#Reach for the Stars

YTUMUN 2025

UNHRC STUDY GUIDE

Agenda Item:
Protecting Digital Rights: Ensuring Privacy and Freedom of Expression in the Age of Artificial Intelligence (AI).

Board Members

Juman Salameh Şira Çavluer Ahmet Arda Yildirim

YTUMUN'25| 26-27-28 December

YTUMUN

Letter from the Secretary-General **2**

Letter from the Committee Board **3**

1. Introduction **4**

1.1. Introduction to the Committee (United Nations Human Rights Council) 4

1.2 Introduction to the Agenda Item: Ensuring Privacy and Freedom of Expression in the Age of Artificial Intelligence 4

2. Key Terminology **5**

3. Background Information **7**

3.1. Evolution of Privacy and Expression in International Human Rights Law 7

3.2 Key International Frameworks and Legal Instruments 8

3.2.1. Universal Declaration of Human Rights (UDHR, 1948) 8

3.3.2. International Covenant on Civil and Political Rights (ICCPR, 1966) 8

3.3.3. European Union Charter of Fundamental Rights (2000; legally binding since 2009) 9

4. Current Situation and Key Challenges **10**

4.1. Disinformation Ecosystems & Manipulation Risks 10

4.2. Threats to Journalists, Activists, and Vulnerable Populations 11

4.3. Algorithmic Suppression, Content Moderation, and Censorship 12

5. Case Studies **13**

5.1. AI Surveillance in Authoritarian Regimes 13

- Russia Clamps Down on Online Searches: **16**

5.2. Regulation of Digital Platforms in Democratic States 18

6. Possible Solutions and Policy Recommendations **19**

6.2. Transparency, Accountability, and Algorithmic Audits 20

6.3 Safeguards Against AI-Enabled Censorship 21

6.4 Protecting Digital Civic Space and Human Rights Defenders 22

6.5 Capacity-Building for Low-Resource States 23

7. References **23**

Letter from the Secretary-General

Dear Esteemed Participants and Guests,

Dear Esteemed Participants and Guests, It is my distinct honor and privilege to welcome you to YTUMUN'25. As Secretary-General, I am thrilled to invite you to what promises to be an enriching experience of debate, diplomacy, and collaboration mixed with unforgettable moments and memories.

Model United Nations is more than just a simulation of the UN; it is a platform where ideas meet action, and where the leaders of tomorrow practice the art of negotiation today. Whether this is your very first conference or one of many in your MUN journey, we are committed to providing you with an environment that challenges you intellectually and inspires you personally.

This year, our Secretariat has worked tirelessly to craft a conference where everyone feels welcomed. We believe that the variety of our topics reflects the complexity of our world and ensures that every delegate finds a space where their voice matters, and that every single participant will leave with amazing moments carved in their memories.

On behalf of the entire Secretariat, I thank you for joining us. We look forward to witnessing the passion, creativity, and leadership that you will bring to the conference. Together, let us make YTUMUN'25 a memorable and transformative experience for all. Let us reach for the stars!

Yours sincerely,
Bilel Elarem
Secretary-General of YTUMUN'25

Letter from the Committee Board

Distinguished Delegates of YTUMUN'25,

It is our distinct honor to welcome you to the United Nations Human Rights Council. We, your Committee Board, are Juman Salameh, Senrat Şira Çavluer, and Ahmet Arda Yıldırım.

We wish to express our appreciation to our Secretary-General, Bilel Elarem and Deputy Secretary-General, Tibet Tuna Topçu for their commitment and effort they have devoted to this conference up to this point.

We highly recommend that you deeply review and examine the study guide. Gaining a grasp of these resources is crucial as it will enhance the effectiveness of our meetings and allow a productive discussion.

We are confident that we will spend three productive and unforgettable days engaging in debate to find solutions to the pressing issues before us.

If you have any questions, please do not hesitate to contact us via email:

sirasenrat@gmail.com

ardayldrm9696@gmail.com

Sincerely,

Juman Salameh, Senrat Şira Çavluer, and Ahmet Arda Yıldırım

Committee Board Members | United Nations Human Rights Council

1. Introduction

1.1. Introduction to the Committee (United Nations Human Rights Council)

The United Nations Human Rights Council (UNHRC) stands as one of the central pillars of the United Nations system in its enduring mission to promote, protect, and uphold the universal principles of human rights. Established to replace the former Commission on Human Rights, the Council embodies the UN's commitment to safeguarding human dignity, equality, and justice in every part of the world. It operates as an intergovernmental body composed of 47 Member States, elected by the General Assembly for staggered three-year terms and distributed among regional groups to ensure fair geographic representation.

Meeting at the United Nations Office at Geneva (UNOG) in Switzerland, the Council convenes regularly throughout the year to deliberate on pressing human rights concerns. It is mandated to address situations of violations, make recommendations to Member States, and foster international cooperation in advancing the protection of fundamental freedoms. The Council's work spans a broad spectrum of issues, encompassing the rights to freedom of expression, belief, assembly, and association, as well as the protection of women's rights, the rights of children, LGBTQ+ persons, and racial and ethnic minorities.

Beyond its thematic focus, the UNHRC also plays a critical role in investigating alleged human rights abuses in Member States through mechanisms such as fact-finding missions, special rapporteurs, and the Universal Periodic Review (UPR) process. By engaging governments, civil society organizations, and international experts, the Council seeks to promote accountability and foster dialogue toward sustainable reform.

1.2 Introduction to the Agenda Item: Ensuring Privacy and Freedom of Expression in the Age of Artificial Intelligence

The long standing human race has been indulged in a variety of industrial revolutions that altered the course of history. starting with the first revolution of Mechanization that introduced steam engines and has had notable effects on the quality of labour and economic growth, then the second where electricity and mass production, and the third/ digital revolution in which the internet, computers, and the early automation in manufacturing were set in motion. Last but not least, the fourth revolution of intelligent automation consists of artificial intelligence, machine learning, robotics, IoT, etc.

