Committee: SOCHUM 1

Topic: The question of oversight and governance of social media

Chair: Alice Sharp

School: Croydon High School

Summary

Social media connects billions and shapes what we see and believe. It helps people organise, learn, and be heard. But it also spreads hate, lies, scams, and pressure on young people, often driven by hidden algorithms. Governments want to keep people safe and protect elections; companies run the platforms and set many of the rules. The challenge is to protect free speech and privacy while reducing real harms. The UN can guide countries and platforms toward simple, fair standards: clear content rules, more transparency, independent checks, better data protection, and child-safety by design, but has to keep free speech laws in mind and not breaking secure messaging services.

Definition of Key Terms

- **Social media / platform** – Apps where people post, share, and message (e.g., TikTok, Instagram, X).
- **Content moderation** – How platforms enforce their rules (remove, label, age-restrict, limit reach).
- **Algorithm / recommender system** – The code that decides what you see next and what gets boosted.
- **Misinformation vs. disinformation** – False content shared by mistake or on purpose to mislead.
- **Hate speech & incitement** – Attacks on people for who they are, or speech that urges real-world harm.
- **Freedom of expression** – The right to share opinions; limits exist only for lawful, specific reasons.
- **Privacy & data protection** – Your control over personal data: what's collected, used, and shared.
- **End-to-end encryption (E2EE)** – Only sender and receiver can read messages, not the platform or state.

- **Child-safety-by-design** – Features that protect young users by default (incl. **age assurance** tools).
- **Transparency report** – A platform's regular update on removals, rule enforcement, and government requests.
- **Risk assessment (systemic risk)** – Platforms checking how their services might cause big harms (elections, youth, health) and how they'll reduce them.
- **Independent audit** – Outside experts' test whether a platform's policies and safety measures actually work.
- **Notice and appeal (due process)** – Telling users why action was taken and giving a fair way to challenge it.
- **Deepfakes / synthetic media** – AI-made audio/video/images that look real but are fake.
- **ICCPR & Rabat Plan** – UN human-rights rules: protect free speech (ICCPR Art. 19) and ban incitement (Art. 20); Rabat explains when speech crosses the legal line.

Background Information

Social media is now one of the main places where people talk, learn news, organise events, and express themselves. It can help during emergencies, amplify voices that are often ignored, and support education and business. At the same time, it can spread hate, harassment, bullying, scams, and false information very quickly. Much of what people see is decided by algorithms, which can boost some posts and hide others. Because these systems are not fully transparent, users and governments often do not know why certain content travels so far or reaches young people.

Who should set the rules is at the heart of this topic. Platforms already have "community guidelines" and teams that moderate content. They also label or remove posts, limit accounts, and cooperate with law enforcement in serious cases. Governments want to protect citizens, elections, national security, and especially children. But state action can also go too far, leading to censorship, surveillance, internet shutdowns, or unfair pressure on companies to remove lawful speech. Approaches differ across the world. Some regions focus on transparency and safety duties for platforms. Others rely more on market pressure or court cases. Many countries are still building the skills and laws they need.

International human rights law offers a shared baseline. The right to freedom of expression protects most speech, with narrow limits for things like direct incitement to violence. The UN's guidance states that the same rights people have offline should also apply online. It also encourages careful tests before restricting speech, and calls for clear rules, transparency, and the chance to appeal decisions. UNESCO

promotes principles for platform governance, and the UN system hosts global discussions that include governments, companies, experts, and civil society.

The main question is how to promote safer, fairer platforms while respecting rights. Possible ideas include encouraging transparency reports, independent audits, and risk assessments on issues such as harm to minors and election integrity. Other options include better data access for trusted researchers, clearer political ad labels, child-safety by design, and media literacy for users. There is also a live debate about keeping private messaging secure with end-to-end encryption while tackling serious crimes. Because the internet crosses borders, international cooperation and shared standards matter.

