Freedom of the Media and Artificial Intelligence
Julia Haas, Office of the OSCE Representative on Freedom of the Media
FREEDOM OF THE MEDIA AND ARTIFICIAL INTELLIGENCE

Acknowledgement and disclaimer: The views and positions expressed in this report are solely those of the author and do not necessarily reflect the views of the Department of Foreign Affairs, Trade and Development or the Government of Canada. The report is in its original language.

Executive summary

This paper addresses how the use of artificial intelligence (AI) affects freedom of expression and media freedom. While AI can improve communication and information access in numerous ways, including through legacy media, this paper focuses on the main concerns when AI is not deployed in a human rights-friendly manner.

AI can be used as a tool to censor the media and unlawfully surveil citizens and independent journalists. Moreover, in today’s online environment, a few dominant internet intermediaries act as gatekeepers in the curation, distribution and monetization of information, including news content. These intermediaries increasingly deploy AI to govern private speech and public discourse.

AI tools, which underpin much of today’s content dissemination, are often embedded in the business model of targeted advertising. The use of AI to distribute content based on the predicted preferences of individuals is based on extensive data-driven profiling. To maximize revenue, intermediaries may prioritize content that increases user engagement over providing access to diverse information of public interest or to independent quality journalism. This may undermine users’ ability to access pluralistic information and bias their thoughts and beliefs.

To police speech, AI is often applied to identify and remove content considered illegal or undesirable, both by states and intermediaries. The vast amount of available content exceeds the ability for human scrutiny. While AI-based filtering of user-generated content may thus be appealing, AI tools are prone to mistakes. In addition to deploying AI themselves, states mandate private actors to monitor and remove content based on vague definitions within strict timeframes. Such outsourcing of human rights protection to revenue-driven private actors may incentivize over-blocking of legitimate speech and raises additional concerns about the rule of law and discrimination.

AI’s potential to facilitate surveillance and censorship for both economic and political reasons poses a threat to the right to seek and receive information, as well as to media pluralism. The power and influence of a few intermediaries, as well as the fact that most AI tools operate opaquely with little regulation or oversight, exacerbates this threat.
This paper also addresses how biases both in datasets and of human developers may risk perpetuating existing inequality, how AI affects legacy media and how the COVID-19 pandemic aggravates the above-mentioned concerns. Providing policy recommendations, this paper concludes that states and the private sector need to guarantee that the design and deployment of AI are grounded in human rights, with transparency and accountability being ensured at all stages.

Freedom of the media and artificial intelligence

Emerging technologies provide unprecedented opportunities for exercising free speech and media freedom. Artificial intelligence (AI) plays an important role in transforming how people communicate and how they consume and engage with media content. AI offers appealing solutions to filter and rank the seemingly infinite user-generated content and information online. As many technological advancements, AI has the potential for good, but can also pose a genuine threat to human rights—in particular, free speech and media freedom.

Despite no universally agreed definition, “AI” is regularly used as an umbrella term for automated, data-driven processes. AI tools may be simple, human-designed instructions; others are more sophisticated and include machine learning. As AI is based on designs and data provided by humans, its outputs are inevitably shaped by cultural values and subjective experiences and beliefs, including inherent biases.

Some states deploy AI to unlawfully surveil citizens and control public communication in ways inconsistent with international human rights law. Enabling unparalleled possibilities for surveillance, AI can facilitate censorship and means to suppress dissent and independent journalism, both online and offline. Consequently, some states use AI to coerce the press and, ultimately, to tighten digital authoritarianism.

Moreover, private actors, in particular providers of search engines and social media platforms, apply AI to filter content in order to identify and remove or deprioritize “undesired” content, known as content moderation, and to rank and disseminate tailored information, referred to as content curation. Both applications regulate speech with the intention to facilitate online communication, provide user-friendly services, and, crucially, increase commercial profit.

AI-powered filtering and ranking of content is enabled by the surveillance of user behaviour at scale. To evaluate and predict the “relevance” of content, AI requires extensive, fine-grained data. These data also facilitate advertising, which is the basis of many internet intermediaries’ business model. Commodified personal data for targeted advertising—which equals profit—incipitizes extensive data collection and processing, a phenomenon described as “surveillance capitalism.” Offering services “for free,” intermediaries profit from profiling and commercializing the public sphere. Being inherently invasive, this also invites potential abuses of power and pervasive state control. While every form of surveillance has a chilling effect on free speech and the media, AI may impose detrimental constraints on investigative journalists and the protection of sources.

Frequently compared to a “black box,” AI is often opaque and its application invisible. This may lead to the mistaken assumption that its output is neutral and an objective representation of reality. Users may not be aware if AI is utilized, how it obtains a search result or how it promotes or removes content. At the same time, it may not be evident when and how AI is deployed to obstruct the media through surveillance or other forms of interference. Opacity and lack of awareness are major flaws of any AI application.

