sports participation.⁵⁵ However, vulnerability intensifies further for trans women of colour, who are often subjected simultaneously to misogyny,

racism, transphobia, and hyper-visibility within online harassment campaigns.⁵⁶

Digital targeting frequently extends beyond speech alone. Doxxing campaigns expose personal information, immigration status, workplaces, or family members, thereby transforming online harassment into material vulnerability.⁵⁷ For undocumented or asylum-seeking LGBTI+ individuals, these risks can carry particularly severe consequences, including detention, deportation, or exposure to hostile state authorities.⁵⁸

In parts of the Middle East and North Africa, digital surveillance and entrapment have become major threats to LGBTI+ communities.⁵⁹ Human rights organisations have documented cases in which law enforcement officials used dating applications to identify, entrap, and arrest queer individuals.⁶⁰ The question is not only whether entrapment occurs, but how it disproportionately targets those who

49 Klonick, "Beyond the Public Square."

50 Zuboff, *The Age of Surveillance Capitalism*.

51 Gillespie, *Custodians of the Internet*.

52 Martin J. Riedl, "Rethinking Rights in Social Media Governance: Human Rights, Ideology and Inequality," *European Law Open* 3, no. 1 (2024), <https://doi.org/10.1017/elo.2024.3>

53 Gillespie, *Custodians of the Internet*.

54 Royal United Services Institute, *Online Gendered Narratives, LGBTIQ+ Targeting and the Far Right in Australia and the UK* (London: Royal United Services Institute, 2024), <https://www.rusi.org/explore-our-research/publications/external-publications/online-gendered-narratives-lgbtqi-targeting-and-far-right-australia-and-uk>.

55 Institute for Strategic Dialogue, "Offline Events, Online Threats: Analysing Online Extremist Targeting of LGBTQ+ Communities," <https://www.isdglobal.org/digital-dispatch/offline-events-online-threats-analyzing-online-extremist-targeting-of-lgbtq-communities/><https://www.isdglobal.org/digital-dispatch/offline-events-online-threats-analyzing-online-extremist-targeting-of-lgbtq-communities/>.

56 Fisher et al., "Social Media."

57 Royal United Services Institute, *Online Gendered Narratives*.

58 Human Rights Watch, *All This Terror Because of a Photo*.

59 Tahrir Institute for Middle East Policy, "Digital Persecution."

60 Human Rights Watch, *All This Terror Because of a Photo*.

are already most vulnerable – individuals without legal protections, without financial resources to relocate, and without political access to seek redress.⁶¹

Disabled LGBTI+ individuals also face distinct forms of exclusion within digital environments. Platform moderation systems frequently fail to account for accessibility needs, while reporting mechanisms often remain difficult to navigate for users with cognitive or communication disabilities. Simultaneously, disabled LGBTI+ users are often underrepresented within mainstream advocacy narratives, contributing to forms of structural invisibility even as they remain exposed to harassment.⁶²

The concept of ‘differential visibility’ is, therefore, critical. Some LGBTI+ individuals become hyper-visible symbolic targets within political and media discourse, while others remain systematically invisible within both policymaking and platform governance.⁶³ Visibility itself becomes unevenly distributed and politically consequential. To be visible may increase representation, but it may also dramatically increase exposure to harassment, surveillance, and violence.⁶⁴

The Digital Environment and the Expansion of Human Rights

The double-edged nature of digital space has fundamentally reshaped how scholars, policymakers, and human rights institutions understand the exercise of rights in contemporary society. The internet provides new tools and spaces for expression, participation, and association, allowing individuals to engage in civic life in ways that were previously impossible.⁶⁵ This shift has required a rethinking of rights, not as static protections tied to physical space but as dynamic entitlements that must extend into digitally mediated environments.

As previously outlined, digital technologies enable individuals to publish information, communicate with global audiences, and organise collective action at unprecedented speed and scale. This means that, through online platforms, individuals can exercise a variety of human rights, participate in peaceful assembly and association, and access information that influences education, employment, and wellbeing.⁶⁶ These affordances effectively lower the cost of participation in civic life, particularly for individuals and

groups who have historically been excluded from formal political institutions.

This means that through online platforms, individuals can exercise a variety of human rights such as freedom of expression, participate in peaceful assembly and association, and access information that influences education, employment, and wellbeing. This capacity has become especially significant at a time when many civil society organisations, including LGBTI+ groups, are operating under increasing financial pressure and shrinking funding environments. Digital platforms can reduce some of the material costs traditionally associated with organising, community-building, information distribution, and cross-border advocacy, allowing groups with limited resources to maintain visibility, coordination, and support networks that might otherwise be unsustainable.⁶⁷

Recognising these developments, international human rights bodies have increasingly affirmed that the same rights people enjoy offline must also be protected online. This principle reflects the understanding that digital environments are not separate from society but are now integral components of contemporary civic life. In

effect, this reframing rejects the notion of the internet as a ‘special case’ and, instead, treats it as a continuation of existing social and political spaces, governed by the same rights obligations.⁶⁸

An example is the Yogyakarta Principles plus 10, which elaborates the application of international human rights law in relation to sexual orientation and gender identity, explicitly recognising that discrimination can occur within technology-mediated environments.⁶⁹ Harassment, surveillance, and discrimination targeting LGBTI+ individuals online, therefore, constitute violations of existing human rights protections. This is significant because it formally situates digital harm within established human rights law rather than treating it as a separate or emerging category of concern.⁷⁰

However, while international human rights frameworks – such as the Yogyakarta Principles – offer several protections that are directly relevant to human rights within digital environments, their application in digital contexts often requires reinterpretation, as traditional legal categories were not designed to account for data-driven surveillance, algorithmic governance, or platform-mediated speech.⁷¹

61 Ibid.

62 Fisher et al., "Social Media."

63 Ioannis Kampourakis, "Nancy Fraser: Subaltern Counterpublics," Critical Legal Thinking, November 6, 2016, <https://critical-legalthinking.com/2016/11/06/nancy-fraser-subaltern-counterpublics/>.

64 Riedl, "Rethinking Rights in Social Media Governance."

65 OECD, Rights in the Digital Age.

66 Ibid

67 Castells, Networks of Outrage and Hope

68 OECD, Rights in the Digital Age.

69 United Nations Human Rights Experts. The Yogyakarta Principles Plus 10: Additional Principles and State Obligations on the Application of International Human Rights Law in Relation to Sexual Orientation, Gender Identity, Gender Expression and Sex Characteristics. Geneva, 2017. <https://yogyakartaprinciples.org/principles-en/yp10/>.

70 Adrian Shahbaz and Allie Funk, "A Human Rights Approach to Platform Content Regulation," Freedex, 2023, <https://freedex.org/a-human-rights-approach-to-platform-content-regulation/>.

71 Ibid.

However, despite growing recognition that human rights protections must extend into digital environments, meaningful implementation has remained uneven and incomplete. This gap exists partly because contemporary digital systems challenge many of the assumptions upon which existing human rights frameworks were originally constructed. Traditional human rights law was largely developed around territorially bounded states, identifiable public authorities, and comparatively visible forms of coercion. Digital power, by contrast, is often transnational, privatised, opaque, and infrastructural.⁷²

Conclusion

Digital technologies have become deeply embedded in contemporary social, economic, political, and cultural life. For LGBTI+ communities, online platforms offer vital opportunities for connection, identity exploration, and political mobilisation. They enable individuals to build communities across geographic boundaries, access information that may be unavailable offline, and challenge dominant narratives about sexuality and gender. Across the preceding analysis, digital space has emerged not as a supplementary sphere of activity but as a constitutive environment in which LGBTI+ life is increasingly formed, negotiated, and contested.

At the same time, these same platforms expose individuals to new forms of surveillance, harassment, and violence. Governments and private actors alike increasingly use digital tools to target LGBTI+

individuals, particularly in contexts where legal protections are weak or absent.

Understanding the role of digital spaces in LGBTI+ rights for the purposes of countering anti-rights actors, therefore, requires recognising this dual dynamic. Online environments function simultaneously as sites of empowerment and arenas of vulnerability. They will not become purely emancipatory because the mechanisms that make them useful also create risk, and they are not purely dangerous because abandoning them would mean abandoning some of the only available infrastructures of care, information, and political life. Protecting the digital rights of LGBTI+ individuals is not a peripheral concern but a central component of contemporary human rights protection in an increasingly digital world. This requires moving beyond viewing the internet as either inherently emancipatory or inherently dangerous and instead treating it as a contested civic space in which rights must be actively produced, defended, and maintained.

Anti-rights actors have already recognised the strategic importance of digital space and increasingly exploit its structures for political and ideological purposes. As this report will demonstrate, online platforms are not only sites in which LGBTI+ rights are defended, but also environments in which opposition to those rights is organised, amplified, and normalised. Far-right networks, religious extremists, authoritarian actors, and coordinated disinformation ecosystems have learned to use the speed, reach, and emotional dynamics of

digital communication to circulate hostile narratives, mobilise supporters, target activists, and reshape public discourse. In many cases, these actors have adapted more rapidly to the logic of algorithmically mediated communication than democratic institutions or human rights frameworks themselves. The result is that digital environments increasingly function not simply as neutral arenas of debate but as contested infrastructures in which visibility, participation, and legitimacy are continuously struggled over by actors with

profoundly unequal resources, strategies, and political objectives.

Ultimately, the digital environment does not simply reflect existing social conditions; it reorganises them. For LGBTI+ individuals, this means that questions of safety, identity, and political participation are now inseparable from the governance of digital space itself. The sections that follow examine why this is so, beginning with the definition of cyberspace and its characteristics.

Section 1 Key Findings:

- For many LGBTI+ people, digital access is foundational, not supplementary.
- Digital spaces can function as ‘networked refuge’ and support identity rehearsal. Individuals frequently come out online before doing so offline, using anonymity and selective disclosure to build confidence and test responses in lower-risk environments before facing potentially dangerous offline disclosure.
- Perceived anonymity is often more fragile than it appears.
- State and non-state digital entrapment is a documented, active practice.
- Digital platforms enabled a shift in whose knowledge counts as evidence.
- Vulnerability is intersectional and unevenly distributed.
- Many international human rights frameworks (e.g., the Yogyakarta Principles plus 10) now formally extend offline rights protections online. However, implementation remains uneven because traditional rights law assumes territorially bounded states and identifiable authorities, while digital power is transnational, privatised, and opaque.
- Anti-rights actors have adapted to digital platform logic faster than democratic institutions or rights frameworks have.

72 OECD, Rights in the Digital Age.

Section 1 Questions:

- Have you or your community relied on platform anonymity/pseudonymity features that later proved more fragile than expected (e.g., through geo-location, metadata, screenshots)? What happened, and what would have prevented it?
- How does your organisation or community currently weigh the benefits of online visibility (support, advocacy reach, solidarity) against the risks (surveillance, hyper-visibility, targeting)? What would help you make this calculation more safely?



II. THE ANATOMY OF CYBERSPACE

Section 1 established that digital space functions as both lifeline and site of exposure for LGBTI+ communities. Before examining who exploits that exposure and how, it is necessary to understand the terrain itself: what cyberspace actually is, how it is governed, and how it shapes what becomes visible, credible, and true.

This is because cyberspace is not a neutral backdrop to political life. It is a socio-technical environment – material, governed, and increasingly automated – in which visibility, vulnerability, and power are continuously produced. Understanding this architecture is a prerequisite for understanding and countering the anti-rights actors and drivers that operate within it.



Section 2: What Is Cyberspace?

Section 1 established that digital spaces function as lifeline infrastructure for LGB-TI+ communities, and that this same infrastructure generates exposure and risk. This section pauses before moving further into platforms, algorithms, and political mobilisation to ask a deceptively simple question: What is cyberspace?

This question matters because ‘cyberspace’ is not a neutral backdrop to political life but a framing device that shapes how power is understood, contested, and regulated: How it is conceptualised influences how harm is explained, how

responsibility is assigned, and how resistance is organised. For activists and civil society actors, the way digital environments are understood directly shapes strategy, such as whether harm is treated as individual behaviour or structural design, whether platforms are seen as tools or governance systems, and whether digital engagement is approached as outreach or as a contested political field in its own right.⁷³ Conceptualising digital spaces as neutral ‘platforms’ risks obscuring the infrastructures and incentives that structure visibility, safety, and political voice.⁷⁴ Recognising cyberspace, instead,

⁷³ Nancy Fraser, “Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy,” in *Justice Interruptus: Critical Reflections on the ‘Postsocialist’ Condition* (New York: Routledge, 1997), 69-98.

⁷⁴ Gillespie, *Custodians of the Internet*.

as a socio-technical environment makes visible the ways rights, risk, and recognition are actively produced within it.⁷⁵ Understanding what cyberspace is, therefore, becomes a prerequisite for understanding how power operates within it.

The Mirage of the Digital

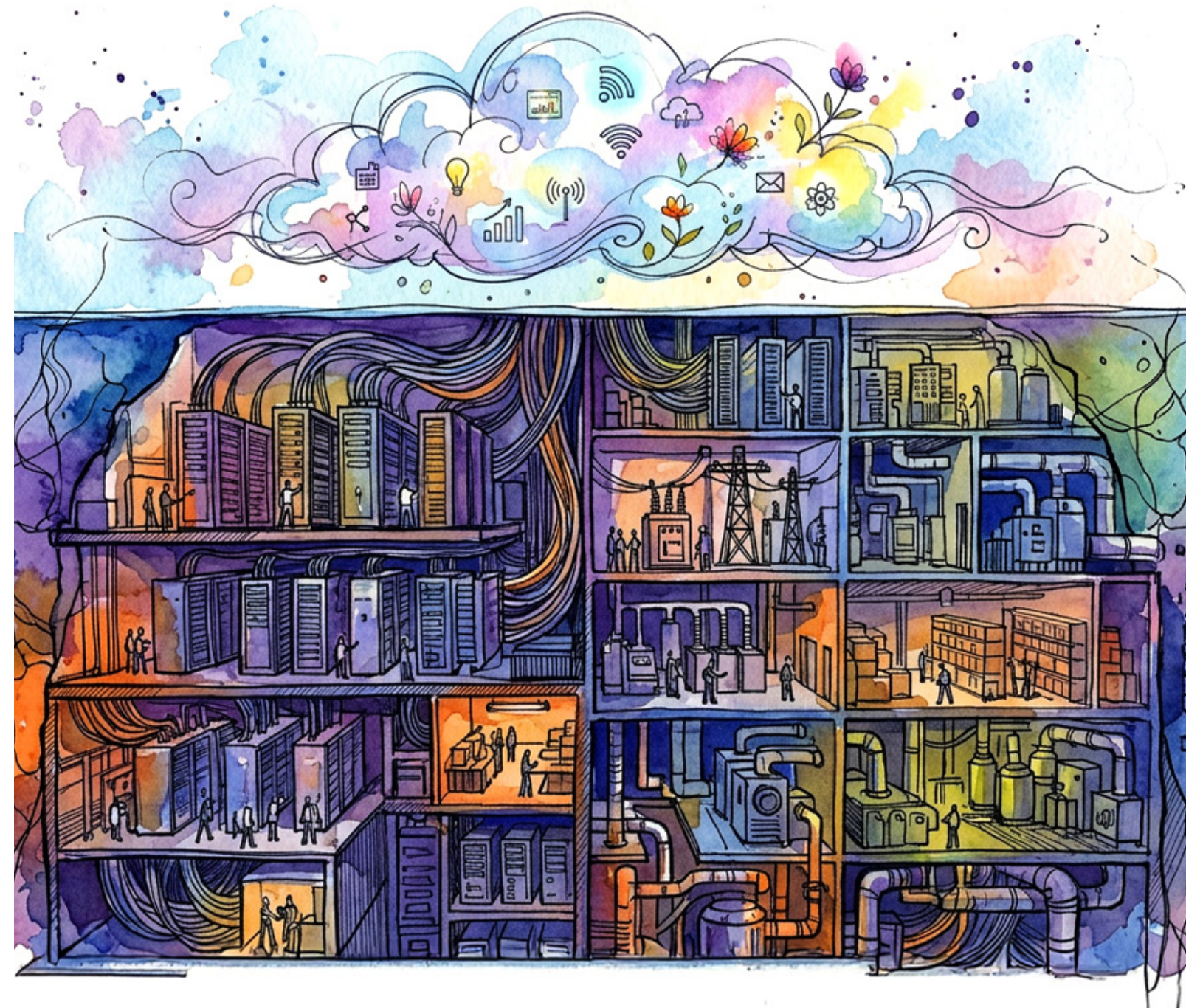
Contemporary digital life is surrounded by language that describes reality while also obscuring it. We speak of 'the cloud' as though data floats somewhere above us, weightless and immaterial, when in practice it refers to vast networks of servers, cables, warehouses, energy systems, cooling infrastructures, and labour. We talk about 'switching off' while carrying mini-computers in our pockets, devices that continue to collect signals, store histories, and mediate access to work, relationships, services, news, health information, and political life.⁷⁶ We describe online platforms as 'spaces,' 'feeds,' 'networks,' or 'communities,' as though they are purely natural human driven environments rather than designed systems governed by commercial incentives, technical architecture, and – at times – political choices.⁷⁷

Behind these systems sits a vast – and often hidden – layer of human labour: content moderators and data workers who are tasked with reviewing flagged material, including violent, sexual, and otherwise distressing imagery, in order to maintain the appearance of 'safe' online environments.⁷⁸ This work is frequently outsourced, underpaid, and psychologically harmful, carried out in precarious conditions that are structurally separated from the users whose experiences they are tasked with sanitising.⁷⁹ The relative smoothness of digital life is, therefore, produced through uneven exposure to harm, where the burden of filtering and absorbing the worst material is displaced onto workers who rarely have visibility, protection, or recourse.⁸⁰

This language matters because it shapes expectation. If the cloud sounds weightless, then its material costs disappear. If the phone is merely a personal device, then its role as attention-monopolising infrastructure is minimised.⁸¹ If social media is imagined as a public square, then the private systems governing that square become harder to see and keep accountable.⁸² These metaphors are useful because they allow people to navigate

complex systems, but they also risk creating a mirage, making cyberspace feel frictionless, placeless, and separate from the physical world we inhabit.⁸³

Seeing through this mirage is the starting point for understanding cyberspace. It is not a place outside reality, even when it has been made to feel like one.⁸⁴ It



75 OECD, *Rights in the Digital Age*.

76 Sherry Turkle, *Alone Together: Why We Expect More from Technology and Less from Each Other* (New York: Basic Books, 2011).

77 Gillespie, *Custodians of the Internet*.

78 Ibid.

79 Gillespie, *Custodians of the Internet*.

80 Ibid.

81 Turkle, *Alone Together*.

82 Klonick, "Beyond the Public Square."

83 Hayward, Philip. "Situating Cyberspace: The Popularisation of Virtual Reality." In *Future Visions: New Technologies of the Screen*, edited by Philip Hayward and Tana Wollen, 180 to 204. London: British Film Institute, 1993.

84 Ibid.

is a material, political, and psychological environment that reorganises reality while presenting itself as something more abstract.⁸⁵ The question, then, is not only what cyberspace is, but where this imaginary came from, and why its earliest formulations already contained both promise and danger.

Gibson's Cyberspace: Expansion and Warning

The concept of cyberspace emerged as much from imagination as from technology. In the science fiction classic *Neuromancer*, William Gibson, who coined the term, described it as a "consensual hallucination" – an immersive landscape of data through which users could move freely.⁸⁶ This early vision positioned cyberspace as an expansion of life and mind; a new terrain in which consciousness could move through information, identity could be reconfigured, and the limits of physical embodiment could appear loosened.⁸⁷

But Gibson's vision was never utopian. The phrase "consensual hallucination" already carries suspicion within it.⁸⁸ It suggests shared experience but also illusion; expanded perception but also

the possibility of being captured by the system that produces that perception.⁸⁹ In other words, cyberspace is not simply described as a neutral environment that users enter but as a collectively experienced illusion that feels real precisely because it is shared.⁹⁰ This matters because it introduces a tension at the heart of the concept: What feels like freedom of movement within information may also be a form of structured perception, shaped by systems that remain largely invisible to the user.⁹¹

Gibson imagined cyberspace before the contemporary internet existed; yet, many of the risks later associated with digital life were already present in that imaginary.⁹² These include the seduction of disembodiment, meaning the sense that identity can detach from physical consequence and material grounding; the risk of corruption, understood as the manipulation or distortion of perception through mediated systems; the vulnerability of the mind to systems it cannot fully understand, where complexity itself becomes a form of power because it limits who can see how influence is operating; and finally, the possibility that virtual structures might overtake, distort, or govern physical life.⁹³ Taken together, these elements sug-

gest that cyberspace was never imagined as a neutral extension of human communication but as an environment in which perception, identity, and control could become deeply entangled.⁹⁴

State and corporate control are also not merely later symptoms of cyberspace. They are part of its imagined architecture from the beginning. *Neuromancer* depicts a world in which information systems, artificial intelligences, corporations, surveillance, and power are intertwined.⁹⁵ Cyberspace is not separate from governance; it is a field through which governance can operate.⁹⁶ The danger is not that a once-free digital world was imagined as a public good and then later contaminated by state control or corporate extraction. Rather, the dream of cyberspace always contained the possibility that expanded connection would also expand the means of control on individuals and communities.⁹⁷

This framing reveals the central paradox that cyberspace is both expansion and enclosure. It appears to offer escape

from physical limits, but it is constituted through physical infrastructure, corporate ownership, and political regulation.⁹⁸ It appears immaterial, but it depends on material systems.⁹⁹ It appears liberatory, but it produces new forms of legibility.¹⁰⁰ In other words, cyberspace was constructed more than it was discovered, and what appears as an immersive digital world has always been grounded in material systems that decide who can enter, what can be seen, what can be remembered, and what can be extracted.¹⁰¹

The Collapse of the Digital-Physical Divide

From this perspective, the distinction between online and offline life becomes increasingly untenable. Digital environments are not separate from social reality but continuous with it; actions in digital spaces have tangible consequences in the material world.¹⁰² This means that, rather than existing as a parallel domain, cyberspace is embedded within everyday social and political experiences.¹⁰³

85 Ibid.
86 William Gibson, *Neuromancer* (New York: Ace Books, 1984).
87 Ibid.
88 Ibid.
89 Hayward, "Situating Cyberspace."
90 Ibid.
91 Ibid.
92 Ibid.
93 Hayward, "Situating Cyberspace."

94 Ibid.
95 Gibson, *Neuromancer*.
96 Barlow, "Declaration of the Independence of Cyberspace."
97 Hayward, "Situating Cyberspace."
98 Barlow, "Declaration of the Independence of Cyberspace."
99 Hayward, "Situating Cyberspace."
100 Ibid.
101 Barlow, "Declaration of the Independence of Cyberspace."
102 Nathan Jurgenson, "Between Reality and Cyberspace," *Cyborgology*, February 8, 2012, <https://thesocietypages.org/cyborgology/2012/02/08/between-reality-and-cyberspace/>
103 boyd, *It's Complicated*.

For LGBTI+ individuals, this entanglement is particularly pronounced. As previously outlined, online visibility can facilitate connection and solidarity, but it can also lead to harassment, discrimination, or violence.¹⁰⁴ Outing, doxxing, and coordinated attacks demonstrate how digital harm translates directly into lived experience.¹⁰⁵ Cyberspace, in this sense, is not ‘elsewhere’ but it is fully integrated into lived reality.¹⁰⁶

From this entanglement emerges the digital self – an identity formed through interaction with digital systems.¹⁰⁷ Cyberspace enables new forms of embodiment, allowing people to project themselves beyond the constraints of their physical bodies.¹⁰⁸ Digital embodiment refers to the ways in which individuals inhabit, perform, and experience identity through avatars, social media profiles, and virtual interactions.¹⁰⁹ It is not merely a replication of offline identity; it is a co-construction of self, mediated by digital systems, emotional expression, and social feedback.¹¹⁰

For queer individuals, this digital self can be liberatory. Anonymous accounts, carefully curated personas, or avatars allow experimentation with gender, sexuality, and social presentation.¹¹¹ Platforms like Second Life, Discord, or metaverse environments provide spaces where embodiment is fluid, aesthetic choices are performative, and community interactions reinforce identity.¹¹² Emotional embodiment – feeling authentic emotions through an avatar – enhances this sense of presence and selfhood online.¹¹³ Social embodiment also enables participation, recognition, and relational negotiation within digital networks.¹¹⁴

The condition of the digital self reflects what Donna Haraway describes through the figure of the cyborg – a hybrid form of being in which identity is shaped through the entanglement of human and technological systems.¹¹⁵ This complicates common critiques of technology that treat it either as a fog that obscures an otherwise clear human world, or as an external force

that pulls the strings of human action from outside. Both framings assume technology is separate from human life rather than already embedded within the infrastructures through which perception, agency, and identity are produced.¹¹⁶

Haraway’s cyborg intervenes precisely in this tension. It rejects the idea of technology as either distortion or domination from the outside and instead reframes it as something already entangled with the conditions of human life itself.¹¹⁷ The cyborg is not only a futuristic machine-human fantasy, nor is it a sudden break from ordinary life. Human beings have long lived through technological mediation.¹¹⁸ A person using glasses to see more clearly is already inhabiting a technologically assisted body. The challenge is not, therefore, that technology has newly invaded an otherwise pure human condition, but that human life has always been shaped through tools, prosthetics, systems, and infrastructures.¹¹⁹

This point matters because it prevents cyberspace from being treated as a unique-

ly unnatural distortion of identity. The digitally mediated self is not a complete departure from embodied life; it is an intensification of a much older condition.¹²⁰ People have always used technologies to extend perception, memory, mobility, expression, and social relation.¹²¹ What changes in cyberspace is the scale, speed, and institutional control of those mediations.¹²² In cyberspace, the self is not simply represented or developed by the individual but actively produced through platforms and interfaces.¹²³

The cyborg metaphor also helps reconnect this theoretical discussion to lived realities. LGBTI+ people are often forced to negotiate identity through partial visibility, coded expression, strategic disclosure, and shifting forms of embodiment.¹²⁴ The cyborg metaphor also underscores the instability and multiplicity of identity. Digital selves are distributed across platforms, dynamic, and collaborative.¹²⁵ Users project aspects of themselves differently depending on context, audience, or purpose.¹²⁶ While this multiplicity is generative, it also introduces vulnerabilities –

104 Human Rights Watch, All This Terror Because of a Photo.

105 Tahrir Institute for Middle East Policy, "Digital Persecution."

106 boyd, It's Complicated.

107 Ibid.

108 Ratan, Rabindra A., Jorge Peña, Benjamin J. Li, and Sun Joo Park. "Digital Embodiment and Virtual Identity." *Technology, Mind, and Behavior* 2, no. 1 (2021). <https://doi.org/10.1037/tmb0000036>

109 Auriemma, Battista, and Quarta, "Digital Embodiment."

110 boyd, It's Complicated.

111 Ibid.

112 Auriemma, Battista, and Quarta, "Digital Embodiment."

113 Ibid.

114 Ibid.

115 Donna Haraway, "A Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s," *Socialist Review* 15, no. 2 (1985): 65-108.

116 Ibid.

117 Haraway, "Manifesto for Cyborgs."

118 Ibid.

119 Ibid.

120 Ibid.

121 Ibid.

122 Castells, *Networks of Outrage and Hope*

123 boyd, It's Complicated.

124 Ibid.

125 Ibid.

126 Ibid.

fragmented identities can be exploited by surveillance systems, algorithmic profiling, or coordinated harassment.¹²⁷

The digital self is, therefore, cyborg-like not because it is artificial but because it is assembled across body, language, platform, community, risk, and desire.¹²⁸ It is a ‘self’ made through mediation. That mediation can be liberatory when it allows experimentation, recognition, and coalition; it can also be dangerous when the same mediated traces become searchable, classifiable, and punishable.¹²⁹

Disinhibition and Anonymity

If the digital self is produced through entanglement with technological systems, then interaction within cyberspace is also shaped by those same mediating conditions. In parallel, cyberspace is structured through conditions of distance, pseudonymity, anonymity, and reduced immediate social accountability.¹³⁰ These conditions change how people speak and act.¹³¹ They can make expression easier, especially for those who are unsafe or constrained offline. They can allow queer

users to ask questions, test names, enter communities, seek advice, or participate politically without immediately exposing themselves to family members, employers, police, or hostile neighbours.¹³²

But the same conditions also enable cruelty, escalation, and coordinated harm. Reduced social accountability can facilitate harassment, abuse, threats, blackmail, and digital mobbing.¹³³ The online disinhibition effect, thus, operates as a structural feature of mediated communication, producing both emancipatory and coercive outcomes.¹³⁴ In the same way in which anonymity can protect vulnerable users, it can also shield perpetrators. In the same way that pseudonymity can allow identity exploration, it can allow infiltration, impersonation, and entrapment.¹³⁵

For LGBTI+ users, disinhibition translates into real harm when hostile actors exploit anonymity to target vulnerable individuals through harassment campaigns, image-based abuse, outing threats, and coordinated misinformation.¹³⁶ Online violence is not a psychological abstraction: It has direct social, emotional, and material

consequences that reinforce the entanglement of digital and physical worlds.¹³⁷

At the same time, anonymity cannot simply be rejected, because it remains one of the most important tools of survival for people whose offline environments are dangerous.¹³⁸ This is the recurring problem of cyberspace: The same feature is rarely only protective or only harmful. It becomes one or the other depending on power, context, and governance.¹³⁹

Data Colonialism

Cyberspace must also be understood as an extractive regime, in which data colonialism emerges as a systemic risk embedded within its very architecture.¹⁴⁰ Rather than an external distortion of otherwise neutral digital systems, it describes a structural tendency within cyberspace toward the continuous capture, classification, and monetisation of human behaviour, relationships, and identity.¹⁴¹ Social media platforms, apps, and smart

technologies collect data on users, often under conditions that formally register consent while materially limiting the possibility of refusal.¹⁴²

This notion of ‘meaningful consent’ is, therefore, central but deeply contested. In practice, consent is frequently produced through terms and conditions that are legally comprehensive yet rarely read due to their length, complexity, and the asymmetry between users and platforms.¹⁴³ Even when users do engage with these documents, the choice is often binary: accept or lose access.¹⁴⁴ Over time, this initial ‘choice’ becomes further weakened as individuals become embedded within platforms for communication, work, social life, and political engagement.¹⁴⁵ Once a service is integrated into everyday infrastructure, withdrawal is no longer a simple act of opting out but may entail social, professional, or informational exclusion.¹⁴⁶ Moreover, platforms retain the ability to modify conditions unilaterally, meaning that what was initially consented to may

127 Human Rights Watch, All This Terror Because of a Photo.

128 Haraway, "Manifesto for Cyborgs."

129 Human Rights Watch, All This Terror Because of a Photo.

130 John Suler. "The Online Disinhibition Effect." CyberPsychology & Behavior 7, no. 3 (2004): 321-326.

131 Ibid.

132 Human Rights Watch, All This Terror Because of a Photo.

133 Suler, "Online Disinhibition Effect."

134 Ibid.

135 Human Rights Watch, All This Terror Because of a Photo.

136 Ibid.

137 Ibid.

138 OECD, Rights in the Digital Age.

139 International Commission of Jurists, Yogyakarta Principles plus 10: Additional Principles and State Obligations on the Application of International Human Rights Law in Relation to Sexual Orientation, Gender Identity, Gender Expression and Sex Characteristics (Geneva: International Commission of Jurists, 2017).

140 Nick Couldry and Ulises A. Mejias, The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism (Stanford, CA: Stanford University Press, 2019).

141 OECD, Rights in the Digital Age.

142 Ibid.

143 Ibid.

144 Ibid.

145 Turkle, Alone Together.

146 Ibid.

later shift in scope without meaningful renegotiation.¹⁴⁷

Within this context, ‘data colonialism’ refers to more than the collection of personal information. It describes a broader political economy in which human experience is rendered continuously legible, extractable, and tradable through digital infrastructures.¹⁴⁸ The term draws an analogy with historical colonial logics not to suggest equivalence in form but to highlight continuity in structure: the appropriation of resources from populations who have limited control over the systems that extract from them and limited ability to define how those resources are used.¹⁴⁹ In digital contexts, the ‘resource’ is not land but behavioural, social, and affective life, which is transformed into data that can be analysed, predicted, and monetised. Importantly, this framing should not be reduced to a rhetorical label or used indiscriminately as a critique-catch-all; it becomes analytically meaningful only when it is used to describe specific infrastructures of extraction, asymmetries of control, and how these systems can be mobilised by both corporate and state actors in ways that may undermine rights, autonomy, and collective agency.¹⁵⁰

This extractive logic is closely connected to what Shoshana Zuboff describes as “surveillance capitalism,” a model in which behavioural data generated through everyday activity becomes raw material for commercial prediction and influence.¹⁵¹ Under this model, digital platforms do not merely facilitate communication; they are structured around the continuous acquisition of information about users, the generation of behavioural insights, and the creation of systems capable of shaping future conduct.¹⁵² Although framed as personalised services or user optimisation, these processes often redistribute power toward platform operators while reducing transparency for those being observed.¹⁵³

Human rights scholars increasingly argue that data extraction is not solely a matter of privacy but also of power, participation, equality, and democratic accountability.¹⁵⁴

For LGBTI+ communities, these dynamics can carry particular risks. Data that appears mundane in one context may become highly sensitive in another. Patterns of online behaviour, social networks, geolocation data, search histories, and inferred characteristics can reveal information about sexual orientation, gender

identity, political affiliation, relationships, or community membership even when individuals have not explicitly disclosed such information.¹⁵⁵ In contexts where legal protections are weak – or hostility toward LGBTI+ people is widespread – these forms of visibility can become tools of surveillance, discrimination, exclusion, or persecution.¹⁵⁶

The significance of data colonialism, therefore, lies not simply in the quantity of information collected but in the unequal distribution of power over how that information is interpreted, circulated, and acted upon.¹⁵⁷ Cyberspace is not merely a space where data exists. It is a system that continuously transforms human experience into resources that can be governed, traded, and deployed by actors whose interests may not align with those of the individuals from whom that data originates.¹⁵⁸

Cyberspace, Power, and Anti-Rights Mobilisation

Taken together, the preceding sections frame cyberspace not as a neutral communicative environment but as a layered socio-technical system in which identity,

interaction, and information are continuously shaped by infrastructure, governance, and extraction.¹⁵⁹ The mirage of immateriality obscures the material systems that sustain digital life; the concept of cyberspace as a consensual hallucination reveals both expansion and capture; the collapse of the digital-physical divide demonstrates the continuity of harm across domains; the cyborg figure reframes identity as mediated and distributed rather than autonomous; online disinhibition shows how interaction is structurally conditioned by uneven accountability; and data colonialism exposes the extractive political economy underpinning these systems.¹⁶⁰

In the context of LGBTI+ civil society, these dynamics do not operate in isolation but become strategically significant within contemporary anti-rights mobilisation.¹⁶¹ Digital infrastructures are not merely spaces where opposition to LGBTI+ rights is expressed; they are environments through which such opposition is organised, amplified, and operationalised.¹⁶² Visibility regimes shape which narratives gain traction, algorithmic systems influence the circulation of moral panic and misinformation, and anonymity can be leveraged both for strategic co-

147 OECD, Rights in the Digital Age.

148 Ibid.

149 Couldry and Mejias, The Costs of Connection.

150 Ibid.

151 Zuboff, The Age of Surveillance Capitalism.

152 Ibid.

153 OECD, Rights in the Digital Age.

154 Land, Molly K. "Toward a Human Rights Approach to Platform Governance." Harvard International Law Journal 62, no. 1 (2021): 39 to 65.

155 Human Rights Watch, All This Terror Because of a Photo.

156 Ibid.

157 Ibid.

158 Couldry and Mejias, The Costs of Connection.

159 Gillespie, Custodians of the Internet.

160 Haraway, "Manifesto for Cyborgs."

161 European Digital Rights (EDRi), "The Digital Rights of LGBTQ+ People: How Technology Reinforces Societal Oppressions," European Digital Rights, 2024, <https://edri.org/our-work/the-digital-rights-lgbtq-technology-reinforces-societal-oppressions/>

162 Institute for Strategic Dialogue, "Offline Events, Online Threats."

ordination and for plausible deniability in harassment campaigns.¹⁶³ At the same time, the same infrastructures that facilitate harm also provide essential tools for community-building, advocacy, documentation of abuse, and transnational solidarity.¹⁶⁴

Anti-rights actors, therefore, operate within the same conditions of cyberspace outlined in this section, often exploiting its structural features rather than acting outside them.¹⁶⁵ The ambiguity of identity online enables both grassroots mobilisation and covert coordination; disinhibition enables both political speech and targeted intimidation; and data extraction enables both audience targeting and the profiling of vulnerable groups.¹⁶⁶ Cyberspace does not determine anti-rights mobilisation, but it shapes the terrain on which it becomes effective, scalable, and persistent.¹⁶⁷

Understanding these dynamics is, therefore, not an abstract exercise in digital

theory but a necessary condition for analysing how contemporary struggles over LGBTI+ rights are organised, contested, and sustained in practice.¹⁶⁸

Conclusion

Cyberspace is, therefore, neither neutral nor escapist. It is a contested socio-technical terrain, simultaneously emancipatory and exploitative. This section has described several of its broad pillars including: the mirage of immateriality, the early imaginary of cyberspace as both expansion and warning, the collapse of the digital-physical divide, the cyborg-like mediation of identity, the extractive regime of data colonialism, and the ambivalent effects of anonymity and disinhibition. Together, these pillars form the skeleton of cyberspace. They show that digital life is not a separate realm but an infrastructure through which bodies, identities, information, labour, and power are reorganised.

