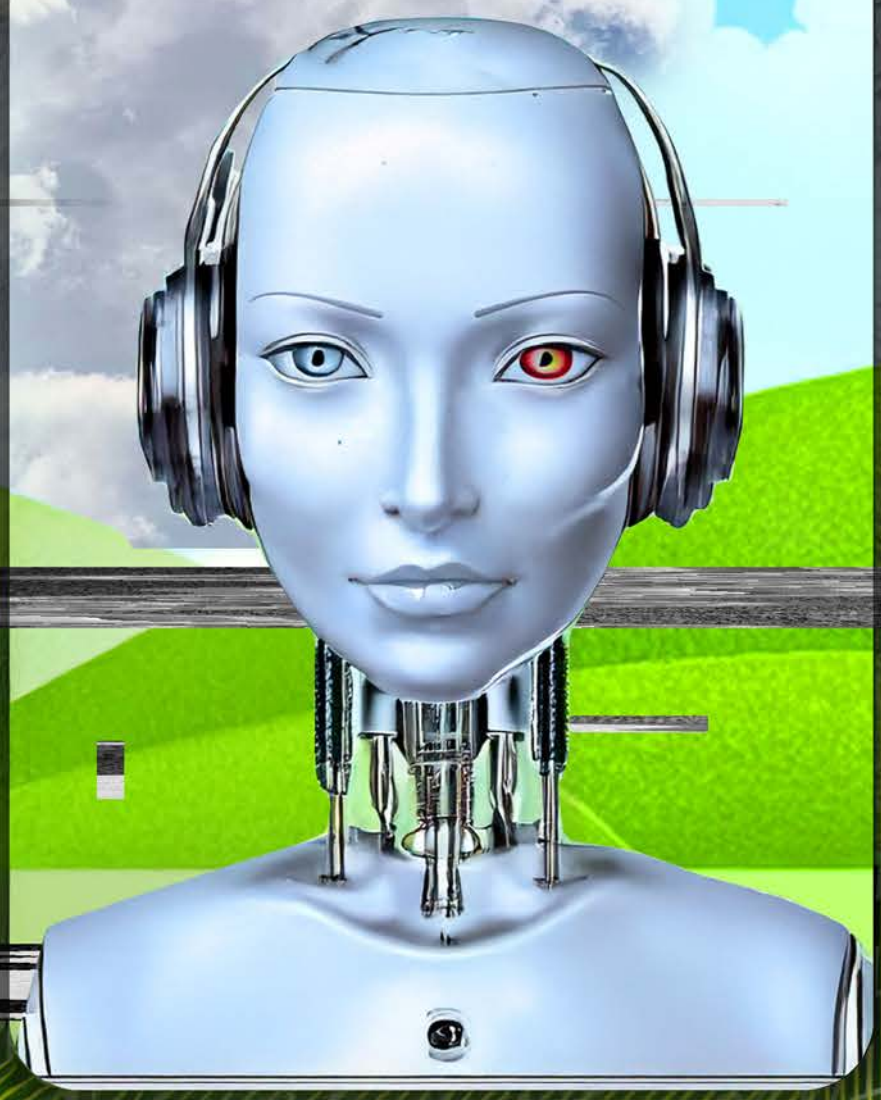
WACC

Media Development

3/2025

Artificial Intelligence: Friend or Foe?



Join WACC

WACC is an international non-governmental organization that promotes communication as a basic human right, essential to people's dignity and community. WACC works with all those denied the right to communicate because of status, identity, or gender. It advocates full access to information and communication, and promotes open and diverse media. WACC strengthens networks of communicators to advance peace, understanding and justice.

MEMBERSHIP OPPORTUNITIES

Membership of WACC provides opportunities to network with people of similar interests and values, to learn about and support WACC's work, and to exchange information about global and local questions of communication rights and the democratization of the media.

WACC Members are linked to a Regional Association for the geographic area in which they are based. They receive regular publications, an annual report, and other materials. Regional Associations also produce newsletters. In addition, members are invited to participate in regional and global activities such as seminars, workshops, and webinars.

Full details can be found on WACC's web site:
www.waccglobal.org

CURRENT ANNUAL MEMBERSHIP RATES

Individual	35 USD
Institutional	120 USD
Student Rate	20 USD

Media Development **is published quarterly by WACC**
80 Hayden Street
Toronto, Ontario M4Y 3G2
Canada.

First floor, 10 Queen Street Place
London EC4R 1BE
United Kingdom.

Editor: Philip Lee
Assistant Editor: Lorenzo Vargas

Subscriptions to ***Media Development***
Individuals worldwide US$40.
Libraries, universities and other institutions (access may be shared with students, staff and users) US$75

Published in Canada
ISSN 0143-5558

Media Development Vol. LXXI 3/2025

In the next issue

The 4/2025 issue of *Media Development* will explore the outcomes and implications of the World Summit on the Information Society (WSIS+20), which took place 7-11 July 2025.

EDITORIAL

The term Artificial Intelligence (AI) encompasses three variations. There is Artificial Narrow Intelligence (ANI) with limited capabilities, e.g. Google translate and Siri. There is Artificial General Intelligence (AGI), which attempts to replicate human capabilities, e.g. chatbots. And then there is Artificial Superintelligence (ASI): machines that are more capable than humans, of benefit to healthcare, scientific research, and the military. Such machines are either the solution or the problem – depending on your point of view.

On the positive side, AI can assist with data analysis, brainstorming, drafting and proofreading of texts. It can help generate social media posts, structure workshops, turn complex descriptions into readable web texts and data into graphics. It can translate and transcribe voice recordings into multiple languages, and offer automated sign language.

On the negative side, AI might sidestep human oversight and become self-aware, leading to unforeseen consequences and even existential risks. The superior cognitive abilities of Artificial Superintelligence could allow it to manipulate systems or even gain control of advanced weapons. Military usages include autonomous warfare systems, strategic decision-making, target recognition, and threat monitoring. Human interventions would be subordinate to "machine thinking".

Consequently, the most important questions surrounding systems based on AI applications are ethical – in terms of their development, application, and impact. In the words of Gabriela Ramos, UNESCO's Assistant Director-General for Social and Human Sciences:

> "In no other field is the ethical compass more relevant than in artificial intelligence. These general-purpose technologies are re-shaping the way we work, interact, and live... AI technology brings major benefits in many areas, but without the ethical guardrails, it risks reproducing real world biases and discrimination, fuelling divisions and threatening fundamental human rights and freedoms."

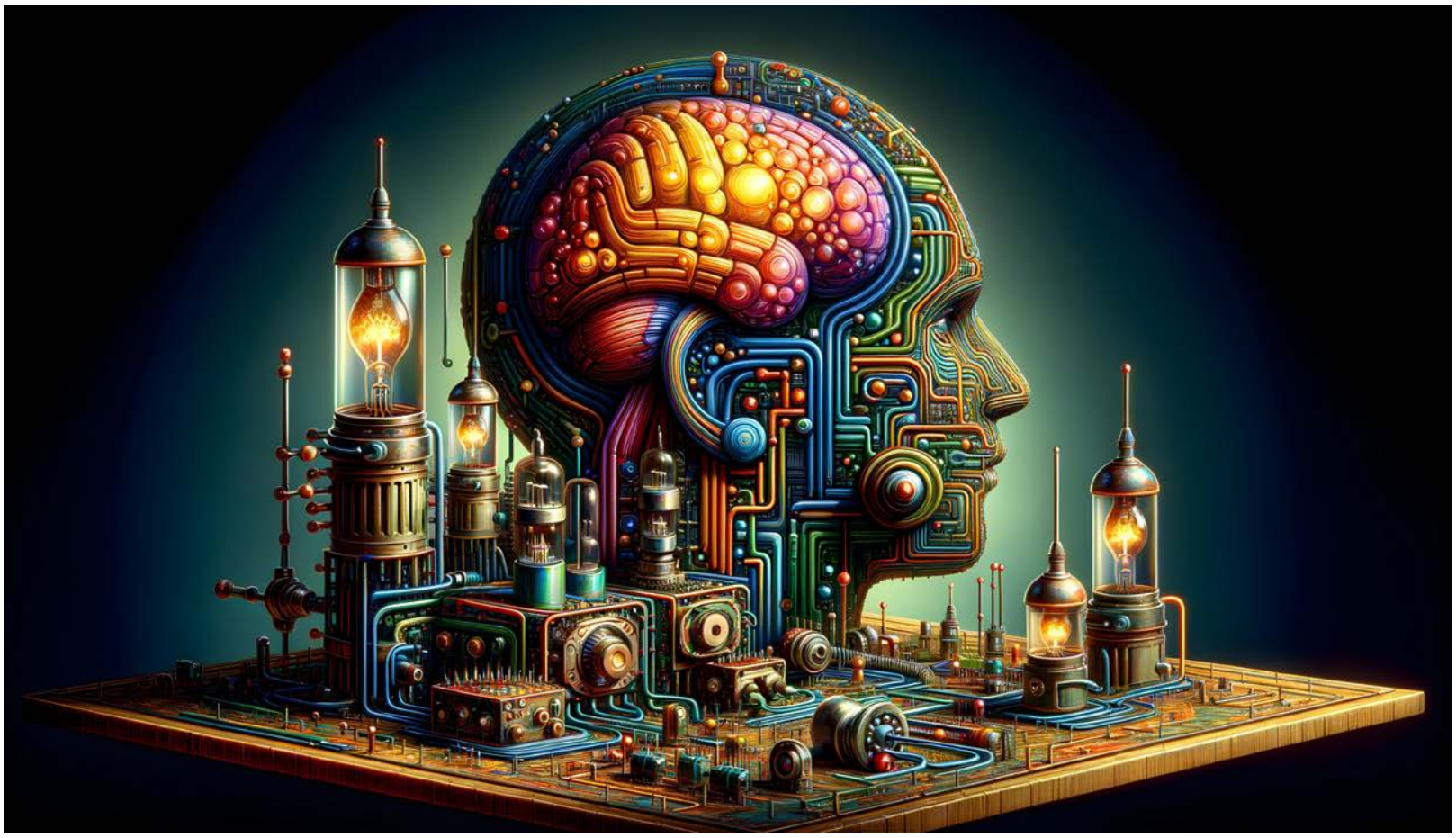
AI is impacting freedom of expression, freedom of information, and public interest journalism in terms of accuracy, authenticity, and trust. At a time when misinformation, disinformation, and fake news bedevil journalism and social media platforms, AI has been seen as a means of dispelling confusion and restoring trust. However, as Julius Endert, senior consultant to the Deutsche Welle (DW) Akademie, points out, if AI is to be used in professional journalism, we need concrete business and editorial decisions that:

> "[R]esult in structures and processes grounded in organizational values, ethical guidelines, and policies. This work must be tailored to the size and scope of each organization and developed incrementally. Crucially, the perspectives of all stakeholders – especially regarding data governance, privacy, and transparency – must be included."

The DW Akademie, which focuses on international media development, journalism training and knowledge transfer, proposes a three-tiered approach to AI governance:

* Ethical foundations: Define ethical reference points and principles as the foundation of an overarching AI strategy. Develop your strategy and guidelines.
* Compliance systems: Establish systems to ensure adherence to legal and other relevant norms.
* Operational implementation: Create and implement responsibilities, processes, and structures according to your AI strategy.

Aside from or in addition to these considerations, there is the issue of public safety. Will these extraordinarily powerful AI systems be subject to oversights that prevent them from behaving in unexpected and potentially catastrophic ways? For decades this has been



seen as a military question: who or what decides to carry out nuclear war? But, today, what if machines get to decide who is a refugee? Or deserving of a heart transplant? Or eligible for schooling?

Beyond such immediate concerns, many people are also worried about the long-term impact of AI on social and cultural identity: the way people understand themselves and others, their sociocultural environments, and the way technologies shape and alter human behaviour.

> "For as the pace of change increases, not just the economy but the very meaning of 'being human' is likely to mutate… Such profound change may well transform the basic structure of life, making discontinuity its most salient feature."[1]

Continuity has always been a measure of stability. For good or ill, it has enabled political, cultural and social identities. Continuity itself relies on a certain tension between the public and the private, in what ways information and knowledge are shared or commoditised or even weaponised. Lemi Baruh and Mihaela Popescu's article in this issue of *Media Development* underlines the dilemma:

> "Privacy as a dynamic process of 'becoming' – essential for shaping our identities and fostering autonomy through an ongoing dialogue with our past, present, and future – faces profound challenges in the age of pervasive artificial intelligence… It's not just about data points being collected; it's about how AI and algorithms actively intervene in our temporal experience, potentially derailing our capacity to make meaningful choices (and learn how to make choices) that we can claim as our own, thereby threatening our journey of becoming who we aspire to be."

Artificial Intelligence is here to stay. Technological development never goes backwards, and its impact is always far reaching and unpredictable. What we must do – and urgently – is to think ethically, to act transparently, and to communicate the implications of AI development widely and intelligibly. Only then can AI serve humanity responsibly. ■

Note

1. Yuval Noah Harari (2018). *21 Lessons for the 21st Century*. Signal/Penguin Random House Canada, pp. 269-270.

Image source Deviant Art.

Temporal selves under siege: Artificial Intelligence and the need for privacy as a right to becoming

Lemi Baruh and Mihaela Popescu

When we talk about online privacy, what comes to mind? For decades, the dominant answer in law and everyday thinking has centred on one idea: control over personal information. This approach, called informational privacy, treats our personal data – our clicks, likes, searches, and purchases – as an extension of us. The core belief is that we, as individuals, should have the authority to determine when, how, and to what extent information about us is shared with others.[1]

This thinking isn't new. Its roots trace to the 1970s with the Fair Information Practice Principles (FIPPs). These guidelines recommended basic rights like knowing what data is collected, seeing and correcting our records, and limiting how companies use information beyond its original collection purpose. These principles spread globally, shaping standards from the OECD to Council of Europe from the 1980s.[2]

Today, this "control over data" thinking is the bedrock of major regulations like Europe's General Data Protection Regulation (GDPR)[3] and state-level laws in the U.S. like the California Consumer Privacy Act (CCPA).[4] These regulations empower individuals with specific rights: the right to access the data held about them, the right to correct inaccuracies, the right to request deletion, and the right to object to or opt-out of certain processing or the sale of our data. The main mechanism? Usually, it's "notice and consent" – companies provide a privacy policy (the notice), and we click "agree" (the consent), theoretically putting us in charge of the data flow.