In spite of the latter's advantages (e.g. Automation of cognitive tasks, Algorithmic decision-making) it has sparked wide-spread debates regarding its safety when it comes to privacy, surveillance, and ethics. Thus, the need for regulations and preventative measures was and still is crucial to ensure a cautious approach to this useful yet not fully controlled trend.

The main concern of our committee revolves around artificial intelligence, a tool that is already fully integrated within societies, relied on in many sectors, and implemented in daily communication transactions. This integration alongside the lack of fundamental knowledge of its threats on personal and communal aspects risks our privacy, integrity and freedom.

Mass surveillance, Global use of facial recognition, data leaks, and narrative censoring (especially on social media) , are all threats imposed by the wrongful usage of what AI offers.

These threats contradict respective articles within the UDHR and ICCPR. the fact that AI could be considered a part of a grey area where many loopholes and voids could be manipulated to serve personal interests due to the shortage of strict regulative measures upon the matter proposes the need for further strict and comprehensive regulations and policies to prevent further violations to privacy & freedom, and to ensure ethical use of AI.

Therefore, as the international community steps deeper into the age of intelligent technologies, the responsibility of safeguarding fundamental human rights becomes more pressing than ever. The challenge before this committee is not to hinder innovation, but to ensure that technological progress does not outpace the legal, ethical, and humanitarian frameworks that protect individuals and societies. Delegates are thus encouraged to examine the gaps in current global governance, identify the risks posed by unregulated AI systems, and propose forward-looking solutions that uphold the universal principles of privacy, dignity, and freedom of expression. Only through a rights-based, precautionary, and collaborative approach can we guarantee that artificial intelligence remains a tool for empowerment rather than a mechanism of control.

2. Key Terminology

Artificial Intelligence (AI): Artificial Intelligence refers to computational systems capable of performing tasks that traditionally require human cognition — such as pattern recognition, decision-making, prediction, and language processing. In governance contexts, AI

encompasses a spectrum from simple automated algorithms to advanced machine-learning models that adapt based on data.

Algorithmic Bias: Algorithmic bias describes systematic and unfair outcomes generated by AI systems when the data, design choices, or training methods embed historical, social, or structural inequalities. It can lead to discriminatory decisions affecting marginalized groups, often in ways that are invisible or difficult to contest.

Surveillance Technologies: Surveillance technologies are tools used to monitor, track, collect, or analyze information on individuals or groups. They range from CCTV networks and biometric systems to predictive analytics, facial recognition, and AI-enhanced monitoring tools. Their human rights implications grow as AI expands the reach and accuracy of surveillance.

Digital Rights: Digital rights refer to the extension of fundamental human rights — such as privacy, freedom of expression, access to information, and non-discrimination — into digital environments. They recognize that online spaces are not separate from real life; they are integral arenas where individuals exercise their autonomy and civic agency.

Data Protection & Privacy: Data protection and privacy involve the safeguarding of personal information from misuse, unauthorized access, or exploitation. It encompasses how data is collected, stored, processed, and shared. In the context of AI, privacy becomes essential for preventing intrusive surveillance, profiling, and manipulation at scale.

Freedom of Expression: Freedom of expression is the right to seek, receive, and impart information and ideas through any media, without interference. Within AI-governed digital platforms, this right faces new pressures — from automated content moderation and algorithmic amplification to targeted censorship and opaque decision systems.

Misinformation vs. Disinformation: Misinformation refers to false or inaccurate information that is shared without the intention to deceive, often spreading through misunderstanding, lack of verification, or rapid online circulation. Disinformation, by contrast, consists of deliberately false or manipulated content created with the purpose of misleading audiences, shaping public opinion, or advancing political or strategic objectives. In the age of artificial intelligence, both forms are amplified: misinformation through automated, fast-moving sharing patterns, and disinformation through sophisticated tools such as deepfakes, bot networks, and algorithmic targeting.

3. Background Information

3.1. Evolution of Privacy and Expression in International Human Rights Law

The right to privacy holds a unique place in the architecture of international human rights law. Remarkably, it was recognised at the international level before any national constitution offered a general guarantee of privacy. Early constitutional protections in the post–World War II era focused narrowly on the inviolability of the home, of correspondence, and freedom from unreasonable searches; yet no state codified privacy as an integrated, overarching right. The Universal Declaration of Human Rights (UDHR), drafted between 1946 and 1948, therefore represented a groundbreaking moment. Article 12 introduced a comprehensive protection against "arbitrary interference" with privacy, family, home or correspondence, long before states themselves had articulated such a broad concept. This international formulation became the normative foundation upon which later treaties would expand.

The International Covenant on Civil and Political Rights (ICCPR, 1966) deepened these guarantees through Article 17, which prohibits unlawful or arbitrary interference with privacy and requires states to provide legal protection against such violations. Over time, the interpretation of Article 17 has become central to debates over digital surveillance, data collection, and new forms of interference enabled by technological advancements. Contemporary controversies—such as the 2013 Snowden revelations and the 2018 Cambridge Analytica scandal brought renewed urgency to the right to privacy, revealing the vast scale of both state-sponsored and corporate surveillance. These events accelerated global calls for transparency, accountability, and a modernised understanding of what privacy means in the digital age.

Yet international law has struggled to keep pace with technological change. The UN's attempts to clarify the right to privacy in cyberspace, including the *UN Resolution on Privacy in the Digital Age*, highlight persistent gaps in interpretation—particularly concerning the extraterritorial application of privacy norms. Existing doctrines, such as the "effective control" test used to determine state jurisdiction, are ill-suited for the borderless nature of cyber surveillance. Scholars now propose the idea of "virtual control," arguing that states exercise human rights obligations when they remotely influence, intercept, or manipulate an individual's communications, even without physical control over the individual. This

approach aligns with emerging jurisprudence from the European Court of Human Rights and could help close the normative gap exploited by intelligence agencies today.