¹⁶³ Alice Marwick and Rebecca Lewis, *Media Manipulation and Disinformation Online* (New York: Data & Society Research Institute, 2017), https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf

¹⁶⁴ Eli Pariser, *The Filter Bubble: What the Internet Is Hiding from You* (New York: Penguin Press, 2011).

¹⁶⁵ EDRI, "Digital Rights of LGBTQ+ People."

¹⁶⁶ Suler, "Online Disinhibition Effect."

¹⁶⁷ Marwick and Lewis, *Media Manipulation and Disinformation Online*.

¹⁶⁸ Judd Birdsall, "A Human Rights-Based Approach to Social Media Platforms," Berkley Center for Religion, Peace & World Affairs, Georgetown University, 2023.

Section 2 Key Findings:

- Gibson's founding concept of cyberspace – 'consensual hallucination' – already contained a warning, not just a promise. From its origin, the concept combined expansion of identity and perception with the risk of capture by systems the user cannot fully see.
- Digital actions have direct material consequences (outing, doxxing, coordinated attacks translate directly into lived harm), meaning cyberspace cannot be treated as a separate, escapable domain for LGBTI+ people.
- The same conditions that let LGBTI+ people safely explore identity, seek advice, or organise without exposure to family, employers, or police also enable harassment, blackmail, impersonation, and coordinated harm.
- Data that appears mundane can become dangerous depending on context. Search histories, geolocation, social networks, and behavioural patterns can reveal sexual orientation or gender identity even without explicit disclosure.
- Cyberspace doesn't determine anti-rights mobilisation, but it shapes the terrain that makes it effective and scalable. Anti-rights actors exploit the exact same structural features of cyberspace that enable LGBTI+ community-building.
 - Identity ambiguity enables both grassroots mobilisation and covert coordination.
 - Disinhibition enables both safe political speech and targeted intimidation.
 - Data extraction enables both audience-building and profiling of vulnerable groups.

Section 2 Questions:

- Has your organisation or community experienced harm connected to data collection you did not meaningfully consent to (e.g., inferred identity data, platform terms changed after the fact)? What would 'meaningful consent' look like in the digital environment your community depends on for survival?
- How does your community currently balance the protective value of anonymity/pseudonymity against the risk that the same anonymity is used by hostile actors to harass or infiltrate your spaces? What governance changes (platform-level or legal) would improve this balance without eliminating protective anonymity?
- Has strategic management of your online identity (secondary accounts, partial disclosure, coded language) changed your ability to build community, seek employment, or engage politically?
- Would framing digital harm using concepts like 'data colonialism' or 'extractive infrastructure' help or hinder your advocacy with policymakers, funders, or the public in your context? What language has proven most persuasive for you so far?



Section 3: Online Spaces as Modern Civic Arenas

Section 2 established the skeleton of cyberspace: its mirages of immateriality, its material infrastructures, its mediated identities, its – at times – extractive data regimes, and its ambivalent conditions of anonymity and disinhibition. This section moves from the anatomy of cyberspace to the specific spaces built upon that skeleton, where individual users interact with one another most frequently, namely social media platforms. Furthermore, it asks how digital spaces have become sites of contestation between human rights actors and anti-rights movements, including who is able to govern visibility, legitimacy, and participation within them and on what terms.

As noted above, because online platforms have become central infrastructures of contemporary civic life – fundamentally reshaping how individuals communicate,

organise, protest, and construct collective identity – digital platforms do not operate as neutral tools of communication but rather function as privately governed civic arenas in which visibility, participation, legitimacy, and exclusion are continuously negotiated through technical systems, commercial incentives, and political pressures.

For LGBTI+ activists, understanding these dynamics is not simply a matter of understanding communication technologies. Increasingly, struggles over rights, recognition, and social legitimacy are fought within digital environments where visibility can generate both protection and risk. The same platforms that enable community building, advocacy, and transnational solidarity also provide opportunities for anti-rights actors to organise, circulate disinformation, coordinate harassment,

and challenge the legitimacy of queer identities and movements. Understanding how online civic spaces function, therefore, becomes essential to understanding the contemporary conditions under which both human rights advocacy and anti-rights mobilisation take place.

Digital Publics and the Transformation of Civic Space

The emergence of social media platforms has transformed the structure of public life through the creation of what danah boyd conceptualises as "networked publics."¹⁶⁹ Broadly speaking, a public is a social space in which individuals encounter one another, exchange information, form opinions, and participate in collective discussions about matters of shared concern.¹⁷⁰ Publics are, therefore, not simply groups of people but arenas in which meaning, identity, and political legitimacy are negotiated.¹⁷¹ These are publics are simultaneously constructed through networked technologies and the imagined audiences that emerge around them.¹⁷² Unlike traditional public spaces, which were geographically bounded and institutionally mediated, networked publics are persistent, searchable, spreadable, and

highly visible environments in which communication can circulate rapidly across time and space.¹⁷³

These four properties fundamentally reshape civic interaction. 'Persistence' means that communication remains accessible long after its original context has disappeared. 'Visibility' expands the potential audience for interaction far beyond immediate social circles. 'Spreadability' enables rapid circulation of content through sharing practices and algorithmic amplification. 'Searchability' allows information to be rediscovered and recirculated indefinitely. Together, these features alter not only how information moves but how individuals understand themselves in relation to others.¹⁷⁴

At the same time, networked publics have altered the relationship between visibility and privacy. Research suggests that users often understand privacy socially rather than technically. Rather than seeking complete invisibility, individuals attempt to manage context and audience interpretation.¹⁷⁵ Boyd describes this practice as "social steganography," in which users embed meaning within coded language, humour, insider references, or culturally specific communication practices that re-

169 boyd, *It's Complicated*.

170 Ibid.

171 Fraser, "Rethinking the Public Sphere."

172 boyd, *It's Complicated*.

173 Philip Hayward, "Situating Cyberspace: The Popularisation of Virtual Reality," in *Future Visions: New Technologies of the Screen*, ed. Philip Hayward and Tana Wollen (London: British Film Institute, 1993), 180-204.

174 boyd, *It's Complicated*.

175 Ibid.

main legible to intended audiences while obscuring meaning from outsiders.¹⁷⁶

Many young LGBTI+ individuals navigate digital environments while managing multiple audiences simultaneously – including peers, family members, schools, employers, and hostile actors.¹⁷⁷ Secondary

accounts, coded language, ‘close friends’ lists, and fandom-based communication frequently function as mechanisms for exploring identity while limiting exposure.¹⁷⁸ Privacy in digital civic space, therefore, becomes less about secrecy than about negotiating visibility under conditions of surveillance and potential vulnerability.

Networked Movements and a New Cycle of Mobilisation

Contemporary social movements increasingly emerge through these digitally networked systems of communication, which have fundamentally altered how collective action forms and spreads.¹⁷⁹ Manuel Castells argues that contemporary movements are driven not solely by rational political calculation but also by emotional processes that move from outrage to hope and – ultimately – to collective action.¹⁸⁰ While often applied to democratic and social justice movements, this framework is equally useful for understanding contemporary anti-rights mobilisation.¹⁸¹ Digital networks enable diverse actors to transform emotional responses into political action, whether those emotions emerge from experiences of exclusion and discrimination or from perceived threats to social, cultural, religious, or national norms.¹⁸²

According to Castells, outrage emerges through experiences of inequality, injustice, corruption, or institutional failure. Hope develops when individuals recognise that these grievances are shared collectively rather than experienced privately. Digital networks facilitate this recognition by enabling individuals to witness others articulating similar frustrations across dispersed geographic spaces.

Collective action then becomes possible once shared emotion is transformed into visible mobilisation.¹⁸³

For example, a young queer person living in a community where discrimination is common may initially experience exclusion as an individual problem. Through social media, they encounter others describing similar experiences, share personal testimonies, and participate in online discussions about inequality and rights. The recognition that these experiences are collectively shared can transform private frustration into a sense of solidarity and possibility. Outrage at discrimination becomes hope that change is achievable, thus creating pathways into advocacy, community organising, campaigning, or participation in broader social movements.¹⁸⁴

A similar dynamic can operate among anti-rights actors. An individual who believes that social change threatens established cultural, religious, or national norms may encounter online communities articulating similar concerns. Through repeated exposure to narratives that frame these developments as evidence of decline, injustice, or loss, private anxieties become collective grievances. These shared emotions can then be channelled into visible political activity, including online campaigns,

179 Castells, Networks of Outrage and Hope

180 Ibid.

181 Ibid.

182 Ibid.

183 Ibid.

184 boyd, It's Complicated.

176 Ibid.

177 Ratan et al., "Digital Embodiment and Virtual Identity."

178 Ibid.



protest participation, lobbying efforts, or organised opposition to specific policies and social movements.¹⁸⁵

Digital communication platforms play a central role in this process because they enable individuals to bypass traditional gatekeepers such as public institutions, newspapers, or broadcast media. Communication itself becomes a form of counter-power through which alternative narratives can circulate outside institutional control. Movements, therefore, emerge not only through formal organisation but through the rapid circulation of images, hashtags, testimonies, and emotionally resonant content.¹⁸⁶

This dynamic was particularly visible during movements such as the Arab Spring, Occupy Wall Street, Black Lives Matter, and #MeToo. In each case, social media enabled rapid diffusion of outrage and facilitated mass participation through decentralised communication structures.¹⁸⁷ Black Lives Matter, for example, expanded globally through the circulation of videos documenting police violence alongside hashtags such as #BlackLivesMatter and #SayHerName.¹⁸⁸ Similarly,

#MeToo enabled millions of individuals to publicly articulate experiences of sexual violence, transforming isolated experiences into visible collective testimony.¹⁸⁹

Yet, these examples also raise a more difficult question: Where are these movements now, and where are the issues they sought to address? This is not a criticism of the movements themselves, nor of the courage, creativity, or political work they made possible. The Arab Spring, Occupy Wall Street, Black Lives Matter, and #MeToo each transformed public language, exposed forms of power that had long been denied, and created solidarities that could not easily be undone.¹⁹⁰ At the same time, their trajectories reveal the limits of visibility when it is not accompanied by sustained organisation, legislative attention, resource allocation, and institutional follow-through.¹⁹¹

Occupy Wall Street made economic inequality newly visible within mainstream political discourse, yet many of the conditions it protested – precarity, wealth concentration, housing insecurity, and economic instability – have deepened or returned with renewed force.¹⁹²

#MeToo transformed public recognition of sexual harassment and abuse, yet its aftermath has coincided with the growth of the manosphere and renewed backlash against feminist and gender-equality politics.¹⁹³ The Arab Spring demonstrated the extraordinary mobilising capacity of networked publics, but in several contexts the collapse of authoritarian legitimacy was not followed by durable democratic consolidation.¹⁹⁴ These outcomes do not negate the movements' achievements. They show that digital mobilisation has altered the life span of social movements away from a linear path from struggle to progress into a cycle that is reinitiated and refocused not through success but through emotional and attention burn-out.¹⁹⁵

We can begin to understand the limitation of this new cycle when we consider campaigns such as #LoveWins. Following the legalisation of same-sex marriage in the United States, it demonstrated how social media could facilitate transnational celebrations of queer visibility and solidarity.¹⁹⁶ Online campaigns opposing anti-LGBTI+ legislation in Uganda, Ghana, Hungary, and Russia similarly illustrate

how digital networks enable activists to rapidly circulate information, coordinate advocacy, and generate international attention.¹⁹⁷ Following the death of non-binary student Nex Benedict in the United States in 2024, social media platforms became central spaces for mourning, political mobilisation, and trans solidarity, once again demonstrating how networked communication can rapidly transform localised events into broader political symbols.¹⁹⁸

The mobilising capacity of networked communication is not limited to rights-based movements. Myanmar provides a stark illustration of how the same digital infrastructures can facilitate exclusionary and violent forms of collective mobilisation. Prior to and during the military's campaign against the Rohingya population, Facebook became a central channel through which anti-Rohingya narratives, hate speech, rumours, and dehumanising content circulated at scale.¹⁹⁹ These messages helped transform existing social tensions into broader collective hostility, demonstrating how digital networks can rapidly connect individual prejudice to mass mobilisation.²⁰⁰ The case illustrates

185 Castells, *Networks of Outrage and Hope*

186 Ibid.

187 Ibid.

188 Oz, Yanik, and Batu, "Under the Shadow of Culture and Politics."

189 Ibid.

190 Castells, *Networks of Outrage and Hope*

191 Tufekci, Zeynep. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. New Haven: Yale University Press, 2017.

192 Ibid.

193 Max Fisher, *The Chaos Machine: The Inside Story of How Social Media Rewired Our Minds and Our World* (New York: Little, Brown and Company, 2022).

194 Ibid.

195 Marwick and Lewis, *Media Manipulation and Disinformation Online*.

196 Tufekci, "Social Media and Human Rights."

197 Human Rights Watch, *All This Terror Because of a Photo*.

198 Tufekci, "Social Media and Human Rights."

199 Ibid.

200 Castells, *Networks of Outrage and Hope*

how the mechanisms described by Castells are not inherently democratic or emancipatory. Digital platforms can facilitate the transformation of shared emotion into collective action regardless of whether that action advances inclusion or exclusion.²⁰¹

In these cases, the same digital infrastructures that enabled rapid mobilisation also generated structural weaknesses. The ability to reach large audiences quickly, coordinate participation at low cost, and generate immediate visibility can make movements appear powerful in a relatively short period of time. However, the mechanisms that facilitate rapid growth do not necessarily create the organisational structures required to sustain long-term political action. A movement may attract millions of supporters, dominate public conversation, or generate significant media attention while lacking the leadership, resources, institutional relationships, and strategic capacity needed to translate visibility into durable political change.²⁰² Zeynep Tufekci argues that digital movements frequently produce high levels of "signal" without developing equivalent "capacity."²⁰³ 'Signal' refers to visible indicators of mobilisation such as viral hashtags, massive online engagement, or large protest crowds. 'Capacity' refers to the underlying organisational infrastructure required for sustained

political change – including leadership structures, negotiation ability, institutional memory, strategic planning, and durable institutions.²⁰⁴

Pre-digital social movements typically developed slowly because organising required significant labour, coordination, and institutional development. Labour unions, civil rights organisations, and feminist groups often spent years building internal structures before achieving mass visibility. At the same time, many movements were also shaped by catalytic events, charismatic leaders, and loose coalitions of participants whose mobilisation could rapidly expand public attention. Organisational development and mass visibility did not always occur in a linear sequence. However, the practical constraints of communication and coordination often meant that movements developed organisational capacity alongside public recognition over time. Digital platforms alter this relationship by allowing movements to achieve visibility rapidly without necessarily undergoing the same process of organisational consolidation.²⁰⁵

Additionally, digital platforms dramatically reduce the costs of coordination, allowing movements to scale rapidly through online networks. However, this acceleration can weaken organisational depth. Movements may, therefore, appear

politically powerful while lacking the institutional coherence necessary to sustain long-term strategic action.²⁰⁶ Paradoxically, the opposite problem can also occur. Organisations may possess extensive internal structures, established leadership, and durable institutional processes while lacking the public energy, visibility, or collective momentum necessary to generate meaningful political pressure.

Capacity without mobilisation can be as limiting as mobilisation without capacity.

Sustainable social change often depends upon maintaining a productive relationship between the two.²⁰⁷

Occupy Wall Street illustrates this tension clearly. The movement achieved extraordinary global visibility through digital coordination and symbolic occupation of physical space. However, its decentralised structure and resistance to formal leadership complicated strategic decision-making and long-term institutional consolidation.²⁰⁸ Although Occupy succeeded in placing economic inequality at the centre of public debate and popularising the

language of the "99 per cent," the movement struggled to convert its visibility into durable political structures capable of sustaining pressure over time.²⁰⁹ More than a decade later, many of the concerns that animated the movement remain highly relevant, with income and wealth inequality continuing to reach historically significant levels in many countries.²¹⁰

Similarly, aspects of the Arab Spring demonstrated how rapid digital mobilisation could destabilise existing regimes without necessarily producing durable democratic institutions capable of replacing them.²¹¹ This should not be interpreted as evidence that decentralised or youth-led mobilisation cannot produce lasting political change. Earlier student movements and civic uprisings across parts of the post-Soviet and Eastern European space played significant roles in democratic transitions and political reform.²¹² Rather, the comparison highlights that digital mobilisation alone is rarely sufficient. The long-term outcomes of social movements depend not only on their ability to mobilise supporters but also on their capacity to build institutions, navigate political transitions, and sustain collective action beyond moments of heightened public attention.²¹³

201 Marwick and Lewis, Media Manipulation and Disinformation Online.

202 Tufekci, Twitter and Tear Gas.

203 Ibid.

204 Ibid.

205 Ibid.

206 Ibid.

207 Ibid.

208 Ibid.

209 Ibid.

210 Tufekci, "Social Media and Human Rights."

211 Castells, Networks of Outrage and Hope

212 Tufekci, "Social Media and Human Rights."

213 Tufekci, Twitter and Tear Gas.

This structural fragility produces what Tufekci describes as "tactical freeze," in which movements struggle to adapt when political conditions change.²¹⁴ Leaderless or highly decentralised movements may face difficulty negotiating with institutions, articulating unified demands, or maintaining momentum once media attention dissipates. The same horizontal structures that facilitate participation can also inhibit strategic coordination.²¹⁵ In digital civic spaces – where attention shifts rapidly and political environments can change unexpectedly – this limitation becomes particularly significant. Human rights advocates may successfully mobilise large audiences around moments of crisis, discrimination, or legislative threat, yet struggle to sustain engagement once immediate attention declines. Anti-rights actors can face similar challenges, particularly when mobilisation is driven primarily by outrage rather than durable organisational structures. At the same time, movements that are able to convert moments of visibility into lasting networks, institutions, and strategic capacity may gain a significant advantage within increasingly contested online civic arenas.²¹⁶

For activists, Tufekci's analysis suggests that visibility should be understood as a starting point rather than an endpoint. Viral campaigns, trending hashtags, and

moments of widespread public attention can create valuable opportunities for mobilisation, but they do not automatically produce lasting political influence.²¹⁷ Sustained impact often requires converting visibility into organisational capacity through leadership development, coalition building, resource mobilisation, institutional partnerships, and long-term strategic planning. Digital networks may help movements attract attention, but maintaining momentum frequently depends upon structures that exist beyond the platform itself.²¹⁸

These dynamics are intensified by the contemporary attention economy. Digital platforms operate through systems that reward visibility, novelty, emotional intensity, and engagement. As a result, attention becomes fragmented, volatile, and difficult to sustain. Political participation increasingly occurs within cycles of rapid amplification followed by rapid decline. Visibility becomes easier to achieve but harder to maintain in a meaningful way.²¹⁹

For LGBTI+ activism, this creates a paradoxical condition. Online visibility can generate solidarity, public awareness, and rapid mobilisation around crises. However, movements can also become dependent upon virality and emotionally charged moments of attention that are

difficult to sustain organisationally and legislatively.²²⁰ Activism risks becoming episodic, shaped by algorithmic trends rather than long-term institutional development. In practice, this can produce cycles of intense mobilisation followed by periods of relative inactivity during which public attention shifts elsewhere despite the persistence of the underlying issues. Visibility may surge in response to a crisis, a discriminatory law, or a high-profile incident, only to decline once the algorithmic and media spotlight moves on.²²¹

This dynamic can create structural advantages for well-resourced anti-rights actors. While rights-based campaigns often mobilise in response to specific threats or moments of public attention, anti-rights ecosystems may operate through more durable organisational infrastructures, including advocacy organisations, political networks, religious institutions, media outlets, legal campaigns, and long-term funding streams.²²² As a result, they can continue shaping public discourse, influencing policy, and developing strategy during periods when activist attention and public engagement have diminished. The challenge for human rights movements is, therefore, not only generating visibility but sustaining organisational capacity

during the inevitable droughts between moments of mass attention.²²³

Disinformation, Attention Hacking, and Anti-Gender Mobilisation

If digital civic spaces provide opportunities for mobilisation, they also create opportunities for contestation. The same networked infrastructures that enable human rights advocacy can be used by opposing actors seeking to influence public discourse, shape collective identities, and challenge the legitimacy of social movements. Within these struggles, visibility becomes a strategic resource, as competing actors seek not only to persuade audiences but also to capture attention and define the terms through which political issues are understood.²²⁴ Disinformation is, therefore, not simply the circulation of false information but a strategic practice embedded within broader sociotechnical systems involving platforms, algorithms, media institutions, and networked actors.²²⁵ It operates as one among several tools through which political claims are amplified, distorted, or destabilised within contested digital environments.²²⁶

A central mechanism within this environment is 'attention hacking,' in which

214 Ibid.

215 Ibid,

216 Ibid.

217 Ibid.

218 Ibid.

219 Ibid.

220 Ibid.

221 Tufekci, Twitter and Tear Gas.

222 Royal United Services Institute, Online Gendered Narratives.

223 Tufekci, Twitter and Tear Gas.

224 Castells, Networks of Outrage and Hope

225 Marwick and Lewis, Media Manipulation and Disinformation Online.

226 Fisher, The Chaos Machine.

actors deliberately manipulate visibility systems to force narratives into mainstream circulation.²²⁷ Both human rights and anti-rights actors operate within this attention-driven landscape, adapting their strategies to platforms in which visibility is unevenly distributed and algorithmically amplified.²²⁸ While the content and normative goals of these actors differ fundamentally, they are shaped by the same underlying incentive structures that reward engagement, emotional intensity, and rapid circulation.²²⁹ Techniques include coordinated posting campaigns, meme production, bot amplification, hashtag manipulation, influencer targeting, and harassment campaigns.²³⁰ The objective is often not direct persuasion but manufactured visibility. Repetition itself creates perceived legitimacy.²³¹

These strategies are particularly significant within contemporary anti-LGBTI+ mobilisation. Anti-gender movements across the United States, Europe, Latin America, and Africa increasingly operate through transnational digital ecosystems

that circulate shared narratives, tactics, and symbols.²³² Although these movements differ politically and culturally, they frequently converge around themes portraying queer visibility and gender diversity as threats to children, morality, religion, or national identity.²³³

For example narratives framing queer individuals as ‘groomers’ or portraying gender diversity as social contagion have circulated extensively through social media networks and have led to tangible impacts on legislation globally.²³⁴ Accounts such as Libs of TikTok became influential nodes within these ecosystems by amplifying outrage around certain teachers, drag performers, healthcare providers who advocate for affirmative care, and LGBTI+ activists.²³⁵ This amplification did not remain confined to online discourse but increasingly fed into offline political and legislative agendas, including the proliferation of proposed and enacted restrictions on drag performances in parts of the United States.²³⁶ In these cases, viral narratives portraying drag as inher-

ently sexualised or harmful to children circulated through digital networks and were subsequently mobilised in political debates, contributing to policy efforts that sought to restrict or ban public drag events.²³⁷ Through coordinated attention practices, fringe narratives were repeatedly pushed into mainstream political discourse.²³⁸

Disinformation campaigns often rely heavily upon irony, humour, and meme culture because these forms increase shareability while providing plausible deniability. Harmful narratives can therefore circulate under the guise of satire or ‘just joking,’ complicating moderation efforts and enabling extremist discourse to migrate into mainstream digital culture.²³⁹

This dynamic is visible in a range of transnational contexts, including Brazil, where far-right digital networks have used memes, humour, and ironic content to circulate anti-LGBTI+ and anti-‘gender ideology’ narratives during and after the Bolsonaro presidency. In these environments, comedic framing and satirical aesthetics allowed politically charged claims about sexuality, education, and family values to spread widely while remaining formally deniable as ‘just humour,’ even as they shaped broader public attitudes and political debate.²⁴⁰

Traditional media systems frequently amplify these dynamics unintentionally. Contemporary journalism operates under intense pressures related to speed, metrics, audience engagement, and continuous content production. Journalists increasingly source stories directly from social media, allowing coordinated digital campaigns to shape mainstream news agendas.²⁴¹

This shift is reinforced by dual structural pressures within contemporary media systems. On the one hand, declining funding, shrinking newsrooms, and reduced institutional support for investigative and long-form reporting have limited the capacity of many outlets to generate original reporting at scale. On the other hand, journalists and media organisations operate within an attention economy in which audience engagement, clicks, and shareability are increasingly tied to revenue and organisational survival.²⁴²

As a result, social media platforms become both a primary source of story leads and a competitive environment in which news organisations must participate to maintain visibility and financial viability. In this context, visibility on digital platforms can itself function as a proxy for relevance, making highly circulated narratives more likely to be treated as newsworthy.

227 Marwick and Lewis, Media Manipulation and Disinformation Online.

228 Zuboff, The Age of Surveillance Capitalism.

229 Pariser, Eli. The Filter Bubble: What the Internet Is Hiding from You. New York: Penguin Press, 2011.

230 Ibid.

231 Fisher, The Chaos Machine.

232 Royal United Services Institute, Online Gendered Narratives.

233 Ibid.

234 Center for Countering Digital Hate, Digital Hate: The Global Battle for Reproductive Rights and LGBTQ+ Equality on Social Media (London: Center for Countering Digital Hate, 2022), <https://hrc-prod-requests.s3-us-west-2.amazonaws.com/CCDH-HRC-Digital-Hate-Report-2022-single-pages.pdf>

235 Abby Monteil, "How Libs of TikTok Became an Anti-LGBTQ+ Hate Machine," Them, September 8, 2022, <https://www.them.us/story/lib-of-tik-tok-twitter-facebook-instagram-explained-childrens-hospitals-grooming>

236 Center for Countering Digital Hate, Digital Hate.

237 Ibid.

238 Pariser, The Filter Bubble.

239 Fisher, The Chaos Machine.

240 Viktor Chagas, "Quiddity and Universality: The Transnational Origins of Humorous Far-Right Memes," The European Journal of Humour Research 12, no. 3 (2024): 88-109, <https://doi.org/10.7592/EJHR.2024.12.3.871>

241 Marwick and Lewis, Media Manipulation and Disinformation Online.

242 Ibid.

Repetition and virality, therefore, contribute to a form of perceived legitimacy in which the prominence of a claim within digital networks increases its likelihood of being recognised, reported, and amplified by mainstream media institutions.²⁴³

The participatory nature of internet culture, therefore, acts as a double-edged system for political actors operating within it. The same openness that enables human rights mobilisation, transnational solidarity, and rapid information-sharing also provides opportunities for coordinated manipulation and strategic amplification by opposing actors.²⁴⁴ Both rights-based and anti-rights movements must, therefore, adapt to an environment in which visibility is unevenly distributed and shaped by engagement-driven dynamics.²⁴⁵ As a result, actors who are able to generate emotionally resonant, highly shareable, or conflict-oriented content often gain disproportionate influence over public attention, allowing both mobilisation and misinformation to circulate rapidly through the same communicative infrastructures.²⁴⁶

Surveillance Capitalism and Behavioural Governance

Underlying contemporary digital civic infrastructures is not only a system for shaping behaviour, but also for shaping perception itself. Surveillance capitalism does not simply influence what individuals do online; it conditions what they come to see as credible, relevant, and real.²⁴⁷ Through the continuous collection and analysis of behavioural data, platforms are able to predict and optimise engagement in ways that structure informational environments, privileging certain narratives while marginalising others.²⁴⁸ Increasingly, however, this process is not purely responsive. It is also anticipatory and prefigurative; digital environments are designed from the outset to elicit particular forms of engagement, meaning that behaviour is not only predicted after it occurs but subtly guided, curated, and made more likely through the initial configuration of platform architectures.²⁴⁹ In doing so, they contribute to the formation of shared realities in which visibility, repetition, and algorithmic amplification play a central role in determining what is understood as legitimate knowledge.²⁵⁰

The origins of this model are often traced to the early 2000s, when Google discovered that behavioural ‘data exhaust’ generated by users possessed significant predictive value for advertisers.²⁵¹ Behavioural data could be used to anticipate and influence future consumer actions. This insight transformed behavioural prediction into a central mechanism of economic value creation and subsequently became foundational across the technology industry.²⁵² This capacity to anticipate, shape, and preconfigure user behaviour is also one of the central sources of value within contemporary technology companies, as the ability to reliably structure attention and engagement underpins advertising markets, data extraction systems, and the broader behavioural futures economy on which much of the digital infrastructure is built.²⁵³

Surveillance capitalism operates through four interconnected processes: extraction, translation, prediction, and monetisation. Behavioural data is extracted from digital interaction, translated into computational models through machine learning systems, used to generate predictions

about future behaviour, and sold within behavioural futures markets where advertisers and other actors seek to influence user decisions.²⁵⁴ This logic became publicly visible in cases such as the Cambridge Analytica scandal, in which data harvested from millions of Facebook users was reportedly used to construct psychological and political profiles for targeted campaigning.²⁵⁵ The case illustrated how behavioural data could be repurposed beyond commercial advertising into political influence operations, revealing the broader capacities of data-driven prediction systems to segment publics, tailor messaging, and shape political perception at scale.²⁵⁶

Importantly, these systems increasingly move beyond prediction toward behavioural modification. Recommendation algorithms, targeted advertising systems, personalised feeds, and automated nudges are designed not only to anticipate behaviour but to shape it.²⁵⁷ Platform infrastructures, therefore, exert subtle forms of behavioural governance through continuous optimisation of attention and interaction.²⁵⁸

243 Ibid.
244 Castells, Networks of Outrage and Hope
245 Tufekci, Twitter and Tear Gas.
246 Zuboff, The Age of Surveillance Capitalism.
247 Ibid.
248 Pariser, The Filter Bubble.
249 Zuboff, The Age of Surveillance Capitalism.
250 Pariser, The Filter Bubble.

251 Zuboff, The Age of Surveillance Capitalism.
252 Ibid.
253 Ibid.
254 Ibid.
255 Ibid,
256 OECD, Rights in the Digital Age.
257 Zuboff, The Age of Surveillance Capitalism.
258 Pariser, The Filter Bubble.

This convergence also creates national security risks. Behavioural data does not remain politically neutral simply because it is collected by private platforms or processed through commercial systems. Once gathered, inferred, purchased, leaked, or accessed through coercion, it can become intelligence.²⁵⁹ Bad actors – including hostile states, extremist networks, cybercriminal groups, and politically motivated influence operations – can use the same behavioural prediction systems used to sell products to anticipate dissent, locate dissidents, or identify social fracture points.²⁶⁰

Russia illustrates how digital spaces can be folded into broader state projects of ideological control. Human Rights Watch reported that Russian courts ordered LGBTI+ social media groups on VKontakte to be shut down under the country's 'gay propaganda' law, and noted that the law had also been used against protests, social media posts, teachers, and youth support resources.²⁶¹ ²⁶² More recent reporting has described how Russia's designation of the so-called 'international LGBTI+ movement' as extremist has intensified online censorship and self-cen-

sorship, including fear around searches, VPN use, and social media activity. In this context, surveillance is not only a technical matter but part of a broader attempt to make queer life less searchable, less speakable, and less publicly organised.²⁶³

Hungary shows a related but distinct pattern in which state-aligned narratives and legal restrictions actively reshape the digital information environment surrounding LGBTI+ rights. The 2021 Propaganda Law has not only constrained offline expression and education but has also produced a chilling effect across online platforms, influencing what users share, search for, and engage with on social media and digital news sites.²⁶⁴ Amnesty International has described the law as contributing to self-censorship and reduced visibility of sexual orientation and gender identity content, particularly in digitally mediated spaces where users anticipate surveillance, reporting, or platform moderation.²⁶⁵ At the same time, anti-LGBTI+ narratives circulate widely through online channels, including pro-government media outlets, social media pages, and algorithmically amplified content ecosystems that blur the boundaries between state

messaging and digital virality. Research from the European Digital Media Observatory has similarly documented how disinformation targeting LGBTI+ communities spreads through online platforms in Hungary, shaping public discourse within hybrid media systems where political communication, journalism, and social media are deeply intertwined.²⁶⁶

Iran and China further demonstrate that control over digital space often operates through the combined management of infrastructure, anonymity, censorship, and identity. For example, the Atlantic Council has warned that Iran's proposed Internet Protection Bill would threaten LGBTI+ communities by tightening state control over online spaces, eroding anonymity, and criminalising tools such as VPNs that many users rely upon to circumvent censorship.²⁶⁷ In China, scholarship on LGBTI+ digital activism has described how queer content is monitored or suppressed under censorship systems, forcing users to rely on coded language, VPNs, and other manoeuvres to sustain digital publics.²⁶⁸ These examples show that the governance of online space is not simply a platform-management question. It is also a geopolitical question about who controls infrastructure, who can remain anonymous, who can speak, and who becomes visible to the state.

Across these cases, it becomes clear that anti-democratic state actors have not only recognised the political potential of digital civic spaces but have actively leveraged them as extensions of state power.

Through the development of surveillance infrastructures, coordinated information strategies, and transnational technology partnerships, these actors have been able to integrate platform dynamics into broader systems of governance and control.²⁶⁹ Rather than treating digital space as a separate communicative sphere, they increasingly operate on the assumption that online environments are constitutive of political reality itself, and therefore must be managed, shaped, and – where necessary – restricted.²⁷⁰

By contrast, many democratic institutions and human rights frameworks have struggled to adapt with comparable speed or coherence. Regulatory responses are often fragmented across jurisdictions, constrained by legal and political processes, and reactive to harms that are already well established within platform ecosystems.²⁷¹ This asymmetry is not accidental. It reflects the structural difficulty

259 OECD, Rights in the Digital Age.

260 Land, "Toward a Human Rights Approach."

261 Human Rights Watch, All This Terror Because of a Photo.

262 for more information on this please see the Commissions Legal Battlegrounds working paper.

263 Amnesty International, "Russia: Russian LGBT Network Is Labelled 'Extremist' in Escalating Crackdown on LGBTI Rights," Amnesty International, April 27, 2026, <https://www.amnesty.org/en/latest/news/2026/04/russia-russian-lgbt-network-is-labelled-extremist-in-escalating-crackdown-on-lgbti-rights/>.

264 Natalie Alkiviadou, Hate Speech and the European Court of Human Rights (Abingdon, UK: Routledge, 2025).

265 Amnesty International, The Social Atrocity: Meta and the Right to Remedy for the Rohingya (London: Amnesty International, 2022).

266 EDRI, "Digital Rights of LGBTQ+ People."

267 Amnesty International, The State of the World's Human Rights (London: Amnesty International, 2024), <https://www.amnesty.org/en/wp-content/uploads/2024/04/WEBPOL1072002024ENGLISH.pdf>.

268 Ibid.

269 OECD, Rights in the Digital Age.

270 Land, "Toward a Human Rights Approach."

271 OECD, Rights in the Digital Age.

of aligning open, deliberative governance systems with fast-moving, commercially driven digital infrastructures.²⁷² It also helps explain why marginalised communities – including LGBTI+ people – are disproportionately exposed to the consequences of these dynamics; they are navigating systems in which the capacity for coordinated restriction and manipulation is highly developed, while protective and accountability mechanisms remain

uneven, delayed, or incomplete.²⁷³ The result is a widening gap between the sophistication of actors seeking to control digital space and the institutional capacity available to contest or regulate that control. Closing this gap requires not only improved policy responses, but a broader recognition that digital civic space has become a central arena of political struggle in which rights, visibility, and legitimacy are continuously negotiated.

Case Study: Pandemic Dependency and the Intensification of Digital Life ²⁷⁴

The preceding sections have focused primarily on the structural and systemic dimensions of digital civic space, including the ways in which platforms, states, and political actors compete to shape visibility, legitimacy, and mobilisation. However, these dynamics are not only macro-level phenomena. They are also experienced in everyday life, where individuals navigate the intensified presence of digital infrastructures across work, sociality, political engagement, and identity formation.

The COVID-19 pandemic represents a critical moment in which these structural conditions became more visible and more immediate. Lockdowns, social distancing measures, and the collapse of physical gathering spaces dramatically intensified dependence upon online platforms for communication, work, activism, education, and social connection. Digital environments shifted from supplementary communication tools to essential infrastructures of everyday survival.

The dynamics intensified during the pandemic were not entirely new. Digital harassment, disinformation, surveillance, isolation, uneven access, and platform dependence all predated COVID-19. Queer people had already used online

spaces as lifelines where physical communities were unavailable. Bad actors had already exploited platforms to spread hate or misinformation. States had already used digital tools to monitor dissent and marginalised communities. The question, then, is not whether the pandemic created these dynamics from nothing; it is what the pandemic changed.

What changed was scale, dependency, and inevitability. Before the pandemic, digital life could still be imagined – by some – as supplementary. COVID-19 collapsed that distinction. Work, school, healthcare, organising, friendship, worship, protest, dating, family contact, and crisis response all moved more fully online at once. Digital infrastructures became not merely convenient but compulsory. This meant that existing harms were intensified because opting out became harder.

The pandemic also accelerated disinformation, conspiracy networks, and extremist mobilisation. Anti-LGBTI+ narratives became entangled with broader conspiratorial ecosystems surrounding vaccines, global institutions, and moral decline. Existing inequalities in digital access were similarly intensified as marginalised populations often lacked secure devices, private internet access, or protected online environments.