While personal control over information sounds appealing, this approach faces challenges in reality. The vast amount of data collection makes meaningful consent impossible. We're asked to agree to lengthy privacy policies, but understanding how our data might be used downstream – combined with other datasets or fed into algorithms – is beyond most people's capability. Often, the choice is simply a "take-it-or-leave-it" clickwrap agreement; refuse, and we lose access to the service entirely.[5] This lack of real choice often leads to resignation rather than genuine consent.[6]

Furthermore, this focus on notice and consent doesn't fundamentally challenge the business model – what Shoshana Zuboff famously termed "surveillance capitalism."[7] Instead of limiting data harvesting, it often legitimizes it. By getting us to "agree," the system shifts responsibility onto us, the individuals, while allowing the large-scale collection and monetization of personal data to continue largely unchecked. Regulations like GDPR, while aiming for protection, may even inadvertently strengthen the dominance of large platforms better equipped to handle compliance costs.[8]

The increased ubiquity of machine learning algorithms and artificial intelligence throws another wrench into the works.[9] These technologies operate on a scale and complexity far beyond simple data sharing. Algorithms don't just store the data we provide; they analyse it to make inferences and predictions about us – our personalities, preferences, vulnerabilities, and future behaviour.[10] These algorithmically generated insights, the "outputs," often exceed the "inputs" we initially consented to share. The complex, pro-

prietary nature of these algorithms – the "black boxes" of the digital age – makes it difficult to understand how these powerful inferences shape our opportunities and experiences.[11] Merely having control over our raw data provides little protection against the privacy harms stemming from how that data is interpreted and used by these increasingly powerful algorithmic systems.

Privacy as a Right to Becoming: Protecting the Development of the Autonomous Self

Given the shortcomings of seeing privacy purely as data control, especially in our algorithmic age, we need a different approach. Instead of focusing narrowly on managing information flows, we propose reframing privacy as essential for protecting and nurturing our capacity for self-formation and autonomous action – a concept we call *privacy as a right to becoming*. This perspective shifts the focus from data points to the person, seeing individuals not just as data subjects, but as socially embedded, temporal beings actively engaged in shaping their own lives.

Central to this idea is autonomy. Not the isolated, purely rational self often imagined in liberal theory, but a *relational* autonomy that recognizes how we develop our sense of self and our ability to make meaningful choices through our connections with others and within specific social and cultural contexts. True autonomy – the freedom to ask what kind of a person one wants to be and what kind of a life one wants to lead – isn't just about being free from external obstacles.[12] It requires conditions that allow us to authentically identify with our own desires, reflect on them without undue manipulation, and form and commit to life goals and projects.[13]

Privacy is crucial for creating these conditions. It provides the necessary space – both literally and metaphorically – for the self-reflection and self-discovery vital for autonomous living. This includes controlling access not just to our data (informational privacy), but also to our physical spaces and our decision-making processes.[14] These dimensions of privacy help us manage our relationships with ourselves and others, allowing us to develop and exercise the competencies needed to guide our lives according to our own values.

We argue that this process of becoming autonomous unfolds across time. Think about how we understand ourselves – as part of a dynamic narrative. Our identity is shaped through internal dialogue spanning past, present, and future: we reflect on memories and experiences, engage with our current context, and project aspirations for who we want to be. This process helps us make sense of where we've come from, where we stand now, and where we're headed. It enables us to assemble a coherent life story, take ownership of our past actions, and act intentionally toward future goals. In doing so, we become people who can make authentic choices and take responsibility for them. This capacity allows us to experience our lives as a meaningful whole, connecting who we were, who we are, and who we hope to be.[15]

Sociologists Mustafa Emirbayer and Ann Mische provide a useful framework for understanding this dynamic engagement with time through their concept of the "agentic triad."[16] They argue agency emerges from the interplay of three temporal dimensions:

* **Practical-evaluative dimension (the present):** Our capacity to make practical and normative judgments among alternative possible trajectories of action at present.
* **Projective dimension (the future):** Our imaginative generation of possible future trajectories and outcomes of action
* **Iterational dimension (the past):** Our selective reactivation of past patterns of thought and action to classify past actions in terms of their similarity to a current situation.

We are constantly engaged in this internal temporal conversation, balancing habits from the past with present realities and future hopes. Learning to manage this internal dialogue effectively is how we learn to be autonomous agents. It's a skill developed through practice – through

reflection, making choices, and even making mistakes within a space protected enough to allow for genuine self-exploration. This isn't about achieving a final, static state of autonomy, but about engaging in the ongoing *process* of becoming.

Privacy as a right to becoming defends our capacity for vital temporal work. Privacy functions as a necessary condition for navigating the agentic triad. It provides quiet space for iteration, allowing us to reflect on past actions and habits without constant external judgment or pressure of immediate reaction. It safeguards our practical evaluation of the present by allowing moments of focused attention, free from manufactured distractions designed to capture immediate impulses rather than long-term goals.

Artificial Intelligence and the Case for Privacy as a Right to Becoming

Privacy as a dynamic process of "becoming" – essential for shaping our identities and fostering autonomy through an ongoing dialogue with our past, present, and future – faces profound challenges in the age of pervasive artificial intelligence. While the traditional "privacy as control" model already struggles with the sheer scale and complexity of modern data practices, the lens of "privacy as a right to becoming" that we propose helps see how these technologies can more deeply undermine the very foundations of self-development. It's not just about data points being collected; it's about how AI and algorithms actively intervene in our temporal experience, potentially derailing our capacity to make meaningful choices (and learn how to make choices) that we can claim as our own, thereby threatening our journey of becoming who we aspire to be.

This framework, which emphasizes protecting our journey towards autonomous selfhood through a rich internal dialogue with our past, present, and future, reveals significant shortcomings in simply viewing privacy as control over data, especially when we consider the impact of artificial intelligence and algorithms. These systems don't just manage information; they actively shape our experiences and, in doing so, can profoundly disrupt the very processes necessary for us to develop and maintain a coherent sense of self.

Consider how algorithms like recommendation systems curate and re-present our histories. They selectively highlight certain memories and behaviours, often without our awareness of their choices. From the perspective of privacy as a right to becoming this isn't just about data accuracy or inference veracity; it's about these algorithmic narratives potentially altering our personal stories and how we view our past and understand ourselves. Our ability to own our narrative, to draw lessons from our past and integrate them into who we are becoming, is crucial for developing autonomy.

Similarly, algorithms shaping our experiences raise concerns beyond loss of information control. Recommendation engines and personalized feeds can create "filter bubbles" that narrow perspectives and limit exposure to diverse viewpoints, which are vital for critical thinking and practical evaluation of agency.[17] This invisible algorithmic steering can undermine our ability to make choices reflecting our values and long-term goals. The privacy that traditionally affords breathing room for independent thought is eroded when our present moments are mediated and manipulated by systems designed to steer attention and influence actions.

Looking to the future, the predictive power of algorithms and the "data-driven personas" they construct can significantly impact our projective capacities – our ability to imagine and strive for different futures. When algorithms engage in anticipatory and pre-emptive governance,[18] using past data to predict and shape future pathways or target us during vulnerable transitional moments. This isn't just about data being used; it's about our potential life trajectories being directed and, perhaps, limited by predictive models. Privacy as a right to becoming is about our ability to defend ourselves and engage authentically and autonomously with our past, present, and future, shielding the core processes of self-development from

these profound algorithmic interventions.

The European Union's AI Act and emerging "neurorights" field indicate growing recognition that privacy challenges brought by AI extend beyond data control issues. These initiatives address privacy as a right to becoming – protecting our autonomous self-development in an algorithmic world. They acknowledge that technological interventions, particularly AI and neurological interfaces, can impact our freedom of thought and ability to make choices free from manipulation, which are vital for our process of becoming.

The EU AI Act aims to address such deeper concerns by, for example, banning AI systems that use "subliminal techniques beyond a person's consciousness to materially distort behaviour" or exploit vulnerabilities related to age or disability.[19] This directly resonates with privacy as a right to becoming by recognizing that certain algorithmic influences threaten our core ability to self-determine. However, the AI Act still struggles to define "manipulation" or impairment of "informed decision making" in complex AI interactions and what makes AI based manipulation different from other manipulative practices.

Here, privacy as a right to becoming offers richer vocabulary. It articulates why AI manipulations are damaging, not just from divergent interests between organizations and data subjects, but because they interfere with our internal temporal dialogue – our ability to reflect on past, evaluate present, and project future authentically. It explains harm through undermining competencies for skilful choice, narrative control, and self-justification. This allows privacy as a right to becoming to be more flexible in adjusting to shifts in the operational logic of future technologies.

Similarly, discussions around "neurorights" arise from the development of neurotechnologies like brain-computer interfaces (BCIs), which promise direct pathways to "reading" and even influencing brain activity. The concern here is profound, touching upon mental privacy, cognitive liberty, and the very integrity of our thoughts and mental states.[20] The privacy as a right to becoming framework emphasizes that our minds are not just static entities needing protection, but are the dynamic seat of self-formation. Interference with our cognitive processes, memories, or future-oriented thought directly impacts our ability to construct and live our life stories.

Privacy as a right to becoming extends the neurorights conversation by highlighting that the harm isn't just about neurological or other biometrical data extraction, but about disrupting the ongoing, temporally-grounded process of self-development that defines us as autonomous individuals. It helps us understand that protecting the "mind" is also about protecting the unfolding narrative of a life. Ultimately, privacy as a right to becoming provides a unifying lens that focuses on the individual's lifelong journey of self-creation, emphasizing the need to safeguard the temporal dimensions of autonomy. ■

Notes

1. Woodrow Hartzog, "The Inadequate, Invaluable Fair Information Practices," *Maryland Law Review* 76 (2017): 952–77.\\uc0\\u8221{} {\\i{}Maryland Law Review} 76 (2017
2. Fred H. Cate, "The Failure of Fair Information Practice Principles," in *Consumer Protection in the Age of the "Information Economy,"* ed. Jane K. Winn (Abingdon, Oxon: Routledge, 2006), 343–79.
3. European Parliament and Council of the European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)," L 119 Official Journal of the European Union § (2016).
4. "California Consumer Privacy Act of 2018," California Civil Code § 1798.100 et seq. (2018).\\uc0\\u8221{} California Civil Code \\uc0\\u167{} 1798.100 et seq. (2018
5. Lemi Baruh and Mihaela Popescu, "Big Data Analytics and the Limits of Privacy Self-Management," *New Media and Society* 19, no. 4 (2017): 579–96, https://doi.org/10.1177/1461444815614001.
6. Nora A Draper and Joseph Turow, "The Corporate Cultivation of Digital Resignation," *New Media & Society* 21, no. 8 (August 2019): 1824–39.
7. Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: Public Affairs, 2019).
8. Damien Geradin, Theano Karanikioti, and Dimitrios Katsifis, "GDPR Myopia: How a Well-Intended Regulation Ended up Favouring Large Online Platforms - the Case of Ad Tech," *European Competition Journal* 17, no. 1 (January 2, 2021): 47–92.
9. Daniel J Solove, "Artificial Intelligence and Privacy," *Florida*

Law Review, 2025.
10. Tal Z Zarsky, "Privacy and Manipulation in the Digital Age," *Theoretical Inquiries in Law* 20, no. 1 (2019): 157–88; Karen Yeung, "'Hypernudge': Big Data as a Mode of Regulation by Design," *Information, Communication & Society* 20, no. 1 (January 2017): 118–36.
11. Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015).
12. Beate Rössler, *The Value of Privacy*, trans. R. D. V. Glasgow (Cambridge, UK: Polity Press, 2005).
13. Rössler.
14. Neil Richards, *Why Privacy Matters* (New York, NY: Oxford University Press, 2022).
15. Marya Schechtman, "The Narrative Self," in *The Oxford Handbook of the Self*, ed. Shaun Gallagher (Oxford, UK: Oxford University Press, 2011), 394–418.
16. Mustafa Emirbayer and Ann Mische, "What Is Agency?," *American Journal of Sociology* 103, no. 4 (1998): 962–1023.
17. Sofia Bonicalzi, Mario De Caro, and Benedetta Giovanola, "Artificial Intelligence and Autonomy: On the Ethical Dimension of Recommender Systems," *Topoi* 42, no. 3 (July 2023): 819–32.
18. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*.
19. "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)," Pub. L. No. 2024, L (2024).
20. Daniel Susser and Laura Y. Cabrera, "Brain Data in Context: Are New Rights the Way to Mental and Brain Privacy?," *AJOB Neuroscience* 15, no. 2 (April 2, 2024): 122–33.\\uc0\\u8221{} {\\i{}AJOB Neuroscience} 15, no. 2 (April 2, 2024): 122-33.

Lemi Baruh (Ph.D., University of Pennsylvania, Annenberg School for Communication, 2007) is a Senior Lecturer at the School of Communication and Arts, The University of Queensland, and the co-director of the Social Interaction and Media Lab at Koç University, Istanbul, Turkey. His research focuses on digital media and communication technologies, exploring issues such as interpersonal relationships, online safety and security, privacy, surveillance, and decision-making.

Mihaela Popescu (Ph.D., University of Pennsylvania, Annenberg School for Communication) is a Professor of Digital Media in the Department of Communication & Media at California State University, San Bernardino, and the Faculty Director of the Extended Reality for Learning Lab (xREAL). Her research and teaching interests include media and communication policies, privacy and surveillance, immersive and algorithmic media, and human-machine communication.

Australasia's AI unveiling: pedagogy, practice and policy

Anne Kruger and Richard Murray

Responses from policymakers, professionals and educators to the urgency of AI draw many parallels to earlier phenomena in the growth of online misinformation and disinformation. A key difference however was the sudden global generative AI product releases that forced stakeholders into the frontiers reckoning with AI generated harms and opportunities. Two years on, Australia's response could be described as purposeful in areas of social media safety by design, but piecemeal in terms of digital literacy and media empowerment.

In navigating the frontiers of AI generated harms, we need to also consider the potential chaos that AI creates for everyday citizens' digital literacy levels. As such, the role of journalists has never been more necessary to support their audiences' critical thinking. While safety standards (discussed below) have understandably been an urgent focus of governments for technological and social media companies, others in the media ecosystem have been left – literally and figuratively – to their own devices.

That was the case for Australia's news organisations. On the one hand, they were faced with how to serve their publics through delivering quality journalism and sift through potential AI powered explosions of mis- and disinformation; while on the other hand, the same organisations

had to consider the opportunities AI may deliver for news production and effective workflows. As this article demonstrates, while news organisations and institutions are still figuring it out via trial and error, there is a clear case for regulations that include AI labelling and mandated media or digital literacy training. These media developments are also areas where journalism and communications focused academics are well placed to provide support.