3.2 Key International Frameworks and Legal Instruments

The following international and regional legal frameworks form the normative backbone of the agenda on artificial intelligence, privacy, and freedom of expression. Together, they establish the minimum human rights standards that should guide the development, deployment, and regulation of AI technologies. Delegates are strongly encouraged to engage with these instruments closely, as they provide both the legal authority and ethical principles upon which potential AI governance mechanisms may be built.

3.2.1. Universal Declaration of Human Rights (UDHR, 1948)

Although not legally binding, the UDHR constitutes the foundational document of international human rights law and has achieved the status of customary international law in many of its provisions. Its relevance to AI governance lies in its articulation of core rights that remain applicable regardless of technological medium.

Article 12 – Right to Privacy

"No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks."

This provision establishes a broad and technology-neutral protection against arbitrary interference, making it directly applicable to modern practices such as digital surveillance, data profiling, biometric identification, and AI-driven monitoring systems.

Article 19 – Freedom of Opinion and Expression

"Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers."

Article 19 explicitly anticipates technological change by protecting expression "through any media," providing a strong normative basis for safeguarding online speech, algorithmically mediated communication, and cross-border information flows in the digital age.

3.3.2. International Covenant on Civil and Political Rights (ICCPR, 1966)

Legally binding on States Parties

The ICCPR transforms the principles of the UDHR into binding legal obligations and is central to assessing state responsibility in the context of AI-enabled surveillance and content regulation.

Article 17 – Right to Privacy

"No one shall be subjected to arbitrary or unlawful interference with their privacy, family, home or correspondence, nor to unlawful attacks on their honour and reputation. Everyone has the right to the protection of the law against such interference or attacks."

Article 17 is the primary legal basis for evaluating the legality of mass surveillance, data interception, and AI-driven monitoring practices. The prohibition of both "arbitrary" and "unlawful" interference requires that any limitation on privacy be lawful, necessary, proportionate, and subject to effective oversight.

Article 19 – Freedom of Expression

1. Everyone shall have the right to hold opinions without interference.
2. Everyone shall have the right to freedom of expression; this includes the freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, orally, in writing or in print, in the form of art, or through any other media of their choice.

3.3.3. European Union Charter of Fundamental Rights (2000; legally binding since 2009)

The EU Charter provides one of the most advanced regional frameworks for digital rights protection and has significantly influenced global AI governance debates.

Article 7 – Respect for Private and Family Life

"Everyone has the right to respect for his or her private and family life, home and communications."

This article extends classical privacy protections to modern communications, including digital correspondence and online interactions, making it directly applicable to AI surveillance and data interception.

Article 8 – Protection of Personal Data

"Everyone has the right to the protection of personal data concerning him or her. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law."

Article 8 uniquely elevates data protection to a standalone fundamental right, reinforcing principles such as purpose limitation, consent, transparency, and accountability—core elements of responsible AI development.

Article 11 – Freedom of Expression and Information

"Everyone has the right to freedom of expression. This right includes freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers."

This provision is particularly relevant to AI-driven content moderation, recommender systems, and digital platform governance, reinforcing the need to prevent unjustified algorithmic suppression of speech.

4. Current Situation and Key Challenges

4.1. Disinformation Ecosystems & Manipulation Risks

The rapid integration of artificial intelligence into information production and dissemination has fundamentally altered the nature of public discourse. Disinformation is no longer limited to isolated falsehoods or manual propaganda efforts; instead, AI enables the construction of coordinated ecosystems in which misleading narratives are generated, amplified, and targeted at scale. Machine-learning systems can produce highly convincing text, audio, and video content, including deepfakes that blur the distinction between authentic and fabricated speech. When combined with algorithmic recommendation systems that prioritise engagement, such content can spread faster and more widely than verified information, shaping perceptions before fact-checking mechanisms can intervene.

These developments pose a direct challenge to the enjoyment of freedom of expression as protected under international human rights law. While freedom of expression includes the right to seek and receive information, the saturation of information spaces with manipulated or deceptive content undermines individuals' ability to form opinions freely and on the basis of reliable information. Disinformation campaigns have been linked to electoral interference, the incitement of hostility against minorities, and the erosion of trust in democratic institutions. The transnational nature of digital platforms further complicates accountability, as influence operations frequently cross borders and fall into legal grey zones between domestic and international jurisdiction. As a result, states face the dual challenge of

countering harmful manipulation without imposing disproportionate restrictions that would themselves violate freedom of expression.

Beyond the sheer volume of false or misleading content, artificial intelligence reshapes how disinformation operates as a system rather than as isolated incidents. AI allows for micro-targeting based on behavioural data, enabling tailored narratives that exploit individuals' fears, identities, or grievances. These personalised manipulation strategies are particularly difficult to detect, as different users may encounter entirely different versions of reality. This fragmentation of the information environment undermines the shared factual basis necessary for democratic deliberation and weakens the collective exercise of freedom of expression.

4.2. Threats to Journalists, Activists, and Vulnerable Populations

Artificial intelligence has profoundly altered the risk landscape for journalists, human rights defenders, activists, and vulnerable communities by enabling forms of surveillance, monitoring, and repression that are both pervasive and difficult to detect. Unlike traditional surveillance, which often required physical presence or targeted warrants, AI-driven tools operate continuously and invisibly, aggregating data from digital communications, biometric systems, social media interactions, and location tracking. This shift erodes the boundary between public and private life, exposing individuals engaged in legitimate expression to constant observation and potential retaliation.

For journalists, these technologies pose an existential threat to the practice of independent reporting. AI-assisted surveillance can identify communication patterns, infer relationships between reporters and sources, and retroactively reconstruct networks of information exchange. The exposure of confidential sources not only endangers individuals but undermines the foundational principles of press freedom and the public's right to receive information. Even the *perception* of surveillance can be enough to deter whistleblowers and foster self-censorship, weakening investigative journalism long before any overt repression occurs.