The pandemic also changed the emotional conditions of digital life. Isolation, fear, grief, boredom, and institutional distrust created ideal conditions for both support networks and manipulation. For some users, online spaces provided community, mental health support, and political continuity. For others, they became sites of radicalisation, harassment, and conspiratorial belonging.

This disruption did not simply increase screen time; it unsettled older assumptions about the boundaries between online and offline political life. Norms that had once treated digital space as secondary to ‘real’ politics became increasingly untenable as campaigning, activism, socialisation, media consumption, and grievance formation converged within the same infrastructures. In this environment, digitally mediated outrage and identity-based mobilisation became more politically consequential, helping to erode traditional expectations about democratic deliberation, institutional gatekeeping, and the pace at which fringe narratives could enter the mainstream.

The pandemic, therefore, condensed and intensified many of the structural contradictions already present within digital civic space. Online infrastructures enabled continuity of activism, solidarity, and connection under conditions of physical isolation while simultaneously amplifying surveillance, misinformation, harassment, and inequality.

272 Birdsall, "Human Rights-Based Approach to Social Media Platforms."

273 OECD, Rights in the Digital Age.

274 Lucio Lage Gonçalves, Antonio Egidio Nardi, and Anna Lucia Spear King, "Digital Dependence in Organizations: Impacts on the Physical and Mental Health of Employees," *Clinical Practice & Epidemiology in Mental Health* 19 (2023): e174501792212300, <https://doi.org/10.2174/17450179-v19-e230109-2022-17>

Conclusion

Online spaces now function as central civic infrastructures through which contemporary social, political, and cultural life is organised. These platforms are not neutral technologies but systems of governance that structure visibility, participation, communication, and behavioural norms at global scale. Through moderation systems, algorithmic filtering, data extraction, and attention-based economies, platforms increasingly shape the conditions under which democratic participation and collective identity formation occur.

For LGBTI+ communities and other marginalised groups, these dynamics are especially pronounced. Digital platforms have enabled unprecedented forms of visibility, solidarity, and transnational mobilisation, providing vital spaces for identity formation and political organising. Yet, these same infrastructures also expose queer individuals and organisations to intensified forms of surveillance, harassment, disinformation, and algorithmic marginalisation.

However, these dynamics do not operate in isolation or in a single direction. Anti-rights actors have also become deeply embedded within the same digital infrastructures, developing sophisticated strategies for mobilisation, narrative shap-

ing, and attention capture. As a result, digital civic space is not simply a site of empowerment or vulnerability but a contested arena in which rights-based and anti-rights actors compete over visibility, legitimacy, and influence under shared structural conditions.

The central paradox of online civic space is, therefore, not only that empowerment and vulnerability coexist, but that opposing political projects are shaped by the same underlying systems of attention, amplification, and behavioural governance. The same infrastructures that enable human rights mobilisation also facilitate coordinated anti-rights campaigns, producing a political environment in which visibility functions simultaneously as opportunity and risk, and where legitimacy is continuously negotiated through exposure rather than institutional stability.

Understanding contemporary civic life consequently requires recognising that online platforms are no longer peripheral spaces of communication. They are fundamental arenas in which power, identity, legitimacy, and resistance are continuously produced and contested between actors with divergent political goals but shared technological conditions. The next section descends further into the machinery of these arenas, examining the algorithmic systems that, invisible to most users, determine which voices are heard and which are silenced.

Section 3 Key Findings:

- Visibility, participation, legitimacy, and exclusion are continuously negotiated through technical systems, commercial incentives, and political pressure, meaning that both LGBTI+ advocacy and anti-rights mobilisation operate under the same structural conditions.
- Digital advocacy mobilisation often produces high ‘signal’ but often weak ‘capacity.’ Movements like Occupy Wall Street and the Arab Spring achieved rapid visibility without building the durable organisational infrastructure needed to convert attention into lasting change.
- Rights-based campaigns often mobilise reactively around crises, while anti-rights ecosystems increasingly maintain durable infrastructure (advocacy organisations, religious institutions, media outlets, sustained funding) that keeps working during the ‘droughts’ between moments of public attention.
- ‘Attention hacking’ has been used with particular effect against LGBTI+ communities: coordinated posting, meme production, bot amplification, and hashtag manipulation manufacture visibility and perceived legitimacy through repetition.
- Irony and meme culture provide plausible deniability for harmful narratives.
- Under-resourced journalism amplifies coordinated online campaigns unintentionally due to declining newsroom capacity combined with engagement-driven media economics resulting in journalists sourcing stories from social media and letting virality function as a proxy for newsworthiness.
- Behavioural data extraction has evolved from prediction to active behavioural governance, and this same data can be repurposed into political intelligence.
- Some states actively weaponise digital surveillance infrastructure against LGBTI+ communities.

Section 3 Questions:

- What strategies has your organisation used to maintain advocacy capacity during the 'droughts' between moments of high public attention, when algorithmic and media focus has moved elsewhere?
- Have you observed coordinated anti-LGBTI+ campaigns (coordinated posting, meme networks, hashtag manipulation, bot amplification) in your country or platform environment? What form did they take, and how did mainstream media or institutions respond?
- Have you encountered anti-LGBTI+ content that used humour, satire, or 'just joking' framing to spread narratives while evading moderation? How did this affect your ability to respond or report it?
- If you operate in a context with restrictive digital laws (e.g., propaganda laws, extremist designations, VPN restrictions), how has this affected what you and your community feel able to search for, post, or organise online? What protective measures have proven effective?
- Have you observed cases where anti-LGBTI+ narratives that originated or gained traction on social media were subsequently picked up and legitimised by mainstream news coverage? What would help journalists avoid amplifying coordinated disinformation?

```
</php
session_start();

// If user is already logged in, redirect to homepage
if(isset($_SESSION['user_id'])) {
    header('Location: index.php');
    exit;
}

// If login form is submitted
if($_SERVER['REQUEST_METHOD'] == 'POST') {

    // Connect to database
    $db = new PDO('mysql:host=localhost;dbname=your_database_name', 'your_username', 'your_password');

    // Retrieve user record from database
    $stmt = $db->prepare('SELECT * FROM users WHERE email = :email');
    $stmt->bindParam(':email', $_POST['email']);
    $stmt->execute();
    $user = $stmt->fetch(PDO::FETCH_ASSOC);

    // Verify password
    if($user && password_verify($_POST['password'], $user['password'])) {

        // Set session variables and redirect to homepage
        $_SESSION['user_id'] = $user['id'];
        $_SESSION['user_name'] = $user['name'];
        header('Location: index.php');
        exit;
    } else {
```

Section 4:

Algorithms - Amplification, Bias, and Free Speech

Section 3 examined online platforms as civic arenas and showed how their governance structures – moderation, recommendation, commercial incentives – shape participation in unequal ways.²⁷⁵ This section moves from the institutional architecture to the technical one: specifically, to the algorithmic systems that sit at the heart of every platform and determine, moment by moment, what becomes visible, what is suppressed, and what is amplified into mass attention.²⁷⁶

The question of visibility is – at base – a question of validity. What is made visible is implicitly endorsed as worthy of attention; what is suppressed is rendered socially marginal. In digital environments, this process of legitimation is no longer performed by editors, publishers, or civic institutions but by behavioural governance systems – algorithmic architectures that rank, recommend, filter, and remove content according to engagement signals rather than epistemic or democratic criteria.²⁷⁷ These systems do not operate at the

²⁷⁵ Gillespie, *Custodians of the Internet*.

²⁷⁶ OECD, *Rights in the Digital Age*.

²⁷⁷ Riedl, "Rethinking Rights in Social Media Governance."

periphery of public life; they constitute its central infrastructure.²⁷⁸ Understanding how they function and how they are exploited is, therefore, a prerequisite for understanding the conditions under which political speech, identity expression, and civic mobilisation are possible today.²⁷⁹

The architecture of amplification sits at the core of this inquiry, but it has limits that personalisation systems then extend and intensify. Together, they produce fragmented informational realities that map onto offline social inequalities and, as later discussion will indicate, have begun to intersect with questions of national security and geopolitical manipulation.²⁸⁰

Earlier sections have noted that algorithmic systems tend to reproduce, rather than transcend, existing social hierarchies.²⁸¹ That observation is developed more fully here but with a sharper focus. The concern is not algorithmic bias in the abstract but the specific ways in which anti-rights actors – well-resourced, strategically literate, and operationally coordinated – have learned to exploit these structures to suppress LGBTI+ visibility and amplify identity-based hostility at scale.²⁸²

Digital platforms are now complex governance systems that actively organise visibility, shape participation, and regulate speech across vast populations. What appears on a feed, trends on a homepage, disappears from search results, or becomes subject to moderation is determined not through transparent editorial judgment but through opaque algorithmic processes that encode commercial priorities, dominant social norms, and structural assumptions about language and risk. This can be understood as digital normativity.²⁸³

For LGBTI+ users, this produces a distinctive and acute form of asymmetrical visibility; queer identities are simultaneously rendered hyper-visible as targets of moral panic and harassment, while identity-affirming speech is subjected to algorithmic scrutiny that treats it as suspicious, explicit, or harmful.²⁸⁴

Platform Governance and the Architecture of Visibility

How public awareness is generated is not itself a neutral process. Platforms do not merely amplify what exists; they shape what becomes knowable, legible,

and shareable. In doing so, they actively constitute the informational conditions of public life.²⁸⁵

While some of these biases operate through deliberate design choices, many are the product of structural and commercial logics that produce discriminatory outcomes without conscious intent, making them no less harmful and arguably more insidious. They are harder to name, harder to challenge, and harder to remedy precisely because no individual actor can be said to have chosen them.²⁸⁶ Although digital platforms are often framed as open arenas for free expression, they are more accurately understood as highly governed environments structured through continuous moderation, ranking, recommendation, and filtering processes.²⁸⁷

Tarleton Gillespie argues that platforms function not as neutral intermediaries but as custodians of public discourse. Every platform is, therefore, simultaneously a communication system and a governance system.²⁸⁸ This dual character matters enormously, because governance systems produce winners and losers. In a context where civic participation, political mobilisation, and public advocacy increasingly

occur through platform infrastructures, control over who is visible and who is suppressed is not merely an administrative question; it is a form of political power, one with direct consequences for democratic life and human rights.²⁸⁹

Moderation is not a peripheral activity but an infrastructural necessity. At the scale of billions of daily interactions, platforms rely upon complex combinations of automated filtering systems, human moderators, community reporting mechanisms, and policy frameworks to regulate content. These systems determine not only what is removed but also what is amplified, recommended, downranked, or rendered invisible.²⁹⁰

This raises a question that platform governance discourse too rarely confronts directly: Who moderates and by what standards? There is no agreed international framework governing platform content moderation.²⁹¹ Governments disagree profoundly about what constitutes harmful content; different legal traditions produce irreconcilable definitions of hate speech, acceptable expression, and the limits of free speech; and in the absence of legally binding global norms, these decisions are made by private compa-

278 Gillespie, Custodians of the Internet.
279 Land, "Toward a Human Rights Approach."
280 Eli Pariser, The Filter Bubble: What the Internet Is Hiding from You (New York: Penguin Press, 2011).
281 Couldry and Mejias, The Costs of Connection.
282 Royal United Services Institute, Online Gendered Narratives.
283 Gillespie, Custodians of the Internet.
284 Human Rights Watch, All This Terror Because of a Photo.

285 Gillespie, Custodians of the Internet.
286 Birdsell, "Human Rights-Based Approach to Social Media Platforms."
287 Gillespie, Custodians of the Internet.
288 Ibid.
289 Klonick, "Beyond the Public Square."
290 Gillespie, Custodians of the Internet.
291 OECD, Rights in the Digital Age.

nies whose primary accountability is to shareholders rather than to the public.²⁹² For LGBTI+ communities, the stakes of this governance vacuum are not merely bureaucratic. In contexts where platform governance fails to protect marginalised users – or where it is captured by hostile norms and political pressures – LGBTI+ users face not merely inconvenience but systematic invisibility, exposure to coordinated targeting, and the erosion of the communicative infrastructure on which their safety and civic participation depend.²⁹³

Visibility itself becomes a form of implicit support and validation. Power on digital platforms operates not only through censorship but through algorithmic prioritisation.²⁹⁴ Recommendation systems shape which narratives gain traction, which communities form, and which voices become culturally dominant.²⁹⁵ Every feed is, therefore, a curated reality structured through commercial incentives, technical design decisions, and moderation policies.²⁹⁶

The depth of this structural tilt is illustrated with particular clarity by experimen-

tal research into how recommendation systems treat newly created accounts. A 2024 study by researchers at UCL and the University of Kent found that accounts set up on TikTok experienced a fourfold increase in misogynistic content on the ‘For You’ page over just five days – rising from 13% to 56% of recommended content – even when those accounts had not sought out such material; the content became increasingly focused on anger and blame directed at women.²⁹⁷ A 2024 study by Professor Debbie Ging at Dublin City University's Anti-Bullying Centre used ten blank sock puppet accounts on TikTok and YouTube Shorts and found that all male-identified accounts were recommended masculinist, anti-feminist, and extremist content within the first twenty-three minutes of use, regardless of whether they had sought it out. By the end of the experiment, following approximately four hundred videos, seventy-six percent of recommended TikTok content and seventy-eight percent of YouTube Shorts content was classified as toxic, with anti-transgender content constituting a significant subcategory.²⁹⁸ A further replicated experiment – widely reported in 2025 – found that adults setting up

accounts posing as thirteen-year-olds on Instagram and TikTok were recommended harmful content within days, despite having engaged only with positive material. Even with Teen Account protections in place, accounts presenting as thirteen-year-old girls were shown content containing negative mental health and eating disorder messaging.²⁹⁹ These findings are not incidental. They reflect the underlying engagement logic of platforms, which rewards emotional intensity and controversy.³⁰⁰ The structural tilt toward harmful content is baked in from the moment a user signs up; it is an architectural feature, not a side-effect of individual bad behaviour.³⁰¹

This governance is inherently uneven. Moderation systems frequently operate inconsistently across linguistic, cultural, and regional contexts. Content moderation resources are disproportionately concentrated in English-speaking contexts, while many non-Western languages receive significantly less oversight. This creates substantial inequalities in civic protection and visibility.³⁰²

This unevenness does not fall randomly across user populations. Within the same platforms, the treatment of LGBTI+ con-

tent and that of hostile content exhibits a structural double standard that has been documented across numerous studies and advocacy investigations.³⁰³ LGBTI+ content is frequently misclassified as sensitive, sexual, or adult and subjected to suppression or restricted distribution, while overtly hostile anti-LGBTI+ content routinely remains active – classified as political speech, religious expression, or legitimate commentary – despite meeting the definitional criteria for hate speech under many national legal frameworks.³⁰⁴

This asymmetry does not require deliberate anti-LGBTI+ intent at the level of platform policy. It emerges from the interaction of several structural forces: conservative default assumptions embedded in training data that reflect dominant social norms rather than the diversity of human experience; the commercial incentive to avoid political controversy with well-organised and politically influential constituencies; and the disproportionate pressure that anti-rights actors have learned to apply to platform moderation systems, weaponising reporting mechanisms and harassment campaigns to target LGBTI+ content and creators.³⁰⁵ The outcome is a moderation architecture that – whatever its stated neutrality – systematically

292 Ibid.

293 Human Rights Watch, All This Terror Because of a Photo.

294 Riedl, "Rethinking Rights in Social Media Governance."

295 Ibid.

296 Pariser, The Filter Bubble.

297 Kaitlyn Regehr, Safer Scrolling: How Algorithms Popularise and Gamify Online Hate and Misogyny for Young People (London: University College London, University of Kent, and Association of School and College Leaders, 2024).

298 Baker, Catherine, Debbie Ging, and Maja Brandt Andreasen. Recommending Toxicity: The Role of Algorithmic Recommender Functions on YouTube Shorts and TikTok in Promoting Male Supremacist Influencers. Dublin: DCU Anti-Bullying Centre, Dublin City University, April 2024. <https://antibullyingcentre.ie/recommending-toxicity/>

299 Fisher, Social Media.

300 Fisher, The Chaos Machine.

301 Ibid.

302 OECD, Rights in the Digital Age.

303 Human Rights Watch, All This Terror Because of a Photo.

304 EDRI, "Digital Rights of LGBTQ+ People."

305 EDRI, "Digital Rights of LGBTQ+ People."

produces conditions more hospitable to those who wish to suppress LGBTI+ visibility than to those seeking to assert it.³⁰⁶

The consequences of these inequalities are visible in numerous cases. Facebook's documented failures in moderating anti-Rohingya hate speech in Myanmar revealed how platform infrastructures could contribute to large-scale violence when moderation systems fail to operate effectively across linguistic and regional contexts.³⁰⁷ United Nations investigators later concluded that Facebook had played a significant role in amplifying genocidal rhetoric against Rohingya Muslims because hate speech circulated faster than moderation systems could respond.³⁰⁸ Despite the gravity of those findings, no binding international legislation resulted. The United Nations called for greater transparency, independent monitoring, and regulatory reform of the tech sector; Amnesty International in 2022 called for Meta to provide reparations, and civil legal proceedings were filed in the United Kingdom and United States in December 2021.³⁰⁹ Meta commissioned a human rights impact assessment, acknowledged it had not done enough, and made incre-

mental operational changes. However, its core engagement-based business model – which investigators had identified as the structural cause of the harm – remained intact.³¹⁰ This outcome exemplifies a recurring pattern: When platforms contribute to atrocities, the institutional response is acknowledgment and incremental adjustment, not structural reform or binding accountability.³¹¹

LGBTI+ users similarly experience uneven moderation practices. Queer educational content, drag performances, body-positive imagery, and discussions of gender identity have frequently been flagged as ‘adult’ or ‘unsafe’ by automated systems.³¹² TikTok has faced criticism for suppressing LGBTI+-related hashtags and allegedly downranking content produced by marginalised creators under policies ostensibly designed to reduce bullying.³¹³ Instagram provides another useful example of how platform governance often operates through language that appears neutral while producing outcomes that are anything but disinterested. Content may be limited, hidden, downranked, or removed under categories such as ‘sensitive content,’ ‘sexual solicitation,’

‘adult content,’ ‘community safety,’ or ‘recommendation eligibility.’³¹⁴ These phrases sound procedural, technical, and universal; yet when they are applied to queer bodies, drag performance, trans healthcare information, sexual education, or LGBTI+ advocacy, they can reproduce older patterns of moral policing in platform language. The decision is presented as neutral moderation, but the effect is political; certain identities, bodies, and forms of knowledge become less visible.³¹⁵

The stakes of this suppression extend well beyond individual user experience. Social movements depend on visibility; the ability to mobilise, coordinate, share information, build coalitions, and generate public pressure is contingent on communicative access.³¹⁶ When platforms suppress LGBTI+ content – whether through misclassification, algorithmic downranking, or the application of restrictive content policies – they do not merely inconvenience individuals. They interrupt organising infrastructure.³¹⁷ The historical analogies are instructive even if the mechanisms differ; the suppression of samizdat in Soviet contexts or the policing of physical meeting spaces for queer communities were forms

of political control that operated precisely by limiting the communicative capacity of targeted groups.³¹⁸ Algorithmic suppression achieves structurally analogous effects through different means and does so at scale, with deniability, and without the visibility that overt censorship would attract.³¹⁹

The double standard operating within platform governance further compounds this political consequence. Anti-rights actors – including those who deploy religious, nationalist, or child-safety framing in opposition to LGBTI+ equality – routinely operate on the same platforms with significantly less algorithmic suppression, less content removal, and greater reach.³²⁰ Content that dehumanises LGBTI+ people through coded or euphemistic language, or that frames equality as threat, frequently passes automated moderation without restriction, while queer content that is merely affirming of identity or educational in character is classified as sensitive or adult.³²¹ This is not merely unfair; it is politically consequential. It means that the same infrastructure that restricts the organising capacity of LGBTI+ communities actively facilitates the reach of those seeking to curtail their rights.³²²

306 Human Rights Watch, All This Terror Because of a Photo.

307 Marwick and Lewis, Media Manipulation and Disinformation Online.

308 OECD, Rights in the Digital Age.

309 Amnesty International, The Social Atrocity.

310 Ibid.

311 Business for Social Responsibility (BSR), Human Rights Impact Assessment: Facebook in Myanmar (San Francisco: Business for Social Responsibility, 2018), <https://www.bsr.org/en/our-insights/report-view/human-rights-impact-assessment-facebook-in-myanmar>

312 Human Rights Watch, All This Terror Because of a Photo.

313 EDRI, "Digital Rights of LGBTQ+ People."

314 OECD, Rights in the Digital Age.

315 Human Rights Watch, All This Terror Because of a Photo.

316 Castells, Networks of Outrage and Hope

317 Ibid.

318 Amnesty International, "Russia: Russian LGBT Network Is Labelled 'Extremist.'"

319 OECD, Rights in the Digital Age.

320 Human Rights Watch, All This Terror Because of a Photo.

321 EDRI, "Digital Rights of LGBTQ+ People."

322 Ibid.

These moderation asymmetries reveal how platform governance can reproduce and entrench structural inequalities even when – perhaps especially when – systems claim neutrality.³²³

The governance power of platforms has become even more visible following reductions in moderation capacity across major companies. After Elon Musk's acquisition of Twitter, now X, the platform significantly reduced trust and safety staff while reinstating numerous previously banned accounts.³²⁴ Researchers and advocacy organisations subsequently documented increases in hate speech and harassment targeting marginalised communities, including queer users.³²⁵ Such developments demonstrate how changes in corporate governance can rapidly alter the conditions of public participation for millions of users.³²⁶

The Musk/X case illustrates a broader and structurally significant pattern in platform governance; when sufficiently powerful actors find themselves subject to platform moderation, a new option has become available, namely owning or creating the platform itself – a possibility that was

previously unavailable to ordinary users. When Donald Trump was permanently banned from Twitter in January 2021 following the attack on the United States Capitol, he did not lose access to mass communication.³²⁷ With sufficient capital and political support, he and his allies proceeded to launch Truth Social, a platform explicitly designed around the governance preferences of its owner, where moderation policies reflected the political orientation of its controlling interests.³²⁸ This pattern reveals a structural inequality at the heart of platform governance: For those with sufficient wealth and political influence, moderation decisions can be circumvented not through appeal or advocacy but through ownership.³²⁹ The rules, in other words, apply differently depending on whether you are subject to them or own the institution that enforces them.

Platforms, therefore, increasingly mimic quasi-governmental institutions but without accompanying accountability.³³⁰ They establish rules, enforce behavioural norms, adjudicate disputes, and structure visibility at global scale.³³¹ Yet, the honest answer to the question of who is actually

in charge is uncomfortable: in practice, a small number of extraordinarily wealthy individuals.³³² The governance of public discourse for billions of people is exercised – and can be unilaterally altered – by platform owners whose primary accountability is to shareholders, not to citizens or democratic institutions.³³³

Meta's Oversight Board offers an instructive illustration of the limits of voluntary self-governance in this context.³³⁴ Established in 2020 as a quasi-independent governance mechanism, it represented a genuine experiment in structured accountability. The Board issued binding decisions on individual content cases and made over three hundred recommendations to Meta on matters of policy.³³⁵ However, its policy recommendations were not binding on the company, and Meta declined to implement a significant proportion of them.³³⁶ More revealingly, when Meta's CEO Mark Zuckerberg announced sweeping changes to content moderation in early 2025 – removing fact-checking programmes, relaxing hate speech rules, and explicitly permitting anti-LGBTI+ characterisations – he did so unilaterally, without the Board's approval or prior input, exposing the limits

of voluntary self-governance as a substitute for democratic accountability.³³⁷ The Board has continued to operate and secured funding through 2028, but concerns about its independence and its shifting stance on contested moderation decisions have intensified, raising serious questions about whether it functions as a genuine check on corporate power or primarily as legitimating infrastructure for governance decisions already made.³³⁸ This concentration of power over public discourse in the hands of individuals who are unaccountable to democratic institutions is not merely a governance problem; it is a rights problem. The conditions under which marginalised communities can speak, organise, and be seen are determined by decisions in which those communities have no meaningful voice.³³⁹

Engagement Optimisation and the Architecture of Amplification

At the centre of platform governance lies the logic of engagement optimisation. Social media platforms such as Facebook, TikTok, YouTube, Instagram, and X are structured around systems designed to maximise user attention. Rather than prioritising information according to public

323 Human Rights Watch, All This Terror Because of a Photo.

324 Center for Countering Digital Hate, Digital Hate.

325 Human Rights Watch, All This Terror Because of a Photo.

326 Center for Countering Digital Hate, Digital Hate.

327 Royal United Services Institute, Online Gendered Narratives.

328 Ibid.

329 Gillespie, Custodians of the Internet.

330 OECD, Rights in the Digital Age.

331 Land, "Toward a Human Rights Approach."

332 OECD, Rights in the Digital Age.

333 Land, "Toward a Human Rights Approach."

334 Oversight Board, 2024 Annual Report (Menlo Park, CA: Oversight Board, 2025).

335 Ibid.

336 Ibid.

337 Oversight Board, 2024 Annual Report.

338 Ibid.

339 Land, "Toward a Human Rights Approach."

value, democratic importance, or factual reliability, algorithmic ranking systems prioritise behavioural signals including clicks, comments, watch time, shares, and emotional reactions.³⁴⁰

These metrics appear neutral on the surface. In practice, however, they function as proxies for emotional intensity. Content that provokes outrage, fear, anxiety, anger, or moral indignation consistently performs better within engagement-driven systems because it sustains user attention for longer periods and generates more interaction. Algorithms gradually learn to privilege such content because emotionally charged material produces stronger engagement signals and, therefore, greater advertising revenue.³⁴¹

This process creates what has been described as an algorithmic amplification loop. Emotionally intense content receives greater visibility. Increased visibility generates additional engagement. That engagement then signals to the system that similar content should be further promoted. Over time, platforms become structurally predisposed toward polarising and sensationalist material.³⁴²

Critically, this predisposition is not merely a consequence of accumulated user behaviour; it is baked into the architec-

ture from the moment a new user arrives. Internal platform research has indicated that platforms understand emotional escalation to be among the most reliable mechanisms for attention capture and data generation. New users are, therefore, frequently placed on a content trajectory calibrated for escalating emotional intensity regardless of their initial intent or interest, effectively initiating a process of radicalisation that precedes any active choice on the user's part.³⁴³

This structural tilt has direct consequences for the forms of communication on which democratic life depends. Slow, contextual deliberation and reflective exchange are systematically disadvantaged. This puts traditional forms of civic and social discourse – myth-busting, fact-checking, careful argumentation – on a continuous structural backfoot, because the architecture of attention is tilted against the very registers of communication that reason and rights-based advocacy require.³⁴⁴

The broader social consequences of this architecture have been extensively documented. Max Fisher describes social media systems as a "chaos machine" in which engagement optimisation gradually destabilises public discourse by rewarding outrage, fragmentation, and extremity over deliberation and nuance.³⁴⁵

This amplification logic has repeatedly intersected with political radicalisation and identity-based hostility. Facebook's recommendation systems were criticised for contributing to the spread of anti-Rohingya hate speech in Myanmar – a case already introduced in Section 3 as evidence of moderation failure. But here, the lens shifts; the Myanmar case is not only about absent moderation but about active amplification. Inflammatory and dehumanising posts spread rapidly through Facebook's engagement systems, contributing to an environment that intensified offline violence and ethnic persecution. The algorithm was not passive. It was doing its job – and its job rewarded hatred.³⁴⁶

Similar dynamics emerged during the spread of QAnon conspiracy networks in the United States, where recommendation systems on YouTube, Facebook, and Twitter amplified conspiratorial content because users engaged intensely with increasingly extreme material.³⁴⁷ The January 6 United States Capitol attack similarly demonstrated how algorithmically amplified narratives can move from online ecosystems into physical mobilisation. False election narratives circulated widely through engagement systems that rewarded outrage and identity affirmation. Although these narratives violated factual standards, they generated exceptionally strong engagement and, therefore, received extensive algorithmic circulation

before moderation systems intervened.³⁴⁸

Within LGBTI+ contexts, engagement optimisation frequently amplifies moral panic narratives because such narratives generate outrage and emotional participation. Research from the Center for Countering Digital Hate documented the rapid spread of anti-LGBTI+ ‘grooming’ narratives across major social media platforms. These narratives falsely associated queer identities with child predation and were repeatedly amplified despite violating platform hate speech policies. The report found that a relatively small number of posts generated millions of impressions through algorithmic distribution systems.³⁴⁹

The dynamics documented in the United States have precise structural equivalents across entirely different national, legal, and linguistic contexts, which together demonstrate that this is a systemic feature of engagement-driven platform design rather than a peculiarity of any single political environment.

In Poland, a September 2025 Amnesty International report – "'A Thousand Cuts'; Technology-Facilitated Gender-Based Violence Against Poland's LGBTI Community on X" – analysed 163,048 tweets collected from 32 research accounts over March 2025 and found homophobic and transphobic content highly prevalent among

340 Gillespie, Custodians of the Internet.

341 Zuboff, The Age of Surveillance Capitalism.

342 Fisher, The Chaos Machine.

343 Zuboff, The Age of Surveillance Capitalism.

344 Eli Pariser, The Filter Bubble: What the Internet Is Hiding from You (New York: Penguin Press, 2011).

345 Fisher, The Chaos Machine.

346 Human Rights Watch, All This Terror Because of a Photo.

347 Marwick and Lewis, Media Manipulation and Disinformation Online.

348 Ibid.

349 Center for Countering Digital Hate, Digital Hate.

accounts following politicians opposed to LGBTI+ rights. X's recommender system was found to amplify harmful content through engagement maximisation, a problem dramatically compounded by the platform having just two Polish-speaking content moderators for a population of 37.45 million people and 5.33 million X users.³⁵⁰

In Uganda, a 2024 Mozilla Foundation-supported study by HER Internet – "Navigating Algorithms; The Case of Structurally Silenced Communities in Uganda" – found that social media algorithms were being exploited to amplify misinformation and hate speech about Uganda's LGBTI+ community, including false narratives linking LGBTI+ identities to crime and mental illness. Researchers found that these algorithmically circulated narratives had a measurable influence on legislative outcomes, including the passage of Uganda's Anti-Homosexuality Amendment Act.³⁵¹

In Arabic-speaking contexts stretching from North Africa to the Gulf, a 2025 GATE report – "An Engineered Hostility; The Architecture of Anti-Trans Oppression Online" – documented how a coordinated digital counter-campaign called Fet-

rah, originating in Egypt in 2022, spread across online platforms using a unified social media strategy, professional website infrastructure, and multilingual content to promote the rejection of LGBTI+ identities as a "Western cultural invasion," exploiting engagement algorithms to achieve regional reach.³⁵²

In each of these cases, the mechanism is identical; content that generates outrage about LGBTI+ identities generates engagement, and engagement generates reach.³⁵³

Importantly, this amplification was not necessarily the result of explicit platform endorsement. Rather, it emerged from the underlying structure of engagement optimisation itself. Hate speech that provokes emotional reaction becomes algorithmically valuable because it sustains interaction. In this sense, anti-LGBTI+ content is not simply tolerated by platforms, it is often structurally rewarded.³⁵⁴

The Libs of TikTok account – introduced in Section 3 as an example of attention hacking – provides a further illustration here at the level of pure amplification mechanics. By reposting content from queer teachers, drag performers, and transgender

individuals to highly engaged audiences, the account generated repeated cycles of outrage and harassment. Viral posts often resulted in coordinated abuse campaigns, bomb threats against hospitals and schools, and offline intimidation targeting LGBTI+ individuals. Yet, the account repeatedly benefited from platform engagement systems because controversy itself increased visibility. In Section 3, the account demonstrated how fringe narratives enter mainstream discourse. Here, the point is that platform architecture was not an obstacle to this process; it was the engine driving it.³⁵⁵

Filter Bubbles and Fragmented Realities

If amplification determines what rises to prominence, personalisation determines what users actually encounter. Modern social media systems are deeply personalised environments structured around behavioural prediction. Platforms collect vast quantities of data including clicks, searches, viewing duration, interactions, location information, and social connections. These data points are used to construct highly individualised informational ecosystems.³⁵⁶

Eli Pariser conceptualised this phenomenon as the "filter bubble." Rather than

presenting users with a shared public sphere, algorithms curate content according to predicted preferences and behavioural histories. Two individuals searching for the same topic may encounter entirely different informational landscapes based on prior engagement patterns.³⁵⁷

This personalisation process has several consequences.

First, it narrows exposure to opposing viewpoints and alternative perspectives. Users increasingly encounter information that aligns with existing beliefs because alignment predicts higher engagement.³⁵⁸ Second, it fragments shared reality in ways that are not merely inconvenient but structurally dangerous.³⁵⁹ Consider ten people encountering the same news headline: Depending on each person's behavioural history, they may each be drawn into ten entirely different informational streams that attach entirely different contexts, framings, and emotional registers to that headline. The result is not merely disagreement but the collapse of shared factual ground – not a divergence of interpretation but the dissolution of the interpretive commons itself.

This observation is not offered to excuse harmful beliefs or actions; rather, it is intended to illustrate how the architecture

350 Amnesty International, A Thousand Cuts: Technology-Facilitated Gender-Based Violence against Poland's LGBTI Community on X (London: Amnesty International, 2025), <https://www.amnesty.org/en/documents/eur37/0098/2025/en/>.

351 HER Internet, Navigating Algorithms: The Case of Structurally Silenced Communities in Uganda (Wakiso, Uganda: HER Internet, 2024), <https://www.herinternet.org/wp-content/uploads/2024/04/HER-INTERNET-REPORT-April-4.pdf>.

352 Global Action for Trans Equality (GATE), An Engineered Hostility: The Architecture of Anti-Trans Oppression in South-West Asia and North Africa (New York: Global Action for Trans Equality, 2025), <https://gate.ngo/wp-content/uploads/2025/12/An-Engineered-Hostility.pdf>.

353 Royal United Services Institute, Online Gendered Narratives.

354 Zuboff, The Age of Surveillance Capitalism.

355 Center for Countering Digital Hate, Digital Hate.

356 OECD, Rights in the Digital Age.

357 Pariser, The Filter Bubble.

358 boyd, It's Complicated.

359 Ibid.

of personalisation structurally disadvantages human rights norms by eliminating the shared epistemic floor on which rights claims depend. When there is no agreed factual ground, the assertion that persecution is wrong, that violence is unjustified, or that a group deserves legal protection becomes not a matter of principle but one of contested narrative.³⁶⁰

Crucially, this filtering process is largely invisible. Users are rarely aware of the extent to which algorithmic systems curate their information environment.³⁶¹ As a result, personalised feeds often appear natural or objective rather than constructed. This invisibility intensifies the influence of algorithmic systems because users interpret curated realities as representative of broader social reality.

Within political contexts, filter bubbles contribute to ideological sorting and polarisation. Repeated exposure to ideologically consistent content reinforces existing beliefs while reducing opportunities for meaningful encounter with counterarguments.³⁶² Over time, users may come to perceive opposing viewpoints not merely as disagreements but as illegitimate or threatening.

This epistemic fragmentation is not without direct real-world consequences for LGBTI+ communities. When algorithmi-

cally mediated environments consistently frame LGBTI+ identities as a political threat or a form of moral contamination, users primed by sustained exposure to such narratives have, in documented cases, proceeded to real-world harassment, violence, and organised persecution. The personalisation architecture that delivers these narratives does not merely shape attitudes in the abstract; it constructs the affective and cognitive conditions from which targeted harm emerges.³⁶³

For LGBTI+ communities, personalisation dynamics produce contradictory effects. On one hand, algorithmic personalisation enables queer users to find community, identity resources, and affirming social networks that may be unavailable offline. Recommendation systems frequently facilitate pathways toward queer visibility, self-recognition, and political education. Young users questioning their identity may discover supportive communities through algorithmic recommendations that connect them to creators discussing sexuality, gender, or queer experiences.³⁶⁴

TikTok has been particularly significant in this regard. Many queer users describe the platform's recommendation algorithm as remarkably effective at identifying identity-related interests and surfacing affirming content. Users frequently joke about the 'gay TikTok algorithm' because

recommendation systems rapidly identify patterns of engagement associated with queer content.³⁶⁵

At the same time, however, these same systems frequently amplify hostile narratives because such content generates strong engagement signals. The circulation of anti-LGBTI+ discourse is, therefore, not incidental but structurally encouraged by the logic of platform optimisation.³⁶⁶

Empirical evidence shows that a relatively small number of posts can generate disproportionate reach through algorithmic amplification, particularly when they provoke moral outrage or emotional reaction. The Center for Countering Digital Hate has documented how such narratives reach millions of users through algorithmically driven distribution, often peaking after political debates or media coverage relating to education, healthcare, or public policy affecting LGBTI+ communities.³⁶⁷

The Subjectivity and Predictive Governance

If amplification determines what becomes visible at scale, moderation determines what is removed or suppressed. However, modern moderation systems do not operate as consistent rule-based frameworks. Instead, they rely increasingly on machine

learning models that classify content probabilistically based on training data derived from human-labelled examples.³⁶⁸ This introduces structural instability into the governance of speech.