An ethical template

Advances in digital technology during the mid 2010s led to a rise in Open Source Intelligence (OSINT) techniques available to the mainstream. Proponents encouraged transparency in their reports that showed the investigative methods they used. This was a new hook given journalism had previously shied away from the inclusion of details in reports showing how the so-called sausage was made. Through OSINT, journalists began to show to their audiences how their use of publicly available digital tools could provide powerful evidence in their investigations – tools that ranged from being able to determine the provenance of images to being able to provide critical evidence to hold "governments and other powerful actors to account" in international courts of law.[1]

In a watershed case, open source investigations by the independent collective of researchers known as Bellingcat[2] uncovered crucial evidence into the origin of the Buk missile launcher that downed Malaysia Airlines Flight MH17 on July 17, 2014 in Ukraine. Investigative methods included as part of Bellingcat reports, provided the basis of a transparent and ethical framework where audiences are clearly aware of what is used – and sometimes not used – in the process of information gathering. The at times step-by-step explanation of the tools and resources used in investigations not only built evidence and credibility but also a new approach to transparency.

The OSINT approach is one that can be used to inform media as they experiment with the uptake of AI to improve productivity, workflows or create content. Frameworks for the mainstream media's ethical and transparent use of tools and processes have undergone an AI transformation. But this work is still in its infancy and Australia has experienced very public missteps along the way. For example, in April 2025 unsuspecting audiences found out that "Thy", a popular radio DJ, who had been on air in Australia for six months, was not a real person, but AI generated. This "triggered an industry-wide discussion on diversity, consent, and AI's place in creative audio environments."[3]

The Thy example revealed a big part of the perceived problem is Australia's lack of AI labelling regulations. There are no specific restrictions on the use of AI in broadcast content, and no obligation to disclose its use. This deficit found its way into the political domain with Australia's May 2025 federal election rife with unlabelled AI enhanced political campaigns; and questions were repeated about Australia's lack of truth in political advertising regulation (discussed below).

Surveys in Australia and globally have revealed newsrooms began experimenting cautiously with generative AI[4] while the development of transparency and ethical use frameworks was slow.[5] Australia's public broadcaster and news organisation, the Australian Broadcasting Corporation (ABC) released its AI principles in June 2024, which "reflect its values and editorial standards and will govern the ways in which it will use AI/ML technologies."[6] Arguably, the ABC's in-house experimentation and trials – such as transcripts to enhance podcast accessibility[7] – has reduced risk compared with more public facing trial and error approaches as the Thy example showed.

Globally, the lead has come from larger organisations such as Thomson Reuters, whose CEO Steve Hasker noted they "used artificial intelligence to radically transform the company into a technology business." That included the Reuters news agency – although somewhat ironically Hasker added the news agency side is "… the smallest part of our business … that accounts for a bit less than 10 per cent of our revenue and a bit less than five per cent of our profit." Hasker

conceded, "In many places, it's the only part of the company that anyone recognises."

Having a large organisation behind a news agency provides a huge advantage in terms of resources to build responsible AI systems and processes. The news agency is transparent about its journalists' use of AI[8] and the company as a whole has developed AI principles.[9] Other notable improvements from newsrooms globally come from Agence France Presse's (AFP) Hong Kong-based standards and ethics director, Eric Wishart.[10] Wishart incorporated uptake of AI in a common sense approach that aligns with his years of journalistic practice and news wire expertise.

Regulations, elections, shortcomings

The Australian Government has been an early leader globally as a pioneer in digital policies addressing issues such as cyberbullying. It established the office of the eSafety Commissioner in accordance with the initial Enhancing Online Safety Act 2015[11] (amended in 2017).[12] This laid a foundation for social media regulation standards and protections in the area of online safety and safety by design. However, the establishment of misinformation and disinformation regulation (which has strong intersections with tech developments such as AI) was slower. Mis- and disinformation regulation eventually arose as something of a by-product from the Australian Treasury's Digital Platform Inquiry that began in 2017.

The central focus of the Digital Platform Inquiry was whether and if large digital platforms and social media organisations that operate in Australia should pay local news publishers for news content. The Digital Platforms Inquiry report was handed down in July 2019.[13] Initially policy makers and stakeholders expected a voluntary code to address the market competition between news and platforms; and a mandatory or co-regulatory code to address disinformation. However, what eventuated flipped these expectations was a mandatory News Media Bargaining Code, (NMBC) that came into effect in March 2021.[14]

And while the Digital Platforms Inquiry recommended a co-regulatory response to combat disinformation, the Australian Government at the time instead asked digital platforms to develop a voluntary code of practice to help address disinformation. This was assigned to the Digital Industry Group Inc. (DIGI), a not-for-profit industry association that advocates for the digital industry in Australia. Eventually, a code that covered both disinformation and misinformation was tasked, under strict guidance notes from Australia's media and communications regulator, the Australian Communications and Media Authority (ACMA). The code, that covered both misinformation and disinformation was groundbreaking given the dearth of similar international regulation, and the heavy-handed approaches by near neighbours in Asia to adopt blunt force legislation.[15]

The explosion of generative AI into the wider information ecosystem in 2023 created a new urgency to ensure safety and information integrity. Australia's government was quick to acknowledge generative AI ushered in "new and emerging harms."[16] Not surprisingly, it responded with announcements of a statutory review of the Online Safety Act, and an intention to amend the Basic Online Safety Expectations (Bose) system, both of which apply to tech and social media companies. This resulted in a series of draft new codes[17] that focus on measures to restrict Australian children from accessing adult content online and other harms. In January 2024 the government also released its findings from consultations with stakeholders on potential AI "guardrails"[18] required throughout industry. However, it was Australia's new legislation that passed through Parliament in December 2024 setting a minimum age limit for social media platforms that attracted the spotlight from international headlines.

Similarly, Australia's near neighbour, New Zealand, adopted a framework of online safety for the Aotearoa New Zealand Code of Practice for Online Safety and Harms.[19] This voluntary

code commits signatories to a set of Guiding Principles and Commitments "that acknowledges the need for flexible responses to ever-changing risks from harm."[20] Online discourse had offline, real life implications in New Zealand during the Covid-19 pandemic. This ranged from the Parliamentary Protests[21] co-opted by online harmful campaigns during Covid-19, to the lessons learned in the need to craft culturally appropriate messages to encourage vaccine uptake among diverse communities.[22] This is important given now the capacity for AI-generated content to take advantage of unsuspecting citizens and would exacerbate the susceptibility of audiences to fall for dangerous mis- and disinformation. This highlights the need for digital literacy in broader society. In terms of AI regulations, the New Zealand Government has introduced a non-legally binding best practice principles-based framework to guide the responsible use of Artificial Intelligence technologies across the public sector.[23]

Generative AI played a role in elections around the world in 2024 – from "voice clones" of imprisoned opposition leader Imran Khan in Pakistan,[24] to Britain's Independent Candidate "AI Steve",[25] and the use of AI in India to translate campaign speeches into multiple ethnic languages.[26] Researchers have described the year as "the good, the bad and the in-between."[27] The downright "absurd" can be added to the list. When outspoken like-minded supporters create and repost AI-generated misinformation, this can galvanise their communities' political views, frame the narratives or online discourse, and often break through into the mainstream media. This was the case in the US when President Donald Trump not only picked up on false online slurs[28] about pet eating immigrants, but decided to give this oxygen during a live, internationally televised Presidential debate.[29]

While the AI Apocalypse or Armageddon that was feared for the global elections may not quite have eventuated, researchers from the Alan Turing Institute's Centre for Emerging Technology and Security noted "deceptive AI-generated content still influenced election discourse, amplified harmful narratives and entrenched political polarisation."[30] This supports research by DIGI in Australia that an individual's perception of what constitutes misinformation can be skewed by their political bias.[31]

Journalists and broadcasters noted the influx of satirical AI generated political campaigns on social media during the 2025 Australian Federal Elections. Broadcaster Sofie Formica noted how the rise in "fake political videos" that can "manipulate the message" makes the job of media more challenging[32] in sifting through what is legitimate, and there is too much of a lag in addressing it. Indeed, there is a new urgency to ensure journalists are equipped with the skills to uphold information integrity through the discernment of a range of online content in order to deliver quality news. Equally, there is a need for sustained, government support of mandatory media and digital literacy curricula for ongoing generations, rather than the piecemeal approach of projects being funded for single digit years.

Many of the AI generated political videos in the 2025 campaign were obvious or satire. But this begs the question – what happens as AI becomes harder to detect, or satire is more nuanced? A combination of AI specific standards and change in political advertising laws could go far in mitigating the effects. This includes standards that require labelling of AI generated content; and laws that address Australia's patchwork of accountability mechanisms to prevent lies in political advertising. The lack of truth in political advertising laws has arguably compromised the Australian public's ability to identify mis- and disinformation.[33] It also did little to assist Australians during the landmark 2023 Voice Referendum which was aimed at giving Indigenous Australians a voice to parliament. Further lessons from the Voice also showed how easy it is for misinformation to spread in more closed spaces such as emails, screenshots in chat apps and via newsletters. [34-35]

While the lack of truth in political laws is a matter for Australian legislators and policy-

makers to address, there is more immediate hope on the horizon in terms of AI labelling. Researchers from the "Partnership on AI" have collated years of case studies that showcase the benefits of labelling AI,[36] where one can reflect with "what happens when you don't"? Further to this, the Coalition for Content Provenance and Authenticity (C2PA) and Content Authenticity Initiative (CAI)[37] have developed an open industry standard for content authenticity and provenance. The C2PA has trademarked a "cr" icon that notifies a consistent standard in provenance credentials that travels with content for use by creators, editors, publishers, media platforms, and consumers. Such tools assist OSINT investigations and if put to wider use in Australia, this could support digital literacy programs, giving audiences a visual cue with the "cr" icon. Additionally current regulation such as the ACPDM could push for the adoption of such labelling in Australia.

A community of practice

The University of Queensland's School of Communication and Arts faculty launched its AI and the Next Generation of Journalists Community of Practice on September 6, 2024. The aim is to encourage and facilitate ongoing discussion, skills building and curricular responses for staff and students alongside industry. The launch heard from learning innovation designers; Adobe solutions experts; and the ABC's Product Strategy Manager. This is an opportune time for us as journalism educators with strong industry links, to deliver experiential learning via the establishment of an ongoing industry-academica-student Community of Practice[38] and bring an ethical lens to practical research outputs.

Through experiential learning and research, students will experience work authenticity adapting substantial developments concurrently with industry leaders.[39] We have selected a cross section of media with leaders working and experimenting with generative AI workflows. Our approach with collaborative industry experts in this time of testing and trialling generative AI workflows continuously and iteratively informs future curriculum design in journalism (and communications) for the decade to come.

Training the trainers

One of the greatest challenges for the Community of Practice is that newsrooms are notoriously competitive ecosystems – both internally and externally. How to get a bunch of editors and journalists in a room sharing potential secrets and even proprietary information? One answer is to approach newsrooms individually and bring them together at key points for seminars with glossy outside experts from tech firms as the guest speakers. Another approach focuses on training. And on the trainers. Shaun Davies, formerly a principal product manager in trust and safety at Microsoft, recently developed and helped to lead the Google News Initiative AI Workshops – a 16 week program for small to medium newsrooms across Australia.[40] Davies also consults throughout Asia and spoke from Japan for this article about how he builds maturity in newsroom AI uptake from "ad hoc vibes usage, to a more structured, formally tested and proven use case." Davies's said his aim is to develop robust processes:

> And the way you do that is not by throwing stuff at the wall and going it feels good. You actually have to make structured tests - say you want [AI] to save time. Well, let's benchmark the amount of time it takes to do this [task].[41]

Davies advises newsrooms to first set a bar where output should be no worse than it currently is now, any use of AI should meet the same quality. "And to do that, you need to define what quality means to your organization, and then you need to develop a test data set that demonstrates what that quality is," he said.

From there, newsrooms are developing more robust solutions that even in test phases can compare outputs of the model. For example, if testing for an AI to provide a quick summary at the top of an article, Davies noted:

> Get a number of experts which could be three of your most senior writers to rate the prompts, and blind test against human written content. In that case you've got the rest of the article so you can test for hallucinations.

Davies' training further provides the prompts and checks of inputs to remove hallucinations.

Meanwhile other platform experts have joined University of Queensland Data Journalism classes to introduce AI "productivity tools". Assessment for the course requires students to produce a long form journalistic report with data, visualisations and interviews. Throughout the course a student chased a local municipal council for environmental data. He had a breakthrough in the form of a "data dump" of technically public, but practically impenetrable, links to files from council meetings. The stash came during a two-hour tutorial workshop, during which the student uploaded the files into Notebook LM, and for comparison Pinpoint. While this took some initial conversion time, both allowed the student to search and organise hundreds of documents, potentially saving weeks of work. And while Notebook LM may have the capacity, we didn't encourage the student to turn the council files into a podcast.

Lessons are taken from the experts to our classrooms. Davies' earlier advice for newsrooms, also extends to strategic communications professionals – and thus an even wider range of students. In terms of quick background research for journalists and communications professionals, Davies sees great potential in the deep research models that can concurrently search the internet:

> You can type a research query in, and Gemini will go out and search the web. It will probably search through about 300 websites, and it will compile the information it's found into a really detailed report for you. You can be quite explicit about what you need to know.[42]

In terms of employment risk and opportunity, Davies noted he tells newsroom staff it's not so much as "an AI [will take] their job", but rather, "somebody could use AI more effectively than they could." As tertiary educators, this feeds into the pedagogical aims, particularly in a journalism program where students need to graduate with an edge, ahead of the AI curve, in terms of skills and ethics.

Conclusion

While Australia has made inroads in coming to terms with AI, the current regulatory framework does not sufficiently cover existing risks from AI. Cyberbullying and safety by design has been led by the eSafety Commissioner. The sense of urgency and the resources provided by the government in that area needs to be extended across the digital ecosystem in order for a proper coordination of risk mitigation. Australia's ACP-DM regulatory code is technology neutral, open to further signatories and allows for agility in addressing AI developments. While the code of practice is in place for platforms and social media organisations to address mis- and disinformation which can incorporate technological developments in AI, this does not include the whole information ecosystem. For example, news media and political parties could each produce codes of practice to address mis- and disinformation. Where a suite of legislation is a blunt force reserved for the highest of risks such as safety and national security; codes of practice have the ability to set standards that can adjust with the pace of technological change.

This article recommends the adoption of regulation and codes of practice from a wide range of stakeholders incorporate transparency, truth, and promote evidence-based information in an era of AI. Voluntary codes provide safe testing grounds that mitigate unintended consequences that may stem from legislation as technology develops. Action is provided with real guardrails and allows time for any consideration of future legislative requirements. News media and political parties should develop industry-wide transparency codes in the uptake and use of AI in content and practices, and these should be explicit for digital platforms via the

current ACPDM. Templates exist in the form of OSINT techniques as transparency and software developments in labelling of AI from the CAI show how this can be implemented at scale. Examples from missteps in Australian media show audiences appreciate transparency and ethical applications of AI in media.

In terms of legislation, urgent work is required in addressing truth in political advertising (which may include labelling of AI moving from regulatory codes to legislation) and mandatory digital literacy curricula. Governments must move fast to introduce a suite of intergenerational, culturally relevant long-term funded media literacy initiatives to enhance critical thinking skills throughout society as standardised education.