Opaque AI that governs information dissemination according to business interests may have severe implications for public discourse, particularly in light of the market dominance of very few intermediaries. Oligopolies have become private arbitrators of speech, setting the terms and conditions for global online communication and access to information. Individuals that want to participate in the online sphere are presented with little choice other than to accept the rules and surveillance of dominant intermediaries. Further, such private AI systems and extensive digital footprints may also facilitate state surveillance and political censorship of the press.

The advertising-driven business models at the core of today’s internet structure have profoundly affected the sustainability of legacy media by structurally shifting power, to the detriment of quality journalism. The use of AI technologies further shifts this imbalance—with
a particular impact in countries with low internet penetration or no strong public service media.  

Any intentional use of AI to interfere with independent reporting—be it through targeted censoring, pervasive surveillance of investigative journalists or using AI-driven bots to attack and silence individual journalists—is a serious threat to media freedom.  

Without bad faith, however, the overall use of AI to monitor speech in order to flag, demonetize, deprioritize or remove certain content, or ban specific accounts, is regularly deployed as pre-moderation in the form of upload filters and to analyze content once it is online or after users have reported it. AI will then either take independent action or final assessments remain subject to human review.  

AI is still limited in its capability to analyze content. Speech evaluation is highly context-dependent, requiring an understanding of cultural, linguistic and political nuances. Consequently, AI is frequently inaccurate. False positives lead to unjustified limitations on speech, and false negatives may cause a chilling effect, leading to self-censorship and silencing marginalized voices.  

Intermediaries’ use of AI to proactively moderate content is a form of self-governance, with AI-driven decisions typically based on terms of services or community guidelines. States increasingly request intermediaries to take down specific posts and mandate them to remove certain categories of content, often based on vague definitions, which may lead to blocking news content of public interest. This outsourcing of law enforcement and judicial responsibilities pressures private actors to deploy AI, especially when strict time limits are instructed. While this raises significant concerns regarding the rule of law and due process, it also results in dependence on a few already powerful companies. Altogether, AI seems to accelerate the trend toward general monitoring of communication, which profoundly affects media freedom.  

During the COVID-19 pandemic, human moderators under lockdown and an increasing demand to tackle disinformation led to states and intermediaries expanding their reliance on AI. The pandemic illustrated the importance of reliable, pluralistic information, and—as errors increased and remedy responses were delayed—highlighted the need to address AI’s own side-effects.  

Content curation  

With an abundance of online content, user attention is becoming increasingly scarce. Internet intermediaries apply AI to disseminate information that is based on the predicted preferences of individual users. These predictions, however, are driven by intermediaries’ intent to monetize data for targeted advertising. Therefore, the AI-fueled curation of newsfeeds and search results seeks to entice users to increase their engagement and time spent on the respective service. Controversial and sensational content can attract more attention, just as misogyny, racism and content instilling fear or hatred can. Hence, AI-powered ranking systems that prioritize “click worthy” rather than newsworthy content may lead to polarization, radicalization and the spread of deceptive or hateful content.  

Moreover, increasingly depending on having their content accessed and shared online, legacy media too have to compete for users’ attention and may thus be compelled to increasingly focus on “infotainment” rather than prioritizing public interest, which additionally pressures quality journalism. The AI-fueled personalization of content, including news, may strengthen users’ pre-existing views, creating “echo chambers” and “filter bubbles” and decreasing the likelihood of individuals’ exposure to diverse media content. Distorting the perception of reality, this may reinforce power imbalances and amplify “otherness,” while seriously threatening media pluralism.
Today’s internet structure provides little economic incentive for intermediaries to offer diversity or, indeed, facts. AI that is designed to serve commercial or political interests will unavoidably be biased toward certain types of content in order to nudge and reorient behaviour to optimize profit or persuade, or to intentionally suppress independent journalism. Authoritarians and others can misuse intermediaries’ AI systems for nefarious purposes, for example through bots to propagate specific messages or drown out the visibility of journalistic content. AI tools can be used to attack journalists with the aim of silencing them, for example by orchestrating harassment campaigns that simulate a grassroots movement. This method is particularly prevalent in targeting women journalists—and AI-driven distribution systems may even reward such attacks with virality.

While the magnitude of AI’s impact on public discourse is still unclear, it is undisputed that AI is regularly deployed to influence people’s perceptions and attitudes. Internet intermediaries have become information gatekeepers that use AI to manage media content and information flows, which inevitably shapes users’ opinions and behaviour. AI structures can be used to censor the press, by enabling both negative control over information in the form of censorship as well as positive control in the form of propaganda or attacks. Without democratic safeguards, AI-powered monitoring of speech and content dissemination jeopardizes media freedom, access to information and free speech, while at the same time raising concerns about rule of law and systemic discrimination.