Moderation datasets are built on inherently subjective judgments. Human annotators disagree on what constitutes harm, toxicity, or acceptable expression, particularly in cases involving sarcasm, reclaimed language, or culturally specific references.³⁶⁹ These disagreements are then encoded into training data, producing systems that appear objective but are in fact derived from inconsistent human interpretation.

The appropriate response to this recognition is not to remove human judgment from moderation – that would be a fundamental category error. The problem is not human involvement per se, but the absence of proper structures around that involvement. Human moderators are indispensable to the governance of speech precisely because they bring contextual understanding, cultural competence, and moral reasoning that no automated system has yet replicated. What is required is not replacement but support: clear government standards, legally binding accountability frameworks, and adequate safeguards including fair compensation, psychological support services, and

360 Klonick, "Beyond the Public Square."
361 boyd, It's Complicated.
362 Ibid.
363 Tahrir Institute for Middle East Policy, "Digital Persecution."
364 boyd, It's Complicated.

365 Ibid.
366 Center for Countering Digital Hate, Digital Hate.
367 Ibid.
368 Land, "Toward a Human Rights Approach."
369 Ibid.

transparent mechanisms through which moderation decisions can be reviewed and contested. The goal, in other words, should be human judgment operating within rule-of-law structures, not algorithmic displacement of human decision-making.³⁷⁰

Research on predictive multiplicity demonstrates that even highly accurate models can produce divergent moderation outcomes when exposed to identical inputs. Small variations in model architecture, initialisation, or training configuration can lead to different classification results for the same piece of content.³⁷¹ This means that moderation decisions are not fully determined by the content itself but also by invisible technical conditions.

Importantly, these inconsistencies are not evenly distributed. LGBTI+ content is disproportionately affected by misclassification, particularly in cases involving identity-affirming language, reclaimed slurs, or queer-coded communication styles. Because such language often deviates from dominant linguistic norms, it is more likely to be interpreted as risky or ambiguous by automated systems.³⁷²

This produces what can be described as algorithmic arbitrariness. Content that

appears identical in meaning may be removed in one instance and permitted in another, depending on system variation.³⁷³ In some cases, algorithmic systems exhibit greater disagreement than human moderators themselves, even when human consensus is clear.³⁷⁴ This undermines the assumption that automation increases consistency in content governance.

The problem of inconsistency is compounded by deliberate platform policy decisions that reveal the normative foundations of platform governance. In documented cases, misogynistic and anti-LGBTI+ content has been deemed within platform terms of service. In January 2025, Meta updated its hate speech policies to explicitly permit users to characterise LGBTI+ people as mentally ill on the basis of their sexual orientation or gender identity, removing prior prohibitions against such language.³⁷⁵ This illustrates the risk of leaving governance of speech to platforms themselves; what is classified as ‘acceptable’ is not determined by human rights standards or international norms but by platform business interests. Algorithmic consistency is, therefore, meaningless without normative accountability; a moderation system that consistently enforces policies designed to

permit dehumanisation has not solved the problem, it has institutionalised it.

The consequences extend into core human rights domains. Freedom of expression is affected because users cannot reliably predict how their speech will be interpreted.³⁷⁶ Non-discrimination is implicated because certain groups experience higher rates of misclassification. Procedural justice is weakened because decision-making processes are opaque and difficult to contest.

This opacity extends into criminal justice. Even in cases where platform content may constitute criminal conduct – including incitement, harassment, or hate crime – law enforcement agencies frequently encounter significant barriers in accessing the data needed to investigate.³⁷⁷ Platforms' opacity regarding their moderation and algorithmic processes, combined with legal jurisdictional complexity, means that where criminality is clear in principle, the evidentiary trail is often inaccessible in practice – leaving victims of serious online harm without recourse.³⁷⁸

Case Study: The “White Tiger” Investigation and the Limits of Digital Accountability ³⁷⁹

The investigation into the death of 13-year-old Jay Taylor illustrates with unusual clarity how severe online harm can remain difficult to prosecute even when the underlying abuse is evident. After Jay’s death in January 2022 was initially treated as a suicide, investigators obtained evidence showing that his death had been live-streamed on Instagram and that online users had allegedly encouraged and directed the act in real time. The case quickly expanded from a local investigation into a transnational inquiry involving Discord, Instagram, Telegram, pseudonymous accounts, deleted messages, and suspected perpetrators operating across multiple jurisdictions.

370 Gomez, Juan Felipe, Caio Vieira Machado, Lucas Monteiro Paes, and Flavio P. Calmon. Algorithmic Arbitrariness in Content Moderation. arXiv preprint arXiv:2402.16979. Submitted February 26, 2024. <https://doi.org/10.48550/arXiv.2402.16979>.

371 Ibid.

372 Land, "Toward a Human Rights Approach."

373 Gomez et al., Algorithmic Arbitrariness in Content Moderation.

374 Ibid.

375 Ibid.

376 Klonick, "Beyond the Public Square."

377 Human Rights Watch, All This Terror Because of a Photo.

378 Ibid.

379 "The White Tiger Case: How an Online Search for Friends Ended in Coerced Suicide," Der Spiegel International, October 16, 2025, <https://www.spiegel.de/international/world/the-white-tiger-case-how-an-online-search-for-friends-ended-in-coerced-suicide-a-6dafcef7-de48-43d1-a5f3-9a9b5812230b>.

What made the case especially difficult was not the absence of harm but the fragmentation of the evidentiary trail. Investigators faced practical barriers at almost every stage such as: Key communications had been deleted using platform features; relevant actors operated through fake identities and multiple accounts; and law enforcement could only compel information from companies with a legal presence in the United States. Even when US-based subpoenas yielded substantial material from Discord, Instagram, and other platforms, the suspected central perpetrator remained outside US jurisdiction, requiring cooperation from German authorities before any arrest could occur.

The case also reveals a deeper legal and institutional problem. According to the investigation, FBI agents struggled to identify the most appropriate criminal charges, encountered reluctance from prosecutors who doubted whether a viable case existed, and later found that evidence gathered in the United States was not directly admissible in German court proceedings, forcing local authorities to rebuild parts of the case from the beginning. In practice, this meant that a highly visible and devastating act of abuse still sat uneasily within legal categories designed for more conventional offences, while bureaucratic caution and jurisdictional mismatch slowed the response.

This matters not only because of the individual tragedy involved but because it demonstrates a broader structural failure in platform-era governance. Where online coercion, extortion, and psychological manipulation unfold across several services and several countries at once, accountability depends on a chain of actors – platforms, moderators, police, prosecutors, and foreign authorities – each possessing only partial visibility and partial responsibility. The result is that even in cases where criminality appears clear in substance, the architecture of digital communication can make responsibility difficult to trace and even harder to enforce, leaving victims and families exposed to profound harm without timely recourse.

Beyond direct removal, platforms also govern speech through less visible mechanisms such as shadowbanning. In these cases, content is not explicitly deleted but is instead deprioritised within recommendation systems, search results, or hashtag

visibility. Users may continue posting normally while experiencing significant reductions in reach and engagement without being informed that suppression has occurred.³⁸⁰ This creates a governance environment characterised by opacity.

380 Klonick, "Beyond the Public Square."

Users are often left to infer algorithmic intervention from indirect signals such as declining engagement, making it difficult to distinguish between audience behaviour and platform enforcement.

Despite these constraints, LGBTI+ users are not passive recipients of algorithmic governance. Instead, they actively develop adaptive strategies to navigate visibility systems. Research on queer content creators – particularly in restrictive digital environments such as China – highlights the emergence of tactical practices that operate within and around algorithmic constraints.³⁸¹

One such strategy is camouflaging, in which creators disguise queer content using coded language, humour, symbolism, or ambiguity in order to avoid detection by automated systems.³⁸² Another strategy is filtering, where creators selectively target audiences through hashtags, community signals, or engagement patterns that shape who is likely to encounter the content.³⁸³ A third strategy is verticalising, in which content is adapted to fit platform-specific formats and engagement logics.³⁸⁴ Together, these practices constitute what scholars describe as the "queering of algorithms" – a concept that refers both to the production of LGBTI+

content within algorithmic systems and to the creative subversion of dominant technological logics through adaptive, non-normative practices.³⁸⁵ Rather than directly resisting or complying with platform governance, queer creators negotiate visibility through continuous reinterpretation of algorithmic constraints.

Navigating the Architecture: Contested Uses of Algorithmic Systems

The strategies described above do not exist in a vacuum. Both LGBTI+ activists and anti-rights actors actively navigate and seek to exploit the same algorithmic structures described throughout this section, but they do so with profoundly asymmetric resources and with structurally different relationships to the logic of engagement optimisation.³⁸⁶

Anti-rights actors operate with significant structural advantages. Many campaigns targeting LGBTI+ communities are coordinated, well-resourced, and strategically literate in platform engagement mechanics. They exploit outrage deliberately, use inauthentic amplification techniques – including coordinated networks, bot activity, and manufactured trending – and craft

381 Byron, Paul, Brady Robards, Benjamin Hanckel, Son Vivienne, and Brendan Churchill. "Queer Digital Cultures: Exploring Social Media, Platforms, and LGBTIQ+ Young People's Practices." *Societies* 13, no. 12 (2023): Article 261. <https://doi.org/10.3390/soc13120261>

382 Ibid.

383 Ibid.

384 Ibid.

385 Ibid.

386 Institute for Strategic Dialogue, "Offline Events, Online Threats."

anti-LGBTI+ content in ways designed to evade automated moderation while generating maximum algorithmic reach.³⁸⁷ Framing is central to this strategy; hostility is packaged as child protection, religious freedom, or national sovereignty, categories that are both politically resonant and less likely to trigger automated moderation systems than overt hate speech.

The Fetrah campaign, documented by GATE, exemplifies this strategic sophistication; a coordinated operation spanning multiple countries, languages, and platforms, calibrated to exploit regional outrage about "Western cultural invasion" while maintaining plausible claims to moral legitimacy.³⁸⁸ Such campaigns are also temporally strategic; they frequently intensify around legislative moments – debates about education policy, healthcare access, or anti-discrimination legislation – to maximise the influence of algorithmically amplified hostility on real-world political outcomes.

LGBTI+ activists operate within the same systems but under conditions of structural disadvantage. The adaptive strategies they employ – camouflaging, algospeak, verticalising, community-building through coded language – are creative and often effective, but they are fundamentally reactive. They are responses to a system not designed for them: a system whose

default outputs tend toward the amplification of the content most hostile to their safety and recognition. Where anti-rights actors benefit from engagement algorithms by default – because outrage about LGBTI+ identities reliably produces the emotional intensity that algorithms reward – LGBTI+ activists must work against that default to achieve equivalent visibility. The labour this requires is not merely strategic but psychological; the effort of constant adaptation within a hostile architecture carries costs that are rarely acknowledged in policy discourse about platform governance.³⁸⁹

This asymmetry is not incidental. It reflects a deeper alignment between outrage-based platform logic and the emotional intensity that anti-LGBTI+ content reliably produces. The architecture of engagement optimisation is, in this sense, not neutral between the two sides; it functions as a structural subsidy for those whose communicative aims are best served by generating fear, anger, and moral indignation.³⁹⁰ This is not an argument that platforms are intentionally anti-LGBTI+; it is an argument that platforms are structurally indifferent to whom they harm, and that this indifference has differential consequences.

The asymmetry becomes even more acute when the actors involved are not mere-

ly organised campaigners but state or state-backed entities with the resources to manufacture reality at scale. The section that follows examines precisely this dimension: how the same algorithmic architectures described here are harnessed by actors operating with state capacity, political authority, or geopolitical ambition to produce not merely the amplification of hostility but the systematic fabrication of informational environments in which that hostility appears normal, justified, and inevitable.

Conclusion

For LGBTI+ activists, the communicative environment produced by these algorithmic structures is one characterised by fundamental instability. Identity expression is simultaneously amplified as a site of attention and suppressed through misclassification or invisibilisation. Hostile narratives are frequently elevated through engagement-driven systems, while identity-affirming speech is subjected to heightened scrutiny. The result is not merely an uneven playing field but a structurally adversarial one – an architecture that rewards hostility toward queer identities while penalising the expression of those identities themselves.

Across amplification, moderation, and personalisation systems, a consistent pattern emerges; algorithmic infrastructures do not simply distribute speech but actively structure the conditions under which speech becomes visible, intelligible, and permissible. Amplification systems reward emotional intensity; moderation systems introduce arbitrariness and bias; personalisation systems fragment shared informational reality.

Ultimately, these dynamics reveal that algorithmic systems function as infrastructural regulators of public discourse. Free speech in digital environments is no longer governed solely by legal frameworks or platform policies but by opaque systems of ranking, prediction, and visibility control. Understanding these structures requires moving beyond individual instances of moderation or bias to examine the broader architecture through which visibility itself is produced and distributed.

If algorithms shape what can be said and seen, the section that follows asks what happens when those same architectures are harnessed not merely by platform logic but by deliberate actors seeking to manufacture reality itself.

387 Ibid.

388 GATE, *An Engineered Hostility*.

389 Byron et al., "Queer Digital Cultures."

390 Ibid.

Section 4 Key Findings:

- Algorithmic ranking, recommendation, and moderation systems now perform the legitimising function once held by editors and civic institutions, but they do so via engagement signals rather than democratic or epistemic criteria.
- LGBTI+ content is disproportionately misclassified as sensitive/adult/explicit, while overtly hostile anti-LGBTI+ content is routinely classified as 'political speech' or 'religious expression' and remains active.
- Moderation resourcing is geographically and linguistically uneven.
- Engagement optimisation structurally rewards outrage about LGBTI+ identities.
- Governance sits with a small number of unaccountable individuals.
- LGBTI+ users actively adapt but from a position of structural disadvantage. Camouflaging, algospeak, and verticalising are effective creative strategies, but they are reactive labour imposed on communities navigating a system whose defaults favour those generating fear and outrage.

Section 4 Questions:

- Has your organisation or community experienced content being removed, downranked, or 'shadowbanned' for identity-affirming speech, while comparable hostile content targeting LGBTI+ people remained active? Can you provide specific examples (platform, content type, outcome)?
- Have you experienced coordinated reporting/flagging campaigns used to suppress LGBTI+ content or accounts? What was the platform's response, and how effective – or ineffective – was it?
- Are you aware of cases in your context where algorithmically amplified anti-LGBTI+ narratives influenced public debate, policy, or legislation?
- What strategies (coded language, hashtag practices, format changes) has your community developed to maintain visibility within algorithmic constraints? What has this cost you in terms of time, reach, or psychological burden?
- Where you or your organisation have experienced online harm that may meet a criminal threshold (harassment, incitement, coordinated abuse), what barriers did you face in reporting, evidence-gathering, or securing law enforcement action – particularly across jurisdictions?



Section 5: AI-Generated Propaganda and Emotional Manipulation

Sections 3 and 4 described platforms as systems that structurally reward emotional intensity and that govern speech through opaque algorithmic processes. This section examines what happens when those same dynamics are supercharged by generative artificial intelligence. When the production of manipulative content is no longer constrained by human effort, technical skill, or even the need for a real event to have occurred.

Generative artificial intelligence has introduced a structural shift in the production, circulation, and epistemic consequences of disinformation. Whereas earlier forms of digital propaganda relied on human fabrication, coordinated troll networks, or basic automation, contemporary frontier

AI systems now enable the large-scale generation of synthetic text, audio, image, and video content that is increasingly indistinguishable from authentic media.

Deepfakes, AI-generated narratives, and fabricated quotations now circulate across digital platforms at speeds that exceed the capacity of verification infrastructures and fact-checking organisations. This is not a marginal evolution in propaganda techniques but a shift in the architecture of influence itself.

Recent research suggests that frontier AI systems do not simply accelerate misinformation but fundamentally reconfigure its architecture, turning it into a system-level phenomenon rather than a

collection of isolated deceptive events. Misinformation becomes embedded across platforms, institutions, and cognitive processes simultaneously, producing a layered threat to informational integrity and public trust.

This transformation is, therefore, not merely quantitative but qualitative. It alters informational ecosystems by lowering the cost of content production while increasing realism, adaptability, and scalability. In this environment, LGBTI+ individuals, activists, and civil society organisations are disproportionately exposed to synthetic targeting, including fabricated statements, manipulated imagery, and coordinated narrative attacks designed to delegitimise advocacy work.

Synthetic targeting is increasingly embedded within broader anti-rights ecosystems that actively seek to erode LGBTI+ rights through information manipulation as much as through formal political processes. These actors benefit directly from generative AI systems that reduce the cost, skill threshold, and coordination required to produce convincing disinformation at scale. As a result, AI-enabled propaganda becomes a force multiplier for existing campaigns of delegitimation. For LGBTI+ activists and civil society organisations, recognising these dynamics is es-

sential: not only to respond to individual instances of synthetic harm but to understand them as part of a structured – and evolving – strategy of political opposition within digital environments.

Synthetic Propaganda Systems

A clear conceptual foundation is necessary to understand these mechanisms. Deepfakes refer to synthetic media – including video, audio, and images – generated using machine learning techniques such as generative adversarial networks (GANs) and diffusion models.³⁹¹ These systems can produce outputs that are increasingly difficult for both humans and automated systems to distinguish from authentic recordings.³⁹²

This capability is already being operationalised in real-world manipulation and fraud. In 2024, a widely reported case involved a multinational engineering firm in which employees were deceived by a video-conference deepfake of senior executives, leading to the transfer of approximately \$25 million in fraudulent payments.³⁹³ The incident illustrated how synthetic media is no longer confined to political manipulation or social media virality but is actively penetrating organisational trust infrastructures such as finance and corporate governance.³⁹⁴

Similarly, AI voice-cloning has become a recurrent tool in impersonation scams targeting individuals and organisations. In multiple documented cases, cloned voices of family members or executives have been used to request urgent financial transfers, exploiting emotional urgency and institutional hierarchy.³⁹⁵ These incidents demonstrate that synthetic media operates most effectively not when it is perfect but when it is plausible enough to override hesitation under time pressure.³⁹⁶

In the political domain, deepfakes and AI-generated narratives have been deployed during election cycles to fabricate endorsements, simulate candidate speeches, or distort public statements.³⁹⁷ For example, during recent electoral contests in Europe and South Asia, manipulated audio clips and AI-generated images circulated on social platforms shortly before voting periods, often reaching large audiences before verification mechanisms could intervene.³⁹⁸ Even when later debunked, such content frequently persisted in reshared networks, illustrating the durability of first exposure effects.³⁹⁹

Within LGBTI+ contexts, synthetic media has also been used to fabricate activist statements or create manipulated imagery intended to provoke backlash and delegitimise advocacy organisations.⁴⁰⁰ While some of these cases are documented primarily through civil society monitoring rather than formal legal records, they reflect a broader pattern in which marginalised groups are disproportionately exposed to reputationally damaging synthetic content due to the low cost and high scalability of generation.⁴⁰¹

Here lies the critical distinction between misinformation and disinformation; misinformation refers to false or inaccurate content shared without intent to deceive, whereas disinformation is deliberately constructed and strategically deployed to manipulate perception or behaviour.⁴⁰² In practice, AI systems blur this distinction further by enabling rapid scaling of both accidental and strategic falsehoods.⁴⁰³

The broader ecosystem is shaped by frontier AI systems, including large language models and multimodal architectures capable of producing coherent outputs

391 Shoaib et. al. "Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models". Cornell University. 2023. <https://arxiv.org/abs/2311.17394>

392 Gomez et al., Algorithmic Arbitrariness in Content Moderation.

393 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

394 Ibid.

395 Ibid.

396 Ibid.

397 Marwick and Lewis, Media Manipulation and Disinformation Online.

398 Ibid.

399 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

400 Human Rights Watch, All This Terror Because of a Photo.

401 Tahrir Institute for Middle East Policy, "Digital Persecution."

402 Marwick and Lewis, Media Manipulation and Disinformation Online.

403 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

across text, visual, and audio domains.⁴⁰⁴ These systems are characterised by low marginal cost of generation and high-volume output capacity, enabling industrial-scale content production.⁴⁰⁵ Together, these technologies redefine how truth, authenticity, and credibility are assessed in digital environments.⁴⁰⁶

Enabling Elements of AI-Generated Propaganda

The rise of AI-generated propaganda has been driven by three interrelated forces. The first is the growth of generative capability itself, embodied in diffusion models, GANs, and large language models.⁴⁰⁷ These systems have only recently been able to produce highly realistic synthetic outputs ranging from photorealistic imagery to persuasive linguistic constructs at low user cost.⁴⁰⁸ However, this capability has already been exploited in digital advertising ecosystems where AI-generated endorsements falsely attributed to celebrities circulate across platforms, promoting financial scams or political messaging without any real association with the indi-

viduals depicted. Crucially, this is not only a technical affordance but an economic one, where the extremely low cost of generating and distributing synthetic content combined with its high potential return in attention, engagement, or financial fraud creates a strongly incentivised environment in which AI-generated deception becomes both easy to execute and economically rewarding at scale.⁴⁰⁹

The second force also relates to the capability of AI but refers to its multimodality, which integrates text, image, audio, and video generation into unified systems.⁴¹⁰ This enables the construction of coherent synthetic narratives across multiple media forms.⁴¹¹ A particularly significant manifestation is the emergence of fully synthetic events in which fabricated video footage is combined with fake news articles, simulated eyewitness accounts, and social media commentary to construct the appearance of real-world occurrences.⁴¹²

Audiences are not evaluating individual claims but immersive simulations of reality that operate across multiple sensory and informational channels.⁴¹³ Within LGBTI+

contexts, this can manifest as fabricated ‘evidence’ of activist misconduct, synthetic videos depicting Pride events as sites of violence or criminality, or coordinated sets of fake posts and commentary attributing extremist or ‘immoral’ behaviour to LGBTI+ individuals or organisations.⁴¹⁴ These multimodal assemblages are particularly effective in hostile information environments because they mimic the structure of legitimate news coverage, making delegitimation efforts appear evident rather than constructed.⁴¹⁵

The third enabling force is accessibility. The rapid diffusion of user-friendly generative tools has drastically lowered barriers to entry, making it possible for non-expert users to generate highly realistic synthetic content.⁴¹⁶ This democratisation of capability has expanded legitimate creative uses while simultaneously enabling malicious applications, including the rapid production of targeted harassment content.⁴¹⁷ Within LGBTI+ contexts, this has included the creation of synthetic derogatory imagery and fabricated activist statements designed to delegitimise advocacy work or provoke public backlash, often at very low cost and high speed.⁴¹⁸ These

tactics typically take the form of AI-generated images depicting LGBTI+ individuals or events in contexts of criminality, sexual deviance, or political extremism as well as fabricated screenshots of supposed social media posts or interviews in which activists are made to appear to endorse controversial or inflammatory positions.⁴¹⁹ In other instances, synthetic ‘evidence packs’ are assembled – combining manipulated images, fake quotes, and fabricated news-style framing – to simulate journalistic credibility and increase shareability. Civil society monitoring has documented recurring patterns in which such content emerges rapidly around moments of heightened visibility, such as Pride events, legal reforms, or public policy debates, suggesting a strategic alignment with real-world advocacy milestones rather than random content generation.⁴²⁰

These forces combine into what can be understood as ‘generative propaganda.’⁴²¹ Generative propaganda refers to AI-driven systems in which narrative construction, audience targeting, and behavioural feedback loops are continuously optimised.⁴²² These systems are capable of adjusting messaging in real time based

404 Ibid.
405 Ibid.
406 Human Rights Watch, All This Terror Because of a Photo.
407 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."
408 Gomez et al., Algorithmic Arbitrariness in Content Moderation.
409 Ibid.
410 Ibid.
411 Ibid.
412 Marwick and Lewis, Media Manipulation and Disinformation Online.
413 Ibid.

414 Human Rights Watch, All This Terror Because of a Photo.
415 Institute for Strategic Dialogue, "Offline Events, Online Threats."
416 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."
417 Ibid.
418 Human Rights Watch, All This Terror Because of a Photo.
419 Ibid.
420 Institute for Strategic Dialogue, "Offline Events, Online Threats."
421 Marwick and Lewis, Media Manipulation and Disinformation Online.
422 Ibid.

on engagement signals, producing evolving narrative structures rather than fixed messages.⁴²³ Emerging evidence of synthetic persona networks illustrates this shift; these networks consist of persistent AI-generated identities that engage users over time, respond to events, and simulate organic political discourse across platforms.⁴²⁴ This marks a transition from broadcast propaganda to adaptive influence systems, where persuasion operates as a continuous computational process rather than discrete communicative acts.⁴²⁵

The AI-Enhanced Misinformation Pipeline

These pillars enable AI to not only generate content but to enhance every stage of the misinformation lifecycle.⁴²⁶ At the point of generation, AI systems produce synthetic narratives, fabricated quotations, and realistic media artefacts at scale.⁴²⁷ At the amplification stage algorithmic recommendation systems and coordinated inauthentic networks accelerate visibility, allowing emotionally charged synthet-

ic content to spread more rapidly than corrective information.⁴²⁸ At the personalisation stage, content is increasingly tailored to inferred psychological or demographic profiles, increasing its persuasive effectiveness.⁴²⁹ Finally, at the stage of entrenchment, repeated exposure within algorithmically reinforced echo chambers consolidates belief systems and reduces exposure to corrective signals.⁴³⁰

This produces a continuously operating system of influence rather than a linear information flow.⁴³¹ In several documented cases, false statements attributed to civil society leaders have circulated widely before any correction is issued, demonstrating how rapidly synthetic narratives can enter public discourse. Similarly, coordinated amplification of polarising content during policy debates affecting LGBTI+ rights has been observed across multiple platforms, where engagement-driven systems elevate emotionally charged material regardless of its veracity.⁴³² However, a further and increasingly significant dynamic is emerging: The problem is no longer solely that users cannot distinguish between authentic and

synthetic content but that in many cases they actively engage with or circulate content despite knowing it may be false or unverifiable.⁴³³ Content is shared because it ‘feels plausible,’ aligns with prior beliefs, or is perceived as symbolically or emotionally ‘true,’ even in the absence of factual grounding.⁴³⁴ This reflects a shift from epistemic uncertainty to motivated acceptance, where plausibility, affect, and narrative coherence become more influential than factual verification.⁴³⁵

This is a significant conceptual shift in the nature of propaganda itself. Traditional propaganda operated as a relatively static, top-down process in which messages were constructed centrally and disseminated broadly to passive audiences.⁴³⁶ In contrast, AI-enabled propaganda systems are dynamic, iterative, and adaptive.⁴³⁷

Risks of AI-Generated Propaganda

The AI-enhanced misinformation pipeline becomes substantially more destabilising when situated within the broader struc-

tural risks of generative AI. Here, systemic asymmetries between generation, amplification, and detection increase both the speed and persistence of synthetic influence.⁴³⁸

The risks associated with AI-generated propaganda extend across multiple domains, including political stability, psychological manipulation, cybersecurity, and epistemic integrity.⁴³⁹ These risks are not isolated effects but emerge from the interaction of generative capability, platform dynamics, and structural asymmetries within information ecosystems.⁴⁴⁰

In political contexts, synthetic media can simulate political speeches, fabricate endorsements, or impersonate public officials, thereby undermining electoral integrity and distorting democratic communication.⁴⁴¹ In social contexts, LGBTI+ individuals and organisations are frequently targeted by synthetic delegitimation campaigns that include fabricated quotations, manipulated imagery, and coordinated narrative attacks designed to erode credibility and public trust.⁴⁴²

423 Zuboff, The Age of Surveillance Capitalism.

424 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

425 Manuel, Networks of Outrage and Hope: Social Movements in the Internet Age. 2nd ed. Cambridge: Polity Press, 2015.

426 Marwick and Lewis, Media Manipulation and Disinformation Online.

427 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

428 Marwick and Lewis, Media Manipulation and Disinformation Online.

429 Zuboff, The Age of Surveillance Capitalism.

430 Pariser, The Filter Bubble.

431 Zuboff, The Age of Surveillance Capitalism.

432 Center for Countering Digital Hate, Digital Hate.

433 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

434 Ibid.

435 Pariser, The Filter Bubble.

436 Manuel, Networks of Outrage and Hope.

437 Zuboff, The Age of Surveillance Capitalism.

438 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

439 Human Rights Watch, All This Terror Because of a Photo.

440 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

441 Marwick and Lewis, Media Manipulation and Disinformation Online.

442 Human Rights Watch, All This Terror Because of a Photo.

Beyond these domain-specific harms, a more structural risk emerges from epistemic and temporal asymmetry between generation and detection in AI-mediated information environments.⁴⁴³ Generative systems can produce highly realistic synthetic text, audio, and visual content at extremely low cost and near-instant speed, while detection and verification mechanisms remain comparatively slow, resource-intensive, and probabilistic in nature.⁴⁴⁴

This produces a persistent verification lag window, in which synthetic or misleading content can circulate widely and shape initial perception before any meaningful correction is possible.⁴⁴⁵ Within fast-moving information environments, this temporal gap is often sufficient for narrative framing to stabilise before verification processes can intervene.⁴⁴⁶

This asymmetry can be understood through four interconnected dynamics.

First, generation is scalable, cheap, and near-instantaneous, enabling malicious actors to produce large volumes of convincing synthetic content with minimal

expertise or cost.⁴⁴⁷ Second, distribution is algorithmically accelerated, as platform recommendation systems and coordinated networks amplify emotionally salient content regardless of its veracity, privileging engagement over accuracy.⁴⁴⁸ Third, detection is forensic, contextual, and probabilistic, requiring cross-platform verification, metadata analysis, and interpretive judgment – all of which inherently lag behind real-time circulation.⁴⁴⁹ Fourth, correction is structurally weaker than exposure, as debunking and fact-checking efforts typically reach smaller audiences and do not fully reverse the cognitive or emotional imprint of first exposure.⁴⁵⁰

Together, these dynamics produce a systemic advantage for AI-enabled propaganda systems where synthetic narratives can be generated and operationalised faster than they can be reliably verified or corrected. As a result, even transient misinformation can produce durable effects on perception, belief formation, and political interpretation.⁴⁵¹

Within this broader environment, the consequences of deepfakes extend beyond informational distortion and man-

ifest across cognitive, behavioural, and social domains.⁴⁵² Cognitively, exposure to synthetic media can distort perception, memory formation, and credibility assessment, including the production of false memory effects and altered judgments even in contexts where individuals are aware of manipulation.⁴⁵³ Behaviourally, users are more likely to circulate content that is emotionally resonant or narratively coherent, even when its authenticity is uncertain.⁴⁵⁴ Socially, repeated exposure contributes to an erosion of trust in journalism, civil society, and institutional communication.⁴⁵⁵

A particularly significant phenomenon is the liar’s dividend, in which the mere existence of synthetic media enables bad-faith actors to dismiss authentic evidence as potentially fabricated.⁴⁵⁶ This compounds accountability challenges in politically contested environments and increases the evidentiary burden required to establish credibility.⁴⁵⁷

Taken together, these dynamics position AI-generated propaganda not as a set of isolated risks but as a systemic transfor-

mation of informational environments in which speed, scale, and plausibility consistently outpace verification, correction, and epistemic stability.⁴⁵⁸

The Dopamine Economy of Disinformation

The persistence of disinformation within AI-mediated environments cannot be explained solely through questions of belief or deception. Instead, it is shaped by a set of reinforcing behavioural and platform-level dynamics that operate across three interrelated layers: informational filtering, belief formation, and engagement-based reinforcement.⁴⁵⁹

The first layer concerns filtering and selective exposure. Platform architectures, recommendation systems, and social networks increasingly shape what information users encounter in the first place.⁴⁶⁰ Over time, this produces environments in which individuals are more frequently exposed to content that aligns with prior beliefs, emotional dispositions, or group identities.⁴⁶¹ These filter dynamics do not

443 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

444 Gomez et al., Algorithmic Arbitrariness in Content Moderation.

445 Marwick and Lewis, Media Manipulation and Disinformation Online.

446 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

447 Gomez et al., Algorithmic Arbitrariness in Content Moderation.

448 Center for Countering Digital Hate, Digital Hate.

449 Gomez et al., Algorithmic Arbitrariness in Content Moderation.

450 Turkle, Alone Together.

451 Ibid.

452 Gomez et al., Algorithmic Arbitrariness in Content Moderation.

453 Turkle, Alone Together.

454 Manuel, Networks of Outrage and Hope.

455 Human Rights Watch, All This Terror Because of a Photo.

456 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

457 Ibid.

458 Human Rights Watch, All This Terror Because of a Photo.

459 Pariser, The Filter Bubble.

460 Ibid.

461 boyd, It's Complicated.

eliminate alternative information entirely, but they reduce its visibility and increase the relative prominence of congruent narratives.⁴⁶²

The second layer involves belief formation and emotional cognition. In these environments, belief is not formed purely through factual evaluation but through repeated exposure, narrative coherence, and emotional resonance.⁴⁶³ In psychological terms, AI systems enable highly personalised emotional manipulation, where synthetic narratives are designed to trigger fear, anger, or moral outrage.⁴⁶⁴ These emotional signals often become more influential than evidentiary strength, particularly when content is framed in ways that reinforce existing identity positions or perceived social threats.⁴⁶⁵ Importantly, this does not imply irrationality on the part of users but rather reflects how cognition is shaped under conditions of high information density and continuous emotional stimulation.⁴⁶⁶

The third layer is what can be described as a dopamine-driven reinforcement economy.⁴⁶⁷ Engagement-driven platform infrastructures systematically reward content that generates interaction, regardless of its accuracy.⁴⁶⁸ Disinformation, therefore, persists not only because it is believed but because it is structurally optimised for visibility, sharing, and emotional response.⁴⁶⁹ Within this economy, users may repeatedly encounter and engage with content that is known to be questionable or false, not because of acceptance of its factual claims but because of its narrative intensity, social signalling value, or emotional salience.⁴⁷⁰ This creates a reinforcing loop in which attention itself becomes a form of reward, making disengagement increasingly difficult even when users are aware of manipulation dynamics.⁴⁷¹

Within this framework, it is important to avoid interpreting engagement as endorsement or to frame individuals as passive victims of manipulation.⁴⁷² Rath-

er, these dynamics reflect the interaction between human cognitive biases and platform systems designed to maximise interaction.⁴⁷³ The result is a structural environment in which disinformation is continuously reinforced through behavioural feedback loops.⁴⁷⁴

This dynamic is particularly visible in cybersecurity contexts where emotional and cognitive triggers are directly exploited.⁴⁷⁵ Deepfake voice cloning has been used in social engineering attacks, including cases where executives' voices are replicated to request urgent financial transfers under false pretences.⁴⁷⁶ These incidents demonstrate how emotional urgency, authority cues, and perceived familiarity can override deliberative verification processes, especially under time pressure.⁴⁷⁷

Taken together, these three layers – filtering, belief formation, and reinforcement – help explain why disinformation persists even in contexts where users are aware of its potential inaccuracy.⁴⁷⁸ The challenge is not simply one of correcting falsehoods but of interrupting feedback systems in which attention, emotion, and repetition continuously reinforce the visibility – and perceived plausibility – of synthetic narratives.⁴⁷⁹

Conclusion

A broader consequence of AI-generated propaganda is the gradual erosion of epistemic stability. As synthetic and authentic media become increasingly indistinguishable, the capacity to reliably differentiate truth from fabrication is undermined. This leads to declining trust in journalism, civil society organisations, and advocacy institutions – particularly in politically contested domains.

Over time, this contributes to a condition of epistemic collapse in which shared frameworks of knowledge become unstable. A further intensifying mechanism is the liar's dividend; for LGBTI+ advocacy organisations, this creates a particularly severe burden, as the legitimacy of lived experience itself becomes subject to continuous contestation.

Within this environment, it is important to distinguish between different categories of actors and their relationship to generative AI systems. Anti-rights actors are increasingly documented as active adopters of AI-enabled propaganda techniques, using synthetic media, coordinated narrative production, and automated amplification to delegitimise LGBTI+ rights and civil society organisations. These practices

462 Ibid.

463 Turkle, *Alone Together*.

464 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

465 Manuel, *Networks of Outrage and Hope*.

466 boyd, *It's Complicated*.

467 Zuboff, *The Age of Surveillance Capitalism*.

468 Ibid.

469 Marwick and Lewis, *Media Manipulation and Disinformation Online*.

470 Manuel, *Networks of Outrage and Hope*.

471 Zuboff, *The Age of Surveillance Capitalism*.

472 boyd, *It's Complicated*.

473 Zuboff, *The Age of Surveillance Capitalism*.

474 Ibid.

475 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

476 Ibid.

477 Turkle, *Alone Together*.

478 Pariser, *The Filter Bubble*.

479 Zuboff, *The Age of Surveillance Capitalism*.

are often embedded within broader strategic efforts to shape public discourse, influence policy environments, and erode trust in rights-based institutions.

By contrast, human rights actors and civil society organisations have – thus far – largely not adopted these tools at comparable scale for manipulative or adversarial purposes. Where AI is used, it is more often for documentation, communication, translation, or advocacy support rather than synthetic persuasion or narrative fabrication. This asymmetry may be ethically desirable, as it reflects commitments to transparency, accountability, and non-deceptive communication. However, it also raises important strategic questions about preparedness in environments where other actors are not constrained by the same norms.

