JOINT DECLARATION ON AI, FREEDOM OF EXPRESSION AND MEDIA FREEDOM

The United Nations (UN) Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression, and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information in Africa ("mandate holders"),

Recalling and reaffirming the Joint Declarations of 26 November 1999, 30 November 2000, 20 November 2001, 10 December 2002, 18 December 2003, 6 December 2004, 21 December 2005, 19 December 2006, 12 December 2007, 10 December 2008, 15 May 2009, 3 February 2010, 1 June 2011, 25 June 2012, 4 May 2013, 6 May 2014, 4 May 2015, 4 May 2016, 3 March 2017, 2 May 2018,10 July 2019, 30 April 2020, 20 October 2021, 3 May 2022, 2 May 2023 and 3 May 2024;

Emphasising that AI systems should strengthen and not undermine freedom of expression and the information ecosystem, and *underscoring* the shared responsibility of States, business enterprises, technology investors, media actors, civil society and academia to ensure that the design, development and deployment of artificial intelligence (AI) systems fully comply with international human rights law, principles and standards;

Noting with concern that voices and perspectives from the Global Majority, as well as of women, minorities and marginalised communities globally are underrepresented in AI across its lifecycle, and *underscoring* the need to close the digital divide and promote an inclusive and intersectional approach in the design, development, training, deployment and governance of AI;

Alarmed that a small number of large corporate actors wield disproportionate influence over research agendas, standard-setting, policymaking and public debate on AI, with little or no transparency or accountability for the social, political and economic impact of their technologies;

Stressing that the benefits of AI must be accessible to all States and the entire population, including women, minorities and marginalised groups, while its risks must be rigorously assessed, mitigated and prevented;

Recognising that the implications of AI for freedom of expression, including media freedom and pluralism, vary according to its use and the broader social, economic and political context;

Mindful that AI is a rapidly evolving field which requires continuous learning and review to ensure that the challenges are accurately understood and assessed and solutions effectively designed and implemented, and *affirming* that human rights cannot be claimed by artificial intelligence;

Adopt, on 24 October 2025, the following Joint Declaration on Artificial Intelligence, Freedom of Expression and Media Freedom.

Risks, challenges and opportunities to freedom of expression arising from the use of AI

A. Impact on access to information

Artificial intelligence has fundamentally reshaped access to information. For instance, AI-powered services like search engines, recommender systems, personalisation and translation, can enable easier and faster access to reliable information, including media content. Generative AI (Gen AI) systems facilitate wider diffusion of content and interactive adaptation of complex issues, breaking down barriers of language, technical knowledge and format. Features such as speech-to-text or image-to-speech expand accessibility, particularly for persons living with disabilities.

At the same time, recommender systems, and other AI-powered curation tools exert a large hidden influence and gatekeeper role over what information and media people access and consume, thereby concentrating power over public debate and knowledge. AI models have the capacity to prioritize engagement over factual accuracy or pluralistic sources, with the potential of flooding information spaces with inauthentic content and impacting the availability and diversity of information online. Gen AI can engage in deceptive "sycophancy" — mirroring beliefs and producing persuasive or flattering outputs — thereby undermining the right to accurate and pluralistic information.

B. Impact on freedom of expression

AI and Gen AI systems have the potential to enhance the space and means for individuals to express themselves, enabling marginalised groups, including racial, ethnic and religious minorities as well as women and LGBTQ+ communities, and those with dissenting and diverse views to engage in outreach, amplify their perspectives, and participate in public life. AI can also contribute to safer digital environments by helping digital platforms detect illegal and/or harmful content such as hate speech, incitement to violence, graphic violence, deepfakes, harassment and abuse.

At the same time, AI-generated disinformation and hate speech have polluted information ecosystems, with severe negative implications for human rights as well as participatory democracy. Deepfakes can be exploited to chill public participation of women, particularly women journalists, human rights defenders and political figures. The use of AI content moderation can lead to over-removal, discrimination and censorship. Reliance on inherently biased datasets and opaque training processes can amplify pre-existing inequalities, risking homogenisation of expression, and erasure of linguistic and cultural diversity. AI-driven surveillance, deployed maliciously or without proper safeguards can target people based on ethnicity, gender or physical appearance.

C. Impact on freedom of opinion

AI and Gen AI can enhance freedom of opinion by expanding access to diverse, multimodal and personalised content, strengthening people's ability to form opinions and participate in public life.

