

7. THE CONTENT: FROM PRIOR CONSENT TO HATE MAILS

LEA BOCK
THE POLITICS OF CONNECTIVITY · TOPIC 7
19.06.2027

FOUR HYPOTHESES

H1: Content control is structurally unchanged, only the actors have changed

H2: Private platforms now make political decision that once required international negotiation

H3: In consolidated authoritarian regimes, content moderation becomes a structural instrument of repression

H4: Ungoverned content management produces measurable political harm

STRUCTURE

- **1. Historical Arc: From NWICO to Platforms**
- **2. Content Control in Authoritarian States**
- **3. Surveillance as Content Control**
- **4. Hate Speech & Citizen Preferences**
- **5. Governance Responses & Gaps**
- **6. Cross-References & Conclusion**

1. HISTORICAL ARC: FROM NWICO TO PLATFORMS (1/2)

1945 – 1984: The original governance conflict

- USA pushes 'free flow of information' globally, primarily to protect its own media exports; states have no right to filter incoming content
- Developing world and USSR coalition demand a New World Information & Communication Order (NWICO) in UNESCO and ITU in the 1970s
- Core demand: prior consent- states must agree before transborder broadcasts reach their territory; directly opposed to US interests
- 1980: MacBride Report documents global media inequality, recommends structural reform; immediately contested by US and UK as 'censorship'
- 1984: Reagan and Thatcher withdraw from UNESCO, citing NWICO as a threat to press freedom , the governance debate collapses

1. HISTORICAL ARC: FROM NWICO TO PLATFORMS (2/2)

2000s – today: The same conflict, new actors

- Content governance is privatised: platforms set global rules without interstate negotiation, without democratic mandate, without international oversight
- The unresolved NWICO conflict resurfaces in every moderation decision: who decides what information crosses borders and on whose terms?
- In 1975 this was debated in a UNESCO session in Paris. Today it is decided in Meta's Trust & Safety team

2. CONTENT CONTROL IN AUTHORITARIAN STATES (1/2)

How the mechanism works

- Authoritarian regimes threaten to block platform operations or impose fines, platforms face a commercial choice, not a legal obligation
- Requests are framed as 'national security' or 'public order' ; vague enough that platforms cannot verify but fear to challenge
- In weak autocracies: social media can still strengthen opposition (Arab Spring 2011)
- In consolidated autocracies: platforms become tools of surveillance, suppression, and propaganda
- Top requesters 2014–2020: Russia, Turkey, Thailand, Vietnam, China
- Categories 'national security' and 'criticism of government' virtually absent in democratic states

2. CONTENT CONTROL IN AUTHORITARIAN STATES (2/2)

Examples

- Saudi Arabia with Snapchat (2017): Al Jazeera removed from Discover section on government request; no legal order, informal pressure sufficient
- Turkey with Facebook (2018): YPG page blocked for Turkish users, internal emails confirm decision driven by fear of market loss, not legal requirement
- Vietnam with Facebook (2020): Zuckerberg personally decided to delete government-critical content to preserve market access
- Russia with Apple & Google (2021): Navalny's election app removed from app stores days before parliamentary elections; at government request
- China with Apple: Koran app and all content referencing Tiananmen, Taiwan independence removed for Chinese users

3. SURVEILLANCE AS CONTENT CONTROL

Empirical findings — bpb APuZ 18–19/2014

- Deep Packet Inspection (DPI): the same technology used by the NSA is used to block YouTube in Turkey and delete VPN apps on government order (Bendrath 2014)
- Only 7.8% of US terror suspects identified through mass data collection, architecture persists because it serves political control, not security (Weidemann 2014)
- Surveillance is embedded in everyday consumption: agencies and corporations share the same categorisation logic (Zurawski 2014)
- Early BND monitored SPD, FDP, media and academia with government knowledge, the line between security and repression was always thin (Henke 2014)

Theoretical frame: Bentham & Foucault

- Panopticon (Bentham): the mere possibility of being watched suffices to discipline behaviour
Chilling Effect: users self-censor not because they are watched, but because they might be
- Conduct of Conduct (Foucault): algorithms govern what is visible, sayable, findable; content moderation is governance without government

4. HATE SPEECH & CITIZEN PREFERENCES

Munzert et al. (2025): ~2,600 participants, ~21,000 synthetic posts, experimental moderation task

- Citizens distinguish carefully: highly differentiated responses based on severity and type , not blanket acceptance or rejection
- Delete, don't punish: post deletion widely supported; harsh sanctions (imprisonment, job loss) find little support -> public preferences more moderate than political rhetoric
- Severity matters most: decisive moderation factors are severity and type —> not sender or recipient identity
- Group asymmetry: hate speech against marginalised groups judged more harshly than against political groups —> content governance is not culturally neutral
- Exposure effect: short-term contact with hate speech reduces tolerance for unpopular opinions and increases acceptance of censorship —> a destabilising feedback loop
- Germany vs. USA: German respondents support stricter regulation; US respondents prioritise free speech —> a single global moderation standard is politically impossible