In summary, while governments in Australia and globally are figuring out the unveiling effects of AI, in the meantime there is a clear case for regulations or legislation that include AI labelling and mandated media or digital literacy training. AI can speed up workflows and so too, spread mis- and disinformation. Journalists therefore need training in both AI production tools and ethical transparency frameworks as well as how to protect themselves and their audiences from AI-enhanced disinformation risks. This enables journalism to support society and our democratic systems. These are media development issues where academia is well suited to support industry and policy, and to provide graduates who are ready to lead. ▪

Notes

1. https://www.icfj.org/news/fundamentals-open-source-intelligence-journalists
2. https://www.bellingcat.com/
3. Lee, N. (2025) ARN's AI voice creator speaks out on 'presenter' Thy as bigger questions remain unanswered. Published April 28, 2025. *Mediaweek*. https://www.mediaweek.com.au/arn-breaks-silence-on-ai-host-thy-but-keeps-silent-on-key-issues/
4. Attard, M., Davis, M., Main, L. (2024) Gen AI and Journalism. Centre for Media Transition. University of Technology Sydney Access via https://www.uts.edu.au/research/centre-media-transition/projects-and-research/gen-ai-and-journalism
5. Henriksson, T. (2023) New survey finds half of newsrooms use Generative AI tools; only 20% have guidelines in place. Survey by World Association of News Publishers. Published May 25, 2023
6. ABC AI Principles Published June 28, 2024. *AB*C. https://www.abc.net.au/about/abc-ai-principles/104036790
7. https://www.abc.net.au/innovation-lab/abc-transcribe/103125708
8. Reuters and AI. Retrieved from https://www.reuters.com/info-pages/reuters-and-ai/
9. Thomas Reuters https://www.thomsonreuters.com/en/artificial-intelligence/ai-principles
10. Wishart, E. (2024) Journalism Ethics: 21 Essentials from Wars to Artificial Intelligence, Hong Kong University Press.
11. Enhancing Online Safety Act 2015 (Cth), https://www.legislation.gov.au/C2015A00024/2017-06-23/text.
12. Enhancing Online Safety for Children Amendment Bill 2017 (Cth), https://parlinfo.aph.gov.au/parlInfo/search/display/display.w3p;query=Id%3A%22legislation%2Fbillhome%2Fr5794%22
13. ACCC, "Digital Platforms Inquiry 2017–19: Preliminary Report," accessed May 15, 2025, https://www.accc.gov.au/by-industry/digital-platforms-and-services/digital-platforms-inquiry-201719/preliminary-report.
14. ACCC, "News Media Bargaining Code," accessed May 15, 2025, https://www.accc.gov.au/by-industry/digitalplatforms-and-services/news-media-bargaining-code/news-media-bargaining-code.
15. https://sites.brown.edu/informationfutures/2022/11/04/information-futures-labs-apac-partner-crosscheck-launches-training-booklet-for-southeast-asia/ Accessed May 22, p 7-8.
16. Butler, J. (2023). Australia to force social media companies to crack down on 'emerging harms' of AI deep fakes and hate speech. Published November 22, 2023, *The Guardian*.
17. https://www.esafety.gov.au/industry/codes/background-to-the-phase-1-standards
18. https://consult.industry.gov.au/supporting-responsible-ai
19. Transparency International New Zealand, "Aotearoa New Zealand Code of Practice for Online Safety and Harms," October 13, 2022, https://www.transparency.org.nz/blog/aotearoa-new-zealand-code-of-practice-foronline-safety-and-harms
20. Kruger, Anne (2024). A review of the landmark Australian Code of Practice On Disinformation and Misinformation (ACPDM). St Lucia, QLD, Australia: The University of Queensland. https://doi.org/10.14264/58981b2 Pages 28-33. P29
21. https://www.theguardian.com/world/video/2022/feb/10/anti-vaccine-protesters-clash-with-police-outside-new-zealand-parliament-video and effects on communities noted here https://www.rnz.co.nz/news/national/483800/impact-of-parliament-protests-still-being-felt-in-the-thorndon-and-pipitea-community
22. Kruger, Anne (2024). A review of the landmark Australian Code of Practice On Disinformation and Misinformation (ACPDM). St Lucia, QLD, Australia: The University of Queensland. https://doi.org/10.14264/58981b2 Pages 28-33.
23. https://www.dlapiper.com/en/insights/publications/2025/02/new-zealands-public-service-ai-framework-guiding-responsible-innovation
24. https://www.theguardian.com/world/2023/dec/18/imran-khan-deploys-ai-clone-to-campaign-from-behind-bars-in-pakistan
25. https://www.ai-steve.co.uk/

26. https://www.techpolicy.press/indias-experiments-with-ai-in-the-2024-elections-the-good-the-bad-the-inbetween/
27. https://www.techpolicy.press/indias-experiments-with-ai-in-the-2024-elections-the-good-the-bad-the-inbetween/
28. https://www.theguardian.com/us-news/article/2024/sep/09/republicans-haitian-migrants-pets-wildlife-ohio
29. https://www.theguardian.com/us-news/article/2024/sep/10/trump-springfield-pets-false-claims
30. https://cetas.turing.ac.uk/publications/ai-enabled-influence-operations-safeguarding-future-elections
31. https://www.uts.edu.au/news/2022/07/australians-and-misinformation
32. https://www.4bc.com.au/podcast/deepfake-surge-sparks-election-fears-most-aussies-cant-spot-ai-generated-content/
33. Anne Kruger and Esther Chan, "Australian Election Misinformation Playbook," First Draft, March 26, 2022, https://firstdraftnews.org/articles/australian-election-misinformation-playbook/. As cited in Kruger, Anne (2024). A review of the landmark Australian Code of Practice On Disinformation and Misinformation (ACPDM). St Lucia, QLD, Australia: The University of Queensland. https://doi.org/10.14264/58981b2
34. https://www.rmit.edu.au/news/crosscheck/voice-misinformation-intervention
35. https://www.rmit.edu.au/news/all-news/2023/jun/croscheck-voice-referendum
36. https://partnershiponai.org/from-deepfakes-to-disclosure-pai-framework-insights-from-three-global-case-studies/
37. https://c2pa.org/
38. Cox, A (2005). What are communities of practice? A comparative review of four seminal works. Journal of Information Science, 31(6), 527–540.
39. Kolb, D. A. 2015. Experiential Learning: Experience as the Source of Learning and Development, (2nd ed.). Upper Saddle River, NJ: Pearson.
40. Google and Bastion roll out AI pilot program for Australian news organisations - Bastion
41. Shaun Davies interview with Dr Anne Kruger, May 16, 2025.
42. Shaun Davies interview with Dr Anne Kruger, May 16, 2025.

Dr Anne Kruger is a member of the governance board for the Australian Code of Practice on Disinformation and Misinformation (ACPDM). She was co-chief investigator in the development of the code, formed under policy directives from the Australian Treasury's Digital Platforms Inquiry. Dr Kruger is currently Convenor of the Bachelor of Journalism and Bachelor of Arts Journalism and Mass Communication Programs at the University of Queensland. Previously she was Asia Pacific Director for online verification NGO First Draft News, and has worked across senior roles in industry and academia in Singapore, Hong Kong and Australia.

Dr Richard Murray is Senior Lecturer & Director of Indigenous Engagement, School of Communication and Arts, Affiliate of Centre for Communication and Social Change, and Affiliate of Centre for Digital Cultures & Societies, and Affiliate of Research Centre in Creative Arts and Human Flourishing, at the University of Queensland, Australia. Dr Murray researches journalism in a time of rapid change. His research specialties include the role law and lawyers play in contemporary journalism, rural, regional and remote journalism, and international journalism with a focus on how South Korea and North Korea are covered and reported on.

Deepfakes, cloned voices, and digital media literacy: AI's role in the misinformation crisis in India

Vamsi Krishna Pothuru

How the advent of generative AI has worsened the misinformation problem in India – and the responses from stakeholders, along with potential solutions for the future.

In recent years, misinformation aided by digital technologies has been a persistent problem. It has been proven to be detrimental to democratic institutions across the world and has been threatening the social fabric of communities. The misinformation problem has been worsened in recent years, and the AI usage has further complicated the issue.

The arrival of generative artificial intelligence (Gen AI) chatbots such as OpenAI ChatGPT, Google Gemini, Microsoft Copilot, and others has democratised the use of artificial intelligence. These chatbots are fuelled by large language models, and require simple text prompts to generate content, including text, images, audio, and video. Deepfake videos, one of the popular variants of AI-generated content, involve techniques such as face swap, lip sync, and puppet master. Similarly, there are textual deepfakes, audio clones, AI-generated images, and others. Internet is flooded with tools and apps that aid in creating these sophisticated forms of altered media.

This synthetic technology is being lever-

aged by actors ranging from large organisations to ordinary citizens for various purposes. The anonymity and scalability of this technology are aiding bad actors in creating more personalised, sophisticated, and convincing mis/disinformation. A few of the prominent examples of deepfakes worldwide include a 2022 video of Ukrainian President Volodymyr Zelenskyy asking his troops to surrender to Russia. Also, AI is being used to resurrect dead personalities; a few examples in recent times include a deepfake video of Indonesia's late president Suharto and a deepfake video of the chief of the now-disbanded LTTE (Liberation Tigers of Tamil Eelam), Velupillai Prabhakaran. The motivations behind AI-generated mis/disinformation range from information warfare, propaganda, election campaigns, financial fraud, and personal vendettas to many others.

Unholy union of AI and misinformation:

Information disorder is the more academically rigorous term compared to the popular term "fake news". The information disorder is categorised into misinformation, disinformation and mal-information. This categorisation is based on two factors, which are the facticity of the information and the intention of the person creating or sharing the particular information. But the arrival of AI led to the creation of a large number of representative images and satirical memes, which are not necessarily factual but still evoke the same emotions as real images. In the traditional sense, to create mis/disinformation, one uses unrelated picture or video along with false narrative to create more virality. However, with the arrival of AI, even ordinary people are creating representative AI images and spreading their false narratives as facts.

AI-generated fake images have the same kind of convincing nature, sometimes more, compared to unrelated images used in a conventional fake news story. Kiran Grimealla and Simon Chauchard argue in *Nature* that even though AI-generated images resemble animation, they still resonate with the emotions of the audience and persuade them to believe in the message.

Similarly, Hany Farid, a professor at the University of California, said in an interview with NPR that these images are designed to push narrative and propaganda rather than being purely deceptive.

Most of the deepfake cases involve famous personalities as their photographs, speeches and videos are available in the public domain. It's easy to create a false narrative around such altered multimedia content which can persuade audiences by resonating with biases and pre-conceived notions. While it is easy to debunk traditional fake news items just by citing the factuality clause, it is difficult to fact-check a representative image or audio clone apart from just pointing out that it is AI-generated. This is true in the context of audio clones generated by AI, which are difficult to fact-check and declare as fake with 100% certainty. Several studies indicate that audio clones

are one of the popular deep fakes encountered across the world.

AI, misinformation and the ordinary citizen in India

In the year 2024, around 50 countries including India had general elections. It was expected that it could lead to the large-scale rise of election related deepfakes and AI-generated mis/disinformation. However, several studies indicated that the concerns about AI-induced misinformation during the election campaigns were overblown. While it is true to some extent, it is important to track the phenomenon and its future implications. Also, the impact of this technology varies across different societies. For example, in India, AI-generated misinformation should be approached very contextually. It must be understood by considering various factors, such as the ordinary people's capacity to use the internet, vulnerability to online harms, awareness about deepfakes, and the array of actors behind misinformation campaigns.

According to the 2024 report by KANTAR and IAMAI (Internet and Mobile Association of India), more than half of the internet users in India are from the rural parts of the country. The Indian rural population has increasingly come online with less to no digital literacy. Their digital vulnerabilities have been worsened in recent years, led by online harms such as mis/disinformation and cybercrimes. "Digital Deception Index: 2024 report on deepfake fraud's toll on India" by Pi-labs has revealed that deepfake-related cybercrimes in India have increased by 550% since 2019. "Digital arrest", deepfake e-KYC, fake trading apps and investment apps endorse celebrity deepfakes are a few of the cybercrimes that have been shaking the country for the last couple of years.

The alarming levels of cybercrime threats persisting in India reveal deep-rooted vulnerabilities among Indians. It coincided with the mis/disinformation crisis over the years, which has resulted in mob lynchings triggered by cow smuggling and child kidnapping rumours, especially in rural parts of the country. It has also resulted in extreme polarisation in society along the lines of religious identity, nationalism, and other ideologies. Upon further probing the potential reasons behind such vulnerabilities, it is evident that a lack of digital literacy and a lack of knowledge about the avenues of authentic information could be a few reasons among many.

According to Vice (2020), one of the first instances of AI usage in an election campaign happened in the 2020 Delhi elections. A deepfake video of Manoj Tiwari, a BJP leader, speaking Hindi and Haryanvi spread across thousands of WhatsApp groups. This seems like a harmless voice clone, but the usage of AI in election campaigns has taken many forms in recent years. In the run up to the general elections 2024, a series of deepfake video of late M. Karunanidhi, of DMK party in Tamil Nadu, a southern state in India, were screened in public events. Similarly, there were deepfake videos on the of Indian actors such as Ranveer Singh and Amir Khan, asking people to vote for the Indian National Congress (INC) party. Similarly, there were several deepfake videos of various celebrities in India endorsing political candidates, promoting fake medicines, and advertising financial scams.

In India, there were several cases of such deepfakes triggering a lot of discourse around the implications of AI. In some cases, these instances even led to punitive actions. Few scholars argue that the fear of legal action has resulted in fewer cases of deepfakes in the recent elections in India. However, in some cases, the state may overstep with regard to freedom of speech and democratic dissent. In a recent instance, Smita Sabharwal, a senior IAS officer in the southern state of Telangana, was summoned and later transferred for sharing an AI-generated Ghibli image about the controversial land auction by the state government at the University of Hyderabad.

Currently, India doesn't have a dedicated fake news law, but often invokes Bharatiya Nyaya Sanhita (BNS), 2023, the earlier Indian Penal Code to punish citizens for creating and sharing mis/disinformation. Similarly, Indian govern-

ment uses Information Technology Act (IT Act) 2000 and its subsequent intermediary guidelines, IT Rules 2021, to address mis/disinformation and deepfakes in India. Along with punishment for individuals for spreading disinformation, this rules also casts specific obligations on social media platforms which are referred to as intermediaries.