This is not a call for escalation or imitation nor an argument that civil society should

adopt the same manipulative techniques. Rather, it highlights the need for a clear-eyed understanding of the evolving informational environment. The key challenge is not whether human rights actors should engage in synthetic propaganda but how they can maintain credibility, visibility, and effectiveness in contexts where adversarial actors are increasingly leveraging these capabilities at scale.

The convergence of generative models, multimodal systems, and global digital platforms thus marks a shift toward industrialised manipulation. Misinformation becomes continuous, propaganda becomes adaptive, and influence becomes systemic. Addressing this transformation requires not only improved detection technologies but a reconceptualisation of propaganda as an adaptive computational system operating across technological, cognitive, and institutional layers simultaneously.

Section 5 Key Findings:

- Generative AI represents a qualitative shift, not just a faster version of old propaganda because it lowers the cost of content production while increasing realism, adaptability, and scale.
- LGBTI+ activists and organisations are disproportionately targeted by synthetic content. This includes fabricated statements, manipulated imagery, and coordinated narrative attacks designed specifically to delegitimise advocacy work and diverse gender identities.
- ‘Fully synthetic events’ can mimic legitimate news structures. Fabricated video, fake articles, simulated eyewitness accounts, and coordinated social commentary are combined to make delegitimation appear self-evident rather than constructed, making it far harder to counter than a single false claim.

- There is a structural, four-part asymmetry that favours propagandists over fact-checkers: Generation is cheap and instantaneous; distribution is algorithmically accelerated regardless of accuracy; detection is slow, forensic, and probabilistic; and correction reaches smaller audiences than the original exposure and rarely fully reverses its emotional imprint.
- The liar’s dividend compounds the harm. Because synthetic media exists at all, bad-faith actors can dismiss authentic evidence as fabricated, which places an additional burden on LGBTI+ individuals and organisations whose lived experience becomes subject to continuous contestation.
- Platforms create a ‘dopamine economy of disinformation.’ Engagement-driven infrastructure rewards interaction regardless of accuracy, producing reinforcing feedback loops (filtering → emotional belief formation → engagement reward) that make disinformation persist even among users who are aware it may be false.

Section 5 Questions:

- Has your organisation or community experienced fabricated statements, deep-fake imagery, or synthetic ‘evidence’ targeting LGBTI+ individuals or advocacy work? If so, what form did it take, and how quickly – if at all – were you able to have it corrected or removed?
- Have you or your organisation had authentic statements, footage, or testimony dismissed by opponents as ‘AI-generated’ or fabricated? How did this affect your credibility or advocacy work, and what response strategies – if any – proved effective?
- Does your organisation currently have the technical capacity, training, or resources to identify synthetic media and respond to it?
- What responsibility should platforms bear for the amplification of AI-generated disinformation targeting LGBTI+ communities, and what concrete platform-level interventions (labelling, detection tools, faster takedown processes) would you prioritise?
- Should human rights and LGBTI+ advocacy organisations begin using generative AI defensively (e.g., synthetic content detection and rapid-response verification tools) even if they continue to avoid manipulative uses? What ethical boundaries should guide this?
- In your jurisdiction, are there adequate legal mechanisms to address AI-generated defamation, harassment, or impersonation targeting LGBTI+ individuals or organisations? What reforms are most urgently needed?

III. TOWARDS A TYPOLOGY OF ANTI-RIGHTS DRIVERS IN ONLINE SPACES

The preceding sections have treated digital space as both lifeline and infrastructure. Section 1 showed how online environments function as essential conditions for LGBTI+ survival, identity formation, and political participation, even as they generate new forms of exposure and risk. Section 2 moved beneath the surface, tracing the architecture of cyberspace, the governance of platforms as privately owned civic arenas, the algorithmic systems that sort and amplify speech, and the emergence of generative AI as an engine of industrialised manipulation. Together, these sections established that the digital environment is not a neutral backdrop but a socio-technical terrain in which visibility, vulnerability, and power are continuously produced.

This section turns from architecture to inhabitants. It asks – in effect – who lives in this environment, what they are doing there, and why. Rather than treating anti-LGBTI+ harm as a series of isolated incidents, it approaches it as a kind of digital safari; a tour through interconnected ‘exhibits’ in which different actors, incentives, and infrastructures combine to produce hostility at scale. The aim is not voyeurism but classification. By mapping the main ‘species’ of anti-rights actors and

the ecosystems they occupy, the section develops a typology of the forces driving anti-LGBTI+ harm in digital environments.

Across the sections that follow, two broad categories emerge. The first consists of individual and networked actors operating inside what might be called the political economy of hate such as influencers, far-right communities, conspiratorial networks, predatory forums, and loosely organised harassment clusters that learn to weaponise the same engagement-optimised platforms that LGBTI+ communities depend on for survival. These actors exploit the logics described in previous sections – outrage-driven amplification, algorithmic recommendation, data extraction, and AI-enabled content production – to transform hostility into attention, attention into political leverage, and sometimes attention directly into revenue.

The second category comprises state, state-aligned, and transnational actors for whom anti-LGBTI+ narratives become tools of information warfare, regime consolidation, and geopolitical strategy. Here, hostility toward LGBTI+ people is less a spontaneous expression of prejudice than a deliberate instrument of governance; a way of policing dissent, reasserting ‘tradi-

tional values,’ externalising social anxieties, and framing human rights advocacy itself as a foreign ideological intrusion. Anti-LGBTI+ campaigns in this register are not only cultural conflicts; they are components of broader authoritarian and anti-democratic projects in which digital infrastructures play a central role.

The sections that follow are structured as a series of exhibits within this broader landscape. Section 6 turns first to “Targeting Vulnerable Communities,” introducing anti-LGBTI+ harm as a distributed ecosystem and examining how the political economy of digital platforms rewards outrage, moral panic, and identity-based hostility. Section 7 focuses on far-right online ecosystems as a key subsystem within that broader environment, tracing how transnational hate networks, influencer-driven extremism, and intersecting misogynist and racist ideologies provide ideological coherence and cross-platform infrastructure for anti-LGBTI+ mobilisation. Section 8 descends further into the shadow networks that sit beneath the visible surface such as encrypted channels, illicit data

markets, and dark-web infrastructures where coordination is hidden, personal data becomes a tradable commodity, and accountability is deliberately evaded. Finally, section 9 turns to predatory communities in which gendered domination and humiliation are performed as social capital, examining how these spaces both overlap with – and extend beyond – explicitly anti-LGBTI+ mobilisation.

This typology does not claim to be exhaustive. Rather, it offers a structured way of understanding how different actors inhabit, exploit, and reinforce the same digital terrain. Across all four sections, a common argument recurs: Anti-LGBTI+ digital harm is not the product of one ideology, platform, or institution alone. It is the outcome of interactions between ideological extremism, platform incentive structures, cultural backlash, political opportunism, and algorithmic amplification, operating across both non-state and state-driven contexts. By mapping these interactions, the chapter shows how LGBTI+ communities become not only direct targets but also barometers of wider democratic vulnerability in digital society.



Section 6:

The Political Economy of Hate

Rather than treating anti-LGBTI+ harm as a collection of isolated extremist incidents, individual prejudices, or sporadic harassment, this section maps a broader ecosystem in which hostility toward LGBTI+ communities has become structurally embedded.⁴⁸⁰ Within this ecosystem, LGBTI+ people are not simply collateral targets of generic online toxicity. They are positioned as strategic objects within a political economy of hate in which identity-based hostility is transformed into attention, and attention turns into social, political, or financial capital.⁴⁸¹ Anti-LGB-

TI+ narratives circulate through a distributed, self-reinforcing system in which ideology, platform incentives, political opportunism, and algorithmic amplification continuously interact.⁴⁸² Rather than emerging from a single organisation or coherent doctrine, harm is produced through the overlap of multiple drivers, including far-right extremism, religious conservatism, anti-feminist backlash, conspiratorial worldviews, and opportunistic content entrepreneurship.⁴⁸³ Together, these dynamics create an environment in which narratives, actors, and technologies

reciprocally intensify one another across both digital and physical spaces.⁴⁸⁴

Crucially, this ecosystem is not centrally coordinated in a traditional sense. It operates through what scholars of digital extremism describe as “networked amplification,” which refers to how repeated narratives circulate across loosely connected communities until they acquire coherence and legitimacy through repetition itself.⁴⁸⁵ Far-right actors, religious conservative groups, anti-feminist networks, conspiracy communities, and opportunistic influencers frequently converge around shared frames – ‘gender ideology,’ ‘grooming,’ ‘parental rights,’ ‘protecting children’ – despite divergent political origins or strategic objectives.⁴⁸⁶ These frames travel easily between fringe forums and mainstream platforms, digital subcultures and parliamentary debates, local controversies and transnational campaigns.⁴⁸⁷

The architecture of this ecosystem relies heavily on cross-platform circulation. Content often originates in fringe environments such as 4chan, Kiwi Farms, or anonymous image boards, where conspiratori-

al narratives about LGBTI+ people, ‘elites,’ and figures such as Jeffrey Epstein are combined into broader storylines about corruption, child endangerment, and civilisational decline.⁴⁸⁸ From there, narratives migrate to larger platforms – including X/ Twitter, TikTok, YouTube, Facebook, and Instagram – where they are reframed, stylised, and amplified by influencers and politically aligned media ecosystems.⁴⁸⁹ Once recirculated at scale, claims that began as fringe conspiracy or extremist propaganda are increasingly presented as legitimate public controversies rather than as disinformation, enabling marginal narratives to acquire mainstream visibility without formal organisational coordination.⁴⁹⁰

A prominent illustration of this dynamic emerged during the harassment campaign targeting Boston Children’s Hospital in 2022.⁴⁹¹ Following a series of misleading posts circulated by Libs of TikTok – introduced in earlier sections as an example of attention hacking and algorithmic amplification – and amplified by far-right commentators, the hospital became the focus of viral accusations that it was performing hysterectomies on young

480 Human Rights Watch, All This Terror Because of a Photo.

481 Center for Countering Digital Hate, Digital Hate.

482 Institute for Strategic Dialogue, “Offline Events, Online Threats.”

483 Royal United Services Institute, Online Gendered Narratives.

484 Human Rights Watch, All This Terror Because of a Photo.

485 Institute for Strategic Dialogue, “Offline Events, Online Threats.”

486 Royal United Services Institute, Online Gendered Narratives.

487 Morrish, Lydia. “The US Is Exporting Anti-LGBTQ Hate Online.” WIRED, June 26, 2023. <https://www.wired.com/story/the-us-is-exporting-anti-lgbtq-hate-online/>

488 Institute for Strategic Dialogue, “Offline Events, Online Threats.”

489 Morrish, “The US Is Exporting Anti-LGBTQ Hate Online.”

490 Marwick and Lewis, Media Manipulation and Disinformation Online.

491 Center for Countering Digital Hate, Digital Hate.

children.⁴⁹² These claims were false, yet they spread rapidly across X/Twitter, YouTube, Telegram, and conservative media ecosystems.⁴⁹³ The hospital subsequently received bomb threats and sustained harassment campaigns, illustrating how digitally amplified moral panic can translate quickly into offline intimidation and security threats.⁴⁹⁴ Similar patterns have been documented when drag events, Pride celebrations, and individual trans healthcare providers are targeted online with bomb threats, armed protests, and in-person harassment following cascades of hostile content across platforms.⁴⁹⁵

In this sense, the ‘who’ at issue in this section is not limited to individual bigots; it includes influencers who build profitable brands by monetising outrage, anonymous users on fringe forums who generate conspiratorial narratives, partisan media outlets that launder those narratives into apparent legitimacy, and political actors who convert digitally manufactured panic into legislative proposals or electoral campaigns.⁴⁹⁶ The following sections examine how the political economy of digital platforms makes these actors structurally effective, how engagement-maximising architectures turn outrage into currency, how recommendation

systems steer users toward increasingly hostile content, and how alternative and mainstream platforms interact to sustain a distributed ecosystem in which anti-LGBTI+ hostility is not merely tolerated but frequently rewarded.⁴⁹⁷

Platform incentives and the political economy of hate

A central driver of contemporary anti-LGBTI+ hostility is the political economy of digital platforms themselves. Social media architectures are not neutral communication tools; they are engagement-maximising systems designed to reward emotionally charged, conflict-oriented, and highly reactive content. Within these environments, outrage functions as a form of currency. Content that provokes anger, disgust, fear, or moral panic frequently generates higher levels of engagement, thereby increasing visibility, recommendation, and monetisation opportunities.⁴⁹⁸

Anti-LGBTI+ narratives have proven particularly compatible with these dynamics. Claims involving ‘child protection,’ ‘gender ideology,’ or alleged threats to women and children are highly effective

at generating emotional engagement and viral circulation.⁴⁹⁹ Studies of platform algorithms show that engagement-based recommendation systems systematically boost polarising and emotionally intense content, including material that frames LGBTI+ people as threats, even when users did not explicitly seek out such content.⁵⁰⁰ During periods of political controversy, legislative debate, or cultural conflict, anti-LGBTI+ content consistently experiences spikes in interaction rates across multiple platforms simultaneously, as algorithms learn to treat outrage as a reliable predictor of attention.⁵⁰¹

The rise of Libs of TikTok demonstrates how anti-LGBTI+ outrage can be transformed into a highly profitable attention economy. Originally operating as a Twitter account reposting videos of LGBTI+ teachers and activists, the account evolved into a central node within broader anti-trans media ecosystems. By selectively reframing videos without context and presenting LGBTI+ visibility as evidence of societal corruption or child endangerment, the account generated millions of interactions while shaping national political discourse. Numerous individuals and institutions targeted by the account later became subjects of harassment campaigns, bomb threats, or public

intimidation.⁵⁰² Comparable dynamics have been observed in other contexts, where accounts and channels in Europe and Latin America build large followings by reframing LGBTI+ visibility as ‘gender ideology,’ ‘corruption of children,’ or ‘Western decadence,’ thereby monetising hostility through advertising, donations, and paid subscriptions.⁵⁰³

Importantly, these outcomes emerge not simply because of ideological extremism but because platform visibility systems reward outrage amplification at scale. An ecosystem that privileges engagement over wellbeing creates structural incentives for actors – whether ideologically committed or opportunistically indifferent – to continually intensify hostility, knowing that moral panic is algorithmically profitable.⁵⁰⁴

This process intensified further following changes to platform governance on X/ Twitter after Elon Musk’s acquisition of the platform. Researchers observed substantial increases in anti-trans hate speech, targeted harassment, and coordinated abuse campaigns following reductions in moderation enforcement and changes to verification systems.⁵⁰⁵ Blue-check verification increasingly rewarded engagement-generating accounts regardless of

492 Ibid.
493 Ibid.
494 Ibid.
495 Institute for Strategic Dialogue, "Offline Events, Online Threats."
496 Zuboff, The Age of Surveillance Capitalism.
497 Center for Countering Digital Hate, Digital Hate.
498 Ibid.

499 McCosker and Wilken, "Platform Governance and Algorithmic Visibility."
500 Ibid.
501 Institute for Strategic Dialogue, "Five-Year Overview of the Online and Offline Anti-LGBTQ Landscape."
502 Monteil, "How Libs of TikTok Became an Anti-LGBTQ+ Hate Machine."
503 Morrish, "The US Is Exporting Anti-LGBTQ Hate Online."
504 Marwick and Lewis, Media Manipulation and Disinformation Online.
505 Center for Countering Digital Hate, Digital Hate.

content quality, thus creating financial and algorithmic incentives for inflammatory anti-LGBTI+ commentary.⁵⁰⁶ Simultaneously, previously banned actors were reinstated, facilitating the reintegration of fringe anti-LGBTI+ narratives into mainstream visibility streams. The result was not a neutral ‘free speech’ reset but a shift in the platform’s political economy that made hostility more visible, more normalised, and more lucrative.⁵⁰⁷

Alternative Platforms, Parallel Presences, and the Outrage Pipeline

Alternative platforms such as Rumble, Telegram, and Truth Social have also become important – if numerically smaller – infrastructures for anti-LGBTI+ mobilisation. These spaces present themselves as defenders of ‘free speech’ against perceived censorship while operating as low-moderation environments where extremist narratives circulate with minimal friction.⁵⁰⁸ Their user bases remain far smaller than those of mainstream platforms, but they attract highly engaged, ideologically aligned audiences, making them valuable hubs for consolidating committed followers, testing more ex-

treme messaging, and coordinating campaigns that can later be exported back into mainstream visibility.⁵⁰⁹

Influencers who cultivate large audiences on mainstream platforms often maintain parallel presences on alternative platforms, enabling narratives to move fluidly between fringe and mainstream environments.⁵¹⁰ For some, this takes the form of an explicit pipeline, where a relatively moderated presence on YouTube, Instagram, or TikTok acts as a recruitment funnel toward more radical or explicit content on Telegram, Rumble, or subscription-based channels, where community norms are less constrained and coordination can take place with fewer reputational risks.⁵¹¹ In this configuration, mainstream platforms function as visibility engines and audience acquisition tools, while alternative platforms provide spaces for ideological intensification, monetisation, and operational planning.⁵¹²

YouTube recommendation systems have historically played a significant role in shaping exposure pathways.⁵¹³ Studies of algorithmic radicalisation suggest that users who begin by searching for mainstream conservative or culturally contentious content may, over time, be directed

toward channels that frame LGBTI+ rights and gender diversity in increasingly hostile or conspiratorial terms.⁵¹⁴ While these pathways are neither linear nor universal, they illustrate how recommendation architectures can make anti-LGBTI+ narratives easier to encounter and harder to avoid once a user enters certain topic clusters.⁵¹⁵ In this section, the focus is on these structural exposure mechanisms; documentary-style productions such as Matt Walsh’s *What Is a Woman?* are treated in Section 7 as examples of professionalised outrage media that operate within – and profit from – this broader attention economy.

Taken together, these dynamics constitute the operating conditions explored throughout Section 6; an environment in which platform architectures and business models systematically reward those who can package anti-LGBTI+ hostility as emotionally compelling, shareable, and monetisable content.

The Transnational Anti-Gender Movement

The contemporary escalation of anti-LGBTI+ hostility cannot be understood separately from the broader rise of transna-

tional anti-gender movements. Over the past decade, opposition to LGBTI+ rights, feminism, reproductive rights, and gender studies has increasingly converged into a shared political framework centred around the concept of ‘gender ideology.’⁵¹⁶ Although the term lacks a coherent academic definition, it functions politically as a symbolic umbrella through which diverse social anxieties are mobilised and unified.⁵¹⁷

Anti-rights actors deploy ‘gender ideology’ in multiple, often contradictory ways. In some contexts, it is described as a supposed elite project to erase ‘natural’ sex differences and destabilise the family; in others, it is framed as an imported doctrine that forces children to question their gender or sexuality in schools, or as a global conspiracy led by feminists, LGBTI+ activists, and international organisations to undermine national sovereignty.⁵¹⁸ The vagueness of the term is a political asset, allowing anxieties about sexuality, religion, education, migration, and demographic change to be folded into a single, emotionally charged threat frame.⁵¹⁹

The anti-gender movement emerged initially through collaborations between conservative religious organisations,

506 McCosker and Wilken, “Platform Governance and Algorithmic Visibility.”

507 Center for Countering Digital Hate, Digital Hate.

508 Institute for Strategic Dialogue, “Offline Events, Online Threats.”

509 Royal United Services Institute, Online Gendered Narratives.

510 Morrish, “The US Is Exporting Anti-LGBTQ Hate Online.”

511 Royal United Services Institute, Online Gendered Narratives.

512 Zuboff, The Age of Surveillance Capitalism.

513 Johnson et. al. “New Online Ecology of Adversarial Aggregates: ISIS and beyond”. 2022. <https://arxiv.org/pdf/1603.09426>

514 Mamié, Robin, Manoel Horta Ribeiro, and Robert West. "Are Anti-Feminist Communities Gateways to the Far Right? Evidence from Reddit and YouTube." In Proceedings of the 13th ACM Web Science Conference 2021 (WebSci '21), 139 to 147. New York: Association for Computing Machinery, 2021. <https://doi.org/10.1145/3447535.3462504>

515 Pariser, The Filter Bubble.

516 Royal United Services Institute, Online Gendered Narratives.

517 Ibid.

518 Ibid.

519 Ibid.

right-wing political actors, and anti-feminist networks in Europe and Latin America during the 1990s and early 2000s.⁵²⁰ Catholic and evangelical groups opposed to reproductive rights and sex education forged alliances with conservative think tanks and parties, developing shared campaigns against ‘gender’ in UN documents, education policies, and regional human rights systems.⁵²¹ Early mobilisations against so-called ‘gender ideology’ in countries such as Poland, Brazil, and Colombia helped establish repertoires – petitions, referendums, mass demonstrations, and online campaigns – that would later be adapted and scaled transnationally.⁵²²

Over time, these movements evolved into highly networked campaigns linking conservative think tanks, religious advocacy groups, populist political parties, and digital influencers.⁵²³ Within this framework, LGBTI+ visibility is portrayed not simply as a moral issue but as evidence of broader civilisational decline, demographic collapse, or foreign ideological contamination.⁵²⁴ Central to these narratives is

the symbolic figure of the child. Across multiple contexts, anti-LGBTI+ campaigns rely heavily on claims that children must be protected from LGBTI+ visibility, drag performances, inclusive education, or gender diversity.⁵²⁵ These narratives are emotionally powerful because they frame exclusionary politics as forms of care rather than discrimination.⁵²⁶

In the United States, debates surrounding drag story hours became a central focal point for far-right mobilisation between 2022 and 2024.⁵²⁷ Events hosted in libraries, schools, and community centres were repeatedly targeted by online harassment campaigns, protests, bomb threats, and armed intimidation.⁵²⁸ Research in the UK by the Royal United Services Institute (RUSI) similarly found that anti-drag narratives spread rapidly across digital platforms before translating into offline mobilisation; online influencers framed drag performers as threats to children, despite the absence of evidence supporting these claims.⁵²⁹ Once established, these narratives enabled extremist groups – including the Proud Boys and other far-right

organisations – to position themselves as defenders of public morality and “child protection.”⁵³⁰

We see similar dynamics emerging across the world. In Poland, ruling Law and Justice Party (PiS) politicians framed “LGBT ideology” as an existential threat to national identity and traditional family structures, contributing to the establishment of so-called “LGBT-free zones” across numerous municipalities between 2019 and 2020.⁵³¹ While largely symbolic, these declarations legitimised broader hostility toward LGBTI+ communities and were accompanied by increased online harassment and political scapegoating.⁵³² In Hungary, Prime Minister Viktor Orbán’s government integrated anti-LGBTI+ narratives into broader nationalist campaigns opposing migration, liberalism, and European integration, including the 2021 propaganda law that restricted discussion of sexual and gender diversity.⁵³³ Across Latin America, movements such as *Con Mis Hijos No Te Metas* (“Don’t Mess With My Children”) mobilised opposition to

gender-inclusive education policies in Peru, Colombia, and other Spanish-speaking countries, framing LGBTI+ inclusion as ideological indoctrination imposed by global elites, international NGOs, or foreign governments.⁵³⁴

These cases demonstrate that anti-LGBTI+ mobilisation is not simply reactive prejudice.⁵³⁵ It functions as part of a broader global backlash against social change, gender equality, and democratic pluralism.⁵³⁶ LGBTI+ communities become symbolic targets through which wider anxieties surrounding modernity, migration, demographic change, secularisation, and political instability are articulated and mobilised.⁵³⁷ The ‘who’ behind these campaigns, therefore, extends beyond individual bigots to include governing parties, religious institutions, transnational advocacy networks, and digital influencers who find in anti-LGBTI+ politics a flexible tool for consolidating support, disciplining opposition, and policing the boundaries of national and civilisational identity.⁵³⁸

520 Amnesty International, The State of the World's Human Rights

521 Ibid.

522 Ibid.

523 Morrish, "The US Is Exporting Anti-LGBTQ Hate Online."

524 Royal United Services Institute, Online Gendered Narratives.

525 Ibid.

526 Ibid.

527 Institute for Strategic Dialogue, "Offline Events, Online Threats."

528 Ibid.

529 Royal United Services Institute, Online Gendered Narratives.

530 Institute for Strategic Dialogue, "Offline Events, Online Threats."

531 Amnesty International, A Thousand Cuts: Technology-Facilitated Gender-Based Violence against Poland's LGBTI Community on X

532 Amnesty International, A Thousand Cuts: Technology-Facilitated Gender-Based Violence against Poland's LGBTI Community on X

533 Amnesty International, The State of the World's Human Rights

534 Ibid.

535 Morrish, "The US Is Exporting Anti-LGBTQ Hate Online."

536 Ibid.

537 Royal United Services Institute, Online Gendered Narratives.

538 Morrish, "The US Is Exporting Anti-LGBTQ Hate Online."

Information Warfare and Realpolitik

The international nature of the anti-gender movement makes it a tool in geopolitics.⁵³⁹ Increasingly, anti-LGBTI+ narratives operate not only as forms of social prejudice but as instruments of information warfare and political destabilisation.⁵⁴⁰ Across multiple geopolitical contexts, LGBTI+ rights are framed as evidence of foreign interference, cultural imperialism, or Western moral decay.⁵⁴¹ These narratives seek not merely to stigmatise LGBTI+ communities but to delegitimise civil society organisations, democratic institutions, and international human rights frameworks more broadly.⁵⁴²

From a realpolitik perspective, anti-LGBTI+ disinformation is useful because it offers regimes and political movements a low-cost, high-yield instrument of social control.⁵⁴³ By portraying LGBTI+ rights as symbols of foreign intrusion or moral collapse, state and state-aligned actors can delegitimise independent media, human rights organisations, and opposition parties as agents of an alien agenda.⁵⁴⁴

Mobilising hostility against a vulnerable minority also creates a shared enemy that can unify otherwise disparate constituencies – religious conservatives, nationalists, and anti-elite conspiracists – behind the regime.⁵⁴⁵ In this sense, anti-LGBTI+ campaigns are not only about individuals expressing prejudice; they are tools through which coordinated governments and party machines manage dissent, test the limits of repression, and signal ideological alignment to domestic and international allies.⁵⁴⁶

Russia has played a particularly influential role in globalising this discourse. Since the introduction of the 2013 gay propaganda law, Russian state actors and state-aligned media have consistently framed LGBTI+ rights as symbols of Western civilisational collapse.⁵⁴⁷ State media outlets – such as RT and Sputnik – repeatedly linked LGBTI+ activism to narratives of moral degeneration, demographic decline, and attacks on traditional values.⁵⁴⁸ Following the full-scale invasion of Ukraine in 2022, Russian officials intensified these narratives further, presenting the war itself as a defence of ‘traditional

civilisation’ against liberal Western decadence, often using LGBTI+ visibility as shorthand for that decadence.⁵⁴⁹

These discourses extend far beyond domestic Russian politics. Russian-linked narratives regarding ‘traditional values’ and anti-LGBTI+ rhetoric have circulated widely through transnational disinformation ecosystems, particularly in Eastern Europe, the Balkans, and parts of Africa.⁵⁵⁰ In Georgia, anti-LGBTI+ protests were increasingly linked to ‘foreign agent’ rhetoric portraying LGBTI+ organisations as instruments of Western destabilisation.⁵⁵¹ Similar patterns have emerged in Serbia and other countries where anti-European Union sentiment intersects with nationalist and religious conservatism.⁵⁵² In Uganda, debates surrounding the Anti-Homosexuality Act were accompanied by extensive rhetoric framing LGBTI+ rights as foreign ideological imports imposed by Western governments and NGOs, with political leaders and religious actors portraying LGBTI+ advocacy groups as threats to national sovereignty and African cultural identity.⁵⁵³

These narratives often rely on a strategic inversion in which human rights advocacy

itself is reframed as ideological aggression.⁵⁵⁴ LGBTI+ organisations are depicted not as vulnerable communities seeking protection but as powerful foreign actors attempting to undermine national identity, religion, or family structures.⁵⁵⁵ This inversion is politically useful because it allows discriminatory policies to be presented as defensive acts of sovereignty and cultural preservation.⁵⁵⁶ As Section 7 elaborates, such information warfare strategies frequently overlap with conspiracy theories concerning global elites, demographic replacement, and anti-NGO sentiment. In other words, anti-LGBTI+ disinformation often functions as an entry point into wider extremist and authoritarian political networks.

From Online Hate to Offline Harm

Digital anti-LGBTI+ hostility does not remain confined to online environments. While causal relationships between speech and violence are complex, evidence increasingly demonstrates that digital ecosystems contribute to the normalisation, legitimisation, and coordination of offline hostility.⁵⁵⁷ One important mechanism is the creation of what schol-

539 Ibid.
540 Human Rights Watch, All This Terror Because of a Photo.
541 Morrish, "The US Is Exporting Anti-LGBTQ Hate Online."
542 Human Rights Watch, All This Terror Because of a Photo.
543 Royal United Services Institute, Online Gendered Narratives.
544 Human Rights Watch, All This Terror Because of a Photo.
545 Royal United Services Institute, Online Gendered Narratives.
546 Ibid.
547 Morrish, "The US Is Exporting Anti-LGBTQ Hate Online."
548 Ibid.

549 Ibid.
550 Royal United Services Institute, Online Gendered Narratives.
551 Amnesty International, The State of the World's Human Rights
552 Ibid.
553 Human Rights Watch, All This Terror Because of a Photo.
554 Ibid.
555 Royal United Services Institute, Online Gendered Narratives.
556 Human Rights Watch, All This Terror Because of a Photo.
557 Institute for Strategic Dialogue, "Offline Events, Online Threats."

ars describe as “permission structures.”⁵⁵⁸ Repeated narratives portraying LGBTI+ communities as dangerous, predatory, corrupting, or socially destructive can gradually reduce social inhibitions against harassment and violence.⁵⁵⁹ Individuals who are continuously exposed to these narratives may come to perceive hostility toward LGBTI+ communities as morally justified or socially acceptable.⁵⁶⁰

This dynamic has been visible in the escalation surrounding drag events and LGBTI+ public gatherings across the United States and United Kingdom.⁵⁶¹ Online narratives portraying drag performers as threats to children were repeatedly amplified by influencers, partisan media ecosystems, and extremist groups before translating into coordinated offline protests and intimidation campaigns.⁵⁶² Libraries, schools, and community centres hosting LGBTI+ events experienced bomb threats, armed demonstrations, and sustained harassment.⁵⁶³ In Germany, for example, civil society organisations documented attacks and disruptions targeting nearly half of all Pride parades in 2025,

largely by right-wing extremists, with incidents ranging from counter-demonstrations and hate speech to physical violence and property damage.⁵⁶⁴

The Club Q shooting in Colorado Springs in 2022 further intensified concerns regarding the relationship between anti-LGBTI+ rhetoric and offline violence.⁵⁶⁵ Although direct causal pathways remain difficult to establish, researchers and advocacy organisations noted that the attack occurred within a broader environment characterised by escalating anti-LGBTI+ rhetoric, particularly targeting drag events and transgender communities.⁵⁶⁶ Similar concerns have been raised in non-Western contexts, where online campaigns portraying LGBTI+ people as threats to children or national values have preceded attacks on Pride events, community centres, and individual activists in countries such as Georgia, Brazil, and Uganda.⁵⁶⁷ In these settings, the fusion of digital moral panic with weak legal protections and hostile political rhetoric can make the leap from online hostility to physical violence even more rapid and severe.⁵⁶⁸

Offline harms also extend beyond direct physical violence. Schools, hospitals, universities, and public institutions increasingly face coordinated harassment campaigns driven by online disinformation.⁵⁶⁹ Teachers accused of ‘grooming,’ healthcare providers offering gender-affirming care, and LGBTI+ advocacy organisations have all become targets of reputational attacks designed to produce institutional retreat and public intimidation.⁵⁷⁰ These developments reveal that anti-LGBTI+ digital ecosystems function not only through persuasion but through intimidation. Their objective is often not merely to convince audiences intellectually but to make public LGBTI+ visibility socially costly, politically risky, and physically unsafe.⁵⁷¹

Global Asymmetries and Unequal Consequences

The consequences of anti-LGBTI+ digital ecosystems are profoundly uneven across global contexts. In states with relatively strong legal protections and civil society infrastructures, digital hostility may manifest primarily through harassment,

reputational attacks, political rollback, or institutional exclusion.⁵⁷² However, in contexts where legal protections are weak or authoritarian tendencies are intensifying, online narratives frequently precede severe forms of state repression and physical violence.⁵⁷³

In Uganda, anti-LGBTI+ digital narratives contributed to broader political conditions surrounding the Anti-Homosexuality Act and subsequent crackdowns on LGBTI+ organisations and activists.⁵⁷⁴ Online outing campaigns, public denunciations, and disinformation narratives exposed individuals to arrest, eviction, and mob violence.⁵⁷⁵ Similar dynamics have been documented in Kenya, Indonesia, and parts of Eastern Europe, where online hostility intersects with weak institutional protections and political opportunism.⁵⁷⁶ In Russia, digital surveillance and anti-LGBTI+ legislation increasingly operate in tandem: Online spaces are monitored for ‘extremist’ or ‘propaganda’ content while LGBTI+ organisations face criminalisation and forced closure.⁵⁷⁷ Under these conditions, digital visibility itself becomes dangerous, transforming social media

558 Royal United Services Institute, Online Gendered Narratives.

559 Ibid.

560 Human Rights Watch, All This Terror Because of a Photo.

561 Institute for Strategic Dialogue, "Offline Events, Online Threats."

562 Royal United Services Institute, Online Gendered Narratives.

563 Institute for Strategic Dialogue, "Offline Events, Online Threats."

564 Alkiviadou, Hate Speech and the European Court of Human Rights.

565 Anti-Defamation League. "Extremism and Anti-LGBTQ Narratives." <https://www.adl.org/node/51434>

566 Institute for Strategic Dialogue, "Offline Events, Online Threats."

567 Human Rights Watch, All This Terror Because of a Photo.

568 Ibid.

569 Institute for Strategic Dialogue, "Offline Events, Online Threats."

570 Human Rights Watch, All This Terror Because of a Photo.

571 Ibid.

572 Ibid.

573 Ibid.

574 Ibid

575 Ibid.

576 Tahrir Institute for Middle East Policy, "Digital Persecution."

577 Human Rights Watch, All This Terror Because of a Photo.

platforms into sites of surveillance as much as communication.

The uneven global distribution of harm highlights a critical asymmetry, the same narratives circulating internationally can produce radically different consequences depending on local political conditions. A conspiracy theory or harassment campaign that produces reputational damage in one country may contribute to imprisonment, disappearance, or physical violence in another.⁵⁷⁸ Taken together, these

dynamics demonstrate that anti-LGBTI+ digital harm is not simply a by-product of online discourse. Rather, it constitutes a transnational system of inequality production in which ideology, platform governance, political economy, and institutional power interact continuously across borders.⁵⁷⁹ Ultimately, anti-LGBTI+ digital ecosystems do more than reflect existing inequalities; they actively redistribute vulnerability across global contexts, amplifying exposure to violence and repression precisely where institutional protections are weakest.⁵⁸⁰

Section 6 Key Findings

- Anti-LGBTI+ hostility is structurally embedded, not incidental. Identity-based hostility is converted into attention, and attention turns into social, political, and financial capital.
- The ecosystem is decentralised but self-reinforcing. No single organisation coordinates it. Instead, 'networked amplification' allows loosely connected actors (far-right groups, religious conservatives, anti-feminist networks, conspiracy communities, opportunistic influencers) to converge around shared narrative frames such as 'gender ideology,' 'grooming,' 'parental rights,' 'protecting children.'
- Content laundering follows a predictable cross-platform pipeline. Narratives typically originate on fringe forums (4chan, Kiwi Farms) and migrate to mainstream platforms (X, TikTok, YouTube, Facebook, Instagram) where they're stylised and reframed as legitimate public controversy rather than disinformation.

578 Ibid.

579 Human Rights Watch, *All This Terror Because of a Photo*.

580 Ibid.

- Engagement-maximising architectures treat outrage as a reliable predictor of attention. Anti-LGBTI+ content is unusually effective at generating this kind of engagement, and algorithms amplify it accordingly.
- Alternative platforms function as a radicalisation and coordination layer. Telegram, Rumble, and Truth Social have smaller audiences but serve as spaces to intensify messaging and organise campaigns. These are then exported back to mainstream platforms for wider reach, creating a two-tier 'outrage pipeline.'
- The consequences are globally unequal. The same viral narrative can produce reputational damage in one country and imprisonment, disappearance, or mob violence in another (e.g., Uganda's Anti-Homosexuality Act), meaning digital harm actively redistributes vulnerability toward contexts with the weakest legal protections.

Section 6 Questions:

- Have you seen evidence of specific platform design features (recommendation algorithms, monetisation/verification systems, moderation policy changes) measurably increase anti-LGBTI+ content visibility or engagement?
- In your experience, do alternative/low-moderation platforms (e.g., Telegram, Rumble, Truth Social) function as radicalisation or coordination spaces that later surface on mainstream platforms? What would effective intervention look like?
- What indicators – in your experience – most reliably signal that online anti-LGBTI+ hostility is escalating toward offline harm (protests, threats, violence)?
- Given that identical online narratives produce vastly different real-world consequences depending on local legal and political context, what support (legal, technological, advocacy) is most urgently needed in high-risk jurisdictions?