However, AI-driven personalisation and "micro-targeting" can manipulate opinion through sophisticated and non-consensual means interfering with individual agency. Gen AI can lead to personalised and interactive persuasion, leveraging individual behaviours and habits to steer exposure to certain information over time. Such practices can sway political perceptions, interfere with public discourse across borders, suppress dissenting views or favour government

narratives, radicalise individuals, or even undermine creativity and cognitive processes such as attention and critical thinking. These technologies can also pose specific risks for children, elderly, and vulnerable groups, harming mental and emotional health and development.

Where AI undermines human autonomy or creativity, it unacceptably intrudes into individuals' absolute right under international law to form their opinions free from non-consensual interference and manipulation. The opacity of AI tools aggravates the risk to freedom of opinion.

D. Impact on media freedom, diversity and pluralism

AI tools are increasingly used by journalists and media actors to increase the efficiency of the newsgathering process, aid in the dissemination of news, and for tasks such as investigative data analysis, fact-checking, automated content generation, translation, transcription, reporting in accessible ways, or for content organisation, customisation and personalisation. As AI becomes embedded in news production and dissemination, it plays a fundamental role in shaping public debate, political agenda and the range of information available to individuals.

Despite these advantages, AI poses significant risks and challenges for journalism. The concentration of power and control among a few large corporate actors for the design, development, training and deployment of AI systems, and growing media dependencies on AI risking to influence editorial choice, raise serious concerns for the diversity, pluralism, inclusiveness, reliability and accountability of the media and information ecosystem more broadly.

AI companies have widely extracted media content, often without consent or compensation, for training their tools, and are increasingly gatekeeping, in particular with the rise of Gen AI-based search, between media and audiences. This raises concerns not only about fair remuneration and economic viability of independent media, but also about the impact of the use of AI on news visibility, pluralism and attribution as well as the overall safety of media actors. This also raises concerns about potential copyright infringement and a lack of linguistic and cultural representation, and reduced access to diverse and local information.

The long-term implications of AI on media freedom and democratic debate are yet to be fully understood.

General principles

States' and other stakeholders' responses to the impact of AI on freedom of expression and media freedom and pluralism should be guided by the following principles:

1. Safeguard freedom of opinion and expression, including media freedom and the right to information

The right to freedom of opinion and expression is integral to human dignity, autonomy and creativity, and must be embedded throughout the lifecycle of AI, including its design, development, training and deployment.

The development, design and deployment of AI must uphold at all times the absolute nature of the right to freedom of opinion and should be guaranteed throughout the AI lifecycle.

While the protection of the right to freedom of expression under international law is broad and inclusive, it is not absolute. When responding to the risks and challenges posed by AI, any

restriction on the right to seek, receive and impart information must respect the standards set out in international human rights law and the human agency of rights holders.

2. Promote a healthy, pluralistic and diverse information environment

The promotion and accessibility of a broad and diverse range of quality and accurate information should be embedded throughout the entire lifecycle of development, training and deployment of AI. AI systems should be built and utilised to enhance, not undermine, public trust in the reliability and integrity of the information order.

The concentration of corporate power in AI technologies is a substantial risk to pluralism and should be mitigated through appropriate human rights-based regulation, to ensure transparency and accountability, and investment in alternative approaches which promote diversity.

3. Respect privacy and data protection

The right to privacy and international data protection standards, including consent, purpose limitation, transparency and accountability should be respected throughout the entire AI lifecycle. The reliance of AI on vast datasets, including personal data, makes robust safeguards essential to uphold these fundamental rights.

4. Guarantee the right to equality and non-discrimination

The right to equality and non-discrimination is a foundation of international human rights law. AI systems should be designed, developed, trained and deployed in ways that mitigate bias and discrimination, ensure inclusivity, reflect diverse cultural contexts and safeguard the equal exercise of freedom of opinion and expression for all communities. It is also essential that the right to freedom of expression is carefully balanced with the right to equality and non-discrimination.

5. Promote equitable access to opportunities

Equitable access to the benefits of AI technology can facilitate the conditions that allow people to meet their labour, economic, cultural and social needs, and the equal enjoyment of the right to freedom of opinion and expression, including media freedom. Equitable access can be guaranteed by closing the digital divide, eliminating the barriers to AI use by vulnerable and marginalised groups, promoting digital literacy and, particularly, by fostering the development and access to inclusive, sustainable, trustworthy and human rights-based AI technologies. Equitable access should be accompanied by the promotion of the knowledge, attitudes and skills needed for people to take full advantage of technological advances, and by enabling people to decide for themselves whether the use of these technologies is helpful to them and in line with their values and needs. In this sense, the Global Majority should not be treated as passive recipients of technology, but as agents with the ability to shape technology to be human rights-compliant.