Gendered and ideological disinformation

"Gendered disinformation" primarily targets women and gender minorities through doctored photos, obscene videos and false narratives to defame their character and undermine their credibility. At the community level, where patriarchal notions prevail, such disinformation has severe implications for women, which may involve physical threats and restrictions on freedom and movement. It also hinders women's access to internet and education. Cheap fakes were one mode of such online harassment, but now AI is aiding this genre of disinformation with more personalisation, sophistication and anonymity. One example of gendered disinformation in India is the plethora of AI-generated soft porn images of Muslim women with Hindu men. This was revealed in an investigative report, "*Zalim Hindu Porn*", by Aditya Menon from Quint, an India-based news platform. These scores of AI-generated images become an ideological tool for digital hate spread across social media in a coordinated manner.

Similarly, an exclusive report by "Decode" of Boom (digital journalism and fact-checking platform in India), revealed how text-to-image tools are being weaponised to generate hateful imagery around certain communities in India. A few such harmful representations in AI generated images include Muslim men as paedophiles, stone pelters and other stereotypes. In a socially diverse country such as India, there is the danger that such hateful trends will be reflected across gender, caste, ethnicity, and other identities.

Response and promises from stakeholders

Fact-checking initiatives have been the first line of respondents to misinformation in India. On the other hand, social media platforms have been collaborating with these initiatives in the areas of debunking claims on their platforms, digital media literacy for their users and capacity building for the factcheckers. Similarly, civil society organisations are working at the grassroots, looking for resources to address the AI variants of mis/disinformation in their communities. More concrete collaborations among these three is essential in the fight against AI variants of mis/disinformation. The following sections will discuss the response from these stakeholders and also potential areas of solutions.

Fact-checking initiatives as front-line respondents

Fact-checking units have been actively responding to the AI-generated misinformation, such as deepfakes and audio clones. One of the biggest challenges for fact-checking initiatives over the years is to make their fact-checking article accessible to the audience. Given the scientific nature of their fact-checking articles, it was difficult for an ordinary person to access or understand these articles. So, they have started converting these long articles into short-form multimedia content such as reels, flashcards, Instagram carousels, and other formats.

One of the reasons behind this is to make people familiar with the fact-checking tools and process, which will serve the digital media literacy purpose in the long run. In a positive trend, fact-checking initiatives in India recognised the importance of digital media literacy among their audience. Along with factchecks, they also started creating explanatory videos, fact-checking tutorials and other educational content for their audience.

In response to AI-induced mis/disinformation, there are a few initiatives that solely address deepfakes in India. Deepfake Analysis Unit (DAU) and Logically Facts in India are active-

ly addressing AI-generated misinformation in India, especially deepfakes. DAU has a dedicated WhatsApp tipline for citizens to report deepfakes. DAU is a part of the Misinformation Combat Alliance (MCA), which is similar in its structure and purpose to the International Fact-Checking Network (IFCN). Most of the fact-checking initiatives in India are signatories of IFCN and the newly established Indian based coalition of MCA. One of the promising trends in the fact-checking ecology in India is the increased collaboration between fact-checkers and news publishers. Similarly, the SHAKTI coalition is one such collaboration that emerged to cover the 2024 general elections in India to combat election related mis/disinformation.

Such collaborations are critical in a multi-language country like India, which can facilitate the transfer of capacities and knowledge within the fact-checker community from different states. Also, collaboration among them could track the AI-generated information in India in real time. This knowledge will help in creating and deploying necessary tools to combat AI-generated mis/disinformation effectively. It also aids in creating digital media literacy resources for the audience and even capacity-building resources for fact-checkers and media professionals. The rapid developments of AI and its usage in the information landscape require fact-checkers and media professionals to be up to date with AI verification techniques. In a way, these collaborations also give a sense of community and equal learning space for the fact-checkers.

Social media platforms and information integrity on online spaces

The recent announcement by Meta to discontinue its third-party fact-checking programme in the USA has sent alarm bells across other countries, including India. Meta has nearly 100 fact-checking partners globally, and in India, it has 12 partners covering 16 languages. On the other hand, the X platform has been watering down "informational harm" policies and relying almost completely on community notes to address mis/disinformation. These developments pose a long-term, severe threat to information integrity on digital platforms.

Also, in recent years, fact-checking initiatives across the world have been facing a credibility crisis where these entities are being falsely appropriated to certain ideologies. Rappler, based in the Philippines, and Alt News in India are two such initiatives among many across the world that have to face threats from both the state and ideological groups. Amidst these conditions, social media platforms as responsible big tech must work with fact-checking community across country and aid in their efforts to fight mis/disinformation.

To address the rise of deepfakes in their platforms, social media platforms must ensure that watermarks or other authenticity indicators are provided on the content generated by AI tools. Such measures by technology companies can help in users distinguish between deepfakes and factual content. It is very important for ordinary internet users to distinguish between original content and content altered by AI on the social media they consume every day. With the tremendous reach to their audience, these platforms should provide AI indicators on deepfakes and incorporate digital media literacy content on their platforms. Such seemingly simple algorithmic measures are a positive step towards information integrity on online spaces.

Technology companies can collaborate with fact-checking initiatives to develop AI tools that can aid fact-checkers. Also, social media platforms should make their platform data more accessible to researchers, fact-checkers, and technology companies to develop free deepfake detection tools for the masses.

Civil society organisations and digital media literacy approach

Just imagine the kind of impact a deepfake video of a local politician inciting religious hatred can create in a small community like a village. How many of these first-time technology users know

about deepfakes or AI? Do they have the capacity to verify or access/read fact-check websites? Do they have avenues of authentic information? Are the existing digital media literacy or AI awareness programmes contextualised for their literacy levels, local culture, and knowledge? These are some important questions that need to be addressed in order to combat AI-generated mis/disinformation at the community level. The proposed technological solutions, such as mere watermarking on AI-generated content, scientific fact-checking articles may not be a complete solution for communities.

Sophisticated misinformation variants such as deepfakes and audio clones may have relatively severe implications for rural communities. AI awareness is the first and most important step in making these communities safe in the current information ecosystem. This basic knowledge among individuals can induce healthy scepticism, which is the first step in this long battle against AI-generated mis/disinformation. Stakeholders must come up with a comprehensive and contextual digital media literacy approach to create resilient communities to these ever-evolving online harms. There are a few civil society organisations that are fighting mis/disinformation at the community level.

Digital Empowerment Foundation (DEF) is a Delhi-based non-profit organisation that enables opportunities for communities through digital tools and digital literacy. They have also been addressing the misinformation problem at the grassroots through digital media literacy toolkits which are contextually developed for the communities they work with. Through their comprehensive "Media Literacy Awareness and Action Plan", they are using culturally relevant, peer-to-peer learning and hands-on training toolkits to make communities resilient against misinformation. They have also incorporated elements of how to recognise misinformation, fact-checking, and the concept of AI in their curriculums to make these communities aware of the implications of AI and misinformation.

Similarly, Ideosync Media Combine, currently running a programme called the "Bytewise Factcheck Fellowship" in partnership with Youth Ki Awaz, a citizen media platform in India. This digital media literacy programme equips school students aged 13-17 years with knowledge about the role of AI in mis/disinformation, digital tools to fact-check and other skills. Comprehensive media literacy training at a young age is a long but promising approach in preserving information integrity on online spaces. Also, Indian schools need dynamic curriculum which can equip students to become aware of mis/disinformation and deepfakes. One such successful programme is "Satyameva Jayate" (Truth triumphs), launched in 2021 by the Kerala government, a southern state of India. This programme aims at infusing responsible digital engagement in students, knowledge about mis/disinformation, and fact-checking skills.

These case studies reveal the importance of digital media literacy in fight against the mis/disinformation. DEF uses socially and culturally appropriated approach to impart digital media literacy for village communities. Apart from contextualising, they also use various strategies such as gamification, peer-to-peer learning which are instrumental in engaging communities in learning activities. Other stakeholders must recognise DEF's approach in realising effective and personalised digital media literacy awareness for citizens.

AI AS A POTENTIAL SOLUTION:

AI detection tools have been emerging in recent years, which can be a potential solution for verifying deepfakes. There are a few publicly available free AI tools for the verification of AI content which are playing a crucial role in debunking scores of AI mis/disinformation generated on a daily basis. They include *Hiya* (deepfake voice detection tool), *The Factual* (source quality checker), *Deepfake-o-Meter*, *Hive Moderation*, *Logically*, *Originality.ai*, *AI or Not*, and others. These tools are helping both fact-checking initiatives and individuals to verify any AI generated information they come across online.

In an interesting case, Factly, a data technology company and a fact-checking initiative based in Hyderabad, has developed its own AI fact-checking tools such as '*Sach*' and '*Tagore AI*' that can assist them in fact-checking. Factly has developed Sach with the support of the Google News Initiative. Fact-checking initiatives have the advantage of observing the creation and spread of deepfakes in real time. Their knowledge about variants of deepfakes and fact-checking techniques are valuable assets in developing AI tools in combating deepfakes and mis/disinformation effectively. Similarly, social and cultural knowledge of civil society organisations at the grassroots can help in further contextualising such AI tools.

Conclusion

Even though the various studies indicate that the concerns around AI-generated misinformation are overblown, it is necessary to foresee the implications of this technology in the future and track its evolution in the present. With the rapid evolution of generative AI, it is imperative that the strategies to combat deepfakes should evolve too. It is also necessary that these strategies and solutions must consider the rural populations due to their increased vulnerability to online harms. The advent of AI has resulted in not only the mass production of deepfakes but also resulting in evolving strategies that can create more sophisticated mis/disinformation.

In this context, technology companies are responsible for sharing the data around emerging AI technologies with researchers and academicians. Also, stakeholder collaborations could lead to effective strategies to create free and easy-to-use AI detection tools for citizens by harnessing AI. Finally, community centric approach of digital literacy programmes, which are socially and culturally contextual, promises resilience against ongoing information crisis.

References

Bond, S. (2024, December 21). How AI deepfakes polluted elections in 2024. NPR. https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections

Christopher, N. (2020, February 18). *We've just seen the first use of deepfakes in an Indian election campaign*. VICE. https://www.vice.com/en/article/the-first-use-of-deepfakes-in-indian-election-by-bjp/

Digital Empowerment Foundation. (n.d.). Media information literacy initiatives. https://www.defindia.org/media-information-literacy-initiatives/

Garimella, K., & Chauchard, S. (2024, June 5). How prevalent is AI misinformation? What our studies in India show so far. Nature. https://www.nature.com/articles/d41586-024-01588-2

Kantar & Internet and Mobile Association of India. (2024). Internet in India 2024. https://www.iamai.in/sites/default/files/research/Kantar_%20IAMAI%20report_2024_.pdf

Menon, A. (2025, March 7). *'Zalim Hindu' porn: How AI is mass producing pornographic images of Muslim women*. The Quint. https://www.thequint.com/news/politics/artificial-intelligence-muslim-women-hindutva-soft-porn-images-facebook-instagram

Press Information Bureau. (2025, April 4). Government of India taking measures to tackle deepfakes. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2119050

Rebelo, K. (2024, October 14). Exclusive: Meta AI's text-to-image feature weaponised in India to generate harmful imagery. BOOM. https://www.boomlive.in/decode/exclusive-meta-ais-text-to-image-feature-weaponised-in-india-to-generate-harmful-imagery-26712

Vamsi Krishna Pothuru is a PhD student in the Department of Communication at the University of Hyderabad, India. Under the supervision of Prof. Kanchan K. Malik, his research examines information disorder in Indian villages and the responses of various stakeholders, including civil society organizations. One key area of his research focuses on a community-centric approach to digital media literacy interventions aimed at addressing misinformation. Before beginning his PhD, he worked as a fact-checker at NewsMeter, an IFCN-certified media house in Hyderabad, India.

Ethical journalism in the age of AI

Hasani Felix

Journalism is a field which entails telling the public what they need to know with accuracy, clarity and simplicity so everyone can understand. The general aim is to provide new and relevant information to an intended audience. Historically, journalists began publishing news utilizing the print medium. However, in the contemporary era information has been shared through more mediums which include radio, broadcast television and social media.

Real journalism requires news to undergo a method of data collection, content organization, editorial assessment and public distribution. The Society of Professional Journalists (2014) states that the codes of ethics governing journalism are seeking the truth and reporting it, minimizing harm with content produced, acting independently with integrity and impartiality as well as being accountable and clear in the choices made in gathering information.

Journalism ethics are standards that safeguard both the public and the reputation of reporters. Nevertheless, applying these ethical considerations is a matter of choice as individuals have the inherent ability of free will. Still, newsroom editors place high levels of trust in reporters hoping that they produce original and factual stories. Presently, in the age of Artificial Intelligence, information is easier to access and there are more avenues to fall into the snares of plagiarism as well as inaccuracy. Yet these issues existed before Artificial Intelligence, which suggests the issue resides with the ethical standards of media practitioners. Maintaining ethical standards in the journalism industry has become increasingly difficult but it can be done by emphasising individual ethical responsibility, promoting the codes of ethics in journalism and employing Artificial Intelligence checkers within the media space.

Of central concern is the necessity to reinforce the importance of individual ethical responsibility. The philosophy of deontological ethics supports this notion in the context of the communication industry. Proponents of deontological ethics believe that adhering to the ethical rules and upholding one's duty is obligatory and should be independent of the potential consequences. Gordon (2011) mentions Immanuel Kant's categorical imperative which suggests that moral rules should be treated as universal laws. This implies that media professionals in the field of journalism have a moral duty to act ethically, to not plagiarize the work of others and to stick to the code of practices for professional journalists.

This ideology suggests that journalists should do the right thing despite the increased difficulty for editors and lawmakers to pinpoint infringements. Reinforcing the importance of individual ethical responsibility in journalists whose thought process aligns with Kant's categorical imperative and by deontological ethics is a realistic measure in maintaining ethical standards. Gordon (2011) states that this ideology appeals to many media professionals as they align themselves with the view of telling the truth, being consistent and not worrying about the consequences. Therefore, reminding these journalists about the ethical principles and practices of journalism is a realistic safeguard to protect the integrity of this profession regardless of the difficulties posed by Artificial Intelligence.

Furthermore, there is evidence to support the existence of journalists who are morally upright despite the potential consequences. Committee to Protect Journalists (2025) comments on the death of Dumesky Kersaint, a journalist who was killed while reporting on the murder of a man in Haiti in 2023. The organization also highlights that a witness told Fabien Iliophène,

the founder of the radio station where Kersaint worked, that Kersaint was killed after refusing to delete photographic evidence of a crime scene he was reporting on. This suggests that Kersaint understood the importance of his work as a journalist and was willing to live for it as much as he died for it. Kersaint's commitment to expose corruption in Haiti can be seen as an example of deontological ethics. The chronology of Kersaint's actions suggests that he valued his moral duty to uncover the truth and hold the murderers accountable for their actions.