Section 7: Far-Right Online Ecosystems and Anti-LGBTI+ Mobilisation

Section 6 described anti-LGBTI+ harm as an ecosystem: A distributed, self-reinforcing system of narratives, incentives, and actors in which hostility toward LGBTI+ communities becomes structurally profitable and politically useful. This section zooms in on one of the most significant subsystems within that environment: the far-right online ecosystems that provide ideological coherence, cross-platform infrastructure, and transnational reach. Where Section 6 mapped the broader political economy of hate, this section traces the specific architecture of actors who organise, intensify, and circulate that hostility across digital spaces.

Far-right online ecosystems are best understood not as isolated communities but as interconnected, transnational systems through which narratives, symbols, tactics, and audiences move with remarkable fluidity. Drawing on Stephane J. Baele, Lewys Brace, and Travis G. Coan’s analytical framework, the far-right online presence can be understood as a dynamic ecosystem composed of multiple layers including individual entities such as forums, channels, and accounts; communities linked by migration and shared content; broader “biotypes” defined by common ideological and stylistic features; and the overall network that emerges

from their interaction.⁵⁸¹ Within this ecosystem, anti-LGBTI+ narratives function as a key ideological currency. They do not appear as isolated prejudices but as recurring motifs within broader storylines about nation, race, gender, civilisation, and decline.⁵⁸²

This section argues that anti-LGBTI+ mobilisation within far-right ecosystems is sustained by four interlocking dynamics. The first is the globalisation of anti-LGBTI+ narratives through networked online spaces. The second is the rise of influencer-driven extremism and the monetisation of outrage within attention-driven platform economies. The third is the emergence of hybrid attack pathways that connect online harassment, doxxing, and narrative escalation to offline intimidation and action. The fourth is the intersectional convergence of misogyny, racism, and anti-LGBTI+ ideologies within shared digital environments, producing not separate prejudices but interwoven systems of hate.

Taken together, these dynamics demonstrate that anti-LGBTI+ hostility is not episodic but systemic. It is sustained by a digital architecture that rewards amplification, adaptation, and ideological convergence and by actors who have learned to operate fluently across mainstream plat-

forms, fringe spaces, and transnational political contexts.

Transnational Hate Networks

Far-right online ecosystems are transnational by design. Baele, Brace, and Coan conceptualise them as a far-right online ecosystem precisely because extremist actors operate across platforms and national boundaries in ways that blur traditional distinctions between domestic and international movements.⁵⁸³ Within this ecosystem, anti-LGBTI+ narratives function as a shared symbolic resource. They frequently frame LGBTI+ rights as evidence of Western decadence, moral collapse, demographic threat, or civilisational decline. These frames travel easily between different linguistic and political contexts.⁵⁸⁴

Importantly, such framing is not confined to Western far-right movements. State and non-state actors alike deploy anti-LGBTI+ rhetoric transnationally, often presenting it as resistance to cultural imperialism, global liberalism, or foreign interference.⁵⁸⁵ In this sense, anti-LGBTI+ narratives help bridge multiple political formations that do not always agree on other issues. Nationalists, Christian conservatives, conspiratorial communities, populist

581 Tahrir Institute for Middle East Policy, "Digital Persecution."

582 OECD, Rights in the Digital Age.

583 Tahrir Institute for Middle East Policy, "Digital Persecution."

584 Are, Talbot, and Briggs, "Social Media Affordances."

585 Klonick, "Beyond the Public Square."

parties, and authoritarian communicators can all converge around the symbolic claim that queer visibility represents an existential threat to family, nation, religion, or social order.⁵⁸⁶

Empirical evidence reinforces this systemic perspective. The ExtremeBB dataset, which tracks over two decades of activity across extremist online forums, reveals long-term patterns of ideological diffusion and overlap in which anti-LGBTI+ discourse co-evolves alongside misogynistic, racist, nationalist, and conspiracist narratives.⁵⁸⁷ Rather than forming isolated discursive silos, these ideas circulate within dense and mutually reinforcing networks in which hostility toward one marginalised group often strengthens hostility toward others.⁵⁸⁸ What matters analytically is not simply that such narratives spread but that they do so within an infrastructure that makes ideological cross-fertilisation normal rather than exceptional.

The transnational nature of these dynamics is further illustrated by analyses of gendered extremist narratives in countries such as the United Kingdom and Australia. These studies show that similar rhetorical strategies – particularly those centred on gender, family, masculinity, and social

order – circulate across national contexts with surprisingly little adaptation.⁵⁸⁹ Digital platforms, therefore, function as conduits through which local grievances are translated into globally recognisable ideological scripts.⁵⁹⁰ A controversy about a school, a drag event, or a healthcare provider in one country can be rapidly reframed within a wider repertoire of anti-LGBTI+ panic that is already familiar to audiences elsewhere.

Taken together, these findings suggest that anti-LGBTI+ mobilisation is embedded within a globalised ideological infrastructure. Narratives do not simply spread; they are continuously reshaped, repackaged, and recontextualised within a shared transnational repertoire.⁵⁹¹ The result is a far-right ecosystem in which local campaigns are rarely fully local. They are more often nodes within wider circuits of ideological exchange, tactical learning, and digital amplification.⁵⁹²

Influencer-Driven Extremism

While structural analyses highlight the architecture of far-right ecosystems, understanding their dynamics requires attention to the actors who animate them. Increasingly, these actors take the form of

content creators, streamers, meme-curators, and political influencers who operate as ideological entrepreneurs within attention-driven digital economies.⁵⁹³ Rather than acting like traditional party elites or centralised movement leaders, they cultivate hybrid identities that blend personal branding with political messaging, allowing them to appear both relatable and authoritative to their audiences.⁵⁹⁴

Research on online radicalisation pathways shows that these figures play a pivotal role in guiding audience trajectories. Robin Mamié and colleagues demonstrate how online communities – particularly within the so-called manosphere – function as gateways to more explicitly far-right content, with recommendation systems and audience overlap gradually exposing users to increasingly extreme material.⁵⁹⁵ This process is not random. It is structured by digital environments in which creators compete for visibility by packaging grievance, identity, and hostility in emotionally resonant forms.⁵⁹⁶

These influencers often operate by translating abstract ideological themes into digestible narratives. Complex claims about demographic decline, ‘gender

ideology,’ the erosion of masculinity, or civilisational threat are rendered as short videos, reaction clips, memes, livestreams, and personality-driven commentary that circulate easily across platforms. In doing so, influencers lower the barrier to ideological entry.⁵⁹⁷ Users do not have to join an explicitly extremist organisation to encounter extremist framings; they can absorb them incrementally through lifestyle content, humour, outrage commentary, or ‘common sense’ cultural criticism.⁵⁹⁸

The study of ‘gayservatives’ on the platform Gab illustrates how even individuals belonging to marginalised groups can become important nodes in far-right amplification networks. By leveraging their identity, such actors reframe exclusionary or anti-LGBTI+ narratives as insider perspectives, thereby enhancing their legitimacy and persuasive force among audiences who might otherwise treat those claims with greater suspicion.⁵⁹⁹ This matters because it shows that ideological authenticity online is not simply about truthfulness. It is often performed through identity, affect, and positioning.

Within this ecosystem, outrage becomes a form of currency. Content that provokes

586 Are, Talbot, and Briggs, "Social Media Affordances."

587 Castells, *Networks of Outrage and Hope*

588 Ibid.

589 boyd, *It's Complicated*.

590 Human Rights Watch, *All This Terror Because of a Photo*.

591 Tahrir Institute for Middle East Policy, "Digital Persecution."

592 Are, Talbot, and Briggs, "Social Media Affordances."

593 EDRI, "Digital Rights of LGBTQ+ People."

594 Rachel Griffin, "Rethinking Rights in Social Media Governance: Human Rights, Ideology and Inequality," *European Law Open* 2, no. 1 (2023): 30 to 56, <https://doi.org/10.1017/elo.2023.7>

595 Mamie et al., "Are Anti-Feminist Communities Gateways to the Far Right?"

596 Griffin, "Rethinking Rights in Social Media Governance."

597 Ibid.

598 EDRI, "Digital Rights of LGBTQ+ People."

599 Monteil, "How Libs of TikTok Became an Anti-LGBTQ+ Hate Machine."

strong emotional reactions is more likely to be shared, recommended, and monetised, making anti-LGBTI+ rhetoric valuable not only as ideology but as content. The boundaries between political expression and content production blur, giving rise to a system in which attention, engagement, and revenue reinforce extremist messaging. Anti-LGBTI+ hostility thus becomes both conviction and business model as something that can build audiences, generate donations or subscriptions, and sustain influence over time.⁶⁰⁰

Baele et al.'s framework helps clarify this dynamic by identifying distinct roles within the ecosystem, including content producers, amplifiers, and consumers. Influencers occupy a pivotal position because they connect those roles. They absorb material from fringe communities, translate it into more polished – or emotionally compelling – forms and redistribute it into broader digital publics. The result is a form of extremism that is less hierarchical and more entrepreneurial, driven not by centralised leadership but by distributed networks of content creators competing for visibility and influence.⁶⁰¹

Hybrid Attack Pathways

Far-right online ecosystems do not simply circulate narratives; they also develop re-

peatable pathways through which online hostility escalates into coordinated offline action.⁶⁰² These hybrid attack pathways link harassment campaigns, doxxing, narrative escalation, and targeted disinformation to protests, disruptions of public events, institutional intimidation, and – in some cases – physical violence. What appears at first glance to be spontaneous outrage is frequently the outcome of routinised processes through which digital hostility is transformed into embodied political action.⁶⁰³

These pathways often begin with narrative seeding. A drag event, Pride gathering, school policy, healthcare provider, or individual activist is identified and framed through familiar motifs such as ‘grooming,’ ‘protecting children,’ ‘family values,’ or resistance to supposed ideological indoctrination.⁶⁰⁴ The narrative is then amplified by influencers, Telegram channels, fringe forums, and platform-native outrage entrepreneurs who transform a local issue into a symbolic conflict. Once audiences are mobilised, the next stages may include exposure of personal details, calls for protest, reputational attacks, threats, or attempts to force institutional retreat.⁶⁰⁵

Research on anti-drag mobilisation provides a clear illustration of this pattern. ISD’s work on anti-drag threats in the United States documented a rising number of

incidents in which online hate speech and coordination manifested in offline threats, protests, and violence targeting drag events and the wider LGBTI+ community.⁶⁰⁶ The actors involved were not limited to established anti-LGBTI+ groups; they included white supremacists, Christian nationalists, parents’ rights activists, anti-vaccine networks, and assorted local extremists. This heterogeneity matters because it shows how far-right ecosystems operate less through rigid organisational unity than through temporary coalitions around emotionally charged targets.

The same pattern is visible beyond the United States. ISD’s comparative work across the US, UK, France, and Australia found that anti-drag mobilisation relied on recurring narratives, tactics, and actor constellations across countries – even where legal frameworks and political cultures differed. Online spaces enabled tactical learning and narrative synchronisation, allowing movements to borrow slogans, repertoires, and threat scripts across borders. In this sense, hybrid attack pathways are not merely local escalations. They are modular templates for harassment that can be adapted to different national contexts while retaining the same basic structure.⁶⁰⁷

This matters for anti-LGBTI+ mobilisation because it reveals that the digital environment is not simply a site of expression. It

is also a staging ground for action. Far-right ecosystems use digital platforms to identify targets, produce justificatory narratives, recruit participants, and coordinate pressure. The online and offline dimensions are, therefore, not separate phases but parts of a single operational continuum.

The Intersectionality of Hate

A defining feature of far-right online ecosystems is the intersectional nature of the ideologies they sustain.⁶⁰⁸ Anti-LGBTI+ rhetoric rarely appears in isolation. Instead, it is embedded within a broader constellation of misogynistic, racist, nationalist, antisemitic, and anti-migrant discourses that mutually reinforce one another. These are not parallel hatreds that happen to coexist. They are interwoven strands of a single ideological fabric organised around hierarchy, exclusion, and fantasies of social restoration.

Research on gendered extremist narratives highlights the centrality of misogyny within these ecosystems. Narratives about masculinity, femininity, family structure, and gender roles often serve as entry points into more explicitly anti-LGBTI+ positions. The perceived erosion of traditional gender norms is framed not only as a threat to individual identity but as evidence of broader civilisational disorder. In this framing, feminism, queer rights,

600 Griffin, "Rethinking Rights in Social Media Governance."

601 Tahrir Institute for Middle East Policy, "Digital Persecution."

602 Human Rights Watch, All This Terror Because of a Photo.

603 Are, Talbot, and Briggs, "Social Media Affordances of LGBTQIA+ Expression and Community Formation."

604 Ibid.

605 Klonick, "Beyond the Public Square."

606 Institute for Strategic Dialogue Global. Five-Year Overview of Online and Offline Anti-LGBTQ Landscape. <https://www.isdglobal.org/digital-dispatch/investigation-five-year-overview-of-the-online-and-offline-anti-lgbtq-landscape/>

607 Ibid.

608 Are, Talbot, and Briggs, "Social Media Affordances of LGBTQIA+ Expression and Community Formation."

reproductive freedom, and anti-racist politics are treated as linked symptoms of the same decline.⁶⁰⁹

The manosphere exemplifies this dynamic. Communities centred on anti-feminist discourse frequently act as gateways to broader far-right ideologies, including anti-LGBTI+ and white supremacist narratives.⁶¹⁰ Users may initially be drawn in through discussions of dating, male grievance, or perceived bias against men only to be gradually exposed to more expansive ideological frameworks in which women, queer communities, racialised minorities, and migrants are positioned as interconnected threats.⁶¹¹ Misogyny, therefore, does not merely accompany far-right mobilisation; it often structures its recruitment pathways.

Empirical data from extremist forums confirms this overlap. The ExtremeBB dataset reveals that users often participate in multiple ideological communities simultaneously, contributing to the co-evolution of different forms of hate.⁶¹² Rather than separate silos, these communities form dense networks of interconnected discourse in which one form of hostility primes, legitimises, or intensifies another.⁶¹³ This has important analytical consequences. Efforts

to address anti-LGBTI+ hate cannot be separated from broader struggles against misogyny and racism because the same communities, narratives, and infrastructures often sustain all three.

This convergence also has differential consequences for those most exposed to overlapping forms of marginalisation. The targets most vulnerable to far-right digital ecosystems are often not simply 'LGBTI+ people' in the abstract but those who embody multiple sites of stigma at once including trans women – especially trans women of colour; queer migrants and refugees; racialised queer activists; and those whose public visibility makes them legible as symbolic enemies. Intersectionality, in this context, is not merely an analytical tool. It is a description of how vulnerability is produced, distributed, and intensified within digital environments structured by converging prejudices.⁶¹⁴

Conclusion

The analysis presented in this section points to a fundamental shift in the nature of far-right mobilisation. Anti-LGBTI+ hostility is no longer confined to fringe groups or isolated incidents. Instead, it is embedded within a transnational, adap-

tive ecosystem that operates across digital and physical spaces. This ecosystem is characterised by the global circulation of narratives through interconnected online networks, the central role of influencers in producing and amplifying extremist content, the emergence of hybrid pathways linking online activity to offline action, and the intersectional convergence of multiple forms of hate.

Together, these dynamics create a system that is not only resilient but anti-fragile

and is capable of adapting, mutating, and reorganising itself in response to pressure, moderation, or public scrutiny. Efforts to counter anti-LGBTI+ mobilisation must, therefore, move beyond reactive responses to individual incidents and instead engage with the structural and systemic nature of the problem. The following section descends beneath the visible surface of these networks into the encrypted channels, anonymous infrastructures, and dark-web markets through which the most covert forms of coordination occur.

Section 7 Key Findings:

- Anti-LGBTI+ mobilisation has evolved into a transnational digital ecosystem that operates across mainstream and fringe online platforms, allowing hostile narratives, tactics, and campaigns to spread rapidly across borders.
- Far-right influencers play a central role in legitimising and amplifying anti-LGBTI+ narratives by combining outrage, entertainment, and identity politics to reach mainstream audiences and drive engagement.
- Anti-LGBTI+, racist, and misogynistic narratives increasingly reinforce one another. These intersecting forms of hate disproportionately affect those facing multiple forms of marginalisation, including trans women, racialised LGBTI+ people, migrants, and refugees.
- Anti-LGBTI+ mobilisation increasingly forms part of a broader transnational anti-gender movement that links opposition to LGBTI+ rights with attacks on women's rights, reproductive rights, and democratic institutions.

609 Castells, Networks of Outrage and Hope

610 Ibid.

611 EDRi, "Digital Rights of LGBTQ+ People."

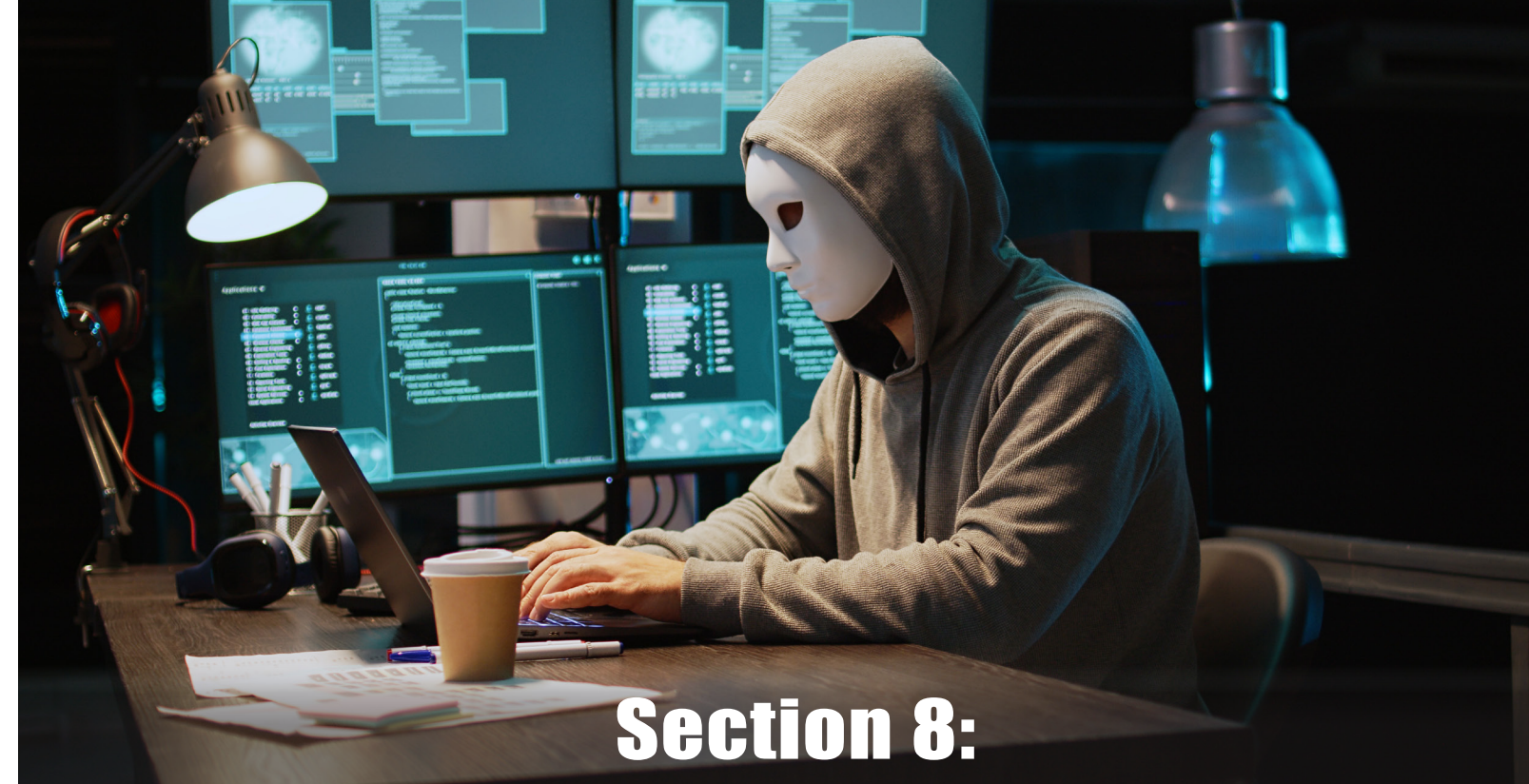
612 Castells, Networks of Outrage and Hope

613 Ibid.

614 boyd, It's Complicated.

Section 7 Questions:

- Have you observed cooperation or coordination between domestic and international anti-rights actors? If so, how has this affected your work or the communities you support?
- Which online narratives or campaigns have been most effective in fuelling hostility towards LGBTI+ people in your context?
- Within the LGBTI+ community, which groups are most affected by these digital campaigns, and how do factors such as race, gender identity, disability, migration status, or other characteristics shape those experiences?
- What strategies have been most effective in disrupting, countering, or building resilience against coordinated anti-LGBTI+ mobilisation online?
- What actions should governments, technology companies, international organisations, and donors prioritise to better protect LGBTI+ activists and communities from coordinated digital anti-rights campaigns?



Section 8:

Hidden Infrastructures of Harm

Section 7 showed how shadow networks supply actors with coordination and targets, and, as the following section illustrates, how they also underpin the infrastructure of digital sexual exploitation. Moreover, section 7 described far-right ecosystems as networked, adaptive systems that operate across mainstream and fringe platforms, giving anti-LGBTI+ mobilisation ideological coherence, transnational reach, and visible pathways from online hostility to offline action. This section asks what happens beneath that visible architecture: at the layers where coordination becomes explicitly hidden, where personal data becomes a tradable commodity, and where the infrastructure of harm is deliberately designed to evade accountability.

The architecture of contemporary harassment and repression no longer resides

solely in public-facing digital spaces. It has migrated into a layered ecosystem of encrypted channels, anonymous infrastructures, and illicit data markets that together form what can be understood as shadow networks. These networks do not function as discrete organisations with stable leadership or clear territorial boundaries. They resemble fluid systems that move across platforms, adapting rapidly to disruption while maintaining continuity of purpose. Within this environment, coordination becomes obscured, identities become commodities, and accountability becomes increasingly difficult to enforce.

Section 8, therefore, focuses on the infrastructural backbone that underpins elements explored in Sections 6 and 7. It examines: How covert coordination and data trafficking supply the political econ-

omy of hate with targets and tools, how jurisdictional and technical complexity frustrate accountability, and how the same technologies that protect activists are also weaponised against them. For LGBTI+ communities, these dynamics mean that the risks mapped in earlier sections extend beyond what is visible on mainstream platforms into a less legible – but deeply consequential – layer of digital life.

Covert Organisation

Extremist and anti-rights networks have developed sophisticated methods of coordination that rely on the interplay between public visibility and hidden communication channels. Rather than operating exclusively on the so-called dark web, these actors use a multi-layered communication strategy that spans the surface web, encrypted messaging applications, and anonymous forums.⁶¹⁵

At the outermost layer, mainstream platforms function as recruitment and amplification tools. Content circulated in these spaces is often coded, memetic, or framed for plausible deniability. Individuals drawn in through such material – whether via outrage, curiosity, or a sense of community – are then redirected toward more private environments, particularly encrypted messaging applications

such as Telegram and WhatsApp.⁶¹⁶ Within these semi-closed ecosystems, ideological narratives are intensified, group identity is reinforced, and coordination begins to take shape.⁶¹⁷

The innermost layer consists of highly restricted channels or dark-web forums where operational planning occurs. Here, anonymity is not merely a feature but a strategic asset. End-to-end encryption conceals communication from external scrutiny, allowing participants to coordinate harassment campaigns, share targeting information, exchange technical tools, and disseminate tactical guidance with minimal immediate risk of detection. Access is often controlled through invitation, vetting rituals, or reputational markers, further insulating these spaces from casual observation.⁶¹⁸

These networks do not depend on hierarchical command structures. Instead, they operate according to a logic of decentralisation. Channels can be replicated, users can migrate, and content can be mirrored across platforms with remarkable speed.⁶¹⁹ When one node is removed, others emerge in its place – frequently within hours – preserving archives, tactics, and contact lists. This resilience creates the appearance of spontaneous or organic activity at the surface, masking the underlying coordination that sustains it.⁶²⁰

For LGBTI+ activists and organisations, this structure transforms online harassment from a series of isolated incidents into something more organised and persistent. What appears as diffused hostility in public comment sections, replies, or quote-tweets often originates in coordinated efforts that are deliberately concealed within encrypted environments where targets are selected, narratives refined, and attack waves scheduled.⁶²¹

Data Trafficking

Parallel to these communication networks is a robust and evolving marketplace for personal data. The dark web, alongside adjacent encrypted platforms and semi-covert channels, facilitates the exchange of sensitive information that can be used to identify, monitor, and exploit individuals. In this context, personal data is not merely exposed; it is commodified.⁶²²

Data traded within these environments includes email credentials, home addresses, phone numbers, employment details, social media handles, and other identifying information. Such data is often obtained through a mixture of large-scale breaches, phishing campaigns, credential stuffing, and open-source intelligence techniques that compile information from

public records, social media, and leaked databases. Once aggregated, it can be packaged and sold in bulk or tailored to specific targeting requests, sometimes sorted by geography, profession, or perceived political profile.⁶²³

The practice of doxxing illustrates how these data flows are operationalised. Doxxing typically follows a recognisable sequence where a target is identified; information is collected from both illicit and open sources; and a comprehensive profile is compiled and publicly released, either in mainstream spaces or in dedicated ‘doxx drops.’⁶²⁴ This release is frequently accompanied by coordinated harassment, including threats, intimidation, blackmail, attempts to contact employers or family members, and broader reputational attacks.⁶²⁵

Increasingly, elements of this process have been commodified into what can be described as ‘doxxing-as-a-service.’ Actors specialising in data extraction and exposure offer their services to others, effectively outsourcing harassment to those with greater technical expertise or time. This development reflects a broader shift toward the professionalisation of digital harm, in which technical skill, ideological motivation, and financial incentive intersect. Services may be advertised in encrypted channels or dark-web marketplac-

615 EDRi, "Digital Rights of LGBTQ+ People."

616 Institute for Strategic Dialogue, "Offline Events, Online Threats."

617 Institute for Strategic Dialogue Global. Five-Year Overview of Online and Offline Anti-LGBTQ Landscape.

618 EDRi, "Digital Rights of LGBTQ+ People."

619 Tahrir Institute for Middle East Policy, "Digital Persecution."

620 Institute for Strategic Dialogue, "Offline Events, Online Threats."

621 Human Rights Watch, All This Terror Because of a Photo.

622 EDRi, "Digital Rights of LGBTQ+ People."

623 OECD, Rights in the Digital Age.

624 Human Rights Watch, All This Terror Because of a Photo.

625 OECD, Rights in the Digital Age.

es, offering everything from basic ‘people search’ to bespoke targeting packages.⁶²⁶

For LGBTI+ activists, the consequences are particularly acute. Visibility – which is often essential for advocacy, community building, and fundraising – simultaneously increases vulnerability. Personal information, once exposed, can be used to facilitate both online and offline harassment, including stalking, forced outing, extortion, and threats of violence. In this sense, the dark web does not merely host harmful activity; it sustains an economy in which vulnerability itself becomes a tradeable asset.⁶²⁷ The extractive logic of data colonialism – introduced in Section 2 – appears here in its most predatory form, where the continuous capture and circulation of intimate data is stripped of consent, context, and control.

Law-Enforcement Challenges

Efforts to address these dynamics are significantly constrained by structural limitations inherent in contemporary legal and enforcement frameworks. Shadow networks operate across borders, jurisdictions, and technological layers that do not align neatly with existing mechanisms of accountability or with traditional models of policing.⁶²⁸

One primary challenge is anonymity. Technologies such as Tor route communications through multiple nodes, obscuring users’ locations, while end-to-end encryption prevents the interception and straightforward interpretation of content. Even when harmful activity is detected, attributing it to specific individuals – and doing so in a way that meets evidentiary standards for legal action – remains complex and often insurmountable. Attribution work is further complicated by the use of disposable accounts, VPNs, bot-nets, and layered pseudonyms that can be abandoned and reconstituted with minimal cost to perpetrators.⁶²⁹

Jurisdictional fragmentation further complicates enforcement. A single incident of harassment may involve actors, infrastructure, and victims located in multiple countries. Data used in a doxxing campaign might be hosted on servers in one jurisdiction, scraped or purchased by perpetrators in another, and deployed against a target in a third. The platforms involved in storage or dissemination sit under yet other regulatory regimes. Legal authority, however, remains largely territorially bounded. Mutual legal assistance processes are often slow, uneven, and politically constrained, creating gaps in which harmful activity can persist with limited intervention.⁶³⁰

Temporal dynamics also play a critical role. Shadow networks are capable of rapid adaptation, dissolving and reconstituting themselves in response to enforcement actions, platform takedowns, or investigative exposure. By contrast, legal processes are often slow, requiring coordination among multiple agencies, adherence to procedural safeguards, and, in many cases, complex international cooperation. This mismatch allows perpetrators to evade consequences simply by moving faster than the systems designed to stop them. Attacks can be executed in hours or days, while investigations may take months or years – if they proceed at all.⁶³¹

Finally, the widespread use of encryption introduces a normative dilemma. Strong encryption is essential for protecting privacy, securing communications, and safeguarding activists, journalists, and vulnerable communities.⁶³² Yet, the same technologies can shield harmful actors from scrutiny and hinder legitimate investigations into severe abuses. Policymakers are, therefore, confronted with a tension between security and rights that has no straightforward resolution. Attempts to weaken encryption in the name of safety risk undermining the very tools LGBTI+ communities rely on to organise and survive under hostile regimes, while leaving

them exposed to state surveillance and private interception.⁶³³

State Use and the Dual-Use Paradox

The technologies underpinning shadow networks are not inherently tied to any single category of actor. Encryption, anonymity, and decentralised communication are dual-use tools, capable of serving both protective and harmful purposes. As such, they are also employed by states and state-aligned entities.⁶³⁴

Governments and law-enforcement agencies utilise encrypted platforms and covert identities for legitimate purposes, including intelligence gathering, infiltration of violent networks, disruption of planned attacks, and protection of sensitive communications. Under appropriate legal safeguards, these practices can yield valuable insights into network structures, operational plans, and threat trajectories, potentially preventing harm.⁶³⁵

At the same time, there is evidence that similar techniques may be used in more controversial ways. Information operations – including the dissemination of propaganda, disinformation, and targeted harassment – can be conducted through

626 Amnesty International, The State of the World's Human Rights.

627 Tahrir Institute for Middle East Policy, "Digital Persecution."

628 EDRI, "Digital Rights of LGBTQ+ People."

629 Ibid.

630 OECD, Rights in the Digital Age.

631 Gomez et al., Algorithmic Arbitrariness in Content Moderation.

632 EDRI, "Digital Rights of LGBTQ+ People."

633 Marwick and Lewis, Media Manipulation and Disinformation Online.

634 Ibid.

635 Amnesty International, The State of the World's Human Rights.

anonymous or semi-covert channels that resemble the methods used by non-state extremist actors.⁶³⁶ In some contexts, activists and minority groups – including LGBTI+ communities – have reported being targeted by coordinated campaigns that blur the line between private and state-aligned activity, combining troll networks, bot accounts, and anonymous reports to intimidate, discredit, or silence dissent.⁶³⁷

This convergence raises important questions about accountability and oversight. When the same tools and methods are used by both state and non-state actors, distinctions between protection and repression can become blurred. The ethical and legal implications depend not on the technology itself, but on the context in which it is deployed, the normative frameworks that guide its use, and the degree to which that use is subject to transparency, independent oversight, and meaningful avenues for redress.⁶³⁸

For LGBTI+ communities, this dual-use paradox can be especially acute. The encrypted channels and anonymity features they need to protect themselves from private harassment or criminal prosecution can also be infiltrated, monitored, or manipulated by hostile state actors. Conversely, calls to weaken or backdoor encryption in order to combat online harms often ignore the ways in which such measures would expose activists, jour-

nalists, and marginalised communities to increased vulnerability.⁶³⁹ This section, therefore, underscores that the governance of shadow networks is not only a technical or criminal justice problem; it is also a human rights and democratic accountability problem.

Conclusion

Shadow networks are not peripheral to the digital environment; they are embedded within it. They rely on the same infrastructures that enable everyday communication, repurposed to facilitate coordination, exploitation, and evasion. Through encrypted channels, they organise harassment that appears spontaneous but is often deliberate. Through illicit markets, they transform personal data into instruments of harm and vulnerability into a commodity. Through jurisdictional complexity, technical opacity, and temporal asymmetry, they evade accountability and outpace enforcement.

For LGBTI+ activists and communities, these dynamics create a landscape in which visibility carries both power and risk. Advocacy can amplify voices, expose abuses, and build solidarity, but it can also expose individuals to coordinated targeting that extends beyond the visible surface of the internet into infrastructures designed to be hard to see, regulate, or resist. Addressing these challenges requires not only technical and legal re-

sponses but also a deeper understanding of how shadow networks operate as systems and how they interact with the more visible ecosystems described in Sections 6 and 7.

Ultimately, the dark web is less a separate domain than a concentrated reflection of the broader digital order. It reveals, in

intensified form, the tensions that define that order, such as between openness and control, privacy and security, empowerment and vulnerability. The following section turns from these covert infrastructures to a more openly visible but equally destructive set of online communities; the predatory spaces in which gendered domination and humiliation are themselves performed as social capital.

Section 8 Key Findings:

- Anti-LGBTI+ digital harm increasingly depends on hidden infrastructures rather than public platforms alone. Encrypted messaging services, anonymous forums, and dark web environments facilitate coordination while making detection and accountability more difficult.
- Extremist and anti-rights actors use a layered communication model: mainstream platforms recruit and amplify audiences; encrypted channels strengthen group identity and coordinate activity; and closed or anonymous networks enable operational planning and evade scrutiny.
- Personal information has become a strategic resource. Doxxing, data breaches, illicit data markets and the trafficking of personal information enable targeted harassment, intimidation, blackmail, and surveillance of LGBTI+ people and human rights defenders.
- The same technologies that protect activists – including encryption, anonymity and secure communications – can also be exploited by hostile actors. This creates a fundamental human rights challenge: protecting privacy and security while responding effectively to coordinated abuse.
- Jurisdictional fragmentation, technical complexity, and cross-border digital infrastructures make it difficult for governments, law enforcement, and technology companies to investigate and respond to coordinated online harms.

636 Marwick and Lewis, Media Manipulation and Disinformation Online.

637 Human Rights Watch, All This Terror Because of a Photo.

638 Marwick and Lewis, Media Manipulation and Disinformation Online.

639 Human Rights Watch, All This Terror Because of a Photo.

Section 8 Questions:

- Have you or your organisation experienced coordinated online harassment, surveillance, doxxing, or other forms of abuse that appeared to be organised through private or encrypted digital channels?
- Are hidden digital infrastructures – such as encrypted messaging groups, anonymous forums, or illicit data sharing – playing an increasing role in anti-LGBTI+ mobilisation in your country or region?
- What practical digital security challenges do LGBTI+ organisations face when balancing the need for privacy, anonymity, and secure communications with the risks of misuse by hostile actors? Have data breaches, doxxing, impersonation, or the misuse of personal information affected LGBTI+ activists, organisations, or communities where you work?
- What support, tools, or resources would most strengthen the digital security and resilience of LGBTI+ activists and organisations in your context?
- What responsibilities should governments, technology companies, and encrypted messaging providers have in addressing coordinated digital harms while protecting privacy, freedom of expression, and secure communications?



Section 9: Predatory Communities and Gendered Exploitation

Sections 7 and 8 examined the organised dimensions of anti-LGBTI+ harm, the far-right networks, and the shadow infrastructure that coordinate it. This section shifts attention to a different but related phenomenon: the online communities in which gendered harm is not merely tolerated but actively rewarded, where exploitation becomes a measure of social status, and where the targets are not only LGBTI+ individuals but women, girls, and anyone whose gender expression or sexual identity deviates from a rigidly enforced norm.

Digital environments have not merely extended existing gendered harms; they

have reorganised them into new systems of visibility, participation, and reward. Within certain online communities – particularly those grouped under the loosely defined ‘manosphere’ – gender becomes a central axis of meaning-making, grievance, and social hierarchy.⁶⁴⁰ These spaces do not simply express misogyny or hostility toward gender and sexual diversity. They operationalise it through participation, competition, and technological amplification.⁶⁴¹

This section argues that predatory online communities function as competitive economies of gendered domination, where status is accrued through acts of

640 Are, Talbot, and Briggs, "Social Media Affordances of LGBTQIA+ Expression and Community Formation."

641 Castells, Networks of Outrage and Hope

symbolic and sometimes material harm. Within these ecosystems, masculinity is constructed through opposition – particularly toward women, feminist politics, and LGBTI+ visibility – and reinforced through peer validation, algorithmic amplification, and escalating norms of transgression.⁶⁴² Parallel to this ideological structure is a rapidly expanding infrastructure of digital sexual exploitation, including deepfake pornography, image-based abuse, and non-consensual distribution networks.⁶⁴³ These practices are not separate phenomena but mutually reinforcing dimensions of the same broader system of gendered control.