5. GOVERNANCE RESPONSES & GAPS

What the EU has done

- Digital Services Act (2023/24): VLOPs (>45M EU users) must conduct risk assessments, audits, transparency reports (over 10 billion removals logged)
- Code of Practice on Disinformation integrated into DSA, February 2025
- AI Act (August 2026) regulates AI used in content moderation
- Brussels Effect: EU regulation shapes global platform behaviour, but weaker for content than for data

What remains ungoverned

- Gsenger (2025): platforms have no unified definition of disinformation so the DSA regulates something that has not been defined
- Global Digital Compact (UN, 2024): consensus between authoritarian and democratic states remains elusive
- Platforms remain de-facto legislators, without democratic mandate

6. TOPIC 7 IN THE BIGGER PICTURE

- Topic 1 (ITU, ICANN, Mega-Monopolies): platforms like Meta and Google ARE the mega-monopolies; ICANN refused to disconnect Russia after Ukraine, same governance vacuum
- Topic 2 (EU Level): DSA is the EU's direct answer to ungoverned content; Brussels Effect weaker for content than for data
- Topic 5 (Networks): Deep Packet Inspection operates at the network layer (e.g. Russia's Sovereign Internet Law controls content by controlling infrastructure)
- Topic 8 (Security/Cyberwar): Russian information warfare in Ukraine = content control as a weapon; disinformation is a security threat (Nocetti 2024)
- Topic 10 (Digital Divide): who gains access immediately enters power structures, connectivity and content control are two sides of the same coin

CONCLUSION (1/2) - HYPOTHESES CONFIRMED

- **H1** confirmed: the structural conflict is unchanged, only the actors moved from UNESCO to Silicon Valley
- **H2** confirmed: platforms resolve in minutes what once took years of interstate negotiation, without democratic mandate
- **H3** confirmed: content moderation in authoritarian states is not an exception, it is a systematic feature
- **H4** confirmed: ungoverned content governance produces measurable harm -> chilling effects, electoral manipulation, hate speech feedback loops

CONCLUSION (2/2)

Who controls what we communicate- and according to whose rules?

- Content governance fails not because of technical problems or bad will, but because economic power and political legitimacy systematically fall apart
- Platforms have the power, but no democratic legitimacy
- Authoritarian states have domestic legitimacy, but abuse it
- The EU has regulatory ambition, but limited global reach
- Nobody has all three: and that is the governance gap

Aus Politik und Zeitgeschichte- Überwachen. (2014). 64. Jahrgang(18–19).

Bradford, A. (2012). The Brussels Effect. In A. Bradford, *The Brussels Effect* (1st edn, Vol. 1, pp. 25–66). Northwestern University of Law. <https://doi.org/10.1093/oso/9780190088583.003.0003>

Europäisches Parlament und Rat der Europäischen Union. (2022, October 27). *Verordnung (EU) 2022/2065 über einen Binnenmarkt für digitale Dienste (Gesetz über digitale Dienste – Digital Services Act, DSA)*, ABl. L 277, 27.

Gsenger, R. (2025). Platform governance under the Digital Services Act: A perspective on disinformation. *Information, Communication & Society*, 1–19. <https://doi.org/10.1080/1369118X.2025.2590561>

Kittel, G. (1997). *Charakteristika internationaler Organisationen und Außenpolitikanalyse- Zwischen konditionaler Kooperation und Austritt: Die Außenpolitik der USA gegenüber der UNESCO und der ITU (1982-1988)*. (Original work published Nomos Verlagsgesellschaft)

Köhler, P., & Voelsen, D. (2022). Content Moderation in autoritären Staaten: Wie soziale Medien zu Komplizen autoritärer Herrscher zu werden drohen. In *SWP-Aktuell 39/2022* (Version 1). Stiftung Wissenschaft und Politik, German Institute for International and Security Affairs. <https://doi.org/10.18449/2022A39>

Munzert, S., Traunmüller, R., Barberá, P., Guess, A., & Yang, J. (2025). Citizen preferences for online hate speech regulation. *PNAS Nexus*, 4(2), pgaf032. <https://doi.org/10.1093/pnasnexus/pgaf032>

Nocetti, J. (n.d.). *A Splintered Internet? Internet Fragmentation and the Strategies of China, Russia, India and the European Union*.

Toptchiyska, A. D. (n.d.). *GOVERNING AI AND CONTENT MODERATION ONLINE: THE EU REGULATORY APPROACH AND GLOBAL MULTISTAKEHOLDER DYNAMICS*.

United Nations. (n.d.). *Global Digital Compact*. Pact for the future. Retrieved 18 June 2026, from <https://www.un.org/pact-for-the-future/en/global-digital-compact>

Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz) (Text von Bedeutung für den EWR) (2024). <http://data.europa.eu/eli/reg/2024/1689/oj>