Artificial Intelligence and Digital Disinformation: Ethical Challenges for Media Literacy and Journalism

MARINA TUNEVA

DOI: 10.64370/PGYX8412

Abstract

This paper explores the dual role of artificial intelligence (AI) in addressing and exacerbating digital disinformation, situating the discussion within the broader context of media ethics and self-regulation. By examining global and regional practices, the research identifies how AI technologies contribute to the rapid dissemination of false information, amplify algorithmic biases and create ethical dilemmas for media. These challenges are particularly pronounced in emerging democracies, where support for ethical journalism remains limited. The study emphasizes the urgent need to adapt self-regulatory frameworks to address the unique risks posed by AI-driven content creation and dissemination. Key recommendations include integrating AI accountability within the media ethics framework, fostering transparency in algorithmic decision-making and providing journalists with targeted training to navigate the ethical implications of AI tools. Comparative findings suggest that aligning these frameworks with international standards could significantly mitigate the impact of disinformation while strengthening trust in the media. This research contributes to the growing discourse on media and information literacy by bridging the gap between theoretical insights and practical applications. It underscores the interdisciplinary nature of addressing digital disinformation, combining technological, sociological and policy perspectives to propose actionable solutions for media sustainability. The paper is particularly relevant for understanding the intersection of AI, critical thinking and ethical journalism in the digital age.

Keywords: Artificial intelligence, digital disinformation, media ethics, media literacy.

1. Introduction

The rapid advancement of artificial intelligence (AI) technologies has significantly reshaped the media landscape, introducing both promising possibilities and complex ethical dilemmas. With society's growing dependence on digital information, advanced AI systems have emerged as powerful tools capable of generating highly convincing forms of disinformation—ranging from deepfakes and synthetic audio to fabricated narratives—at an unprecedented scale. These phenomena have serious implications for journalistic integrity, public trust, and democratic resilience (Westerlund, 2019; Germani et al., 2024).

Digital disinformation, defined as the deliberate dissemination of false or misleading content intended to manipulate public perception, is significantly amplified by algorithmic infrastructures that prioritize user engagement rather than factual accuracy. As research from Northwestern University (2023) reveals, social media algorithms tend to elevate emotionally charged or bias-confirming content, regardless of its factual reliability, thereby accelerating the viral spread of misinformation. This phenomenon is not merely a technical flaw but also presents profound epistemic and ethical challenges. The unchecked proliferation of AI-generated disinformation further threatens journalism's foundational role as a guardian of truth, transparency, and accountability. These concerns are particularly acute in transitional democracies and fragile information ecosystems, where media literacy frameworks are underdeveloped and self-regulatory institutions often lack the capacity to respond effectively (Tuneva, 2023).

Yet, paradoxically, AI also holds potential as a counter-disinformation tool. It offers scalable solutions for automated fact-checking, semantic analysis, and pattern recognition that can support journalists and media organizations in identifying and responding to falsehoods. Still, the application of these tools is not neutral; it raises concerns over algorithmic transparency, the epistemic boundaries of automated systems, and potential infringements on rights to free expression (Lange & Lechterman, 2021; Bontridder & Poullet, 2021).

This paper explores the dual role of AI in the production and mitigation of digital disinformation, with a focus on its ethical implications for media literacy and journalism. Drawing on both empirical findings and normative analysis, it argues for the integration of AI literacy and ethical frameworks into media self-regulation, professional standards, and public policy. Furthermore, the study draws from global and regional case studies—including comparative insights from press councils across Europe to propose a set of actionable recommendations that reinforce democratic integrity and media credibility.

2. 2. Theoretical and Conceptual Framework

Пресекот The intersection of artificial intelligence (AI), digital disinformation, and media ethics necessitates an interdisciplinary framework—one that draws from media studies, information science, communication ethics, and critical algorithm studies. At the heart of this analysis is the recognition that disinformation is not solely a technological issue, but a socio-political and epistemological phenomenon—deeply embedded in how information is produced, circulated, and legitimized within digital environments.