Recommendations

People have repeatedly turned to technology to resolve societal challenges. Yet, matters that have long been controversial cannot be resolved solely by outsourcing decision-making processes to AI. Beyond that, technologies can serve as tools for tracking, censorship and repression of the media at an unprecedented scale. While many of the above-mentioned concerns are not unique to AI, its use exacerbates existing threats to free speech and media freedom. To address them effectively, it is crucial to consider the sociotechnical context in which AI is deployed, by whom it is used, and for which purposes. While there can be no one-size-fits-all solution, AI’s impact cannot be assessed or addressed in any meaningful way without transparency and accountability.

Having a positive obligation to protect freedom of expression and media freedom, states should promote an environment enabling pluralism. When public authorities deploy AI themselves, they must abide by international human rights standards, ensuring that any restriction of speech or the media is necessary and proportionate. Excessively collecting or merging data in public-private partnerships does not fulfil these criteria. Instead, it often facilitates digital authoritarianism to employ mass surveillance and targeting of individuals and journalists as well as unparalleled censorship. States should not exploit AI to manipulate public opinion, harass journalists or for other repressive ends, but rather to determine acceptable limits on the use of these technologies.

Regulatory measures and AI-related policies should be evidence-based and must not have an adverse impact on media freedom. States should refrain from indiscriminately delegating human rights protection to AI. Furthermore, all endeavours need to be integrated in strong data-protection rules. Consenting to intrusive surveillance practices should not be a precondition to participation in online public discourse.

Companies, too, have a responsibility to respect human rights. They should thwart the misuse of their AI systems to suppress dissidents and the press. While many companies commit themselves to “ethics,” these are not necessarily in line with human rights. Nevertheless, private initiatives on AI ethics are important, and codes of ethics play a crucial role in corporate social responsibility. Yet, such codes and principles typically lack democratically legitimated safeguards as well as enforcement regimes, and thus they alone cannot provide effective protection.

Transparency is a basic requirement for any public scrutiny. Individuals should know how decisions that affect their lives were produced and which data were processed for what purpose. Regulators and the broader society should know about AI’s effects on the media and public discourse. Due to the profound information asymmetry, however, the field remains grossly understudied. Independent research on AI’s societal implications should thus be encouraged. To enable scrutiny, AI
needs to be explainable and interpretable.\textsuperscript{60} Hence, states should consider making the disclosure of the utilization of AI and its underlying functions mandatory, while being transparent about their own AI deployment. Such requirements could be tiered depending on the specific purpose, actor’s role and phase of AI development or application, as well as its risk of violating human rights.\textsuperscript{61} Further, clear rules should ensure that information on AI deployment is comparable\textsuperscript{62} and that privacy is protected at all phases.\textsuperscript{63}

Transparency should go hand-in-hand with increased user agency. Users should have a choice and control over collection, monitoring and analysis of their data for customized content and over intermediaries’ interface design.\textsuperscript{64} In order to empower users and strengthen citizen’s resilience, increased digital literacy is needed.\textsuperscript{65}

Transparency is required to know which AI tools are deployed and how automated decisions are made. It is also needed to challenge problematic processes. Those benefiting from AI should be responsible for any adverse consequence of its use. To achieve accountability, strict standards on governance are crucial. Rules should ensure that corporate accountability is tied to companies’ profits and that decision makers can be held responsible.\textsuperscript{66} States should consider establishing a tiered AI oversight structure,\textsuperscript{67} and explore self- and co-regulation models, along with dispute resolution mechanisms, social media councils or e-courts to rapidly determine violations.\textsuperscript{68}

To ensure independent scrutiny, national human rights institutions should be empowered to also supervise AI. An important tool is robust human rights impact assessments, which should be conducted periodically throughout the entire AI life cycle and provide publicly available analyses.\textsuperscript{69} Moreover, AI tools should be audited regularly and independently,\textsuperscript{70} and include careful analysis of whether AI is misused to interfere with the press.