Major Countries and Organizations Involved

1. **European Union (EU)** – Sets wide-ranging rules like the Digital Services Act (DSA) and GDPR that push platforms to be more transparent, assess risks, and protect user data across 27 countries.
2. **United States** – Home to many big platforms; debates focus on free speech, child safety, privacy, and foreign-owned apps. Federal rules are patchy, but state laws and court cases shape the space.
3. **United Kingdom** – The Online Safety Act gives the regulator (Ofcom) powers to set safety codes, require risk assessments (especially for minors), and enforce penalties.
4. **India** – One of the largest user bases in the world. Strong takedown powers and local compliance rules make it a key test case for platform–government relations.
5. **Brazil** – Active in tackling disinformation and platform responsibility (especially around elections), with ongoing proposals to increase transparency and duty of care.
6. **Australia** – Early mover on online safety (eSafety Commissioner) and news bargaining rules; pushes platforms on rapid removal of harmful content.
7. **China** – Tight state oversight of platforms, strong data/algorithm controls, and limits on foreign services; influential model for state-led governance.
8. **United Nations system** – **UNESCO**, **OHCHR**, and the UN Secretary-General promote human-rights–based standards (free expression + limits on incitement), transparency, and multistakeholder dialogue.
9. **Major platforms** – **Meta (Instagram/Facebook)**, **Google/YouTube**, **TikTok (ByteDance)**, **X**, **Snap**, **Reddit**. Their policies, algorithms, and enforcement choices directly shape online speech and safety.
10. **Civil society & researchers** – Groups like ARTICLE 19, Access Now, EFF, and academic labs press for rights protections, transparency, independent audits, and better data access to study platform impacts.

Timeline of Events

Date	Description
2012	UN Rabat Plan of Action sets a high bar for restricting speech that incites hatred/violence.
2018	EU GDPR takes effect, raising global standards for privacy, data control and ad profiling.
2019	Christchurch Call launches after the NZ attacks; platforms tighten livestream and extremist-content rules.
2020	Meta Oversight Board begins (private oversight model for content decisions).
2020-2021	COVID-19 “infodemic” prompts stronger misinformation labels/policies across platforms.
2022	EU adopts the Digital Services Act (DSA); framework enters into force and sets platform duties.
2023	First EU designations of “very large” platforms under the DSA; toughest rules start applying to them.
2023	UK Online Safety Act becomes law; regulator (Ofcom) starts multi-year rollout.
2023	UNESCO Guidelines published, giving soft-law, human-rights–based advice for platform governance.
2024	UN adopts the Pact for the Future with the Global Digital Compact annex (shared digital principles).
2024	Murthy v. Missouri (U.S. Supreme Court) narrows lawsuits over government pressure on platform moderation (no standing).
2024-2025	UK OSA codes: illegal-content rules laid (Dec 2024) and enforced from March 2025; more child-safety codes follow.
2024-2025	U.S. TikTok divest-or-ban law enacted (Apr 2024) and upheld by the Supreme Court (Jan 2025).
2024-2025	EU Political Advertising Regulation starts applying EU-wide (labels, archives, targeting rules).

Relevant UN Treaties and Events

1. **ICCPR (International Covenant on Civil and Political Rights) – 1966**: Protects free expression (Art. 19) and bans incitement (Art. 20); core test for any limits on speech.
2. **CRC (Conventions on the Rights of the Child) – 1989; General Comment No. 25 – 2021**: Children's rights online (safety, privacy, participation); practical guidance for states and platforms.
3. **UN Guiding Principles on Business & Human Rights (UNGPs) – 2011**: Companies must prevent and fix human-rights harms (applies to platforms).
4. **HRC Res. "Human rights on the Internet" – 2012**: Confirms the same rights online as offline; anchors rights-based approaches.
5. **Rabat Plan of Action – 2012**: Clear criteria for when speech crosses into illegal incitement (context, intent, likelihood, etc.).
6. **UNGA Res. "Right to privacy in the digital age" – 2013**: Basis for tackling surveillance, data misuse, and profiling.
7. **UN Strategy & Plan of Action on Hate Speech – 2019**: System-wide UN approach to counter online hate while upholding rights.
8. **UNESCO Guidelines for the Governance of Digital Platforms – 2023**: Non-binding, practical standards on transparency, due process, and media freedom.
9. **Global Principles for Information Integrity – 2024**: UN principles to address mis/disinformation without over-censorship.
10. **Global Digital Compact (annex to the Pact for the Future) – 2024**: High-level commitments for a safe, open, human-rights-based digital space.