2.1 Artificial Intelligence and Information Disorder

AI technologies increasingly shape the architecture of information environments. Machine learning models, particularly large language models (LLMs) and Generative Adversarial Networks (GANs), exhibit powerful capabilities to generate, personalize, and disseminate synthetic content that closely mimics human expression—raising significant concerns about the authenticity of digital information (Shoaib et al., 2023). These systems, trained on vast datasets, are also vulnerable to the flaws and biases of their human developers; such biases can infiltrate the data and algorithms, influencing how content is produced, ranked, and disseminated (Kertysova, 2018).

The concept of information disorder, as developed by Wardle and Derakhshan (2017), provides a foundational framework for understanding the different dimensions of false or harmful information in the digital age. Their typology distinguishes between disinformation (false content created with the intent to cause harm), misinformation (false content shared without harmful intent), and malinformation (genuine information used maliciously to inflict harm). Artificial intelligence is implicated in all three categories—by facilitating the automated creation and dissemination of disinformation, enabling algorithm-driven amplification of misinformation, and supporting data-mining techniques used in targeted malinformation campaigns. The creation, production, and distribution phases of these information types are increasingly complex and opaque due to AI's scale and speed, making journalistic verification and accountability significantly more difficult (Wardle & Derakhshan, 2017). Moreover, the speed, scale, and opacity of AI-driven information systems make it increasingly difficult to verify the origin, intent, and credibility of digital content—posing significant challenges for journalistic fact-checking and accountability (Buchanan et al., 2021).

2.2 Media Ethics and the Challenges of AI

Media ethics provides the normative foundation for journalistic responsibility, rooted in principles such as truthfulness, fairness, accountability, and the minimization of harm. However, these principles are increasingly strained in the age of algorithmic mediation and automated content generation.

The rise of AI-generated journalism challenges conventional notions of authorship and attribution. When large language models or automated systems contribute to the creation of news content, the boundary between human and machine authorship becomes blurred. As Bontridder and Poullet (2021) point out, AI tools used in content production can obscure the line of editorial responsibility, raising concerns about transparency and accountability. Similarly, Lange and Lechterman (2021) emphasize that the absence of clear disclosure regarding AI involvement in the journalistic process can create ethical ambiguity about content provenance, leading to potential erosion of public trust.

In response to these concerns, new models of attribution and transparency are being proposed. Some researchers advocate for the implementation of labeling systems—such as the Artificial Intelligence Attribution (AIA) framework—that mark the extent of AI involvement in content creation (Bertolini et al., 2023). These frameworks suggest visual cues or metadata indicators that distinguish between human-authored, AI-assisted, and fully AI-generated content, offering a clearer ethical structure for digital journalism.

Moreover, the speed and scale at which algorithmically curated content is disseminated—often driven by engagement-based metrics—can lead to the prioritization of sensational or emotionally charged material over fact-based reporting. This dynamic further complicates the ethical obligations of journalists to provide accurate, balanced information in a digitally mediated environment.

Overall, the integration of AI into media production demands a re-evaluation of ethical frameworks, with greater emphasis on transparency, attribution, and editorial oversight in the digital age.

Moreover, algorithmic curation on platforms such as Facebook, YouTube, and TikTok is governed by engagement-based metrics rather than editorial standards. This prioritization of emotional or polarizing content over balanced reporting undermines the gatekeeping function traditionally performed by journalists. Gerbaudo (2023) observes that TikTok's architecture constructs "clustered publics" by algorithmically grouping users around shared preferences, thereby deepening filter bubbles and reducing content diversity.

The spread of synthetic media further erodes epistemic trust—defined as the public's confidence in the reliability and authenticity of information sources. As Twomey et al.

(2023) demonstrate in their analysis of social media discourse during the Russo-Ukrainian war, deepfake content has the capacity to sow confusion and distrust, particularly when audiences are unable to distinguish manipulated content from reality. This epistemic crisis threatens democratic processes based on informed deliberation, consent, and accountability.