Access to remedies and redress need to be ensured, both for journalists and individual users whose content was restricted by AI, for those who report content, as well as for those harmed by AI-driven interface designs.\textsuperscript{71} Remedies must be dealt with in a timely manner and built on sufficient resources.\textsuperscript{72} For some automated decision-making processes, human involvement, review and reversibility must be ensured.\textsuperscript{73}

Good practices from other fields, including legacy media, can provide lessons to address transparency and accountability.\textsuperscript{74}

In addition, the persistent threat of discrimination in both the design and deployment of AI needs to be addressed.\textsuperscript{75} Effective responses require a holistic and interdisciplinary approach. Discussions should involve all stakeholders and layers of society, including affected end users,\textsuperscript{76} civil society, academia and the media.\textsuperscript{77}

Most of the challenges are closely interrelated with the fact that a few dominant companies have significant power and control over the online information ecosystem. A concentration of power, be it by a state or corporation, always entails a risk of far-reaching restrictions of freedoms.\textsuperscript{78} States should ensure a competitive AI market to create incentives for alternative business models for intermediary services.\textsuperscript{79} Supporting the development of AI tools that are not built on a system of data exploitation and targeted advertising could nurture market pluralism, democratize AI and foster public value-oriented online spaces.\textsuperscript{80}

Finally, given the intertwined and transnational nature of AI challenges, it is crucial to join efforts and aim for global solutions. There are various important initiatives, such as those by the Organization for Security and Co-operation in Europe, UNESCO, the Council of Europe or the European Union.\textsuperscript{81}

AI is neither a magical bullet for society’s challenges nor should it take the blame for all challenges to free speech or media freedom. AI ought not to facilitate digital authoritarianism or high-tech repression of the media. If AI is to enable, rather than undermine, freedom of expression, access to pluralistic information and media freedom, it is imperative for all stakeholders to ensure a human rights-based framework for transparent and accountable AI. As AI increasingly affects every aspect of our communication and media consumption, it is long overdue to embed safeguards in its development and application so that media freedom can thrive.
The right to "... freedom to seek, receive and impart information and ideas ... through any media" is enshrined in Article 19 of the International Covenant on Civil and Political Rights (ICCPR) as well as in the UN Universal Declaration on Human Rights and various regional human rights frameworks. The UN Human Rights Council (UNHRC) affirmed that the same rights that people have offline must also be protected online; see resolution on "the promotion protection and enjoyment of human rights on the Internet", A/HRC/RES/20/8, 2012, https://undocs.org/A/HRC/RES/20/8.


19 See, e.g., OSCE project on the safety of female journalists online, https://www.osce.org/fom/sofo; Report of the UN Special Rapporteur on the promotion and protection on right to freedom of opinion and expression (UN Special Rapporteur), A/HRC/38/35, 2018, https://undocs.org/A/HRC/38/35; J. W. Penney, "Internet Surveillance, Regulation, and Chilling Effects Online: A


25 These are not necessarily in line with human rights standards and can be changed in an opaque way. See, e.g., UN Special Rapporteur, A/HRC/38/35, 2018

26 Various legislation introduce very short time periods; for example the EU Directive on copyright and related rights in the Digital Single Market, (EU) 2019/790, or the German Network Enforcement Act, 2017.


65 See, e.g., Council of Europe, Recommendation CM/Rec(2019)10 of the Committee of Ministers to member States on developing and promoting digital citizenship education, 2019, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=090000168098de08.


69 The UN Guiding Principles on Business and Human Rights require human rights due diligence, and that actual and potential human rights impacts are identified and addressed. See also D. Reisman et al., “Algorithmic impact assessments: a practical framework for public agency


72 For remedies and redress to be effective, resources and leadership commitment is crucial; see, e.g., C. Silverman et al., “‘I Have Blood on My Hands’: A Whistleblower Says Facebook Ignored Global Political Manipulation,” BuzzFeed News, 2020, https://www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo.


76 A recent study by BEUC, the European Consumer Organisation, found that users do not trust AI applications with regard to their privacy and are concerned AI is used to manipulate their decisions. Only 20% found existing legislation adequate. See BEUC, “Artificial Intelligence: what consumers say,” 2020, https://www.beuc.eu/publications/beuc-x-2020-078_artificial_intelligence_what_consumers_say_report.pdf.


78 Regarding legacy media, see, e.g., UNHRC, General comment No. 34 on preventing “undue media dominance or concentration of privately controlled media groups in monopolistic situations that may be harmful to a diversity of sources and views,” UN, OSCE, OAS, ACHPR, “20th Anniversary Joint Declaration: Challenges to Freedom of Expression in the Next Decade,” 2019; Z. Teachout, Break ‘Em Up: Recovering Our Freedom from Big Ag, Big Tech, and Big Money, New York: St. Martin’s Press, 2020.


the ethics of artificial intelligence,"
https://en.unesco.org/artificial-intelligence/ethics, Council of
Europe, "Council of Europe and Artificial Intelligence,"
https://www.coe.int/en/web/artificial-intelligence; European
Union, "Artificial Intelligence," https://ec.europa.eu/digital-
single-market/en/artificial-intelligence; ITU, "Artificial
intelligence," https://www.itu.int/en/ITU-T/TAIPages/default.aspx; and OECD, "OECD Principles on