Human rights defenders and political activists face similar dangers, particularly in contexts where dissent is already criminalised or delegitimised. AI-enabled profiling tools allow authorities to classify individuals based on behaviour, associations, or ideology, often without transparency or legal safeguards. These systems can be used to predict participation

in protests, identify organisers, or justify pre-emptive restrictions on movement and expression. Such practices raise serious concerns under international human rights law, as they shift enforcement from punishing actions to controlling perceived intent.

Vulnerable populations are disproportionately affected by these developments. Migrants, refugees, ethnic and religious minorities, women, and LGBTQ+ individuals are more likely to be subjected to heightened scrutiny through biometric databases, predictive policing systems, and automated risk assessments. Because AI systems are often trained on historically biased data, they may replicate and intensify existing patterns of discrimination, producing outcomes that appear technologically neutral but are deeply unequal in practice. For individuals with limited digital literacy or legal access, challenging such decisions is frequently impossible.

Crucially, existing accountability mechanisms are often insufficient to address these threats. Many states lack clear legal frameworks governing the deployment of AI surveillance tools, and oversight bodies may not have the technical expertise or independence required to scrutinise their use. At the international level, the extraterritorial deployment of surveillance technologies and spyware complicates responsibility and remedy. As a result, violations of privacy and freedom of expression frequently occur without effective avenues for redress, reinforcing a climate of impunity.

4.3. Algorithmic Suppression, Content Moderation, and Censorship

Artificial intelligence now plays a central role in governing online speech, as digital platforms increasingly rely on automated systems to moderate, rank, recommend, or remove content. These systems operate at a scale that makes human oversight difficult, if not impossible, and their decisions directly shape which voices are heard and which are marginalised. While AI-based moderation is often justified as necessary to combat harmful content, its growing influence raises profound concerns regarding transparency, accountability, and the protection of freedom of expression under international human rights law.

One of the primary challenges lies in the opacity of algorithmic decision-making. Content moderation algorithms frequently function as “black boxes,” offering little insight into how decisions are reached or which criteria are prioritised. Users whose content is removed or demoted are often provided with vague or automated explanations, if any at all,

and meaningful avenues for appeal are limited. This lack of procedural transparency undermines the right to an effective remedy and prevents individuals from challenging decisions that may be arbitrary, discriminatory, or politically motivated.

Algorithmic suppression does not always take the form of explicit removal. Increasingly, platforms engage in subtler practices such as downranking, shadow-banning, or deprioritising content within recommendation systems. These measures can significantly reduce the visibility of certain narratives without notifying users, making suppression difficult to detect and contest. Minority viewpoints, content produced in less-resourced languages, and politically sensitive material are particularly vulnerable to such forms of invisibilisation, resulting in unequal access to the digital public sphere.

The interaction between state authority and private platform governance further complicates the landscape. Governments may exert formal or informal pressure on platforms to remove or restrict content, especially during periods of political unrest, elections, or security crises. In authoritarian contexts, AI moderation tools can be directly integrated into state censorship regimes, enabling large-scale, real-time suppression of dissent. In democratic states, compliance with restrictive laws or fear of regulatory penalties may incentivise platforms to over-remove content, leading to a phenomenon often described as “collateral censorship.”

5. Case Studies

The last decade witnessed the boom of Artificial Intelligence, the versatility of its services ranged from righteous to frowned upon ones.

The lack of supervision and binding regulations set the scene into chaos and resulted in a fragile system that could go wrong at any moment without accountability mechanisms to keep things in order. This defect could be seen in many aspects of our daily life, or on a bigger scale, in the international arena, cases of mass surveillance, blackmail cases, censorship over certain political views, etc. all are a proof of how pressing is the need for immediate action to protect basic privacy and freedom of speech rights. Here are a few real-life examples to help you further grasp the gravity of the situation.

5.1. AI Surveillance in Authoritarian Regimes

Predicted in Orwell’s 1984, in a land where every aspect of people’s lives were recorded and analysed via a TV centered in the living room, watching you all day long for any

mistake that would get you “vaporized”. Critics claim that this book reflects reality now more than ever and they aren't wrong.

Mass surveillance can subject a population or significant component thereof to indiscriminate monitoring, involving a systematic interference with people’s right to privacy and all the rights that privacy enables, including the freedom to express yourself and to protest. It can include: **communications** (calls, messages, emails), **Online behaviour** (search history, browsing data, etc.), **location tracking**, **facial recognition**, and so on. The word “mass” comes from it being applied to millions of people rather than only targeting specific individuals, as well as the fact that it happens without warrants, consent or suspicion especially that it is now integrated with daily used AI services.

States and different bodies advocate it claiming that it is vital for national security, counterterrorism, and behaviour monitoring in case of illegal tendencies. However, the lack of regulatory laws concerning it, positions this specific usage of AI in a grey area where these justifications could open the path for what is beyond righteous and lawful. especially that it usually comes accompanied with privacy violations, discrimination, authoritarianism, and lack of accountability. This threatens human rights, state sovereignty, and liberty.

Israeli mass surveillance over Palestinians

The Israeli government closely monitors the communications and movements of millions of Palestinians in the occupied territories (West Bank, East Jerusalem, Gaza), who have lived under Israeli military rule since 1967, and inside Israel, making Palestinians one of the most surveilled people in the world.This system of control includes restricting the physical movement of people and goods, dividing and isolating Palestinian cities and towns from one another and the outside world, and suppressing freedom of speech on and offline.

Israel uses cameras and facial recognition technology to record and track Palestinians, including at military checkpoints they’re forced to pass through in the occupied West Bank.