In this context, ethical journalism must not only continue to verify content rigorously but also increase transparency around editorial practices and the involvement of AI technologies. Paik (2023) calls for a reconceptualization of journalistic ethics to reflect the algorithmic conditions of contemporary media. He proposes a model of distributed accountability—one that acknowledges the collective ethical responsibility of journalists, technologists, and platform designers in upholding truth and public trust.

2.3 Media Literacy and AI Literacy as Democratic Imperatives

In the digital age, Media and Information Literacy (MIL) frameworks must evolve to encompass Artificial Intelligence (AI) literacy—the capacity to comprehend, critically evaluate, and responsibly utilize AI technologies. This includes recognizing algorithmic biases, understanding the role of data in shaping AI outputs, and developing critical perspectives on the authenticity of digital content. Integrating AI literacy into existing MIL programs is essential for equipping individuals with the skills to navigate and critically assess information in AI-driven environments (Ndungu, 2023).

The importance of MIL is particularly pronounced in emerging democracies and post-conflict societies, where media systems are often fragile, and public trust may be diminished. As Tuneva (2024) highlights, the absence of robust media education and weak self-regulatory institutions can render societies especially susceptible to disinformation campaigns, both foreign and domestic. Therefore, embedding AI and data ethics into MIL curricula is not merely a pedagogical concern but serves as a vital safeguard for democratic integrity.

Incorporating AI literacy into MIL empowers citizens to critically engage with AI-driven content and fosters resilience against misinformation and manipulation. This integration is crucial for maintaining informed public discourse and upholding democratic processes in an era where AI technologies increasingly influence information dissemination.

2.4 Self-Regulation and the Ethics of AI Use in Journalism

Traditional self-regulatory mechanisms—such as codes of ethics, editorial guidelines, and press councils—must evolve to meet the emerging ethical challenges posed by AI integration in journalism. Comparative analysis shows that a growing number of ethical bodies in Europe, including those in Finland, Belgium, and Armenia, have begun updating their frameworks to address issues such as algorithmic transparency, automated news production, and the ethical implications of user-generated content (Tuneva, 2024). These emerging standards emphasize the disclosure of AI involvement in content creation, the necessity of human oversight in editorial processes, and safeguards against algorithmic bias and discrimination.

While recent updates to ethical codes and self-regulatory frameworks represent meaningful progress, many remain voluntary in nature and depend on the strength and credibility of the media that uphold them. In contexts where media ecosystems face political pressure, economic vulnerability, or fragmented professional standards, the practical implementation of these principles can be inconsistent. At the same time, the rapid development and deployment of AI technologies—often shaped by global tech actors—pose new challenges that require self-regulatory bodies to be more agile, better resourced, and technologically informed. Strengthening the legitimacy and adaptability of self-regulation is therefore not only possible, but essential to ensuring ethical standards evolve alongside technological change.

3. Methodology and Analytical Approach

This study adopts a qualitative, interdisciplinary research design aimed at exploring the ethical implications of artificial intelligence (AI) in journalism and media literacy, with a particular focus on the phenomena of digital disinformation. The approach reflects a commitment to both analytical rigor and contextual relevance, recognizing AI's complex dual function as both a driver of disinformation and a potential instrument for its mitigation. By combining comparative analysis, semi-structured expert interviews, and desk-based research, the study captures a multi-layered view of how media systems respond to AI-driven ethical challenges within different regulatory, professional, and socio-political contexts.

The core of the analysis is a comparative review of ethical and self-regulatory frameworks across more than 20 countries, including Finland, Belgium, Germany, Armenia, and North Macedonia.

This review examines how diverse press councils address several pressing issues: algorithmic transparency and accountability, the use and labeling of AI-generated content, institutional strategies for countering digital disinformation, editorial

responsibility in algorithmically curated environments, and the protection of vulnerable or marginalized groups. National and international documents—spanning codes of ethics, internal newsroom guidelines, and formal recommendations of press councils - were analyzed.

The study incorporates qualitative data from 12 semi-structured interviews with representatives of national ethics councils and self-regulatory bodies affiliated with the Alliance of Independent Press Councils of Europe (AIPCE). These interviews were designed to probe the practical challenges related to AI adoption in the media sector.