Additionally, it suggests that he acted regardless of the consequence of losing his life, which aligns with the principles of deontology. Specifically, Kersaint's actions are in harmony with Kant's theory of ethics and the codes of ethics governing journalism as he exhibited plans of seeking the truth and reporting it to the public. Ultimately, Kersaint's resolve during the age of Artificial Intelligence suggests that it is realistic to maintain the ethical standards of journalism in this era.

Virtue ethics

Another ethical theory which supports the perspective of emphasising individual ethical responsibility is virtue ethics. Tilak (2020) suggests that virtue ethics is in harmony with important principles and practices of journalism such as trustworthiness, respect, responsibility, fairness, truth and self-restraint. Professional journalists who align with the moral standards of virtue ethics are not influenced by consequences such as fear of detection nor is their integrity bound to duty and the rules against unethical conduct. Rather they perceive morally right decisions as intrinsic.

Journalists who subscribe to the tenets of virtue ethics maintain the responsibility to use Artificial Intelligence ethically regardless of the lack of rules governing its usage in the newsroom. Furthermore, proponents of virtue ethics who subscribe to the concept of Aristotle's Golden Mean – which describes a character state that emphasises making moderate choices – find a balanced approach between two extremes (Gordon, 2011). While some believe that Artificial Intelligence should not be used in journalism and others exploit its capabilities, virtue ethicists upholding Aristotle's Golden Mean would aim to thoroughly investigate and verify facts they obtain from Artificial Intelligence and review the information generated from it to report truthfully and accurately.

Another issue of great importance is the need to promote the codes of ethics in journalism from a pragmatic perspective, especially in the age of Artificial Intelligence. Pragmatism focuses on implementing practical approaches that will allow journalism to be more functional, ethical and sustainable despite the current challenges presented by technological advancements.

There are cases where Artificial Intelligence has been exploited and potentially has negative repercussions. Ortiz (2024) states that a Wyoming reporter used fake quotes in his news story that were created by Artificial Intelligence, and he failed to interview the people he quoted. This led to the editor having to apologise for not detecting this error and affected the reputation of the media house. Furthermore, this infringes on the codes and ethics of journalism as a profession.

However, the issues of plagiarism and drifting away from the ethical codes of conduct started before Artificial Intelligence's introduction to the newsroom. Kurtz (1996) mentions former journalist Janet Cooke who won the 1981 Pulitzer Prize for a news story about an 8-year-old heroin addict which was eventually proven fake. This bogus story was not written with Artificial Intelligence, but it still lacks integrity, fails to tell the truth and causes harm to the public.

Additionally, through the lens of pragmatism it can be suggested that a practical measure to combat unethical conduct and plagiarism is to allow the use of Artificial Intelligence while implementing rules which align with the codes of ethical journalism. Aftonbladet and VG, which are two news outlets owned by the news corporation Schibsted, have similar codes and guidelines as they state that all published materials

should be reviewed by humans and all generative AI content must be manually approved before being disseminated to the public respectively (Cools, 2023). These are some practical steps to ensure that Artificial Intelligence is not abused but rather ethically included. Cools (2023) also states that both Aftonbladet and VG mention that when they use AI-generated material, the content is to be labelled clearly so that everyone receiving the information understands that Artificial Intelligence was included.

This aligns with the pragmatic ideology of implementing practical steps and focuses on the journalistic ethical aspect of transparency. (Kodilinye, 2009) states that media houses should prioritize being transparent and ensure that they implement rigorous fact-checking. These instructions are applicable when utilizing pragmatic thinking to combat the imperfections exhibited in gatekeeping. In essence, Artificial Intelligence can be utilized in journalism, but it must be done ethically, transparently and with clear guidelines governing its incorporation.

Another integral necessity in maintaining the ethical integrity of journalism is implementing software in media houses that are capable of tracking and detecting the usage of Artificial Intelligence.

From a utilitarianism perspective – which states that an action is morally correct if it maximizes happiness for most people while minimizing harmful outcomes – implementing AI trackers can help prevent plagiarism and uphold ethical reporting by maximizing the public trust and accuracy. Utilitarianism justifies using AI tools to detect and prevent violations of accuracy and credibility of news as this overall benefit to society exceeds any inconvenience placed on reporters. Tilak (2020) emphasizes the importance of avoiding plagiarism and maintaining the integrity of work.

Safeguarding reputations

Artificial Intelligence tools can effectively scan large amounts of text and identify similarities to uphold ethical standards. These tools can act as a safeguard which protects media houses reputation and their authentically produced content. Furthermore, Artificial Intelligence could be trained to identify reporting practices that are violations as mentioned by Tilak (2020) such as being biased, stereotyping or invading privacy.

This creates the opportunity to bolster the ethical nature of journalism and is of considerable benefit to the public as Artificial Intelligence acts as a watchdog who ensures reporting is done with fairness, responsibility and integrity. Drozdowski (2023) highlights Turnitin's new Artificial Intelligence detector stated to have 98% accuracy in spotting AI generated content. He also mentions that Turnitin was able to pinpoint the difference in three essays he submitted: one without any AI, another fully written by AI and also one partially written by AI, with substantial accuracy. This suggests that it is feasible for similar detectors to be used in the field of journalism to detect plagiarized articles and pinpoint which sections of the article are not written by a human.

In totality, maintaining ethical standards in journalism amid the rise of Artificial Intelligence is important and requires a multilayered approach. By reinforcing individual ethical responsibility through deontological ethics, journalists can adhere to moral principle despite the potential consequence. The case of reporter Dumesky Kersaint exemplifies how professional integrity can withstand external pressures in the face of personal risks. Furthermore, promoting journalistic codes and ethics by utilizing pragmatic solutions ensures that Artificial Intelligence is used at a toll for enhancing the field of journalism rather than deception of the public.

Newsrooms need to create clear guidelines as portrayed by media houses owned by Schibsted where human oversight and transparency are necessary factors when utilizing AI-generated content. Moreover, including AI tracking software aligns with utilitarianism by minimizing misinformation and maximizing societal trust. Turnitin's AI detection technology is an example that using such systems can maintain authenticity in journalism. Virtue ethics encour-

ages journalists to find a balance between AI usage and ethical reporting. Ultimately, ethical journalism is a goal that requires constant vigilance, accountability and unwavering commitment to accuracy. All things considered; these measures can uphold the ethical standards within the media room despite the growing challenge faced by Artificial Intelligence. ▪

References

Committee to Protect Journalists. "Dumesky Kersaint - Committee to Protect Journalists." *Committee to Protect Journalists*, 3 Mar. 2025, cpj.org/data/people/dumesky-kersaint/. Accessed 2 Apr. 2025.

Cools, Hannes , and Nicholas Diakopoulos. "Writing Guidelines for the Role of AI in Your Newsroom? Here Are Some, Er, Guidelines for That." *Nieman Lab*, 11 July 2023, www.niemanlab.org/2023/07/writing-guidelines-for-the-role-of-ai-in-your-newsroom-here-are-some-er-guidelines-for-that/.

Drozdowski, M. J. (2023, April 26). *Testing Turnitin's New AI Detector: How Accurate Is It? | BestColleges* (D. Earnest, Ed.). Www.bestcolleges.com. https://www.bestcolleges.com/news/analysis/testing-turnitin-new-ai-detector/

Gordon, David, and John S Armstrong. *Controversies in Media Ethics*. Third ed., New York, Routledge, 2011.

Kodilinye, G. (2009). Commonwealth Caribbean Tort Law. In Routledge eBooks. Informa. https://doi.org/10.4324/9780203874233

Kurtz, Howard. "JANET COOKE'S UNTOLD STORY." *Washington Post*, 9 May 1996, www.washingtonpost.com/archive/lifestyle/1996/05/09/janet-cookes-untold-story/23151d68-3abd-449a-a053-d72793939d85/.

Ortiz, Aimee. "Wyoming Reporter Resigns after Using A.I. To Fabricate Quotes." *The New York Times*, 14 Aug. 2024, www.nytimes.com/2024/08/14/business/media/wyoming-cody-enterprise-ai.html.

Society of Professional Journalists. (2014, September 6). *SPJ Code of Ethics*. Society of Professional Journalists; Society of Professional Journalists. https://www.spj.org/ethicscode.asp

Tilak, Dr. Geetali. "The Study and Importance of Media Ethics." *Research Gate*, 2020, www.researchgate.net/publication/349685937.

Hasani Felix is a Trinidadian reading for the Bachelor of Arts degree in Journalism at The University of the West Indies, Jamaica.

Needed: An antidote to misinformation in the Caribbean

Ricardo Brooks

It was with some interest that I noted that the 2025 Global Risk Report by the World Economic Forum identified misinformation and disinformation as the most pressing risk facing the world over the next two years.

Admittedly, at first glance one could be forgiven for questioning that declaration. After all, there is the evolving threat of climate change, increasing levels of economic insecurity, and the ever-present threat of war and geopolitical conflict. Yet despite these very real concerns, there is a certain wisdom in seeing misinformation and disinformation as existential threats to human interaction and, more specifically, public discourse.

I do not mean to be hyperbolic, but insofar as increasing levels of misinformation and disinformation risk polarising society, I believe we ought to take the threat seriously. This is particularly true for small societies in the Caribbean that have been bedevilled by political systems that have not always delivered in the way they ought to for the people of the region.

That failure has left generations of citizens disaffected, disillusioned, and dissatisfied. The consequence of this cynicism and malaise is a population of fertile minds in which the seeds of misinformation and disinformation can take root and grow.

It is exactly because Caribbean people are so vulnerable to the distortions of information that we need to be most vigilant against this threat. Already we are seeing the polarising effect

misinformation can have on the body politic and society at large.

In Jamaica, for example, there has been an explosion of vloggers on YouTube. While not all of them are seeking to poison minds, a great deal of them are quite content to share political conspiracies, health myths, and security lies. These outcomes threaten the credibility and reliability of many of society's most cherished institutions. We should take it seriously.

Unfortunately, many governments, civil society actors, and the media are only now waking up to the potency of this threat because it is effectively sidelining them in the battle for ideas. Many of them are losing what the Jamaican Prime Minister, Andrew Holness, has identified as the battle for the mind.

Incumbent governments are becoming prone to sustained misinformation campaigns that imperil their re-election prospects, civil society advocates are being burdened by conspiracy theories about their funding and motivations, and the media are facing a financial fight for eyeballs and ears in an increasingly splintered environment that seems to prioritise the sensational over the factual. In that sense, bad faith actors are incentivised to dabble in the cesspool of misinformation and disinformation.

Havens of misinformation and disinformation

The situation is further complicated by United States' technology giants, which see value in maintaining on their platforms online communities where lies flourish and truth dies. Some of these communities have become havens of misinformation and disinformation. Instead of having to confront facts, already disaffected and disenchanted citizens can now wall off themselves in these spaces and wallow in their misery and discontent, all the while being reinforced in their beliefs by misinformation.

That is a dangerous cocktail for any society. We have already seen in sections of Europe where that kind of outcome leads to violent radicalisation and in some instances anti-immigrant nationalism. Other developed societies have also seen the rise of fringe groups who are sometimes led to extremes by the misinformation they consume.

The Caribbean has not been immune from these risks. To that end, Facebook's recent announcement that it was abandoning independent fact-checkers in favour of user moderation is but another step along the path of disregarding the legitimacy of the truth. These technology giants are stepping back from their duty to combat misinformation and disinformation at precisely the point where there's a need for more vigilance, not less. The rise of artificial intelligence will result in an entirely different beast, with as many threats as opportunities.

Despite the bleak outlook, I do not believe all is lost. In fact, it requires an acceptance that if so-called "Big Tech" is not going to assist us to combat bad faith actors who weaponize information, it is important for societies across the region to start finding ways of at least arming our populations to be more discerning in the information they consume.

We need to see proper, structured media literacy campaigns as not just nice-to-haves, but need-to-haves. We need to start training our people from a young age, particularly those who are digital natives and must navigate technology's many complexities, how to think critically about information and content they consume.

Caribbean societies must see the threat of misinformation and disinformation for what it is and combat it now. Regional communication scholars need to do more to make their scholarship relevant to the deficits faced by our people. It must be an all-of-society approach.

I'm not convinced that the proliferation of misinformation and disinformation has to be the norm. I'm not convinced that generations of young people have to grow up being strangers to the truth. That does not have to be our rule. In fact, I would much rather it be the exception. ▪

Ricardo Brooks is a Jamaican journalist.

Power, responsibility, and trust: A framework for communication governance in the digital age

Cordel Green

In a world where the flick of a thumb can transmit ideas across continents, ignite social movements, or unleash torrents of disinformation, the question before us is urgent: how do we safeguard the public interest in this vast and often uncharted communication space? How do we ensure that the environment in which ideas, news, and opinions flow remains a force for truth, inclusion, and democracy – rather than division, manipulation, and harm?

This is not a challenge for regulators alone. It requires action by politicians, policy makers, power brokers, technology leaders, educators, journalists, students, and every citizen who participates in public discourse. It is a challenge that will define the integrity of democracy and the strength of our social fabric for generations to come.

The Balance of Power Has Shifted

The reality is stark. The balance of power in the information ecosystem has shifted dramatically. Today, global digital platforms wield influence that surpasses the resources and reach of most nation states. Their algorithms and infrastructures, designed to maximise engagement and profit, shape what billions see, hear, and believe – often with little transparency or accountability.

The cracks in our digital realities are evident through the convergence of nanotechnology, bio-technology, information technology, neuro-technology, artificial intelligence, and the socio-cultural undercurrents they set in motion: legal systems that strain to keep up; trust under stress; creativity being liberated – and constrained; the blurred lines between creation and control; the hidden biases in algorithms; and the silent exclusions in access.

Meanwhile, users, no matter how digitally literate or well-intentioned, navigate an environment where even the most sophisticated critical thinkers are outmatched by the speed, scale, and opacity of these systems. What was once a manageable national space of broadcasters and newspapers has become a borderless, high-velocity sphere in which truth competes with falsehood, trust is fragile, and harm can spread at unprecedented rates.

This is the definition of the modern mission: to ensure that human values – not just data and algorithms – shape the systems and societies we build.

Responsibility Must Follow Power

Faced with these realities, we must abandon outdated notions of equal responsibility. In this era, it must be proportionate to power and capacity. The greatest responsibility falls to those who have the greatest capacity to shape our information environment: the digital platforms whose systems determine what content is amplified, what is suppressed, and what is monetised. They bear a systemic duty of care, an obligation to design systems that do not simply profit from attention but protect against foreseeable harms.

Governments must act as stewards of the public interest, setting standards that reflect our values, convening diverse voices to co-create solutions, and safeguarding the rights of citizens. An independent regulator remains crucial, not as

an enforcer of outdated rules, but to build frameworks that enable a healthy, trusted communication space. Content creators, too, must play their part by embracing ethical norms of accuracy, fairness, and transparency. And citizens, supported by strong national digital, media and information literacy initiatives, must be equipped to engage critically and responsibly in this dynamic information environment.