6. Ensure transparency, accountability, and effective remedies

Transparency and accountability, including the availability of effective remedies, are key principles to guarantee a human rights-centric approach to AI. Respect for transparency requires providing meaningful information and access to data, appropriate to the context, to foster general understanding around how AI impacts access to information and freedom of expression, including media freedom, and enable accountability.

7. Promote multistakeholder and multilateral cooperation

Given the cross-border and global impact of AI, multilateral cooperation and multi-stakeholder participation in AI governance are vital, especially at a time when civic space and multilateral institutions are under severe attack. All relevant stakeholders should consider the impacts of AI on the expression of vulnerable and marginalised, including racial, ethnic and religious minorities, women and girls, indigenous communities and persons with disabilities, and meaningfully include their views. Multilateral organisations and governance frameworks should support AI-related capacity building in the Global Majority countries while simultaneously addressing potential negative impacts on freedom of expression around the world.

Recommendations

States

Under international law, States are obliged to respect freedom of expression, including media freedom, and to create an enabling environment for these freedoms to thrive. All States should:

a. Use AI responsibly, transparently, in line with their international human rights obligations and in good faith to *inter alia* facilitate access to information, counter the spread of disinformation and other harmful expression, and enable inclusive participation in public life. States should promote timely access to quality information especially during critical moments such as elections, conflicts or crises. States, at the same time, should not over-rely on AI for disseminating information, and implement safeguards to ensure information provided with the support of AI is accurate, factual and verifiable.

b. Promote equitable access to the benefits and opportunities of AI while addressing its risks and challenges, including through adopting policies and digital education programs ensuring that the benefits of AI are meaningfully accessible in rural, low-income and disconnected areas, and by all communities, including women, racial, religious and linguistic minorities, indigenous populations, and the elderly.

c. Ensure, when responding to the risks and challenges posed by AI, that any restriction on the right to seek, receive and impart information conform to international standards on freedom of expression, specifically that such restriction is prescribed by law, pursue legitimate aim and is necessary and proportionate. Such restrictions should not result in an illegitimate suppression of political commentary, satire or other legitimate forms of expression that are permitted by international law.

d. Ensure that the use of AI by public authorities for content creation or dissemination contributes to a pluralistic and healthy information environment. Public authorities and state officials must refrain from using AI to spread hate speech, disinformation, misinformation and propaganda. In particular, they should not use AI to foment hatred and discrimination against minorities to deny their participation in public life, spread disinformation to vilify journalists or political opponents, or undermine access to reliable information particularly in the context of electoral processes. AI should never be used as an instrument of propaganda for war.

e. Refrain from using digital surveillance technologies against human rights defenders, journalists and civil society as a tool to silence the media and discourage criticism of the government, and more in general refrain from developing, deploying or permitting the use

of AI systems in any way that can clearly undermine freedom of expression and media freedom, whether within their jurisdiction or extraterritorially.

f. Require conducting human rights due diligence as a necessary condition for public procurement, private public partnership and similar forms of cooperation with private companies with regards to AI.

g. Put in place safeguards against bias, discrimination and gender-based violence**.** Among others, they should:

   i. Undertake pre-emptive and proactive efforts to address the structural and institutionalised roots of AI-generated online violence against women, minority groups and other groups.

   ii. Adopt and/or implement legislation to prohibit, investigate and prosecute AI generated online violence against women, minorities and/or other groups at risk, and ensure proper implementation of this legislation by law enforcement authorities, judiciary and private actors. Such legislation should be grounded in international human rights standards.

   iii. Designate independent oversight bodies, with expertise in the areas of equality and non-discrimination, to monitor and address unequal or discriminatory effects of AI on minority groups and other groups at risk, and support them as well as independent researchers and civil society in this area.

h. Guarantee that any public investment in the design, development, training and deployment of AI systems is founded in human rights. This includes, among others, ensuring transparency and public consultation, evaluation of human rights impacts, including on freedom of expression and media freedom, and independent oversight and accountability.

i. Recognise the cross-border nature of the impacts of AI on freedom of expression, including media freedom. To that effect, support multi-stakeholder and inclusive AI governance with participation from civil society, academia, and vulnerable communities and minorities, and foster international cooperation and coordination of findings and observations of AI impacts on freedom of expression and media freedom, in view of addressing them jointly.