Recent academic studies, institutional reports, and policy briefs published by organizations such as UNESCO, the Council of Europe, the European Data Protection Board, and international media development bodies were also analyzed.

4. Results and Discussion: The Dual Role of AI in the Age of Disinformation

The findings of this study reveal the ambivalent nature of artificial intelligence (AI) in today's media landscape. AI simultaneously empowers malicious actors with tools to scale and personalize disinformation, while also offering journalists and platforms advanced methods for detecting and countering falsehoods. This duality presents both ethical risks and strategic opportunities—especially in fragile media environments where regulatory safeguards and digital literacy are still evolving.

4.1 AI as a Catalyst for Digital Disinformation

Artificial intelligence (AI) systems have significantly amplified the production and dissemination of disinformation. These technologies can fabricate entire news articles, synthetic images, or deepfake videos that are nearly indistinguishable from authentic content. Westerlund (2019) discusses the emergence of deepfake technology and its potential to create convincing fake media content, posing significant challenges for information integrity. Similarly, research by Vinay et al. (2024) demonstrates how large language models can be manipulated through prompt engineering to generate persuasive disinformation, highlighting the ease with which AI can be exploited to produce misleading narratives.

During the COVID-19 pandemic, AI-generated health misinformation about vaccines, false cures, and infection rates proliferated across social media platforms, eroding public trust and undermining health policy efforts. Menz et al. (2023) demonstrated that within 65 minutes, a publicly available large-language model generated 102 distinct blog articles containing over 17,000 words of disinformation related to vaccines and vaping.

In political contexts, AI tools have been employed to create microtargeted propaganda based on psychological profiling, posing significant risks to democratic integrity. Research indicates that AI-enhanced microtargeting allows political campaigns to deliver personalized and persuasive messages to specific voter segments, subtly influencing their behavior—a tactic that leverages psychological vulnerabilities to shape electoral outcomes (Ahmad et al., 2024). These practices are increasingly documented as tools for manipulating public opinion and destabilizing trust in democratic institutions.

These developments highlight a shift from traditional disinformation tactics to what Buchanan et al. (2021) term "automated persuasion at scale." When combined with algorithmic recommendation systems that prioritize engagement, this creates a volatile feedback loop where falsehoods are not only produced more rapidly but also disseminated more effectively than truths. Such systems, driven by engagement metrics, tend to amplify sensationalist content, thereby exacerbating the spread of disinformation.

4.2 Ethical Dilemmas and Algorithmic Opacity

The integration of artificial intelligence into media environments presents a range of ethical challenges that transcend the traditional dichotomy of truth and falsehood. Central among these is the issue of algorithmic opacity. Many AI-driven systems operate as "black boxes," offering limited insight into how editorial decisions are made—such as which content is prioritized, demoted, or removed. This lack of transparency undermines accountability and makes it difficult for both journalists and audiences to evaluate the integrity of information flows.

A related concern is the bias embedded in AI training data. As Kertysova (2018) emphasizes, algorithms often inherit the social and cultural biases present in the datasets on which they are trained, leading to the perpetuation of stereotypes and the marginalization of vulnerable communities. When left unexamined, these biases risk entrenching inequality and misrepresentation in both automated content generation and algorithmic curation. Equally problematic is the absence of clear disclosure regarding AI-generated content, which can erode informed consent and blur the boundaries of editorial responsibility.

Insights from expert interviews conducted with representatives of media ethics councils across Europe further reinforce these concerns. While some countries—such as Finland, Belgium, and France—have begun developing ethical guidelines that address the use of AI in journalism, most self-regulatory bodies still operate without binding standards or adequate resources to monitor compliance. As one interviewee aptly remarked, "We are updating the rulebook while the game is already being played—with more powerful

players and less control." This comment captures the institutional urgency and asymmetry that self-regulatory bodies face in responding to the rapid evolution of AI technologies in newsrooms and media platforms.