In November 2021, the Washington post reported that Israel escalated its monitoring of Palestinians in the West Bank over the past two years with a “broad surveillance effort” including use of a facial recognition technology called **Blue Wolf,** which one former Israeli soldier called the Israeli army’s secret “Facebook for Palestinians.” Soldiers were encouraged to take photos of Palestinians, including children and the elderly, for the database, with prizes awarded to units that gathered the most. It also reported that the Israeli army installed face-scanning cameras in Hebron, the largest city in the occupied West Bank, to identify Palestinians before they show their IDs at

checkpoints, part of a "wider network of closed-circuit television cameras, dubbed 'Hebron Smart City,' [that] provides real-time monitoring of the city's population and, one former soldier said, "we can sometimes see into private homes." Israel has also installed an extensive network of cameras with a facial recognition system in the Old City of occupied East Jerusalem.

- Israel uses military drones and balloons extensively to monitor Palestinians in the occupied territories, particularly in besieged Gaza.

Combined with the fact that Israel also uses drones to tear gas and kill Palestinians, their ubiquitous presence overhead causes widespread anxiety and fear among the population.

- Israel uses powerful spyware technology like pegasus, developed by the Israeli military and private companies run by former soldiers, to spy on the smartphones of Palestinians.

Pegasus allows users to spy on smartphones, accessing the targeted individual's encrypted communications, audio and video files, photos, location data, camera and microphone. In early November 2021, it was revealed that Israel had been using Pegasus to spy on six leading Palestinian civil society organizations, including human rights defenders documenting Israeli abuses. The Palestinian Authority said Pegasus spyware had also been found on the phones of three senior Palestinian diplomats. On November 3, 2021, the Biden administration imposed sanctions on the Israeli firm that sells Pegasus (with approval from Israel's government), the NSO Group, because the spyware has repeatedly been used by authoritarian governments to "maliciously target" journalists, human rights defenders, political dissidents, and others.

- Israel uses collaborators to spy on other Palestinians and to incite conflict among Palestinians, often pressuring people into cooperating with violence, imprisonment, and blackmail based on personal information acquired from surveillance or other collaborators.

Since it began its occupation of the Palestinian territories during the June 1967 war, Israel's military and secret police, the Shin Bet, have used collaborators to spy on the occupied Palestinian population. Frequently, Palestinians are coerced into collaborating with physical violence, imprisonment, threats to publicly expose potentially embarrassing personal information, or to deny them or their loved ones the ability to travel to undergo medical treatment not available in the West Bank or Gaza.

- **Israeli restrictions on the physical movement of Palestinians**

- Israel uses biometric ID Cards, travel permits, and control of the population registry in the occupied territories to monitor Palestinians and to limit where and with whom they can live and where they can travel.

All Palestinians in the occupied territories are required to have Israeli-issued ID cards that are color-coded, affecting everything from freedom of movement to family unity. Palestinians in the West Bank and Gaza have green IDs and Palestinians in East

Jerusalem have blue IDs. Palestinians with green IDs are barred from entering Jerusalem or other parts of historic Palestine inside Israel's internationally recognized pre-1967 borders without special permission, which is rarely granted. All Palestinians in the occupied territories require Israeli permission to travel abroad. Israel's control of the population registry provides the Israel army with a huge database of information covering every Palestinian in the West Bank and Gaza, allowing it to control their movements and residency rights.according to Human Rights Watch.

Russian mass surveillance over Ukrainians

- **Russia Clamps Down on Online Searches:**

Following its full-scale invasion of Ukraine in February 2022, Russia has moved aggressively to tighten its control over both the digital and physical realms, accelerating a broad shift toward wartime censorship and enhanced state surveillance. Online, authorities have blocked access to independent media, human rights organizations, political opposition platforms, and numerous foreign websites that fail to comply with increasingly restrictive regulations governing online activity.

As these limitations expanded, many Russians turned to VPNs to bypass state controls. Although Russian law already prohibited VPN providers from offering access to blocked websites—and hundreds of such services have been banned—individual VPN users had not previously faced legal consequences. A new bill alters this landscape by not only targeting VPN use directly but also banning the sharing of SIM cards and online accounts among individuals. These measures significantly strengthen the state's capacity for persistent user identification, reinforcing an already tightening surveillance environment.

At the same time, Russia has rapidly expanded its vast facial recognition infrastructure. According to The Moscow Times' Russian service, which analyzed thousands of procurement documents, the government has accelerated investment in surveillance systems since the start of the war. While officially promoted as tools for crime prevention, experts note that these systems have been most effective in identifying anti-war activists and draft evaders. Authorities reportedly employ an unpredictable pattern of surveillance-related detentions to cultivate fear and uncertainty among the population.

Digital rights activist Sarkis Darbinyan describes the system as still incomplete but advancing at high speed. Since 2022, Russian government agencies have spent 30.7 billion rubles (approximately **$330 million) on video surveillance data storage alone** - representing

one-third of all such spending in the past twelve years. This surge in investment underscores the Kremlin's broader effort to consolidate control over information, movement, and individual identity in the post-invasion security landscape.

More than **half a million facial recognition** surveillance cameras are now installed across Russia. Moscow alone accounts for around 40% of them, with over 200,000 cameras, and its surveillance model is rapidly spreading to other regions. Beyond the Moscow region—which has 80,000 cameras—and St. Petersburg with 67,000, the republic of Tatarstan ranks as the most heavily surveilled area, with 32,000 cameras. At the other end of the spectrum, the remote Chukotka autonomous district in the Far East has only 15 cameras, making it the least monitored region in the country.

Russia's state-owned telecom company, Rostelecom, serves as the main provider of surveillance technology and services nationwide, either directly or through its subsidiaries. According to The Moscow Times' Russian service, Rostelecom received 11.1 billion rubles ($119 million) from the 25.5 billion rubles ($273 million) allocated to the federal "Safe Region" and "Safe City" surveillance programs, underscoring its central role in expanding the country's monitoring infrastructure.