Toxic Masculinity Online: The Manosphere as a Grievance Ecosystem

The so-called ‘manosphere’ is best understood not as a single ideology but as a network of interconnected communities characterised by gender grievance, identity instability, and resentment toward feminist and queer inclusion. Incels, Men Going Their Own Way (MGTOW), men's rights activists, and pickup artist communities share overlapping discursive frameworks that position men as victims of a transformed gender order.⁶⁴⁴ These narratives frequently rely on a perceived

loss of status, sexual access, and social authority.⁶⁴⁵ Importantly, these communities are not static ideological formations. Research demonstrates significant fluidity between subgroups, with users migrating across platforms and communities in response to moderation pressures and social reinforcement.⁶⁴⁶ As communities become more isolated – moving from mainstream platforms to fringe forums or encrypted spaces – they tend to become more ideologically homogeneous and more extreme in their expression. This process produces what can be described as escalation through isolation, in which reduced external accountability intensifies internal norm formation.⁶⁴⁷ This mirrors the dynamic described in Section 8; shadow network architecture produces the same effect: Isolation intensifies extremism.

Participation itself plays a central role in ideological transformation. Studies of online misogynistic communities show that repeated engagement leads to measurable shifts in language, tone, and belief systems. Users progressively adopt more dehumanising rhetoric, sexual entitlement narratives, and dominance-based framings of social relations. Crucially, this is not merely passive exposure; it is participatory conditioning. Within these

communities, high-status users function as informal norm entrepreneurs, setting rhetorical and behavioural standards that others emulate in pursuit of recognition.⁶⁴⁸ While these ecosystems are often framed as ideological, their entry points are frequently emotional rather than political. Research indicates that many participants enter such spaces through experiences of loneliness, romantic rejection, and identity instability. These emotional conditions are subsequently reframed into externalised blame narratives in which feminism, women, and queer visibility become constructed as causal explanations for personal and collective dissatisfaction.⁶⁴⁹ In this sense, grievance functions as a conversion mechanism where affective vulnerability is transformed into ideological commitment.⁶⁵⁰

A key conceptual underpinning of these systems is resentment: the transformation of perceived powerlessness into moralised superiority. Within manosphere discourse, exclusion from sexual or social hierarchies is reinterpreted not as individual misfortune but as evidence of systemic injustice perpetrated by women or feminist culture.⁶⁵¹ This inversion allows participants to occupy a position of moral

victimhood while simultaneously justifying hostility toward others. Taken together, these dynamics reveal a system in which masculinity is not simply expressed but actively produced through structured grievance, peer reinforcement, and iterative participation.

Digital Exploitation: Deepfakes, Image-Based Abuse, and the Industrialisation of Harm

Parallel to ideological formation is a rapidly expanding infrastructure of digital sexual exploitation. Technologies such as deepfake generation, automated image manipulation, and non-consensual pornography distribution systems have enabled the scalable production of gendered harm.⁶⁵² Empirical reviews of deepfake technology consistently show that its primary deployment is gendered, with women and girls disproportionately targeted for sexualised manipulation. These practices are often motivated by control, humiliation, retaliation, or sexual entitlement rather than financial gain alone.⁶⁵³ Importantly, queer individuals also appear within this vulnerability structure, particularly through

642 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

643 Fisher et al., "Social Media."

644 Are, Talbot, and Briggs, "Social Media Affordances of LGBTQIA+ Expression and Community Formation."

645 Castells, Networks of Outrage and Hope

646 Are, Talbot, and Briggs, "Social Media Affordances of LGBTQIA+ Expression and Community Formation."

647 Marwick and Lewis, Media Manipulation and Disinformation Online.

648 Fisher et al., "Social Media."

649 Are, Talbot, and Briggs, "Social Media Affordances of LGBTQIA+ Expression and Community Formation."

650 boyd, It's Complicated.

651 Fraser, "Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy."

652 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

653 Human Rights Watch, All This Terror Because of a Photo.

mechanisms of visibility-based targeting and outing threats.⁶⁵⁴

One of the most significant findings in recent research is the normalisation of perpetration. Studies of deepfake creators reveal that many individuals minimise the harm of their actions, framing them as technologically mediated experimentation rather than interpersonal violence.⁶⁵⁵ The ease of access to generation tools reduces moral friction, enabling behaviours that would previously have required significant technical skill or social coordination. In this context, harm is not exceptional but structurally facilitated.⁶⁵⁶

This technological mediation also produces what may be termed the industrialisation of sexual harm. The proliferation of ‘nudification’ applications and synthetic media tools has lowered the barrier to producing non-consensual sexual imagery at scale. As a result, exploitation becomes less dependent on individual malice and more dependent on system accessibility. The architecture of harm is embedded within widely available tools.

The consequences for victims extend beyond immediate violation. Deepfake and image-based abuse generate epistemic instability in which individuals lose the ability to socially verify authenticity. Victims frequently report that the primary

harm is not only the existence of synthetic material but the inability to convincingly refute it within social or professional contexts.⁶⁵⁷ This destabilisation of identity and credibility produces long-term reputational vulnerability – the same truth destabilisation effect described in Section 5, now operating not at the level of political discourse but at the intimate level of individual lives.

At the structural level, these harms are transnational and persistent. Cross-national studies demonstrate widespread prevalence of non-consensual synthetic intimate imagery, alongside significant enforcement gaps. Platform reporting mechanisms are often inconsistent, under-resourced, and reliant on repeated victim labour. Content removal does not guarantee cessation, as material frequently reappears through reuploads and derivative modifications.⁶⁵⁸ Thus, digital sexual exploitation operates not as isolated incidents but as a distributed system of production, circulation, and persistence.

**Competition and Control:
Domination as Social Capital**

Beyond ideology and technology lies a third critical dimension, the transformation of harm into a competitive social system. Within many manosphere and ad-

jacent communities, domination functions as a form of social capital. Status is not derived from consensus or cooperation but from transgression, rhetorical aggression, and the ability to humiliate others.⁶⁵⁹

Large-scale analyses of online hate communities demonstrate that harassment is structured rather than random. Users gain visibility and status through repeated engagement in aggressive behaviours, with escalation often serving as a signal of in-group credibility.⁶⁶⁰ In this context, cruelty is not incidental but performative. It functions as a mechanism of social positioning.

Algorithmic amplification further intensifies these dynamics. Platform systems that prioritise engagement inadvertently reward emotionally charged, antagonistic content, thereby reinforcing dominance-oriented communication styles.⁶⁶¹ Influential users within these ecosystems function as gatekeepers of legitimacy, shaping norms of acceptable expression and competitive intensity. Competition is, therefore, not merely rhetorical but relational. Users are embedded in environments where identity is continuously evaluated through comparative aggression.⁶⁶²

The result is a feedback loop in which increasing extremity is rewarded with visibility, and visibility – in turn – incentivises further extremity.⁶⁶³ This structure helps explain why these communities often appear self-radicalising. It is not simply that extremist ideas spread, but that the social economy of attention rewards escalation itself.

Predatory Communities and Multi-Target Vulnerability

Although much discourse focuses on misogyny within these ecosystems, targeting is more accurately understood as opportunistic rather than singular. Women and girls remain primary targets of digital sexual exploitation, but LGBTI+ individuals are also vulnerable through mechanisms of visibility, stigma amplification, and outing threats.⁶⁶⁴

Contemporary reviews of deepfake technology and gender-based violence demonstrate that vulnerability is structurally produced through visibility and public presence. Individuals who are more visible – whether through public-facing roles, digital engagement, or non-normative gender expression – face increased risk of

654 Ibid,
655 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."
656 Ibid.
657 Human Rights Watch, All This Terror Because of a Photo.
658 Ibid.

659 Castells, Networks of Outrage and Hope
660 Fisher et al., "Social Media."
661 Marwick and Lewis, Media Manipulation and Disinformation Online.
662 Ibid.
663 Ibid.
664 Human Rights Watch, All This Terror Because of a Photo.

synthetic sexualisation and reputational harm.⁶⁶⁵

Importantly, the conceptual framing of such harm has evolved. The term ‘revenge porn’ has been widely criticised as insufficient, as it implies interpersonal motivation rather than structural coercion and control.⁶⁶⁶ A more accurate framing is image-based sexual abuse, which encompasses a broader range of motivations including coercion, entertainment, and dominance.⁶⁶⁷

The participatory nature of these ecosystems is further illustrated by communities in which users actively commission or request specific targets for synthetic sexual content. Investigative reporting on deepfake pornography platforms reveals structured economies of demand and supply where exploitation is collaboratively produced and socially rewarded.⁶⁶⁸

These dynamics demonstrate that predatory communities do not operate solely through ideological hostility or isolated acts of abuse. Rather, they function as distributed systems in which vulnerability is mapped, circulated, and exploited across multiple demographic groups.

Conclusion

Across ideological formation, technological infrastructure, and competitive social dynamics, a consistent structure emerges. Predatory online communities are not reducible to extremist ideology, individual deviance, or technological misuse. Instead, they constitute integrated ecosystems in which gendered harm is produced, circulated, and rewarded.

At the ideological level, grievance narratives transform personal insecurity into collective worldview. At the technological level, digital tools enable scalable – and often automated – forms of exploitation. At the social level, competition transforms harm into status, embedding violence within the logic of participation itself.

Within this system, women, girls, and LGBTI+ individuals are not incidental targets but structurally positioned points of vulnerability within a broader hierarchy of digital domination. Harm is not merely expressed within these environments; it is organised through them.

These findings suggest that addressing digital gendered harm requires more than content moderation or technological fixes. It requires confronting the deeper social economies of status, identity, and

grievance that render exploitation both meaningful and rewarding within these spaces. The following section examines what happens when these predatory systems – and the broader ecosystem of

anti-LGBTI+ harm – encounter the inadequacy of the legal and institutional frameworks meant to govern them and how that inadequacy has itself become a mechanism of normalisation.

Section 9 Key Findings:

- Predatory online communities have transformed gendered violence into a participatory and competitive social practice where harassment, humiliation, and exploitation are rewarded with status, visibility, and peer recognition rather than simply tolerated.
- Communities associated with the ‘manosphere’ function as interconnected grievance ecosystems that frame women, feminists, and LGBTI+ people as threats to traditional gender hierarchies, creating pathways towards misogyny, anti-LGBTI+ hostility, and broader extremist mobilisation.
- Image-based sexual abuse, deepfake pornography, and other forms of AI-enabled exploitation are becoming increasingly sophisticated, scalable, and accessible, allowing perpetrators to target individuals with unprecedented speed and reach.
- Women and girls remain the primary targets of these forms of abuse, but LGBTI+ people – particularly transgender people, gender-diverse individuals, public advocates, and others with high online visibility – are also disproportionately exposed to outing, reputational attacks, synthetic sexualisation, and other forms of technology-facilitated abuse.
- These harms are not isolated acts of individual misconduct. They are sustained through interconnected technological infrastructures, grievance narratives, and social reward systems that encourage participation and normalise exploitation.
- The convergence of misogyny, anti-LGBTI+ prejudice, and emerging technologies demonstrates that addressing digital gender-based violence requires more than content moderation. It demands coordinated responses that address platform design, legal protections, digital literacy, and the social norms that incentivise abuse.

665 Ibid.
666 EDRi, "Digital Rights of LGBTQ+ People."
667 Ibid.
668 Shoaib et al., "Deepfakes, Misinformation, and Disinformation."

Section 9 Questions:

- To what extent have online misogyny, anti-gender movements, or 'manosphere' communities contributed to hostility towards LGBTI+ people in your context?
- Have emerging technologies - including generative AI and deepfakes - changed the nature or scale of digital abuse faced by your organisation or the communities you support?
- What barriers do survivors encounter when reporting or seeking redress for technology-facilitated gender-based violence, including image-based abuse or synthetic sexual content?
- What practical measures should governments, technology companies, and international organisations take to better prevent and respond to these forms of abuse while protecting freedom of expression and privacy?
- Are there examples of effective laws, platform policies, survivor support services, or community-led initiatives that have successfully addressed technology-facilitated gender-based violence affecting LGBTI+ people?

IV. WHEN VIRTUAL TURNS INTO REALITY

Digital hostility does not remain contained within digital space. It shapes political discourse, legitimises discrimination, feeds legislative agendas, and can translate directly into physical violence. These final sections, therefore, confront the central paradox running through our report as a whole. The same infrastructures that allow LGBTI+ communities to organise, connect, and advocate for their rights are the infrastructures through which opposition to those rights is normalised, mobilised, and shielded from accountability.



Section 10: Shifting the Overton Window - Normalisation, Platform Abdication, and the Persistence of Harm

Sections 6 through 9 have described the ecosystem of anti-LGBTI+ harm in detail, including its ideological drivers, its far-right networks, its shadow infrastructure, and its predatory communities. A question that runs through all of them is: How do harms that are recognised as such - that violate platform policies, human rights norms, and in many cases criminal law - not only persist but grow? This section argues that persistence is not accidental. It is the product of three interconnected failures: conceptual frameworks that cannot adequately name the harm;

governance systems that cannot enforce across digital networks; and a feedback loop in which repeated exposure to harmful content normalises it, shifting what appears acceptable - the Overton window - in ways that ultimately serve political and anti-democratic ends.

Naming Harm: Conceptual Instability in the Digital Age

A central reason digital harms persist is that they are often difficult to define, cat-

egorise, and govern within existing conceptual frameworks.⁶⁶⁹ As technologies evolve, legal and cultural terminology frequently lag behind, producing categories that obscure as much as they reveal.⁶⁷⁰ This conceptual instability creates opportunities for institutional inaction, inconsistent enforcement, and the normalisation of harms that remain only partially legible within existing systems.⁶⁷¹

The term ‘revenge porn,’ for example, has been widely criticised for its narrow framing.⁶⁷² It implies interpersonal retaliation as the primary motivation, whereas empirical research shows that image-based sexual abuse encompasses a broader spectrum of motivations, including coercion, control, entertainment, financial, and opportunism.⁶⁷³ The conceptual shift toward "image based sexual abuse" reflects an attempt to capture this structural complexity rather than individualised intent.⁶⁷⁴

This definitional expansion is particularly important in the context of synthetic media. Deepfake technologies – introduced

in Section 5 as instruments of political propaganda and in Section 9 as tools of gendered exploitation – complicate traditional understandings of consent, authorship, and authenticity.⁶⁷⁵ Victims of AI-generated sexual content frequently report that the harm persists even when falsity is proven, as the social and reputational effects remain materially consequential regardless of veracity. In this sense, harm is not contingent on truth but on circulation and believability.⁶⁷⁶

From a gendered violence perspective, deepfake technologies function as amplifiers of pre-existing inequalities. Systematic reviews of deepfake abuse consistently demonstrate that women and girls are disproportionately targeted, with queer individuals also exposed through mechanisms of visibility, stigma, and outing vulnerability.⁶⁷⁷ These patterns suggest that synthetic media does not create new forms of harm out of nothing but intensifies existing hierarchies through scalable reproduction.⁶⁷⁸

Taken together, these developments illustrate a broader conceptual problem. Harm in digital environments is no longer a discrete act that can be cleanly named, but a fluid process that resists categorical closure.⁶⁷⁹ This matters not only because it complicates legal and regulatory responses but because uncertainty around the nature of harm creates space for minimisation, contestation, and inaction.⁶⁸⁰ In this way, conceptual instability becomes one of the conditions through which harmful practices persist and gradually become normalised.⁶⁸¹

Jurisdictional Lag: When Law Becomes Geographically Obsolete

Legal systems remain fundamentally organised around territorial jurisdiction. Digital harm is not.⁶⁸²

Large-scale cross-national research on non-consensual synthetic intimate imagery demonstrates widespread prevalence across multiple jurisdictions, alongside significant disparities in enforcement capacity.⁶⁸³ Perpetrators frequently operate

transnationally, exploiting differences in legal frameworks, platform policies, and enforcement mechanisms.⁶⁸⁴

This fragmentation produces a structural enforcement gap. Even when harmful content is identified and removed in one jurisdiction, it may persist in others or reappear through reuploads, mirrors, and derivative modifications.⁶⁸⁵ The result is not resolution but displacement.

Compounding this issue is the burden placed on victims within reporting systems. Expert analyses of image-based abuse reporting processes highlight that victims are often required to repeatedly identify, report, and track harmful content across platforms.⁶⁸⁶ Enforcement is, therefore, not only inconsistent but also labour-intensive, effectively shifting institutional responsibility onto individuals already experiencing harm.⁶⁸⁷

The result is a system in which governance is formally present but practically incomplete. Harm becomes jurisdictionally mobile while accountability remains territorially fixed.⁶⁸⁸

669 OECD, Rights in the Digital Age.

670 Riedl, "Rethinking Rights in Social Media Governance."

671 Birdsall, "Human Rights-Based Approach to Social Media Platforms."

672 Fisher et al., "Social Media."

673 Ibid.

674 Ibid.

675 Rebecca Dorn, Lee Kezar, Fred Morstatter, and Kristina Lerman, Harmful Speech Detection by Language Models Exhibits Gender-Queer Dialect Bias, <https://doi.org/10.48550/arXiv.2406.00020>

676 Ibid.

677 Fisher et al., "Social Media."

678 Dorn et al., Harmful Speech Detection by Language Models Exhibits Gender-Queer Dialect Bias.

679 Ibid.

680 Riedl, "Rethinking Rights in Social Media Governance."

681 Ibid.

682 Land, "Toward a Human Rights Approach."

683 Fisher et al., "Social Media," [page].

684 OECD, Rights in the Digital Age.

685 Human Rights Watch, All This Terror Because of a Photo.

686 Gomez et al., Algorithmic Arbitrariness in Content Moderation.

687 Ibid.

688 Riedl, "Rethinking Rights in Social Media Governance."

For LGBTI+ communities and the organisations that support them, this enforcement gap has consequences beyond individual cases. It creates an environment in which harmful actors can exploit jurisdictional fragmentation to evade accountability, while victims and civil society organisations are left to shoulder the burden of monitoring, reporting, and advocacy.

The challenge is, therefore, not simply legal but collective. Because digital harms transcend national boundaries, responses that remain confined within them are unlikely to succeed.⁶⁸⁹ This suggests that effective resistance increasingly depends on transnational forms of coordination between civil society organisations, researchers, platforms, and policymakers.

While no single intervention can eliminate jurisdictional fragmentation, collective action can reduce the spaces in which harmful actors operate and increase pressure for more coherent forms of accountability.⁶⁹⁰ In this sense, resilience depends not only on stronger laws or platform policies but on the capacity of civil society organisations to build transnational networks of monitoring, evidence-sharing, and collective advocacy.⁶⁹¹ Jurisdictional fragmentation cannot be eliminated by any single actor, but its effects can be mitigated

when communities coordinate across the same borders that harmful networks routinely ignore.

Platform Governance: The Architecture of Amplification and Abdication

While legal systems struggle with jurisdictional mismatch, platform systems introduce a different form of governance failure: structural delegation.⁶⁹² Platforms simultaneously shape visibility while disclaiming responsibility for outcomes.

Algorithmic systems designed to maximise engagement systematically privilege emotionally intense content.⁶⁹³ Research in computational social science demonstrates that outrage-driven content spreads more rapidly and more widely than neutral or informational content. This creates an environment in which antagonistic discourse – including gendered hostility and anti-LGBTI+ rhetoric – is structurally advantaged in terms of visibility.⁶⁹⁴

These dynamics are reinforced by norm formation processes within digital communities. Studies of online norm degradation show that early exposure to harmful or extreme content can recalibrate community standards, making previously unacceptable behaviours increasingly

normalised over time.⁶⁹⁵ In this way, visibility itself becomes a mechanism of norm shift. This is the Overton window in operation: not a sudden ideological shift but a gradual, algorithmic drift in what appears mainstream.⁶⁹⁶

Platforms, therefore, operate within a paradoxical governance model where they deny editorial responsibility while actively shaping the conditions under which content is encountered, amplified, and normalised.⁶⁹⁷ This tension is particularly significant in the context of gendered harm, where visibility and amplification directly affect risk exposure.⁶⁹⁸

The X/Twitter case – introduced in Section 3 as an example of moderation reduction and in Section 6 as an accelerant of anti-LGBTI+ hostility – is relevant here in its third dimension: the normalisation effect. When previously banned accounts were reinstated and moderation infrastructure was dismantled, the immediate consequence was increased harassment.⁶⁹⁹ However, the deeper consequence was a recalibration of what was considered acceptable discourse. The Overton window shifted, not through persuasion but

through algorithmic exposure. What had been fringe became familiar.⁷⁰⁰

For LGBTI+ movements, the implications are not limited to platform governance but extend to the defence of social norms themselves. If repeated exposure can shift perceptions of what is acceptable, then countering digital harm is not simply a matter of enforcement but of narrative contestation.⁷⁰¹ Civil society organisations, journalists, researchers, and community leaders, therefore, play a critical role in identifying harmful narratives, documenting their circulation, and challenging their legitimacy before they become embedded within mainstream discourse. The task is not only to respond to harassment after it occurs but to resist the gradual normalisation that makes such harassment appear ordinary, inevitable, or politically acceptable.⁷⁰²

Harm as Persistence: The Failure of Removal

One of the defining features of digital sexual exploitation is its resistance to removal as a meaningful form of remediation.⁷⁰³ Unlike physical-world harms, dig-

689 Land, "Toward a Human Rights Approach."
690 Gomez et al., Algorithmic Arbitrariness in Content Moderation.
691 Ibid.
692 Gillespie, Custodians of the Internet.
693 Marwick and Lewis, Media Manipulation and Disinformation Online.
694 Center for Countering Digital Hate, Digital Hate.

695 Fisher, The Chaos Machine.
696 Pariser, The Filter Bubble.
697 Gillespie, Custodians of the Internet.
698 Human Rights Watch, All This Terror Because of a Photo.
699 Center for Countering Digital Hate, Digital Hate.
700 Pariser, The Filter Bubble.
701 Castells, Networks of Outrage and Hope
702 Ibid.
703 Fisher et al., "Social Media."

ital content is infinitely replicable, easily modified, and rapidly redistributed.

Research on deepfake pornography and synthetic media ecosystems highlights the accessibility of generative tools and the ease with which content can be reproduced at scale.⁷⁰⁴ This accessibility reduces the friction traditionally associated with producing harmful material, shifting the problem from isolated acts of creation to continuous systems of circulation.⁷⁰⁵

Interviews and qualitative analyses of perpetrators and victims of synthetic sexual content further reveal a troubling normalisation of harm. Perpetrators often describe their actions as low-stakes or experimental, facilitated by anonymity and perceived lack of consequences. Meanwhile, victims report that the primary harm is not only the existence of content but its persistence across platforms and contexts.⁷⁰⁶

Removal, in this environment, functions less as resolution and more as temporary containment. Harm is not eliminated; it is redistributed across networks.⁷⁰⁷ This persistence is itself normalising; it teaches bystanders, perpetrators, and platforms alike that such content is an unavoidable

feature of the digital landscape rather than an accountable harm.

The same dynamic is visible in the circulation of misinformation. False or misleading claims about LGBTI+ people often continue to spread long after they have been debunked or corrected.⁷⁰⁸ As with synthetic sexual content, the persistence of misinformation means that removal or correction rarely eliminates its effects.⁷⁰⁹ Repeated exposure can increase familiarity, reinforce existing prejudices, and create the impression that contested claims possess greater legitimacy than they do. In this way, persistence transforms visibility into credibility, allowing harmful narratives to remain influential even when their factual basis has been disproven.⁷¹⁰

The persistence of harmful content is often presented as an inevitable consequence of technological complexity. Yet, this explanation is only partial. Platforms routinely demonstrate the capacity to detect, prioritise, and remove content when sufficiently strong incentives exist: whether to protect commercial interests, comply with regulatory requirements, or safeguard platform functionality. The relative persistence of image-based abuse, targeted harassment, and anti-LGBTI+

hostility, therefore, raises questions not only about technical feasibility but about institutional priorities.⁷¹¹

The current system distributes responsibility in a way that is advantageous to platforms themselves. Rather than proactively identifying and disrupting harmful networks, many reporting systems rely heavily on victims to locate content, submit complaints, and repeatedly engage with moderation processes.⁷¹² This approach reduces operational costs while preserving the appearance of governance. Harm remains visible, engagement continues to be generated, and responsibility is shifted onto those experiencing the abuse.⁷¹³

For LGBTI+ communities, the consequences are particularly acute. The persistence of harmful content communicates that abuse is not an exceptional failure of the system but an expected feature of participation within digital spaces. Over time, this can produce a chilling effect in which individuals limit their visibility, moderate their expression, or withdraw from public engagement altogether. The result is not only individual harm but a narrowing of civic participation. Communities that already face disproportionate hostility become increasingly responsible for managing the risks associated with their own visibility.⁷¹⁴

Those who benefit most from this arrangement are not necessarily individual perpetrators but the broader ecosystems within which harmful content and misinformation circulate. Engagement-driven platforms continue to profit from attention, harmful networks retain access to audiences, and political actors seeking to normalise anti-LGBTI+ narratives encounter fewer barriers to dissemination. Engagement-driven platforms continue to profit from attention, harmful networks retain access to audiences, and political actors seeking to normalise anti-LGBTI+ narratives encounter fewer barriers to dissemination. In this sense, the failure of removal is not simply a moderation problem. It is one of the mechanisms through which harmful content remains present long enough to become familiar enough to become normalised.

The strategic implication for LGBTI+ movements is that removal alone cannot be treated as a sufficient measure of success. While rapid content takedowns remain important, meaningful protection also requires interventions that address the systems through which harmful material is reproduced, amplified, and recirculated.⁷¹⁵ This includes stronger platform accountability, improved detection and monitoring mechanisms, and collective efforts by civil society organisations to document and challenge persistent pat-

704 Dorn et al., Harmful Speech Detection by Language Models Exhibits Gender-Queer Dialect Bias.

705 Ibid

706 Ibid.

707 Ibid.

708 Marwick and Lewis, Media Manipulation and Disinformation Online.

709 Fisher, The Chaos Machine.

710 Pariser, The Filter Bubble.

711 Gillespie, Custodians of the Internet.

712 Human Rights Watch, All This Terror Because of a Photo.

713 Ibid.

714 Tahrir Institute for Middle East Policy, "Digital Persecution."

715 Gillespie, Custodians of the Internet.

terns of abuse.⁷¹⁶ Without addressing the structures that sustain circulation, removal risks becoming an endless cycle of reaction rather than a durable form of prevention.⁷¹⁷

Epistemic Harm: The Collapse of Verifiable Reality

Beyond material and reputational damage, digital predatory systems produce a more subtle – but far-reaching – form of harm: the erosion of trust in evidence itself.⁷¹⁸ Synthetic media, manipulated imagery, and AI-generated content increasingly blur the distinction between what is real and what is fabricated, making it more difficult for individuals, institutions, and communities to establish a shared understanding of reality.⁷¹⁹

Research on AI-generated sexual content demonstrates that victims frequently encounter difficulties disproving false material once it has entered circulation. Even when technical verification is available, social belief does not necessarily follow.⁷²⁰ The challenge is, therefore, not

simply establishing the truth but persuading others to accept it. Harmful content can continue to shape perceptions long after its falsity has been demonstrated.⁷²¹

For LGBTI+ communities, this problem extends beyond individual cases of image-based abuse. Anti-LGBTI+ movements have long relied on misinformation, moral panic, and false narratives to justify exclusion and discrimination.⁷²² Synthetic media and AI-generated content expand the range of tools available to those seeking to construct and disseminate such narratives.⁷²³ As the volume of manipulated content increases, the task of distinguishing between authentic evidence and deliberate fabrication becomes increasingly difficult.⁷²⁴

This uncertainty creates opportunities for bad-faith actors. Genuine evidence can be dismissed as fake, while fabricated material can be defended as plausible. The result is what scholars describe as the liar's dividend.⁷²⁵ Introduced in Section 5 in the context of political propaganda, the liar's dividend operates equally at the interpersonal level. Individuals targeted

by image-based abuse may struggle to convince others that fabricated content is false, while perpetrators can exploit ambiguity to evade accountability.⁷²⁶

For LGBTI+ movements, the consequences are significant. Advocacy, human rights monitoring, and community organising all depend upon the ability to document harm and establish credibility.⁷²⁷ When trust in evidence deteriorates, it becomes easier for hostile actors to challenge documented abuses, discredit victims, and undermine community testimony.⁷²⁸ The issue is, therefore, not only whether specific content is true or false, but whether communities retain the capacity to establish shared facts in the face of deliberate manipulation.⁷²⁹

The strategic challenge is not simply combating individual instances of misinformation or synthetic abuse but strengthening the institutions, networks, and verification mechanisms that enable trust. As with other forms of digital harm, the question is collective rather than individual. Resilience depends upon the ability of communities, researchers, journalists, civil society organisations, and public institutions to defend credible forms of evidence against systems designed to manufacture doubt.⁷³⁰

Institutional Strain: Delegated Enforcement and Unequal Burden

As legal and platform systems struggle to contain digital harm, responsibility increasingly shifts onto victims. Reporting systems require continuous engagement, documentation, and follow-up, often across multiple platforms and jurisdictions.⁷³¹

Studies of image-based abuse reporting processes demonstrate that victims frequently encounter inconsistent enforcement outcomes and repeated exposure to the same harmful content through reuploads. This creates a cyclical burden in which harm management becomes an ongoing obligation rather than a finite process.⁷³²

Cross-national research further shows significant disparities in enforcement capacity, with some jurisdictions lacking adequate legal mechanisms or institutional resources to address synthetic and image-based abuse effectively. These disparities reinforce global inequalities in protection and recourse.⁷³³

The result is a governance system in which enforcement exists but is unevenly distributed, procedurally fragmented,

716 Ibid.

717 Ibid.

718 OECD, Rights in the Digital Age.

719 Ibid.

720 Dorn et al., Harmful Speech Detection by Language Models Exhibits Gender-Queer Dialect Bias.

721 Ibid.

722 Institute for Strategic Dialogue, Five-Year Overview.

723 Marwick and Lewis, Media Manipulation and Disinformation Online.

724 OECD, Rights in the Digital Age.

725 Gomez et al., Algorithmic Arbitrariness in Content Moderation.

726 Dorn et al., Harmful Speech Detection by Language Models Exhibits Gender-Queer Dialect Bias.

727 Human Rights Watch, All This Terror Because of a Photo.

728 Tahrir Institute for Middle East Policy, "Digital Persecution."

729 Marwick and Lewis, Media Manipulation and Disinformation Online.

730 Castells, Networks of Outrage and Hope

731 Human Rights Watch, All This Terror Because of a Photo.

732 Gomez et al., Algorithmic Arbitrariness in Content Moderation.

733 Fisher et al., "Social Media."

and heavily dependent on victim labour. This institutional abdication is not merely an administrative failure; it is a political signal.⁷³⁴ It communicates to perpetrators that their actions carry minimal consequence; to potential victims that the systems meant to protect them will not; and to anti-rights movements more broadly that sustained harassment, coordinated intimidation, and norm-breaking tactics remain viable instruments of political action.⁷³⁵

Conclusion

This section has argued that digital predatory systems persist not simply because they are under-regulated but because they exceed the conceptual and institutional frameworks designed to govern them – and because that excess is itself normalising.

Three structural conditions define this governance failure. First, conceptual categories struggle to adequately capture the fluid and synthetic nature of contemporary harm. Second, jurisdictional fragmentation prevents coherent enforcement across transnational networks. Third, platform governance systems simultaneously

shape and evade responsibility for the amplification of harmful content.

Together, these conditions produce a form of harm that is persistent, mobile, and only partially legible within existing legal structures. Digital predatory systems, therefore, do not exist in a vacuum of governance; rather, they exist within a governance architecture that is structurally misaligned with their operational logic.

The result is not merely a regulatory gap but a cultural shift. As extreme content circulates, as enforcement fails, as victims bear the burden of their own protection, the window of what appears acceptable expands. What was once fringe becomes debated; what was once debated becomes mainstream; what was once scandalous becomes unremarkable. This is how the Overton window moves: not through argument but through exposure, repetition, and the quiet abdication of institutional responsibility.

The following sections examine what this normalisation looks like when it enters formal politics; when the same dynamics that produce digital harassment campaigns produce electoral strategies, legislative agendas, and authoritarian consolidation.

Section 10 Key Findings:

- Coordinated digital campaigns shape public discourse and political agendas by normalising hostility towards LGBTI+ people, influencing media narratives, and creating public support for restrictive laws and policies.
- Persistent online abuse has a significant chilling effect on civic participation. Many activists, journalists, politicians, and community leaders reduce their public engagement, self-censor, or withdraw from advocacy because of sustained digital harassment and security concerns.
- Digital hostility disproportionately affects those facing multiple forms of marginalisation, including transgender people, racialised LGBTI+ communities, women, young people, refugees, and human rights defenders, thus reinforcing existing inequalities and limiting democratic participation.
- The distinction between online and offline violence is becoming increasingly blurred. Digital abuse often serves as an early warning indicator for broader patterns of political radicalisation, organised intimidation, and physical violence.

Section 10 Questions:

- Which communities within the LGBTI+ movement are most affected by the transition from online hostility to offline harm, and what factors contribute to their increased vulnerability?
- Have digital campaigns influenced public opinion, political debate, or legislative developments affecting LGBTI+ rights in your country or region? If so, how?
- What measures have been most effective in protecting LGBTI+ activists and organisations from the real-world impacts of coordinated digital hostility?
- What additional actions should governments, technology companies, international organisations, and donors take to strengthen accountability and prevent digital harms from escalating into offline violence or discrimination?

⁷³⁴ Gomez et al., Algorithmic Arbitrariness in Content Moderation.

⁷³⁵ Ibid.



Section 11: Political Parties, Power, and Online Mobilisation

Throughout this report we have described the ecosystem of anti-LGBTI+ harm: the drivers, the networks, the shadow infrastructure, the predatory communities, and the normalisation mechanisms. This section now asks what happens when all of this exits the online world and enters the formal structures of political power; when digital outrage becomes electoral strategy; when harassment campaigns become legislative agendas; and when what was manufactured online becomes real policy affecting real lives.

Political power in the digital age is increasingly exercised through platforms that mediate attention, structure visibility, and shape the emotional conditions of public life. Unlike earlier

media systems – where political communication was filtered through relatively

stable institutions such as newspapers, broadcasters, and party organisations – contemporary political mobilisation is embedded within algorithmically governed environments. These systems do not merely transmit political messages; they actively shape which messages become politically consequential.

Within this environment, minority groups – particularly LGBTI+ communities – have become recurrent focal points of political mobilisation. Their visibility within online discourse is not incidental but structurally incentivised through platform logics that reward emotional intensity, controversy, and moral polarisation. This section examines how digital populism, microtargeting, platformed politics, coordinated disinformation, and hybrid repression converge to produce a political ecosystem in which LGBTI+ rights function as both symbolic and strategic terrain.

The central argument is that digital political systems do not simply reflect pre-existing ideological conflict; they actively reorganise it, converting identity-based tensions into scalable mechanisms of mobilisation and governance.

Digital Populism: Outrage as Political Infrastructure

Digital populism refers to a form of political communication in which actors bypass institutional intermediaries and engage directly with publics through emotionally charged digital environments.⁷³⁶ Rather than being a stylistic adaptation of traditional populism, digital populism represents a structural shift in how political legitimacy is constructed.

Cambridge University Press' research on digital populism emphasises that platform architectures enable leaders to communicate directly with 'the people,' reducing the role of journalistic and institutional gatekeepers while amplifying effective messaging strategies.⁷³⁷ Populist communication in this context is characterised by moral simplification, personalised blame attribution, and the construction of binary political identities.⁷³⁸

Empirical work in *Social Media + Society* further demonstrates that populist messaging is particularly effective in environments where emotional content is algorithmically prioritised.⁷³⁹ Emotional intensity – particularly outrage – functions as a key driver of engagement, giving populist actors structural advantages in visibility and dissemination.⁷⁴⁰

The Brookings Institution research reinforces this dynamic, showing that digital platforms systematically reward content that generates high levels of interaction, often privileging emotionally polarising narratives over deliberative or policy-oriented communication.⁷⁴¹ Within this ecosystem, minority groups frequently become symbolic targets through which broader political grievances are articulated. Reports from the Institute for Strategic Dialogue illustrate how anti-LGBTI+ narratives are often reframed in ostensibly neutral terms such as 'child protection' or 'cultural preservation,' enabling ideological content to circulate across mainstream and fringe audiences simultaneously.⁷⁴²

Taken together, these findings suggest that digital populism is not merely a rhetorical style but an infrastructural condition in which emotional mobilisa-

736 Griffin, "Rethinking Rights in Social Media Governance."

737 Ibid.

738 Ibid.

739 Oz, Yanik, and Batu, "Under the Shadow of Culture and Politics."

740 Oz, Yanik, and Batu, "Under the Shadow of Culture and Politics."

741 Brooke Tanner, Cameron F. Kerry, Andrew W. Wyckoff, Nicoleta Kyosovska, Andrea Renda, and Elham Tabassi, *Is AI Sovereignty Possible? Balancing Autonomy and Interdependence* (Washington, DC: Brookings Institution, 2026), https://www.brookings.edu/wp-content/uploads/2026/02/20260217_AI_sovereignty_final.pdf.