Big Idea: The Public Interest Compact

But the pressing question remains: how can a small nation like Jamaica, with limited jurisdictional leverage, inspire these global actors to act in our shared interest? A proposed solution is the Public Interest Compact.

The Public Interest Compact is an invitation to digital platforms to enter into voluntary, public agreements that affirm their commitment to transparency, safety, inclusion, and support for national digital trust initiatives. These compacts are not about coercion or levies that we cannot realistically enforce. They are about partnership. They are about positioning the state as a principled, collaborative actor that offers Big Tech the opportunity to demonstrate leadership, rather than resist regulation.

Through these compacts, platforms would commit to sharing data on disinformation and harmful content trends specific to a country, supporting digital and media literacy efforts through tools, content, and micro-grants, and working to co-design standards that reflect values of fairness, truth, and inclusion.

Turning Power into Partnership

What makes the Public Interest Compact powerful is that it leverages what Big Tech will grow to value more: public trust, reputational legitimacy, and the opportunity to showcase good corporate citizenship. It allows a small state to punch above its weight, not by matching the might of these companies, but by offering a credible, constructive path that aligns with their interests and ours. The Compact offers a realistic, positive way forward at a time when adversarial regulation would achieve little more than performative resistance.

A Call to Shape the Future Together

The time to act is now. The health of our democracy, the resilience of our society, and the strength of our civic discourse depend on the choices we make today. The Public Interest Compact is not just a policy tool. It is a statement about commitment to shaping a communication environment where truth flourishes, inclusion is real, and trust can be restored. It is an opportunity for the government to lead, not by force, but by example – to show that small states can be innovators in digital governance, principled in purpose, and bold in vision.

Jamaica has long been admired for the strength of its voice and the clarity of its principles on the world stage. It could now be the nation that helps chart a new course for communication governance: a model of power shared responsibly, of trust rebuilt collaboratively, and of hope renewed through action. ▪

Cordel Green is Executive Director of the Broadcasting Commission of Jamaica. His other affiliations include being Vice-Chairman of the UNESCO Information For All Programme (IFAP), Chairman of the UNESCO IFAP Working Group on Information Accessibility and Member of Jamaica's National AI Task Force.

Can machines think? What feminism can teach us about ethical AI development beyond de-biasing

Laine McCrory

A significant catalyst for the development of artificial intelligence (AI) research occurred when computer scientist Alan Turing asked the question: can machines think? In his research, Turing evaluated whether users could tell the difference between two different terminals – one of which was controlled by a human, and the other controlled by a machine. A machine passed this test (deemed the "Turing Test") as long as it could convince a human user that it was interacting with a human rather than a machine.

Following this question, the next decades of computer science research examined the idea of how human intelligence can be replicated in algorithms. In 1956, the term "artificial intelligence" was coined as part of the Dartmouth Summer Research Project on Artificial Intelligence, a conference of leading computer scientists studying machine learning. From the Dartmouth Summer Research Project, numerous developments in computer science and engineering have occurred, highlighting that the seemingly rapid development of AI tools is the result of decades of research, theorization, and development that still aims to answer the question: *can machines think?*

Currently, a commonly accepted account of AI is found in the OECD's formal definition, which was passed in March 2024 and forms the basis of many policy and research objectives:

> An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment (OECD, p.4).

This definition recognizes the variety of ways in which AI can be developed, as well as alluding to the potential outputs that the systems create, which may lead to harmful impacts. Most commonly, the harms of AI are presented as individual threats to autonomy and privacy that may lead to reputational damage, physical injury, economic loss, or discrimination. We have seen this proliferate with the risk of deep fakes, chatbots and misinformation. As a way to address these harms, initiatives from both industry and policymakers have worked to build ethical AI systems.

Building ethical AI

Within industry, ethical AI development focuses on designing systems according to the principles of Fairness, Accountability, Transparency and Ethics ('FATE'), with industry actors taking the lead on how to design standards and practices that align with the principles. Industry approaches to ethical AI development can often occur as cooperative agreements between corporations – such as the Montreal Declaration on Responsible AI – or as individual values within companies themselves. However, these guidelines are often described as voluntary commitments, as they encourage responsible development without placing requirements and penalties for failing to develop ethical systems. They argue that because

engineers are the most informed about the technical architecture of the systems, they are capable of addressing the individual harms that may arise.

In addition, advocates for industry-led approaches to ethical AI development highlight how the rapid pace of technological development often outpaces policy. Yet, these industry-led initiatives have faced criticism for relying on voluntary principles and framing harms as issues that can be solved through technical fixes, such as the FATE principles. Without a framework that requires compliance, critics argue that it is difficult to ensure that these AI systems will be built ethically.

On the other hand, there has also been an increase in policies that promote ethical AI development, both at national and international levels. The EU AI Act, which was passed in March 2024, has quickly gained prominence as a guideline for risk-based AI policies. The EU AI Act differs from voluntary agreements like the Montreal Declaration, to both promote innovation and minimize the risks of AI development. The Act assigns AI systems according to four different categories of risk: unacceptable-, high-, limited- and minimal-risk applications. By promoting different categories of risk, the Act imposes requirements for trust, transparency, and accountability upon the systems.

This differs from the voluntary approaches, as AI developers are required to comply with the Act, rather than able to choose to adhere to the goals of the Act. Those who support the AI Act argue that this requirement could lead to an increase in public trust of AI. However, certain critics are sceptical of the AI Act's role in increasing public trust, as it does not address the immaterial harms caused by AI and the need for user agency.

Beyond De-Biasing Systems

While they are different in who leads, both of these approaches focus on promoting innovation while preventing individual risks. They reiterate the practice of debiasing as a process of mitigating biases in algorithms to align with certain principles, preventing the potential for harm. Yet in focusing on the risks at an individual and technical level, these approaches do not address the systemic harms ingrained in the technological and policy development processes. Rather than adhering to the definition of AI as proposed by the OECD, scholar of critical AI Kate Crawford argues that we must understand artificial intelligence as "both embodied and material, made from natural resources, fuel, human labor, infrastructure, logistics, histories, and classifications," (Crawford, p. 8). She sees AI as impacting more than individuals, as these systems are designed as structures of power and control.

Crawford also argues that AI research is built on a myth that intelligence is something that can be quantified, captured, and measured independently from social and political realities. The harms of AI systems extend far beyond the individual, as these systems have been used to facilitate digital redlining of marginalized communities, misclassify people who do not conform to the gender binary, reinforce ableist stereotypes of disability, and deepen racial inequalities. As these systems are built on data that does not recognize the systemic biases inherent in it, AI technologies often reiterate harmful historical practices of marginalization and discrimination.

When such deep-rooted biases are only addressed through individual fixes in policy and technical design, they fail to recognize the pattern of collective harm. Thus, it is important to consider the limits of AI ethics discourses, and instead talk about power.

What can feminism teach us about ethical AI development?

In order to address the systemic harms caused by AI, there is a need to understand AI itself as a system of power, as described by Kate Crawford. Systemic biases cannot be "programmed out" as they are interconnected with every aspect of daily life, even if they have become invisible or normalized. Feminist research is particularly important in this field, as feminists argue that *the personal is political*, meaning that politics play a

role in our lives in many different ways. Kimberlé Crenshaw, a legal scholar who focussed on the ways that Black women were marginalized within the legal system due to their existence at the intersection of race and gender, coined the term "intersectionality" in 1989 to describe how systems of overlapping power mean that different groups experience harm along multiple axes. Feminist research operates with a goal of exposing and challenging structures of power, a practice which has been embraced by feminist tech researchers such as Catherine D'Ignazio and Lauren Klein.

In their book *Data Feminism*, D'Ignazio and Klein highlight how tech development comes from histories of counting and classification that sought to control marginalized groups. They argue that feminist research on the histories and futures of AI must challenge these practices and build collective agency. A focus on intersectionality argues that as harm comes from multiple different axes, addressing one source of oppression – for instance, gender inequality – will not lead to justice for everyone, as there are many people who are oppressed along multiple lines. As such, approaches to addressing the specific risks of AI must attend to these intersectional harms by respecting and cultivating a diverse range of agency and expertise outside of the purely technical spheres.

One way to work towards this range of expertise is to use policy to promote digital citizenship as a way for users to actively engage with and impact the future of AI development. For Engin Isin and Evelyn Ruppert, the digital citizen is a distinct role that comes about through an individual acting as part of a broader collective to both learn about digital systems, and promote digital rights. Digital citizenship emphasizes how one should be actively involved in the creation of rights and responsibilities, rather than have perspectives universally imposed upon them. Therefore, for digital citizens to actively participate, policy must develop meaningful methods of engagement.

Digital citizenship involves two processes: enactment and inscription. Inscription involves how users claim rights through legal processes, while enactment refers to the role that users have in defining what these rights are. While inscription is particularly important, focussing only on the ways that digital citizens can be impacted by a system does not give them agency. Rather, there is a need to embrace the importance of digital citizenship as a participatory project, where policy supports enactment processes.

These enactment processes see individuals and collectives as essential to the policymaking process, arguing that if policy wants to address the systemic harms felt by marginalized groups within AI systems, these groups must be considered as central actors in the policymaking process through a process of meaningful engagement. Meaningfully engaging groups involves critically examining relationships of power and access, to highlight that within the AI ecosystem certain technical and policy voices are privileged over others.

In its current structure AI policymaking often happens within high-level groups consisting of a network of experts with very little representation from those outside of technical, industry and academic spheres, or within consultations and forums that are riddled with power imbalances. A feminist approach to building digital citizenship in the age of AI involves building from the participatory methods that promote equitable involvement and collective ownership, such as data trusts and stewardship, as well as participatory methods including mini publics, citizens juries, and community oversight. These processes involve putting digital citizens at the heart of decision making, while focussing on addressing barriers to participation through an equity-focussed lens that accounts for histories of marginalization.

Feminism can teach us a lot about who goes noticed and unnoticed in both AI development and policy. In her book *Living a Feminist Life*, Sarah Ahmed demonstrates how some of the most important work done towards social change can be generated by using feminism

to call attention to problems that go unnoticed and connecting them to systems of power. Much of this work is done not by academics or engineers, but those who have lived experience and the knowledge to know what a community needs to flourish. In proposing a feminist account of the limits of ethical AI development as well as a method that these limits can be challenged, I reiterate the importance of critically analysing how our policy frameworks address the harms posed by AI. Embracing the importance of feminist research *and* participatory digital citizenship recognizes the importance of not only examining power but of developing strategies to shift power as well. ■

Laine McCrory is a Master's student in the Joint Program in Communication and Culture at Toronto Metropolitan University and York University, and an incoming doctoral student at New York University in Media, Culture and Communication. Her research focuses on the intersection of feminism, policy and artificial intelligence. In addition to her academic work, Laine has been a Digital Policy Hub fellow with the Centre for International Governance Innovation, and is the founder of the Techno-Feminist AI Syllabus.

A digital milestone: New resolution on human rights defenders and new technologies adopted by the UN Human Rights Council

Francia Baltazar and Paula Martins

The Association for Progressive Communications (APC) welcomes the landmark resolution on human rights defenders and new and emerging technologies adopted by consensus at the 58th session of the UN Human Rights Council on 4 April 2025. This resolution addresses many of the concerns that APC and our partners have been advocating for in recent years and represents an important step forward for the protection of human rights defenders (HRDs), including women human right defenders (WHRDs), in the digital age.

Led by Norway and co-sponsored by over 50 countries, this resolution signals strong international commitment to ensure the protection of HRDs worldwide. APC was actively involved throughout the drafting process, providing written inputs and meeting with delegations. The final text incorporates recommendations from civil society, including points that had been raised by APC.

The path to consensus was not easy: the negotiation process required 12 informal sessions to agree on the text, with some actors attempting to weaken provisions through amendments. Despite these challenges, the resolution was ultimately adopted without a vote, representing a significant victory for the HRDs movement.

Key achievements in the resolution

The resolution introduces commitments for both governments and private companies on a range of digital issues, including the following:

1. Biometric mass surveillance

For the first time in a UN resolution, governments are called upon to ensure that biometric identification and recognition technologies, including facial recognition, are not used for mass surveillance by public or private actors. This is a significant advancement in protecting HRDs from emerging surveillance threats.

2. Internet shutdowns and connectivity

The resolution explicitly addresses not only internet shutdowns but also other practices that impede connectivity, including filtering and throttling measures. It encourages diverse technological solutions to advance connectivity, including creating an enabling environment for small, non-profit and community-centred internet operators.

3. Spyware and surveillance

The resolution calls on states to "refrain from or cease the use or transfer of new and emerging technologies, including artificial intelligence applications and spyware" that cannot be operated in compliance with international human rights law. It also calls for mechanisms to provide appropriate remedies for victims of surveillance-related violations.

4. Business responsibility

The text underscores the responsibility of business enterprises, particularly those in the tech industry, to respect human rights, including the rights of HRDs. It encourages companies to implement human rights due diligence throughout the life cycle of their products and services.

5. Strategic lawsuits against public participation (SLAPPs)

The resolution calls on governments to adopt and implement laws and policies that discourage strategic lawsuits against public participation targeting journalists, media outlets and HRDs.

6. Online attacks against women

The resolution condemns online attacks against women and girls, including gender-based violence and abuse facilitated by technology, such as doxxing, deepfakes and online harassment. It calls for gender-responsive approaches to address these particular forms of online discrimination.

Looking forward

UN resolutions are not binding documents; however, they represent a strong political commitment by states to act in line with their international human rights obligations and carry significant normative weight. APC will continue to work with partners to ensure that these commitments translate into concrete action at the national level to protect human rights defenders in the digital age.

While this resolution represents significant progress, some important issues remain unaddressed. Initially, the resolution included a reference to transnational repression, but this was removed in the final version, though indirect references remain. The resolution also lacks explicit recognition of the positive role of child HRDs in the digital space and their specific protection needs.

As part of the resolution's implementation, the Office of the UN High Commissioner for Human Rights (OHCHR) is mandated to convene three regional workshops to assess risks created by digital technologies for HRDs and identify best practices to address these concerns.

The OHCHR will also prepare a report with recommendations for improved responses to these risks, to be presented at the 63rd session of the Human Rights Council.

The adoption of this resolution is a testament to the power of collective civil society advocacy. It provides a framework for holding states and companies accountable for protecting those who defend human rights in an increasingly digital world. Now, implementation will pose us with the challenge of transforming these commitments into meaningful protections for HRDs around the world and across regional and local contexts. *Source: APC, 24 April 2025*.

A human rights approach to AI

UNESCO

Ten core principles lay out a human rights centred approach to AI ethics.

1. Proportionality and Do No Harm
The use of AI systems must not go beyond what is necessary to achieve a legitimate aim. Risk assessment should be used to prevent harms which may result from such uses.