Palestine

Israel systematically censors the Palestinian narrative in the digital space through coordinated state mechanisms and the cooperation of major social media platforms, particularly Meta. Palestinian journalists, activists, and content creators face widespread post deletions, account suspensions, livestream bans, and reduced reach, especially during periods of heightened violence such as May 2021 and the ongoing genocide in Gaza. These restrictions occur precisely when Palestinians rely on digital platforms to document human rights violations, as Gaza's communications infrastructure is destroyed and international journalists are barred. In contrast, Israeli content containing explicit hate speech and calls for violence against Palestinians is frequently left online, revealing clear double standards in content moderation.

This censorship is enabled by opaque tools such as Meta's secret "Dangerous Individuals and Organizations" blacklist, which includes dozens of Palestinian entities and automatically suppresses their content without transparency or due process. Israeli authorities, particularly the Cyber Unit within the Attorney General's Office, submit thousands of takedown requests to social media companies each year, with platforms complying in the vast

majority of cases despite minimal evidence of incitement. Additional pressure is exerted through mass-reporting networks, pro-government digital campaigns, and Israeli legislation such as the "Facebook Law," which forces platforms to remove content deemed threatening to Israeli security under vague definitions. Together, these mechanisms criminalize Palestinian expression, distort public discourse, and transform social media from a space of free expression into a tool for silencing the Palestinian narrative.

Ukraine

Following Russia's full-scale invasion of Ukraine in February 2022, the Russian government significantly intensified restrictions on media freedom and independent journalism. The state media regulator, Roskomnadzor, imposed strict censorship rules prohibiting the use of terms such as "war," "invasion," or "attack," requiring media outlets to rely solely on official government sources. Numerous independent and foreign-linked media outlets were blocked or prosecuted for allegedly disseminating "false information," including reporting on civilian casualties and military losses. Social media platforms such as Twitter and Facebook were partially restricted, while journalists covering anti-war protests faced arrests, detention, harassment, and professional sanctions. These measures have been widely described by press freedom organizations, including Reporters Without Borders (RSF), as an effort to control public discourse, suppress dissent, and consolidate state narratives during wartime. As a result, Russia has experienced a sharp decline in press freedom and independent reporting, reinforcing concerns about authoritarian information control during armed conflict.

5.2. Regulation of Digital Platforms in Democratic States

Digital platforms such as social media networks, search engines, and messaging applications have become essential spaces for political discussion, information sharing, and democratic participation. They influence public opinion, shape political agendas, and play a growing role in elections and social movements. However, the rapid expansion of these platforms—especially with the integration of artificial intelligence—has also created serious challenges for democracy. Disinformation, foreign interference, hate speech, and AI-generated content have spread more easily, deepening social divisions and undermining trust in democratic institutions. At the same time, a small number of private technology companies now exercise significant control over online speech, often through opaque

moderation policies that raise concerns about transparency, accountability, and freedom of expression.

Democratic states are increasingly struggling to find the right balance between protecting free speech and privacy while preventing online harm and manipulation. Different approaches have emerged across regions, ranging from strict regulatory frameworks to voluntary self-regulation by platforms. These differences reflect broader debates over how much authority governments should have over digital spaces without sliding into censorship or authoritarian control. As artificial intelligence continues to transform how information is created and distributed, ensuring that digital platforms support democratic values rather than weaken them has become a central challenge for policymakers worldwide.

Germany

Germany has taken steps to regulate large technology companies in order to protect user privacy and freedom of expression in the digital age. The German government works closely with the European Union to limit the power of major platforms such as Google, Meta (Facebook), Amazon, Apple, and Microsoft, especially where artificial intelligence and large-scale data collection are involved. These platforms hold vast amounts of personal data and use algorithms that can influence what people see, say, and share online.

German authorities have intervened when companies combined user data without clear consent, favored their own services over competitors, or used non-transparent algorithms that could restrict fair access to information. By requiring clearer data consent, improving interoperability between services, and preventing unfair platform practices, Germany aims to reduce the risk of mass surveillance, data misuse, and manipulation of online discourse. This approach shows how democratic states can regulate AI-driven platforms while still supporting innovation, protecting privacy, and preserving open and diverse online spaces.

6. Possible Solutions and Policy Recommendations

At its foundation, a human rights–based approach situates existing international human rights law—rather than technological innovation or market efficiency—as the primary reference point for AI governance. This includes binding treaties such as the International Covenant on Civil and Political Rights, as well as interpretive guidance from UN human rights bodies. Importantly, this framework rejects the notion that AI requires entirely new ethical paradigms; instead, it emphasizes that established rights to privacy, freedom of expression, non-discrimination, and access to remedy remain fully applicable

regardless of technological medium. Artificial intelligence does not exist in a legal vacuum, and its governance must therefore be anchored in obligations that states have already voluntarily accepted.

Central to this approach is the principle of state responsibility, even where AI systems are developed or operated by private actors. While much of the AI infrastructure governing surveillance, content moderation, and data processing is controlled by corporations, international human rights law affirms that states have positive obligations to protect individuals from rights violations by third parties. Governments therefore cannot evade accountability by outsourcing core governance functions to technology companies. Instead, they are required to regulate, supervise, and constrain AI deployment in ways that prevent arbitrary interference with fundamental rights.

Equally important is the recognition that accountability must extend across the entire AI lifecycle. Human rights–based governance rejects static or one-time regulatory models and instead emphasizes continuous oversight from design and development through deployment and post-deployment monitoring. AI systems evolve through machine learning, data accumulation, and contextual adaptation, meaning that risks to privacy and expression may intensify over time. Preventive mechanisms such as human rights due diligence and impact assessments are therefore essential. These processes require states and companies to identify foreseeable risks, assess disproportionate impacts on vulnerable groups, and implement mitigation measures before harm becomes systemic. In doing so, human rights–based AI governance shifts the focus from reactive enforcement to structural prevention, reinforcing the primacy of human dignity, legality, and democratic accountability in the digital age.