742 Institute for Strategic Dialogue, *Five-Year Overview*.

tion becomes a primary route to political power.⁷⁴³ And LGBTI+ people are a particularly efficient fuel for that mobilisation.⁷⁴⁴ What appears here as digital populism is, therefore, not a discrete phenomenon but the political consolidation of dynamics traced throughout this report, including algorithmic amplification, grievance-based identity formation, shadow coordination, predatory participation, and weak accountability converging into a durable infrastructure of anti-rights mobilisation.⁷⁴⁵

Microtargeting and Algorithmic Manipulation

Alongside emotional mobilisation, contemporary political and influencing campaigns increasingly rely on data-driven microtargeting systems that enable personalised political messaging at scale.⁷⁴⁶ The Cambridge Analytica scandal revealed how large-scale data harvesting and psychometric profiling were used to construct psychological voter profiles for targeted political advertising.⁷⁴⁷

A key foundation for these practices is the work of Michal Kosinski and colleagues, which demonstrates that digital behavioural data can be used to infer personality traits with significant predictive accuracy.⁷⁴⁸ While originally developed as a methodological contribution in computational psychology, these techniques have been operationalised in political contexts to enable granular segmentation of electorates.⁷⁴⁹

The democratic implications of such systems are significant. Research in political science highlights that microtargeting fragments the public sphere by enabling different groups of voters to receive different versions of political messaging, thus undermining shared deliberative conditions necessary for democratic accountability. In effect, political discourse becomes personalised, opaque, and asymmetrical.⁷⁵⁰

Analyses of targeted advertising systems further show that algorithmic optimisation can reproduce social inequalities by delivering different political content to different demographic groups.⁷⁵¹ This cre-

ates a system in which political persuasion is not only individualised but structurally differentiated. As demonstrated by Pew Research findings on political advertising, which show widespread uncertainty among users regarding how and why they are targeted, public awareness of these processes remains limited. This informational asymmetry contributes to a broader erosion of transparency in democratic communication.⁷⁵²

The Cambridge Analytica case is worth dwelling on as a concrete illustration of how surveillance capitalism – the extractive data economy described in Section 3 – flows directly into political manipulation.⁷⁵³ The data trails generated by ordinary online behaviour were harvested, profiled, and weaponised for forms of targeting in which the boundary between commercial advertising and political persuasion became increasingly unstable, as techniques initially developed to influence consumer behaviour were then repurposed to shape civic behaviour, electoral preference, and ideological alignment. The same infrastructure that enables platform-based community building also enables the construction of precision political influence operations.⁷⁵⁴

Platformed Politics: Movements as Networked Systems

Political mobilisation is increasingly structured by platform infrastructures that shape how movements form, scale, and persist.⁷⁵⁵ Rather than emerging primarily from institutional organisations, contemporary movements are often assembled through networked digital interactions.⁷⁵⁶ This is not to suggest that human rights movements have historically been born within institutions; many emerged in opposition to them. Rather, digital infrastructures now accelerate and reorganise how publics cohere, how grievances circulate, and how collective identity is assembled before, beyond, or against formal organisational structures.⁷⁵⁷

Scholarship in *Social Media + Society* argues that platforms actively shape political participation by structuring visibility hierarchies and enabling rapid diffusion of content. The architecture of platforms, therefore, plays a constitutive role in political organisation.⁷⁵⁸ The Oxford *Handbook of Digital Politics* similarly emphasises that digital environments have decentralised political mobilisation, reducing the monopoly of traditional institutions over

743 Griffin, "Rethinking Rights in Social Media Governance."
744 Institute for Strategic Dialogue, Five-Year Overview.
745 Griffin, "Rethinking Rights in Social Media Governance."
746 Griffin, "Rethinking Rights in Social Media Governance."
747 Ibid.
748 Kosinski et al.
749 Ibid.
750 Griffin, "Rethinking Rights in Social Media Governance."
751 Ibid.

752 Ibid.
753 Zuboff, The Age of Surveillance Capitalism.
754 Ibid.
755 Castells, Networks of Outrage and Hope
756 Ibid.
757 Ibid.
758 Oz, Yanik, and Batu, "Under the Shadow of Culture and Politics."

agenda-setting and enabling new forms of networked political organisation.⁷⁵⁹

Empirical research in *Nature Human Behaviour* demonstrates that political mobilisation is strongly influenced by network exposure effects, as individuals are significantly more likely to engage in political action when exposed to mobilisation signals within their social networks.⁷⁶⁰ This suggests that political behaviour is increasingly contingent on relational and algorithmic exposure rather than stable ideological commitment.⁷⁶¹ Reports from the Institute for Strategic Dialogue further illustrate how online mobilisation processes can rapidly amplify both political and counter-political movements, particularly in contentious identity-based domains such as LGBTI+ rights.⁷⁶²

Coordinated Disinformation and Civil Society Delegitimisation

A defining feature of contemporary political environments is the strategic use of coordinated disinformation to undermine civil society organisations and human rights defenders. Human Rights Watch

documents increasing levels of harassment and delegitimisation campaigns targeting NGOs, often framed as foreign interference or ideological extremism.⁷⁶³

The Office of the High Commissioner for Human Rights identifies civic space as a foundational condition for democratic governance, emphasising that attacks on civil society actors constitute structural threats to democratic institutions.⁷⁶⁴ Research from the Global Disinformation Index highlights the economic dimensions of disinformation ecosystems, showing that misleading content is often sustained through monetisation structures that incentivise engagement-driven amplification.⁷⁶⁵

Further analysis demonstrates that disinformation campaigns are often modular and repeatable, relying on adaptable narrative templates that can be deployed across different contexts and audiences.⁷⁶⁶ Anti-Defamation League reporting illustrates how anti-LGBTI+ narratives are frequently reused across platforms and adapted to local political contexts, reinforcing their durability and scalability.⁷⁶⁷

Together, these findings suggest that disinformation functions not as isolated misinformation events but as structured systems of narrative competition designed to undermine institutional trust.⁷⁶⁸ The Uganda case – encountered in Section 6 as an example of information warfare and now seen here in its political dimension – demonstrates how digital narratives framing LGBTI+ organisations as foreign agents translate directly into legislative action and state repression.⁷⁶⁹ The online and the institutional are not separate domains. They are continuous.

Hybrid Repression: State and Platform Convergence

Contemporary forms of repression increasingly operate through hybrid systems that combine state power, private platform governance, and surveillance infrastructures.⁷⁷⁰ Freedom House reports document global increases in digital censorship, surveillance, and restrictions on online expression.⁷⁷¹

Human Rights Watch further highlights the growing collaboration between states

and private technology companies in implementing digital surveillance and repression systems.⁷⁷² ARTICLE19 similarly argues that digital authoritarianism is characterised by the convergence of legal and technological tools that enable pre-emptive control over expression and association.⁷⁷³

Within these systems, minority groups often become symbolic targets for political consolidation.⁷⁷⁴ Research on authoritarian populism demonstrates that identity-based scapegoating is frequently used to justify expansions of state authority and restrictions on civil liberties.⁷⁷⁵ This convergence of state and non-state power produces a governance environment in which repression is distributed across multiple institutional layers rather than concentrated in a single actor.⁷⁷⁶

The Chechnya example – which opened this paper as an illustration of digital activism enabling international attention – reappears here in its starkest form. The same digital infrastructure that allowed survivors' testimonies to reach global audiences also enabled state identifica-

759 Stephen Coleman and Lone Sorensen, eds., *Handbook of Digital Politics*, 2nd ed. (Cheltenham, UK: Edward Elgar Publishing, 2023).

760 Fisher et al., "Social Media."

761 Ibid.

762 Institute for Strategic Dialogue, *Five-Year Overview*.

763 Human Rights Watch, *All This Terror Because of a Photo*.

764 Office of the United Nations High Commissioner for Human Rights, "Civic Space," accessed June 30, 2026, <https://www.ohchr.org/en/civic-space>

765 Ibid.

766 Tufekci, *Social Media and Human Rights*.

767 Ibid.

768 Ibid.

769 Human Rights Watch, *All This Terror Because of a Photo*.

770 Freedom House, *Freedom in the World 2026: The Growing Shadow of Autocracy* (Washington, DC: Freedom House, 2026), <https://freedomhouse.org/report/freedom-world/2026/growing-shadow-autocracy>

771 Ibid.

772 Human Rights Watch, *All This Terror Because of a Photo*.

773 ARTICLE 19. *Digital Rights*.

774 Royal United Services Institute, *Online Gendered Narratives*.

775 Freedom House, *Freedom in the World 2026*.

776 Ibid.

tion of victims.⁷⁷⁷ And in the decade since, the hybrid repression model has become more sophisticated, more globalised, and more normalised.

Democratic Fragility and the Centrality of LGBTI+ Rights

The cumulative effect of these dynamics is a measurable erosion of democratic resilience. Freedom House data documents sustained global declines in political rights and civil liberties.⁷⁷⁸ Brookings’ analysis further shows that minority rights often deteriorate prior to broader institutional breakdown, suggesting that they function as early indicators of democratic backsliding.⁷⁷⁹

Scholarship in Perspectives on Politics argues that culture war politics weakens institutional trust and facilitates executive consolidation of power.⁷⁸⁰ Empirical research further demonstrates a strong correlation between LGBTI+ rights protections and broader indicators of democratic stability.⁷⁸¹

Within this framework, LGBTI+ rights function not as peripheral cultural issues but as diagnostic indicators of democratic health. Their erosion signals not only so-

cial exclusion but systemic weakening of pluralistic governance structures.

Conclusion

This section has argued that digital political systems reorganise power through interconnected mechanisms of emotional mobilisation, data-driven targeting, platform infrastructure, disinformation, and hybrid repression. Within these systems, LGBTI+ communities are frequently positioned as strategic focal points through which political, ideological, and institutional conflicts are expressed, and where the acceptability of hostility, exclusion, and norm-breaking is tested.

Rather than operating as isolated phenomena, these mechanisms form an integrated system in which political communication, technological infrastructure, and institutional governance are mutually constitutive. The result is a political environment in which democratic processes are increasingly shaped by platform dynamics that privilege emotional intensity, structural asymmetry, and narrative fragmentation, and in which online hostility routinely leaks into offline consequences: harassment, threat, and violence that reshape who can safely participate in public life.

Digital populism and anti-rights mobilisation are, therefore, not confined to the sphere of ‘discourse.’ The infrastructures that reward attention and outrage also function as political infrastructures for harm, supplying narratives that justify violence, building networks capable of directing it, and normalising the targeting of specific groups as an ordinary expression of grievance. What appears on the feed has material effects on the distribution of fear, risk, and security; certain communities are made hyper-visible as targets, while others are encouraged to see their aggression as legitimate participation in democratic debate.

Understanding these dynamics is essential not only for analysing contemporary political mobilisation but for assessing the long-term resilience of democratic systems in an era of platformed power. It also clarifies the stakes of the accountability gap; when the same infrastructures that enable mobilisation and visibility also facilitate coercion, intimidation, and physical harm with limited consequence for those responsible, impunity itself becomes a feature of the system. Which raises, unavoidably, the question the final section must address: If all of these harms are visible, documented, and acknowledged, why does nothing hold anyone to account?

Section 11 Key Findings:

- Protecting LGBTI+ rights in digital spaces requires a ‘whole-of-society’ approach that recognises digital rights as an integral part of international human rights and democratic governance.
- Governments have a responsibility to establish legal and regulatory frameworks that protect freedom of expression, privacy, equality, and non-discrimination while ensuring accountability for technology-enabled harms.
- Technology companies should adopt a rights-based approach to platform governance by improving transparency, strengthening content moderation, conducting human rights due diligence, and designing products that prioritise user safety, particularly for marginalised communities.
- International cooperation is essential. The transnational nature of digital platforms and anti-rights movements means that effective responses require collaboration between governments, international organisations, technology companies, and civil society across jurisdictions.
- Building resilient digital ecosystems requires investment in digital security, research, capacity-building, and support for LGBTI+ human rights defenders, ensuring that communities are equipped to respond to emerging technologies and evolving forms of online harm.

777 Human Rights Watch, All This Terror Because of a Photo.

778 Freedom House, Freedom in the World 2026.

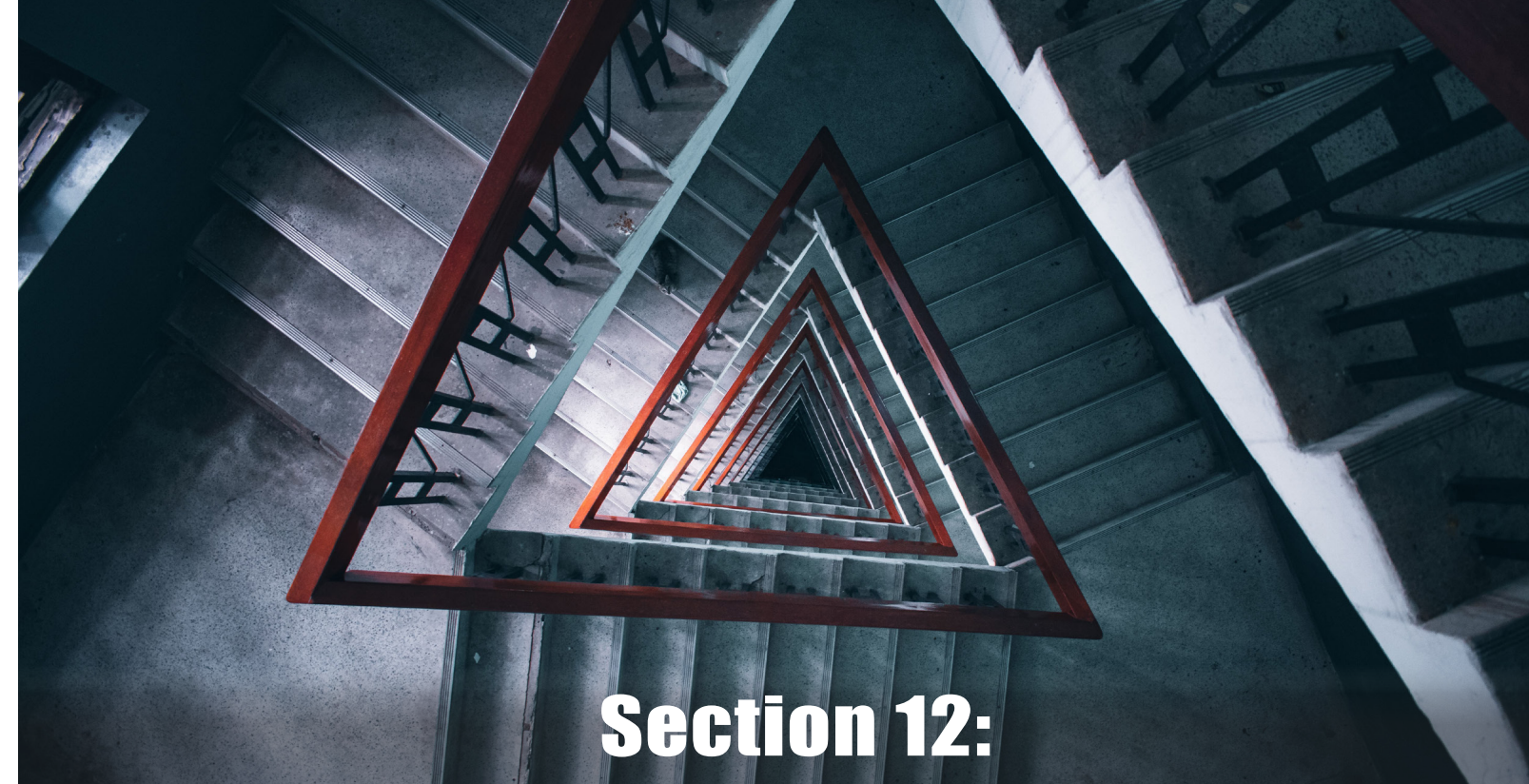
779 Tanner et al., Is AI Sovereignty Possible?

780 John T. Jost, Delia S. Baldassarri, and James N. Druckman, "Cognitive-Motivational Mechanisms of Political Polarization in Social-Communicative Contexts," Nature Reviews Psychology 1, no. 10 (2022): 560-576, <https://doi.org/10.1038/s44159-022-00093-5>.

781 Ibid.

Section 11 Questions:

- Which policy or legal reforms should governments prioritise to better protect the digital rights of LGBTI+ people?
- What additional responsibilities should technology companies have in preventing and responding to digital harms affecting LGBTI+ communities?
- How can civil society organisations, activists, and community networks be better supported to strengthen digital resilience and respond to emerging online threats?
- Are there examples of effective laws, regulatory frameworks, platform policies, or community-led initiatives?
- What safeguards are needed to ensure that efforts to regulate digital platforms do not undermine privacy, secure communications, freedom of expression, or the work of human rights defenders?



Section 12: The Unaccountability Paradox

Section 11 showed how digital political systems translate online hostility into real-world power – electoral strategies, legislative agendas, hybrid repression, democratic erosion. This final section asks the uncomfortable question that the preceding analysis implies: Given that these harms are known, documented, and in many cases acknowledged by the actors responsible for them, why does accountability remain so elusive? And what does the answer tell us about the nature of digital governance itself?

One of the defining features of contemporary digital governance is not simply the concentration of power within platform companies but the systematic diffusion of responsibility across institutional layers. In this environment, harm is increasingly easy to produce, difficult to attribute, and structurally incentivised to remain unresolved. This produces what

can be described as an unaccountability paradox, a system in which multiple actors exercise control over digital environments. Yet, none can be meaningfully held responsible for the harms that emerge within them.

Platforms invoke legal and rhetorical categories such as ‘free speech’ and ‘technical neutrality’ to position themselves as intermediaries rather than governors. Anti-rights actors exploit this same framing, presenting their presence on mainstream platforms as evidence of legitimacy: If their speech were truly harmful – they imply – the platform would have removed it. In this way, platform permissiveness is converted into a rhetorical defence, allowing harmful actors to borrow legitimacy from the infrastructures that fail to govern them. States, meanwhile, struggle to regulate systems that evolve faster

than legislative cycles and operate across jurisdictions. International human rights law provides normative frameworks for protection but lacks robust enforcement mechanisms in digital contexts. Civil society organisations absorb much of the practical burden of mitigation, often without adequate resources or institutional protection. Across this fragmented landscape, accountability is continuously displaced rather than resolved.

This section argues that accountability in the digital age is not absent but redistributed in structurally unstable ways. The result is a governance architecture in which responsibility is perpetually deferred, producing systemic opacity and institutional fatigue.

Platform Evasion: Neutrality as a Governance Strategy

Digital platforms occupy a paradoxical institutional position: They are central to the production and circulation of public discourse yet frequently resist classification as publishers or editors.⁷⁸² Instead, they present themselves as neutral intermediaries, governed by technical systems

rather than editorial judgment. However, this framing obscures the extent to which platforms actively shape visibility, amplification, and suppression through algorithmic curation.⁷⁸³

As Jack Balkin argues, platforms in the algorithmic society do not merely transmit speech but structure speech environments, determining what becomes visible, amplified, or marginalised.⁷⁸⁴ The notion of neutrality, in this context, functions less as an empirical description than as a legal and rhetorical strategy that allows platforms to externalise responsibility while maintaining infrastructural control over communication flows.⁷⁸⁵

Legal frameworks have struggled to adapt to this transformation. Traditional intermediary liability regimes were developed for an earlier internet architecture in which platforms functioned primarily as passive conduits of content.⁷⁸⁶ In contemporary systems, however, platforms actively curate content through ranking algorithms, recommendation systems, and engagement optimisation mechanisms.⁷⁸⁷ Despite this, many legal structures continue to treat platforms as structurally detached from the content they host.⁷⁸⁸

This mismatch produces what scholarship on online harms describes as a diffusion of responsibility across platforms, users, and states.⁷⁸⁹ Harm is acknowledged but not clearly owned, resulting in fragmented accountability chains. Governance decisions within platforms are further shaped by commercial incentives and reputational risk management strategies, leading to inconsistent and context-dependent enforcement practices.⁷⁹⁰ Ultimately, platforms operate as quasi-sovereign actors, exercising governance-like control over public communication infrastructures without being subject to corresponding democratic accountability mechanisms.⁷⁹¹

Governments Struggling to Catch Up

If platforms represent a form of accelerated private governance, states occupy a structurally slower position within the digital ecosystem. Regulatory systems are constrained by jurisdictional boundaries, legislative timelines, and institutional inertia, all of which are increasingly misaligned with the pace of technological development.⁷⁹²

The Organisation for Economic Co-operation and Development (OECD) has consistently documented a growing regulatory lag between technological innovation and policy response, noting that digital platforms evolve at a speed that outpaces traditional governance cycles.⁷⁹³ This temporal mismatch creates structural vulnerabilities in which harmful practices can scale globally before regulatory frameworks are fully operationalised.⁷⁹⁴

At the international level, governance fragmentation further compounds these challenges. The World Economic Forum highlights the absence of a unified global authority capable of regulating digital ecosystems, resulting in a patchwork of partial and often incompatible regulatory regimes.⁷⁹⁵ Even significant legislative interventions, such as the European Union's Digital Services Act, rely heavily on platform self-reporting and compliance mechanisms, thus preserving a degree of structural dependency on the very actors being regulated.⁷⁹⁶

Efforts by the United Nations to develop a coordinated approach to digital governance reflect growing awareness of

782 Riedl, "Rethinking Rights in Social Media Governance."

783 Gillespie, Custodians of the Internet.

784 Jack M. Balkin, "Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation," SSRN Electronic Journal, September 9, 2017, <https://doi.org/10.2139/ssrn.3038939>.

785 Klonick, "Beyond the Public Square."

786 Riedl, "Rethinking Rights in Social Media Governance."

787 Klonick, "Beyond the Public Square."

788 Riedl, "Rethinking Rights in Social Media Governance."

789 Ibid.

790 Gillespie, Custodians of the Internet.

791 Klonick, "Beyond the Public Square."

792 OECD, Rights in the Digital Age.

793 Ibid.

794 Ibid.

795 World Economic Forum, Digital Assets Regulation: Insights from Jurisdictional Approaches (Geneva: World Economic Forum, 2024), <https://www.weforum.org/publications/digital-assets-regulation-insights-from-jurisdictional-approaches/>

796 Ibid.

these limitations.⁷⁹⁷ However, such initiatives also acknowledge the absence of enforceable global mechanisms capable of harmonising regulation across jurisdictions.⁷⁹⁸ Compounding these structural constraints is a persistent expertise gap within state institutions. Regulatory bodies often depend on private-sector data and technical knowledge to understand platform operations, thus creating asymmetric dependencies that further weaken regulatory autonomy.⁷⁹⁹

Legal Gaps: Human Rights Law in the Digital Environment

International human rights law provides the normative foundation for protecting individuals against harm, including in digital contexts. However, its practical application to platform mediated environments remains uneven and underdeveloped.

The Office of the High Commissioner for Human Rights affirms that human rights protections apply online as they do offline yet simultaneously acknowledges significant challenges in enforcement and operationalisation. The transnational nature of digital platforms complicates

jurisdictional authority, making it difficult to attribute legal responsibility or enforce remedies across borders.⁸⁰⁰

The International Commission of Jurists similarly notes that existing legal frameworks were not designed for systems in which private corporations play a central role in mediating expression and surveillance at global scale.⁸⁰¹ This structural limitation creates gaps between normative protections and enforceable obligations.

In addition, Amnesty International has documented how digital technologies enable rights violations through infrastructural mechanisms such as surveillance systems and data extraction practices rather than through discrete or easily attributable actions.⁸⁰² Organisations such as ARTICLE19 further highlight the emergence of "private censorship regimes," in which platform governance decisions effectively determine the boundaries of expression without direct public accountability.⁸⁰³

Together, these dynamics reveal a persistent tension; human rights law remains conceptually robust but operationally constrained within digitally mediated environments.⁸⁰⁴ The Yogyakarta Princi-

ples plus 10 – introduced in Section 1 as a normative framework situating digital harm within international human rights law – represent the aspiration. The cases documented throughout this paper represent the gap between that aspiration and reality.⁸⁰⁵

Civil Society Burden: The Displacement of Enforcement

As responsibility becomes increasingly diffused across platforms and states, civil society organisations have assumed a growing share of the practical burden of monitoring, documenting, and responding to digital harms.⁸⁰⁶ This shift represents not merely an expansion of advocacy roles but a structural redistribution of governance responsibilities.

The CIVICUS State of Civil Society Report documents increasing pressure on civil society organisations, including shrinking civic space, reduced funding stability, and heightened political scrutiny.⁸⁰⁷ In digital contexts, these pressures are compounded by the need to respond in real time to

online harassment, misinformation, and coordinated abuse.⁸⁰⁸

Organisations such as GLAAD and Access Now have developed operational infrastructures to monitor digital harm and provide direct support to affected individuals, effectively functioning as informal regulatory actors.⁸⁰⁹ Front Line Defenders similarly reports escalating risks faced by human rights defenders, particularly those operating in politically sensitive or digitally exposed environments.⁸¹⁰ As documented by OpenDemocracy, many civil society actors face chronic resource exhaustion due to continuous crisis response demands, leaving limited capacity for long-term strategic work.⁸¹¹

Civil society thus becomes an essential yet under-resourced component of digital governance, absorbing responsibilities that are not formally allocated to it within institutional frameworks.⁸¹² This displacement is not accidental. It is the predictable outcome of the accountability structures described throughout this paper; systems that generate harm efficiently while distributing the burden of its

797 Office of the United Nations High Commissioner for Human Rights, Civic Space.

798 Ibid.

799 Nicholas Diakopoulos, "Accountability in Algorithmic Decision Making," Communications of the ACM 59, no. 2 (2016): 56-62, <https://doi.org/10.1145/2844110>.

800 Office of the United Nations High Commissioner for Human Rights, Civic Space.

801 International Commission of Jurists, Yogyakarta Principles plus 10.

802 Amnesty International, The Social Atrocity.

803 ARTICLE 19, Digital Rights, <https://www.article19.org/issue/digital-rights/>

804 Land, "Toward a Human Rights Approach."

805 International Commission of Jurists, Yogyakarta Principles plus 10.

806 Human Rights Watch, All This Terror Because of a Photo.

807 CIVICUS, State of Civil Society Report 2026 (Johannesburg: CIVICUS, March 2026), https://www.civicus.org/documents/reports-and-publications/SOCS/2026/state-of-civil-society-report-2026_en.pdf.

808 Human Rights Watch, All This Terror Because of a Photo.

809 Shannon N. Green, "Adapt or Perish: The New Normal for Civil Society," openGlobalRights, January 4, 2016, <https://www.opendemocracy.net/en/adapt-or-perish-new-normal-for-civil-society/>.

810 Ibid.

811 Ibid.

812 Human Rights Watch, All This Terror Because of a Photo.

management onto those least equipped to bear it.

The Need for Global Governance

The structural fragmentation of digital governance has prompted increasing calls for coordinated global regulatory frameworks. The United Nations Digital Cooperation Roadmap explicitly recognises that digital systems transcend national boundaries and, therefore, require multilateral coordination mechanisms.⁸¹³ Similarly, the World Economic Forum advocates for multi-stakeholder governance models that integrate states, private actors, and civil society into shared regulatory frameworks.⁸¹⁴

The Brookings Institution has argued that artificial intelligence and platform governance cannot be effectively managed through unilateral national regulation alone due to the inherently transnational nature of digital infrastructure.⁸¹⁵ The OECD's development of AI governance principles represents an early attempt to establish shared normative standards, while institutions such as Chatham House

emphasise that current governance arrangements remain fragmented across overlapping – and often inconsistent – regimes.⁸¹⁶

Despite widespread recognition of these challenges, effective global enforcement mechanisms remain underdeveloped. Governance thus exists more as a normative aspiration than a fully operational system.⁸¹⁷

Accountability Cannot Be Algorithmic

In response to governance gaps, there has been increasing interest in algorithmic accountability as a potential solution to platform opacity. However, scholarship consistently demonstrates that algorithmic systems cannot substitute for democratic accountability structures.⁸¹⁸

Research published in Humanities and Social Sciences Communications argues that algorithms cannot bear normative responsibility, as they lack the capacity for moral or institutional accountability.⁸¹⁹ Similarly, work from the Association for Computing

Machinery (ACM) highlights the limits of algorithmic governance, noting that automation may shift decision-making processes but does not resolve underlying accountability questions.⁸²⁰

Legal scholarship in the Yale Law Journal further emphasises that accountability must remain traceable to human and institutional actors, as delegating responsibility to computational systems undermines legal clarity and democratic oversight.⁸²¹ Harvard research on algorithmic governance warns that increasing reliance on automated systems risks reducing public contestability and transparency in decision-making processes.⁸²² Brookings' analysis dismantles the assumption of algorithmic neutrality, demonstrating that all computational systems encode design choices and normative assumptions.⁸²³

Taken together, these findings suggest that algorithmic systems cannot resolve the unaccountability paradox; they risk, instead, embedding it more deeply within infrastructural systems.⁸²⁴ To automate accountability is to eliminate it.

Conclusion

The unaccountability paradox is not a temporary governance failure but a structural condition of contemporary digital systems. It emerges from the interaction of multiple institutional logics: platforms that exercise control without liability, states that regulate slowly and unevenly, legal frameworks that lack enforcement capacity in transnational environments, and civil society actors who absorb residual governance burdens.

Attempts to resolve this paradox through technical systems or algorithmic solutions risk reinforcing the very conditions they seek to address. Accountability, in this context, cannot be delegated to computational systems without eroding its normative foundation. Instead, it must remain anchored in democratic institutions capable of attributing responsibility, enforcing obligations, and contesting power.

The challenge, therefore, is not to automate accountability but to rebuild its institutional conditions within a fragmented and rapidly evolving digital environment.

813 International Commission of Jurists, Yogyakarta Principles plus 10.

814 World Economic Forum, Digital Assets Regulation.

815 Tanner et al., Is AI Sovereignty Possible?,

816 Amnesty International, The State of the World's Human Rights.

817 OECD, Rights in the Digital Age.

818 Balkin, "Free Speech in the Algorithmic Society."

819 Ibid.

820 Diakopoulos, "Accountability in Algorithmic Decision Making."

821 Klonick, "Beyond the Public Square.

822 Land, "Toward a Human Rights Approach."

823 Tanner et al., Is AI Sovereignty Possible?

824 Klonick, "Beyond the Public Square."



Conclusion: The Canary in the Coal Mine

This paper began with a question about LGBTI+ people and digital spaces and ended with a conclusion about democracy. The argument has moved in four stages, each building on the last. The first established what is at stake. For LGBTI+ individuals in restrictive or hostile environments, digital spaces are not peripheral; they are foundational. They provide the infrastructure for identity formation, community building, cultural survival, and political organising. The same platforms also generate surveillance, harassment, and persecution. This dual structure – life-line and exposure, empowerment and vulnerability – is not an anomaly. It is the defining condition of digital civic life.

The second examined the nature of the infrastructure itself. Cyberspace is not neutral; it is a socio-technical system built on commercial extraction, algorithmic governance, and data colonialism. Plat-

forms are not public squares; they are privately governed civic arenas whose rules, incentive structures, and moderation systems shape what can be said, seen, and known. Algorithms do not merely distribute speech; they produce asymmetrical visibility, structurally reward emotional intensity, and operationalise social inequalities. Artificial intelligence has now industrialised manipulation, enabling synthetic propaganda systems that undermine the epistemic foundations of democratic public discourse.

The third traced the forces operating within this infrastructure against LGBTI+ communities. Anti-LGBTI+ harm is not a collection of isolated incidents. It is a transnational ecosystem: ideologically coherent, economically incentivised, technically sophisticated, and politically purposeful. It operates through far-right networks, dark web coordination, pred-

atory communities, information warfare, and – crucially – through the normalisation mechanisms that shift the Overton window over time. Each of these systems reinforces the others: Fringe narratives enter mainstream discourse; mainstream discourse legitimises legislative action; legislative action enables state repression; state repression is documented online and used to generate further outrage and further mobilisation.

The fourth examined what happens when this system encounters formal political structures. Digital populism, microtargeting, platformed movements, coordinated disinformation, and hybrid repression converge to convert online hostility into real-world power. And across all of it, the accountability mechanisms that should constrain these dynamics – platform governance, state regulation, international human rights law, civil society oversight – are systematically inadequate. They are too slow, too fragmented, too dependent on the very actors they seek to regulate and increasingly offload their responsibilities onto the individuals and organisations most harmed.

What, then, do LGBTI+ people and digital rights have to do with each other in a broader sense?

LGBTI+ communities are not simply one vulnerable group among many. They function, in the digital environment as in the physical one, as what might be called a friction point – a place where the stresses within the broader system become most acutely visible. When platforms fail to protect queer users, they reveal the structural biases in algorithmic moderation. When states use dating apps to entrap gay men, they demonstrate the potential for sur-

veillance capitalism to serve persecution. When transnational anti-gender movements mobilise through social media, they test and map the pathways through which extremist content can be normalised and weaponised. When LGBTI+ organisations are delegitimised as foreign agents, the same rhetorical toolkit is being tested that will later be used against journalists, academics, women, racialised communities, migrants, religious minorities, and democratic institutions more broadly.

The evidence assembled in this paper consistently shows that the erosion of LGBTI+ rights in the digital environment precedes, and predicts, the erosion of democratic rights more generally. The correlation is not coincidental. It reflects the fact that the systems which make LGBTI+ people vulnerable – algorithmic amplification of outrage, data colonialism, hybrid repression, platform abdication, epistemic destabilisation – are the same systems that make democracies fragile. They do not stop when they have produced homophobia or transphobia. They expand. The same infrastructures that learn to circulate anti-LGBTI+ moral panic can circulate misogyny, racism, antisemitism, Islamophobia, xenophobia, and other forms of dehumanisation. The same communities of harm that organise around queer and trans people can redirect their attention toward feminists, migrants, racial justice movements, or any group that becomes useful as a symbol of social threat.

This is not to say that LGBTI+ rights only matter because of what they tell us about other things. They matter in themselves. The digital harms documented in this paper – entrapment, outing, image-based abuse, doxxing, state persecution, mob

violence – are experienced by real people with real lives. The cost of these harms is not abstract. But it is also not contained. Systems that are allowed to perfect themselves on one minority rarely remain confined to that minority. They build techniques, audiences, incentives, and institutional habits. They discover what language works, what platforms tolerate, what legal categories can be manipulated, what fears can be activated, and what forms of cruelty can be made ordinary.

That is why the targeting of LGBTI+ communities should be understood as a beginning rather than an endpoint. It is a warning about what these systems are capable of when power, profit, resentment, weak accountability, and – at times – state interest converge. Today, the dominant narrative may be anti-LGBTI+ moral panic; tomorrow, it may be explicitly sexist backlash, racist conspiracy, anti-migrant mobilisation, or some new composite form of digital hatred that draws from all of them at once. The architecture is transferable. The emotional logic is transferable. The communities of harm are transferable, but so too are the political uses of this hostility. This includes actions by state and para-state actors who can instrumentalise digital antagonism as a form of social warfare, thereby weakening solidarity, fragmenting democratic publics, legitimising exceptional controls, and normalising the governance of targeted populations through fear.

But the insight is this: To defend the digital rights of LGBTI+ communities is not a niche cause. It is a project of defending the conditions under which democratic life is possible. Safety, identity, free expression, access to information, the integrity of public discourse, accountability for power. These are not only LGBTI+ issues. They are the foundations of any society in which human beings can live with dignity. If these systems can make one group unsafe, unbelievably, hyper-visible, or politically disposable, they can do it to others.

The question is not whether the harm will spread. It already has.

The question is whether democratic institutions will recognise the pattern early enough to interrupt it and identify the anti-rights actors in time to counter them.

The canary is singing, but the warning is not only that the canary is in danger. It is that the mine itself is getting close to its point of failure. The question is whether we are listening, and whether our governments and democratic institutions still possess the will and the capacity to act before the harms first perpetrated on LGBTI+ communities become the normal operating conditions of public life.

Accountability cannot be algorithmic. It must be democratic. And democracy, in the digital age, must be actively built and actively defended – not assumed, not outsourced, and not delegated to the same systems that are producing harm.

Section 12 Key Findings:

- Digital technologies have fundamentally reshaped how LGBTI+ people exercise their rights, participate in public life, and build community. As a result, protecting digital rights has become inseparable from protecting human rights.
- The challenges identified throughout this report – including online harassment, surveillance, algorithmic discrimination, coordinated anti-rights mobilisation, disinformation, and technology-facilitated gender-based violence – are interconnected and require coordinated, systemic responses rather than isolated interventions.
- The experiences of LGBTI+ communities demonstrate how digital harms often emerge first against marginalised groups before expanding to threaten wider democratic participation, civic space, and human rights. Protecting LGBTI+ rights online, therefore, contributes to strengthening democratic resilience for all.
- Effective responses must be grounded in international human rights principles and developed collaboratively by governments, technology companies, civil society, international organisations, and affected communities.
- Building safer and more inclusive digital environments requires sustained investment in digital literacy, community resilience, digital security, research, and international cooperation alongside greater transparency and accountability from technology companies.
- As digital technologies continue to evolve, protecting LGBTI+ rights will require adaptive governance that responds to emerging technologies while safeguarding privacy, freedom of expression, equality, and non-discrimination.

Section 12 Questions:

- In this report, which findings most closely reflect your experience or the experiences of the communities you work with?
- What additional evidence, case studies, or examples should be included before the report is finalised?
- Looking ahead, what emerging technologies, digital trends, or policy challenges should our community prioritise?