4.3 AI as a Tool for Detection and Defense

Despite the substantial ethical risks posed by AI, these technologies also present significant opportunities for enhancing the resilience of journalism and mitigating the spread of disinformation. In recent years, fact-checking organizations and newsrooms have increasingly adopted AI-driven techniques, particularly those based on natural language processing (NLP) and machine learning, to detect and analyze misleading content with greater efficiency and scale.

Shaik (2023) highlights how machine learning models can identify recurring patterns in deceptive narratives, including emotional manipulation and sentiment polarization—features commonly associated with disinformation campaigns. In parallel, Santos (2023) demonstrates the ability of AI systems to cross-reference claims in real time with trusted databases, thereby accelerating the fact-checking process and improving accuracy. Additionally, Buchanan et al. (2021) explore how AI-powered monitoring systems can detect coordinated disinformation efforts across linguistic and geographic boundaries, offering a transnational view of influence operations and networked propaganda.

Such AI-based tools are particularly valuable during sensitive periods, such as election cycles or public health emergencies, when the speed and volume of content production can overwhelm human verification efforts. However, these technological advancements must not replace human editorial judgment. As emphasized by several interviewees in this study, human oversight remains essential to ensure that automated systems do not inadvertently reinforce dominant ideologies, overlook contextual nuance, or suppress legitimate speech. Maintaining this balance is vital for preserving journalistic standards and upholding the principles of free expression in AI-assisted media environments.

4.4 Emerging Gaps in Media Ethics and Self-Regulation

Comparative analysis conducted for this study reveals a growing disconnect between the pace of technological innovation and the slower evolution of ethical media standards. Most existing codes of ethics were developed prior to the rise of algorithmic content generation and therefore do not adequately reflect the challenges introduced by artificial intelligence. This misalignment has resulted in a number of critical normative gaps that remain insufficiently addressed across national and institutional contexts.

One major area of concern is the lack of explicit guidance on algorithmic transparency. Ethical frameworks rarely provide clarity on whether, or how, AI-generated content should be labeled for the public, raising questions about editorial accountability and

informed consent. Equally problematic are the limited provisions related to data privacy and user consent, particularly in the context of personalized content delivery and behavioral targeting—practices that increasingly rely on opaque algorithmic profiling. Additionally, there is a notable absence of strong protections for vulnerable and marginalized groups, who are often the primary targets of digitally amplified hate speech, stereotyping, or manipulation.

Nevertheless, a few press councils have begun to proactively adapt to these emerging issues. Examples from Finland, Catalonia, and Belgium show how updated codes of ethics can include AI-specific clauses that address transparency requirements, define standards for editorial responsibility when using automated tools, and promote user education as a key component of media literacy. These forward-looking models offer important insights for other self-regulatory bodies seeking to align journalistic ethics with the realities of digital and algorithmic media environments.

4.5 Regional Vulnerabilities: The Case of Emerging Democracies

In transitional media systems such as those found in the Western Balkans, the ethical challenges associated with AI are magnified by long-standing structural and institutional weaknesses. These environments are often characterized by politicized media ownership, which fosters biased reporting and undermines editorial independence. Furthermore, levels of media and AI literacy among the general public remain low, limiting citizens' ability to critically engage with algorithmically generated content or identify manipulative narratives.

Self-regulatory bodies in these regions frequently operate with limited financial and human resources, and their mandates often lack the authority or enforcement capacity needed to effectively oversee compliance with ethical standards. Additionally, institutional collaboration—particularly between media regulators, civil society organizations, and technology platforms—tends to be fragmented or underdeveloped, further weakening the media ecosystem's resilience to AI-driven disinformation.

As documented in the practice of media self-regulatory bodies (Tuneva, 2024), ethical violations are both systemic and widespread. These include the use of manipulative headlines, the failure to mark sponsored or political content transparently, and persistent patterns of misrepresentation or exclusion of vulnerable social groups. In such fragile contexts, the unchecked adoption of AI technologies—without adequate ethical safeguards, journalist training, or public education—risks exacerbating existing social inequalities and reinforcing dynamics of information disorder.