6.2. Transparency, Accountability, and Algorithmic Audits

Transparency and accountability constitute indispensable pillars of rights: respecting artificial intelligence governance, particularly where AI systems are deployed in contexts that directly affect privacy, freedom of expression, and democratic participation. Algorithmic systems increasingly mediate access to information, determine the visibility of political speech, and guide surveillance and law-enforcement practices. Yet the logic, data sources, and decision-making processes of these systems often remain opaque to individuals, regulators, and even the institutions deploying them. This opacity undermines core procedural guarantees embedded in international human rights law, including the right to an effective remedy and the principle that interferences with rights must be lawful, necessary, and proportionate.

From a human rights perspective, transparency is not merely a technical preference but a legal prerequisite. When individuals are subject to algorithmic decision-making—such as content removal, account suspension, predictive monitoring, or biometric identification—they must be able to understand the basis on which these decisions are made. Without access to meaningful information about how AI systems operate, affected individuals are effectively denied the ability to contest decisions or seek redress, rendering rights protections illusory in practice. International human rights bodies have increasingly emphasized that transparency obligations extend to both states and private actors when their systems exercise governance-like power over public discourse or personal data.

Accountability mechanisms are essential to ensuring that transparency translates into enforceable responsibility rather than symbolic disclosure. Human rights law requires that violations be attributable to identifiable duty-bearers and that remedies be available when harm occurs. In the context of AI, this necessitates clear lines of responsibility between governments, technology developers, and platform operators. States cannot rely on claims of technical complexity or proprietary secrecy to avoid accountability for systems used in surveillance or content moderation. Instead, they must establish regulatory frameworks that clarify liability, mandate compliance with human rights standards, and provide oversight bodies with the authority to investigate and sanction violations.

Algorithmic audits have emerged as a central tool for operationalizing accountability in practice. These audits involve systematic assessments of AI systems to evaluate their compliance with legal standards, detect discriminatory outcomes, and identify disproportionate impacts on fundamental rights. Independent audits are particularly important where AI systems are used at scale, as small design choices or data biases can produce widespread and cumulative harm. From a rights-based perspective, audits should assess not only technical performance but also social and legal impact, including effects on marginalized communities, political participation, and access to information.

6.3 Safeguards Against AI-Enabled Censorship

The increasing reliance on artificial intelligence for content moderation, information filtering, and platform governance has introduced new and complex risks to freedom of expression. While AI systems are often justified as necessary tools to combat disinformation, hate speech, or violent extremism, their deployment has frequently resulted in the over-removal of lawful content and the disproportionate suppression of

political and dissenting voices. Automated systems lack the contextual, cultural, and linguistic sensitivity required to assess speech in line with international human rights standards, leading to errors that scale rapidly across digital platforms.

In many regions, AI-enabled censorship has become intertwined with state pressure on technology companies to restrict specific narratives, particularly during periods of political instability, armed conflict, or electoral competition. This dynamic risks normalizing algorithmic censorship as an administrative function rather than a rights-restricting measure subject to strict legal scrutiny. To address this, safeguards must ensure that AI systems do not independently determine the legality of expression. Human oversight, transparency in moderation criteria, and effective appeal mechanisms are essential to prevent automated enforcement from replacing lawful adjudication. Without these protections, AI threatens to transform freedom of expression from a guaranteed right into a conditional privilege governed by opaque algorithms.

6.4 Protecting Digital Civic Space and Human Rights Defenders

Digital technologies have reshaped civic space, enabling unprecedented levels of participation, mobilization, and transnational advocacy. At the same time, artificial intelligence has intensified the surveillance and control of these spaces, exposing journalists, activists, and human rights defenders to heightened risks. AI-driven monitoring tools—such as facial recognition, predictive analytics, and large-scale data aggregation—are increasingly used to map social networks, track online behavior, and identify individuals engaged in dissent or advocacy, often without judicial oversight.

This expansion of digital surveillance has a chilling effect on civic participation, discouraging individuals from engaging in lawful expression due to fear of monitoring or retaliation. Human rights defenders, in particular, face compounded risks as AI systems are used to discredit, harass, or silence their work through coordinated reporting, algorithmic suppression, or targeted disinformation campaigns. Protecting digital civic space therefore requires more than technical safeguards; it demands legal and institutional protections that limit surveillance practices, safeguard anonymity, and uphold the confidentiality of journalistic and activist communications. Ensuring the safety of those who defend human rights is essential to preserving democratic accountability and pluralism in the digital age.

6.5 Capacity-Building for Low-Resource States

The governance challenges posed by artificial intelligence are global, yet the capacity to regulate AI is unevenly distributed across states. Many low-resource countries face significant barriers to implementing effective AI oversight, including limited technical expertise, weak regulatory institutions, and constrained access to independent auditing mechanisms. As a result, AI technologies are often adopted without adequate assessment of their human rights impact, leaving populations vulnerable to unregulated surveillance, data exploitation, and algorithmic discrimination.

This disparity risks creating a two-tiered system of AI governance in which robust safeguards exist in some regions while others become testing grounds for invasive or experimental technologies. Capacity-building is therefore a critical component of equitable AI governance. International cooperation, technical assistance, and knowledge-sharing initiatives can support low-resource states in developing legal frameworks, regulatory bodies, and enforcement mechanisms aligned with international human rights standards. Strengthening capacity not only protects individuals within these states, but also reinforces the universality of human rights by ensuring that technological advancement does not deepen existing global inequalities.