State Regulation of Companies

States hosting or otherwise having jurisdiction over companies that develop or deploy AI have a responsibility to introduce and enforce regulations to mitigate their negative impacts, in line with international human rights standards. States seeking to regulate companies should be guided by the following recommendations. They should:

a. Introduce regulations to protect freedom of opinion and expression, including media freedom from violation by private actors, including companies, which, as a minimum:

   i. Require that human rights due diligence is conducted systematically, iteratively and robustly at every stage of the AI lifecycle. The due diligence should specifically include the assessment of implications for freedom of expression, including media freedom and be subject to external, independent and transparent scrutiny. The due diligence should include a specific assessment for marginalised persons or groups,

more susceptible to negative mental health impacts, opinion manipulation, and online violence and harassment.

ii. Require the timely development and implementation of mitigation measures when adverse impacts of AI on freedom of expression or media freedom are detected, and provide for or cooperate in their remediation through legitimate processes.

iii. Impose meaningful transparency requirements across the AI lifecycle, to help identify and correct biases, identify harms, attribute liability, and provide for effective remedies.

b. Establish robust institutions for oversight and accountability of AI. This includes, but is not limited to:

i. Set oversight and testing mechanisms to observe the impact of AI, and especially Gen AI, on freedom of expression and on media freedom and ensure the availability and accessibility of the detailed findings, as well as of mitigation strategies, in order to increase human oversight, and to help raise awareness amongst stakeholders and users, acting as a means of epistemic counterpower (that is, strengthening the ability to question, verify and challenge dominant narratives and sources of information).

ii. Include human rights experts, civil society and academia in the development of AI policies and AI regulatory frameworks (including due diligence) and strengthen their capacity to contribute to oversight by providing support for independent research, capacity building and awareness raising.

c. Adopt strong data protection laws and other relevant laws that enable greater transparency in data collection, use and access, and limit extractive data mining practices and pervasive tracking and targeting of individuals and their activities online. Strong data protection, copyright and similar rules should ensure data extraction for AI training is based *inter alia* on consent and compensation.

d. Address concentration of market power and take measures to guarantee a competitive field in the AI domain, with the aim to promote plurality, diversity, inclusiveness and innovation that promotes freedom of expression, including media freedom.

e. Fund and support the development of AI models that prioritize linguistic and cultural diversity, which requires ensuring that AI assets are accessible to researchers and developers focusing on minority languages.

International inter-governmental, multilateral and multi-stakeholder organisations

a. Ensure that human rights, particularly freedom of opinion and expression, including media freedom, are incorporated and upheld in frameworks, best practices and policies governing AI's use in the information ecosystem.

b. Encourage and support States and companies to ensure that AI systems are designed and developed with the input of underrepresented and vulnerable groups and communities, prioritize the protection of those most at risk from AI-related harm, and protect and promote cultural and linguistic diversity.

c. Work with States and companies to identify and mitigate specific risks of AI for freedom of expression, including media freedom. These risks include, for example, the use of Gen

AI to perpetuate disinformation during periods of transition, crisis or conflict, or in situations of emergency.

Private sector

Companies designing, developing, training or deploying AI, as well as financial institutions and investors financing their activities, should adopt a human rights-centred approach, and respect and comply with the UN Guiding Principles on Business and Human Rights, to prevent the prioritisation of profit maximisation at the expenses of human rights and democratic values. This includes, but is not limited to:

a. Conduct systemic, iterative and robust human rights due diligence with specific attention to freedom of expression, including media freedom, at every layer of the AI lifecycle. This assessment must be made publicly available in accessible formats to facilitate monitoring and oversight by independent experts, civil society and academia. It should also be subject to external, independent and transparent scrutiny.

b. Hire diverse staff to ensure that diverse voices are being heard through the AI lifecycle which can contribute to lessening bias and guarantee that concerns from different societal groups are considered during AI design, development, training and deployment.

c. Train relevant staff on international human rights standards, and particularly those related to freedom of expression, including media freedom.

d. Ensure that AI systems are developed and trained in line with human rights law and standards and in full respect of copyright and data protection.

e. Engage meaningfully and transparently with civil society and human rights experts to develop and implement industry initiatives (strategies, industry code of practices, standards etc.) incorporating human rights by design and by default. These guidelines and standards should translate principles into concrete measures and include, among others, best practices and standards for fair data extraction, attribution and labelling of synthetic content, transparency, accountability and human oversight, as well as red lines for situations where the use of Gen AI or AI conflicts with human rights law and standards.