2. Safety and Security
Unwanted harms (safety risks) as well as vulnerabilities to attack (security risks) should be avoided and addressed by AI actors.

3. Right to Privacy and Data Protection
Privacy must be protected and promoted throughout the AI lifecycle. Adequate data protection frameworks should also be established.

4. Multi-stakeholder and Adaptive Governance & Collaboration
International law & national sovereignty must be respected in the use of data. Additionally, participation of diverse stakeholders is necessary for inclusive approaches to AI governance.

5. Responsibility and Accountability
AI systems should be auditable and traceable. There should be oversight, impact assessment, audit and due diligence mechanisms in place to avoid conflicts with human rights norms and threats to environmental wellbeing.

6. Transparency and Explainability
The ethical deployment of AI systems depends on their transparency & explainability (T&E). The level of T&E should be appropriate to the context, as there may be tensions between T&E and other principles such as privacy, safety and security.

7. Human Oversight and Determination
Member States should ensure that AI systems do not displace ultimate human responsibility and accountability.

8. Sustainability
AI technologies should be assessed against their impacts on 'sustainability', understood as a set of constantly evolving goals including those set out in the UN's Sustainable Development Goals.

9. Awareness & Literacy
Public understanding of AI and data should be promoted through open & accessible education, civic engagement, digital skills & AI ethics training, media & information literacy.

10. Fairness and Non-Discrimination
AI actors should promote social justice, fairness, and non-discrimination while taking an inclusive approach to ensure AI's benefits are accessible to all.

Source: https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

Indigenous Peoples and the Media

UNESCO

"In a world increasingly influenced by media narratives, the representation of Indigenous Peoples in the media has far-reaching implications for their rights, cultural and linguistic preservation, economic empowerment, well-being and inclusion in society," states the introduction to a new publication from UNESCO. Based on extensive research, it concludes with 12 key recommendations for future action.

1. Ensuring rights, freedom of expression and access to media

Indigenous Peoples have the right to establish their own media and access non-Indigenous media platforms – radio, television, print and digital – without discrimination. Yet, this right is not fully realized, threatening pluralism, diversity, reconciliation and peaceful co-existence. Ensuring freedom of expression and access to information and media development is essential for well-being, education and full participation of Indigenous Peoples in society. All parties – duty bearers, rights enablers and rights holders – shall ensure compliance with UNDRIP Article 16 to uphold human rights and accountability.

2. Advancing media research and policy development

Media research generates essential knowledge for informed policy, decision-making and innovation. It should be evidence-based, grounded in a human rights-based approach and gender equality principles, and include Indigenous perspectives. Ethical, respectful and meaningful research practices, including Indigenous data sovereignty, intersectional gender analysis and disaggregated data, are crucial for effective data collection, planning and monitoring.

3. Strengthening legal and institutional frameworks

There is an urgent need to revise or develop new media laws and policies to support both Indigenous and non-Indigenous media. Equitable allocation of broadcast spectrum for Indigenous media, especially community radio, shall be mandated and supported by national legislation. The participation of Indigenous media professionals in policy and decision-making processes is imperative. Some countries provide legal frameworks for Indigenous media, yet global disparities remain significant and should be urgently addressed.

4. Promoting equitable editorial policies

Adopting equitable editorial policies ensures that Indigenous and non-Indigenous media serve diverse, often underrepresented audiences. Indigenous Peoples remain underrepresented in non- Indigenous media organizations and decision-making bodies, leading to marginalization and stereotyping. Ensuring independent and impartial Indigenous media and preventing interference – including from tribal councils – upholds freedom of expression and access to reliable information and media.

5. Ensuring fair representation in media content

A balanced portrayal of Indigenous Peoples in media shall be underpinned by a human rights-based and gender equality framework. Recognizing Indigenous Peoples as information sources fosters their accountability in the media. Editorial independence shall be strengthened to prevent harmful stereotypes and unfair representation in content, as well as the illicit trafficking of Indigenous cultural goods.

6. Improving working conditions for Indigenous media professionals

Indigenous media professionals, particularly women, shall have access to employment in the media industry under fair and non-discriminatory working conditions. All media professionals shall be ensured safety and non-violence, as well as equal access to management positions, training, capacity-building programmes, and career advancement opportunities. Non-Indigenous media professionals reporting on Indigenous affairs shall also be protected from threats and persecution.

7. Overcoming financial and structural constraints

Indigenous media organizations face challenges including limited human resources, infrastructure and technical support. High licensing fees, insufficient public funding and restrictive advertising regulations hinder their operational sustainability. In accordance with the UNDRIP, governments shall take measures to ensure that state-owned media adequately reflect Indigenous cultural diversity.

8. Addressing digital challenges and opportunities

Digital platforms and AI-based tools present both opportunities and challenges. While digital tools enhance audience engagement, their adoption is hindered by limited internet access, bias in AI based solutions, underrepresentation of Indigenous languages online, gender-based violence and digital literacy gaps. Investing in infrastructure and developing guidelines and digital tools helps to bridge these gaps and support Indigenous media in the digital space.

9. Strengthening media as a platform for public discourse

Media foster public discourse and participation among Indigenous Peoples. Indigenous media serve as platforms for sharing experiences, mobilizing action and shaping narratives. Ensuring Indigenous Peoples' participation in non-Indigenous media programming and content production contribute to inclusive dialogue in society.

10. Promoting Indigenous languages in media

Language profoundly shapes how information is perceived and conveyed. Ensuring access to media content in Indigenous languages is vital for cultural and linguistic preservation, education and broader social inclusion. Developing language tools and resources will strengthen Indigenous language use in media, education, science and technology. Public service and community media shall be supported in fulfilling this mandate.

11. Integrating gender, disability and youth approaches, and crisis preparedness in media

Intersectional approaches addressing gender equality, disability inclusion and youth participation in media shall be prioritized. Efforts should be taken to ensure inclusive coverage of all social groups, support Indigenous women in the media, promote safety of media professionals and address gender-based and disability-related discrimination in media. Developing content and programming relevant to young people would encourage their participation and public engagement. Furthermore, emergency and crisis preparedness for Indigenous media are also essential. Structural barriers shall be dismantled to enable Indigenous media to implement sustainable and impactful initiatives in these areas.

12. Strengthening partnerships and professional networks

New multistakeholder partnerships and professional associations are needed to enhance collaboration between Indigenous and non-Indigenous media. Ensuring that Indigenous media workers, particularly women, have mean-

ingful participation in global journalism will legitimize their role in the media landscape. Knowledge transfer, ethical guidelines, mentorship programmes and advocacy networks should be promoted to support media sustainability and fair working conditions, especially for Indigenous women in the media industry.

Source: UNESCO.

ON THE SCREEN

Oberhausen (Germany) 2025

At the International Short Film Festival Oberhausen 2025, the Ecumenical Jury of the International Competition awarded its Prize to *Dear Leo Sokolsky* directed by Weronika Szyma (Poland, 2024).

Motivation: By employing minimalist animation with live action resurrected from family archives this film transports the viewer through Ansbach and the inner space of one woman's journey to find her great-grandfather who went through the labour camps of the Second World War. Making the past present, and bringing together history, documentary and critical personal reflection, *Dear Leo Sokolsky* is a wondrous cinematic diary that gives us the rare opportunity to pull back the curtain and peer into the depths of a human soul.

In addition, the jury awarded a Commendation to *Nocturne* directed by Sol Muñoz and Ana Apontes (Argentina, 2025).

Motivation: For inviting us into the atmospheric care-free world of childhood, a film which captures two sisters walking through the night as their father works as a security guard for an affluent neighbourhood. Contrasting the loneliness of locked up condominiums with the freedom of sisters discovering an open and wide world, *Nocturne* is a delicate work of social critique and an existential reminder of what it really means to live life embraced by love.

The Ecumenical Jury of the Children's and Youth Film Competition awarded its Prize to *Autokar* directed by Sylwia Szkiłądź (France, Belgium, 2025).

Motivation: Agata's journey from Poland to Belgium offers a universal perspective on leaving one's homeland and explores identity and tradition with sensitivity. In times of political and social polarisation. The film confronts the fear of otherness and provides a valuable insight into human connection that resonates across all age groups.

In addition, the jury awarded a Commendation to *Happy Snaps* directed by Tyro Heath (United Kingdom, 2024).

Motivation: The film offers a sensitive portrayal of a friendship that must face the challenges of separation and loss over the course of the film. It provides an important and inclusive perspective on relationships shaped by the human need to hold on and the emotional process of coping with change.

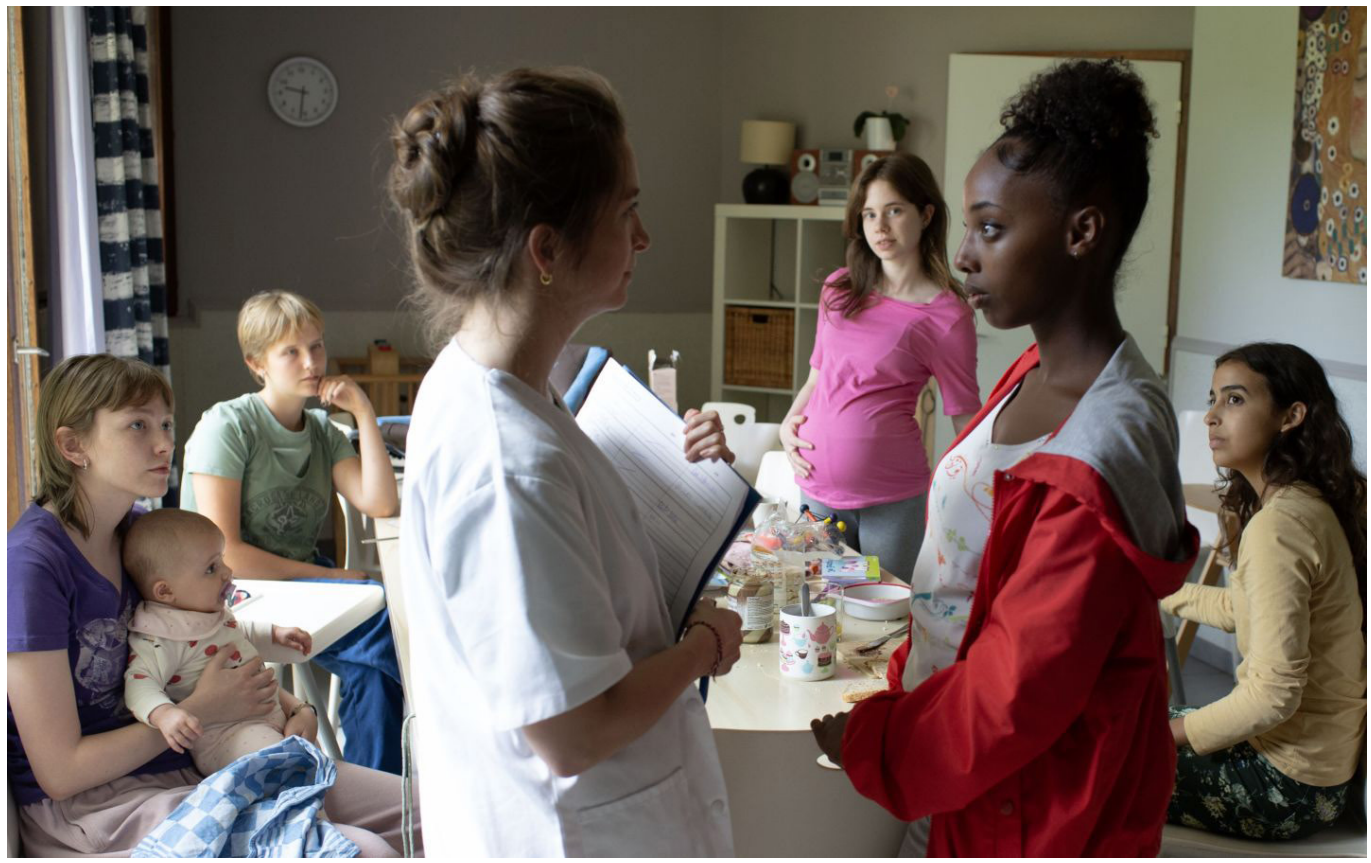
Cannes (France) 2025

The 2025 Ecumenical Jury awarded its Prize to *Jeunes Mères* (still on following page) directed by Jean-Pierre and Luc Dardenne.

Jessica, Perla, Julie, Ariane and Naïma are staying at a maternity home to help them with their lives as young mothers. Five teenagers with the hope of achieving a better life for themselves and their child.

Motivation: The Ecumenical jury gave its prize to a film about the troubles of teenage mothers in a dedicated motherhouse. It finds ethic not on grand gestures, but in quiet acts of care. It is a smoothly told story in the best tradition of its authors who one again are able to add new elements to their refined style.

The film explores the first and utmost im-



portant relationship of every human life, which is motherhood. It touches a profound truth : love can endure even when family – the basic social structure – fails, when circumstances are unfair, when youth is burdened with adult responsibilities. The film proved than even small yet persistent acts of love and care of individuals and institutions can heal the deepest wounds.

Members of the 2025 Jury: Lukas Jirsa (Czech Republic); Arielle Domon (France); Anne-Cécile Antoni (France); Thomas. D. Fischer (Germany); Roland Wincher (Germany); Milja Radovic (UK).

Zlín (Czech Republic) 2025

The Ecumenical Jury, appointed by INTERFILM and SIGNIS, at the 65th International Film Festival for Children and Youth in Zlín (29 May – 4 June, 2025), awarded its Prize to *Kevlar Soul* directed by Maria Eriksson-Hecht (Sweden, 2025).

Motivation: A film about two brothers struggling to find their way in life, dealing with problems of violence and alcoholism in the family with the older brother taking on parental responsibility. It tells the story with consistent visual language with pictures that speak, even when the characters are silent. The film gives an example of compassion for people in need rather than judgment, letting the audience go through the difficult journey together with the main characters. *Kevlar Soul* comes alive with the authentic performance of the actors.

In addition, the jury awarded a Commendation to the film *Nawi: Dear Future Me* directed by Kevin Schmutzler, Toby Schmutzler, Apuu Morine, Valentine Chelluget (Germany, Kenya, 2024).

Eight camels, sixty sheep, one hundred goats – that is how much Nawi is worth. The family is in debt, so she has to marry, even though she scored the highest marks in the final exams and wants to go on with school. Courageously the young Kenyan girl tries everything to avoid having to marry a much older man. Even though Nawi can't escape her destiny in the end, the film leaves the audience with hope for a better future. The film's narrative is both playful and clear, using voiceover with visuals to underline the world of its protagonists. *Nawi* helps us become aware of a problem which affects a huge number of girls around the world, who just want to live their lives and follow their dreams. ▪