Previous Attempts to solve the Issue

EU Digital Services Act (DSA)

Creates duties for big platforms to assess and reduce systemic risks (harms to minors, disinformation), publish transparency reports, open data to vetted researchers, and undergo independent audits. It also requires clearer ad and recommendation rules. The goal is safer feeds and more accountability without directly writing platform algorithms for them.

EU General Data Protection Regulation (GDPR)

Sets strict rules on how companies collect and use personal data: consent, data minimisation, access/erase rights, and heavy fines for violations. Although not a "speech law," it reshaped social media by limiting tracking, profiling, and targeted ads, pushing platforms toward clearer privacy controls and standards used well beyond Europe.

UK Online Safety Act (OSA)

Gives the regulator (Ofcom) power to set safety codes and require risk assessments, especially to protect children. Platforms must tackle illegal content, improve age assurance, give users tools, and explain decisions. Supporters say it strengthens safety-by-design; critics worry about privacy and how rules apply to encrypted messaging.

Australia's eSafety framework

Through the eSafety Commissioner and Online Safety Act, Australia can issue removal notices for seriously harmful content, require faster takedowns, and set "basic online safety expectations." It also targets abuse like image-based harassment. The model aims for quick, practical enforcement and has influenced other countries' thinking on safety.

India's IT Rules (2021, amended)

Require large platforms to have local compliance officers, respond quickly to takedown requests, and provide grievance redress for users. Authorities can order removal of unlawful content, and traceability has been debated for messaging apps. India's approach tests how far governments can push for accountability while safeguarding free expression.

UNESCO Guidelines for the Governance of Digital Platforms (2023)

A global, non-binding playbook that urges transparency, due process (notice and appeal), protection of journalists and researchers, and multistakeholder oversight. It doesn't force one legal model; instead, it offers principles countries and companies can adapt. Many UN discussions and national proposals now reference these guidelines.

Christchurch Call (2019, ongoing)

A joint commitment by governments and tech firms after the Christchurch attacks to reduce terrorist and violent-extremist content online. It promotes faster response to crisis events, safer livestreaming, better detection and sharing of hashes of illegal content, and cooperation with researchers, without mandating broad censorship.

Meta Oversight Board (since 2020)

An independent body that reviews some of Meta's toughest content cases and publishes public decisions and policy recommendations. It's an experiment in private oversight: giving users an appeal route outside the company and creating precedent-like guidance. While limited to Meta's platforms, it has influenced wider debates.

EU Political Advertising Regulation (applies 2025)

Requires clear labels on political ads, public ad libraries, and stricter rules on targeting (especially sensitive data). The aim is to make election-related messaging more transparent so users can see who paid, why they saw an ad, and how campaigns target different groups, reducing hidden influence operations.

Possible Solutions

1. **Transparency & independent audits:** Require platforms to publish risk assessments, detailed transparency reports, and open data to vetted researchers, with regular third-party audits.
2. **Child-safety-by-design (with privacy):** Set safer defaults for teens, use proportionate age-assurance, and protect data, no blanket scanning of private messages.
3. **Political ads integrity:** Label all political ads, keep public ad libraries, and limit micro-targeting using sensitive personal data.
4. **Protect encryption, enable targeted enforcement:** Keep end-to-end encryption intact while using warrants, metadata limits, and safety tools that don't break secure messaging.
5. **Co-regulation with due process:** Combine platform duties and public oversight—clear rules, notice-and-appeal for users, and voluntary reporting to an UN-backed observatory.

Bibliography

https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights

https://www.ohchr.org/en/documents/outcome-documents/rabat-plan-action

https://ap.ohchr.org/documents/dpage_e.aspx?si=a/hrc/res/20/8

https://documents.un.org/doc/undoc/gen/n13/449/47/pdf/n1344947.pdf

https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation

https://www.unesco.org/en/internet-trust/guidelines

https://www.un.org/en/information-integrity/global-principles

https://www.un.org/en/summit-of-the-future/pact-for-the-future

https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en

https://commission.europa.eu/law/law-topic/data-protection_en

https://www.ofcom.org.uk/online-safety

https://www.esafety.gov.au/newsroom/whats-on/online-safety-act

https://www.meity.gov.in/content/information-technology-intermediary-guidelines-and-digital-media-ethics-code-rules-2021

https://www.christchurchcall.org/

https://publicadministration.desa.un.org/capacity-development/igf