f. Provide meaningful information about how the system works, to enable users to interpret the system's output and use it appropriately. This includes information that allows appropriate traceability and explainability, as well as duly informing deployers of the capabilities and limitations of the system and affected persons about their rights.

g. Adopt clear policies to guarantee data fairness and inclusiveness, and the respect of privacy, data protection and the rights of others. These policies should be compliant with international human rights law and standards, easily accessible and understandable and applied and enforced consistently across geographical area, taking into account the particularity of each context.

h. Provide clarity, explainability and accessibility on the use of AI for content moderation, content curation and targeted advertising and provide an avenue for appealing content governance decisions. This mechanism should provide a prompt and specific time limit for resolution. When the review requires greater contextual analysis due to the quality of the content, the sociopolitical or cultural conditions, or the severity of the measure imposed, companies should provide the opportunity to have the remedy review be conducted with human oversight. This is without prejudice to the user's ability to avail

themselves of the courts to determine the compatibility of companies' measures with the right to freedom of expression, including media freedom.

i. Adopt and maintain measures to ensure the availability of accessible and effective remedies for freedom of expression, including media freedom violations resulting from the activities within the lifecycle of AI systems.

j. Proactively respond to concerns and comply with liability regimes at local, regional and international levels.

k. Refrain from exporting surveillance technology if there is significant risk it will be used to commit human rights violations and jeopardise freedom of expression.

In addition, and due to their role, financial institutions and investors, should:

a. Carry out human rights' due diligence process when funding and incentivising development and deployment of AI and Gen AI technologies, including assessment of judicial oversight, privacy protections, and compliance with international human rights law and standards.

b. Do not fund or invest in AI and Gen AI projects and companies that lack adequate human rights safeguards, are linked to systematic human rights and freedom of expression violations and/or whose activities are incompatible with international human rights law and standards.

Moreover, technology providers should respect professional journalistic ethics, editorial autonomy and media independence. To this end, they should:

a. Understand the specific needs and heightened risks faced by journalists and media actors in terms of human oversight of the whole media production chain, and of how their output is interpreted.

b. Refrain from practices that exploit structural dependencies of media on big tech and AI companies, and promote cooperation that supports media sustainability and editorial independence.

c. Recognise and account for the fact that even small changes or adaptations to key AI infrastructure and software can sometimes have substantial consequences for the news organisations' editorial autonomy, their realisation of professional values and their ability to deliver on their mission. Technology providers should therefore provide in situations of such changes or adaptations ample warning time and information, as well as reasonable alternatives.

d. Refrain from utilising the content of journalists and media organisations, such as in creating AI-generated summaries, without consent, attribution or compensation.

Civil society and human rights experts

Civil society and human rights experts can support both States and private companies as well as individuals and groups impacted by AI and the broader public in better understanding trends and developments, and possible preventive and reactive measures. Among others, they should:

a. Continue playing their vital watchdog role, holding governments and corporate actors alike accountable for applying international human rights law and standards on freedom of expression, including media freedom.

b. Sensitise companies and technical experts on international human rights law and standards and help them to adopt a human rights centric approach.

c. Contribute to monitoring, testing and observatory mechanisms for AI systems, to ensure a better understanding and assessment of their impact and risks.

d. Support States and private sectors in the development of codes of practice and other tools to clarify transparency obligations, oversight systems and accountability requirements.

e. Support individuals and groups negatively affected by AI with legal and psychological assistance, guidance on safety and security, and mechanisms for reporting and escalation.

f. Promote information literacy to help people to recognise and differentiate between reliable, verified information and unverified information and to better comprehend how AI works, how it is embedded in our lives and how it can shape our opinions and ideas.

Journalists and media actors

The decision to implement AI systems in the newsroom is a strategic choice with important consequences for internal processes and workflows, as well as for the diversity, quality and independence of the content produced. It constitutes an editorial decision insofar as it is critical to the realisation of the editorial mission and the professional values of a news organisation. Journalists and media actors should use AI systems responsibly and in accordance with professional journalistic ethics. In particular, they should:

a. Identify someone in the organisation who is clearly accountable for the implementation and outcomes of using AI systems.

b. Perform a systematic risk assessment as a necessary pre-condition for the responsible deployment of AI systems for journalistic purposes.

c. When choosing a particular AI technology provider, consider the extent to which the technology provider has made efforts to ensure the responsible use of data. To this aim, data diversity, fairness and quality should be rigorously evaluated, as well as the compliance with privacy and data protection.

d. Work with independent regulatory authorities or self-regulatory bodies to develop guidelines and standards for responsible deployment of AI systems in journalism.