7. References

1. IMEU. (n.d.). *Fact sheet: Israeli surveillance and restrictions on Palestinian movement*. Institute for Middle East Understanding. https://imeu.org/resources/key-issues/fact-sheet-israeli-surveillance-restrictions-on-palestinian-movement/128

2. Vinograd, C. (2021, November 5). *Israel uses facial recognition to track Palestinians in West Bank. The Washington Post*. https://www.washingtonpost.com/world/middle_east/israel-palestinians-surveillance-facial-recognition/2021/11/05/3787bf42-26b2-11ec-8739-5cb6aba30a30_story.html

3. Privacy International. (2021). *Counterterrorism and biometrics: Israel/Palestine*. https://privacyinternational.org/sites/default/files/2021-06/PI%20Counterterrorism%20and%20Biometrics%20Report%20Israel_Palestine%20v7.pdf

4. Harel, A. (2021, April 28). *Israeli police using drones to drop tear gas on Palestinian demonstrators. Haaretz*. https://www.haaretz.com/israel-news/2021-04-28/ty-article/.premium/israeli-police-using-drones-to-drop-tear-gas-on-palestinian-demonstrators/0000017f-f49d-d887-a7ff-fcfd91cd0000

5. Pegasus Project. (2021, July 18). *What is Pegasus spyware and how does it hack phones? The Guardian*. https://www.theguardian.com/news/2021/jul/18/what-is-pegasus-spyware-and-how-does-it-hack-phones

6. Benner, K., & Perlroth, N. (2021, November 3). *U.S. blacklists Israeli spyware firm NSO Group. The New York Times*. https://www.nytimes.com/2021/11/03/business/nso-group-spyware-blacklist.html

7. Harel, A. (2008, September 4). *Shin Bet to Palestinian: Collaborate or go to jail. Haaretz*. https://www.haaretz.com/2008-09-04/ty-article/shin-bet-to-palestinian-collaborate-or-go-to-jail/0000017f-dbe3-d3a5-af7f-fbef33f40000

8. Kershner, I. (2014, September 13). *Elite Israeli officers decry treatment of Palestinians. The New York Times*. https://www.nytimes.com/2014/09/13/world/middleeast/elite-israeli-officers-decry-treatment-of-palestinians.html

9. The Intercept. (2017, December 30). *Facebook says it is deleting accounts at the direction of the U.S. and Israeli governments*. https://theintercept.com/2017/12/30/facebook-says-it-is-deleting-accounts-at-the-direction-of-the-u-s-and-israeli-governments/

10. IMEU. (n.d.). *Fact sheet: Is Israel an apartheid state?* Institute for Middle East Understanding. https://imeu.org/resources/key-issues/fact-sheet-is-israel-an-apartheid-state/214

11. Halbfinger, D. M., & Rasgon, A. (2021, May 15). *Israel-Gaza fighting escalates amid civilian casualties. The New York Times*. https://www.nytimes.com/2021/05/15/world/middleeast/israel-palestinian-gaza-war.html

12. Dwoskin, E. (2021, May 28). *Facebook struggles to police content related to Israel and Palestine. The Washington Post*. https://www.washingtonpost.com/technology/2021/05/28/facebook-palestinian-censorship/

13. The Guardian. (2021, May 26). *Pro-Palestine content censored on Facebook and Instagram*. https://www.theguardian.com/media/2021/may/26/pro-palestine-censorship-facebook-instagram

14. Human Rights Watch. (2012). *"Forget about him, he's not here": Israel's control of Palestinian residency in the West Bank*. https://www.hrw.org/report/2012/02/05/forget-about-him-hes-not-here/israels-control-palestinian-residency-west-bank-and

15. Human Rights Watch. (2025, July 24). *Russia clamps down on online searches*. https://www.hrw.org/news/2025/07/24/russia-clamps-down-on-online-searches

16. Human Rights Watch. (2022, March 7). *Russia criminalizes independent war reporting and anti-war protests*. https://www.hrw.org/news/2022/03/07/russia-criminalizes-independent-war-reporting-anti-war-protests

17. Human Rights Watch. (2022, March 16). *Russia, Ukraine, and social media and messaging apps*. https://www.hrw.org/news/2022/03/16/russia-ukraine-and-social-media-and-messaging-apps

18. The Moscow Times. (2023, August 17). *Mass surveillance in Russia expands rapidly since Ukraine invasion*. https://www.themoscowtimes.com/2023/08/17/mass-survellience-in-russia-expands-rapidly-since-ukraine-invasion-mt-russian-a82151

19. United Nations Human Rights Committee. (2011). *General Comment No. 34: Article 19 – Freedoms of opinion and expression* (CCPR/C/GC/34). United Nations.

20. Office of the United Nations High Commissioner for Human Rights. (2018). *The right to privacy in the digital age* (A/HRC/39/29). United Nations.

21. UNESCO. (2023). *Guidance on the governance of digital platforms: Protecting freedom of expression and access to information in a digital age*. United Nations Educational, Scientific and Cultural Organization.

22. Nashif, N. (2023, June 15). *Digital warfare: Israeli censorship of Palestinian content*. Institute for Palestine Studies. https://www.palestine-studies.org/en/node/1653973

23. 7amleh – The Arab Center for the Advancement of Social Media**.** (2024, December 18). *Erased and suppressed: Palestinian testimonies of Meta's censorship*. https://7amleh.org/post/erased-and-suppressed-palestinian-testimonies-of-meta-s-censorship-en

24. Reporters Without Borders (RSF). (2022, March 1). *Russian regulator censors Ukraine war coverage, reporters told to toe Kremlin line*. https://rsf.org/en/russian-regulator-censors-ukraine-war-coverage-reporters-told-toe-kremlin-line

25. International Bar Association (IBA). (n.d.). *Digital platform regulation in Germany and the Digital Markets Act*. https://www.ibanet.org/digital-platform-regulation-germany-dma

26. Canadian Lawyer Magazine. (n.d.). *A look at the laws governing freedom of expression on the internet*. https://www.canadianlawyermag.com/resources/legal-education/a-look-at-the-laws-governing-freedom-of-expression-on-the-internet/390412