5. Conclusion and Recommendations

The evolving digital information landscape—shaped profoundly by the integration of artificial intelligence (AI), presents both critical challenges and transformative possibilities for journalism, media ethics, and public trust. As demonstrated throughout this study, AI technologies, particularly those utilizing deep learning and natural language generation, hold a dual role. On one hand, they enable the mass and rapid production of disinformation at a scale previously unimaginable. On the other, they offer advanced tools for detecting, verifying, and mitigating the very same falsehoods they help propagate. Navigating this paradox requires a coherent ethical response grounded in transparency, accountability, and critical engagement.

The risks associated with AI are especially pronounced in media systems that are institutionally fragile, as is often the case in emerging democracies. Limited AI and media literacy, weak regulatory infrastructures, and politicized media environments form an ecosystem in which algorithmically amplified disinformation can thrive. In such contexts, traditional self-regulatory mechanisms—though essential for defending press freedom and professional standards—are frequently outdated and ill-equipped to respond to the ethical complexities introduced by algorithmic technologies.

In light of these findings, this paper proposes a multi-level strategy to promote ethical resilience in journalism and fortify democratic information ecosystems. This strategy rests on four interlinked pillars: strengthening ethical and self-regulatory frameworks; promoting AI literacy within both media and the public sphere; enhancing intersectoral collaboration; and reimagining journalism itself to meet the epistemic demands of the AI era.

First, ethical and self-regulatory frameworks must be updated to address AI-specific risks, including issues of algorithmic transparency, synthetic content disclosure, and editorial responsibility in the deployment of automated tools. Press councils and media regulatory bodies should adopt clear, enforceable guidelines that are informed by international best practices, such as those already emerging in Finland, Belgium, and Germany. Furthermore, alignment with global initiatives—such as UNESCO's Recommendation on the Ethics of Artificial Intelligence and the European Commission's AI Act—can help ensure consistency and legitimacy across national contexts.

Second, AI literacy must become an integral component of media and information literacy (MIL) efforts. Educational programs should prepare citizens to critically engage with algorithmically curated and synthetic content, while journalists and editors should receive targeted training on the ethical implications of AI-based tools, including those used for content generation, audience targeting, and automated fact-checking. Transparency by technology companies is also essential; platforms should be required

to explain the logic of their algorithms and provide access to data for independent scrutiny and verification.

Third, sustained collaboration between key stakeholders is essential. Ethical benchmarks and safeguards must be co-developed by journalists, academics, civil society organizations, and technology providers, ensuring that ethical frameworks are both contextually grounded and technically informed. Support for cross-border research and regional knowledge exchange—particularly between countries with strong ethical practices and those in transition—can foster innovation and harmonize standards. Donors and international partners should prioritize capacity building for media regulators and press councils in under-resourced regions.

Finally, journalism must evolve not only technologically but ethically. In an age where authenticity is increasingly questioned and synthetic narratives are indistinguishable from reality, the profession must embrace radical transparency regarding the use of AI in editorial workflows. Credibility will depend not merely on verification but on openness about the tools and logic used in content production. Ethical journalism must also confront the inclusivity gap by promoting diverse representation in AI training datasets and editorial perspectives.

Ultimately, the integrity of journalism in the digital era will depend not only on adaptation to emerging technologies but on the cultivation of ethical resilience—the capacity of individuals, institutions, and societies to uphold truth, accountability, and human dignity in the face of technological disruption and epistemic uncertainty.

6. Limitations and Directions for Future Research

While this study provides a robust exploration of the ethical challenges surrounding AI and disinformation in journalism, several limitations should be acknowledged. First, the research is predominantly qualitative and relies on expert interviews and comparative analysis. While these methods yield rich insights, they may lack the generalizability of quantitative approaches. Second, the evolving nature of AI technologies means that some findings may soon require revision, particularly as regulatory frameworks and AI applications continue to advance rapidly.

Future research should explore the quantifiable impacts of AI literacy programs on user resilience to disinformation, as well as comparative analyses of enforcement mechanisms in self-regulatory bodies. Further interdisciplinary studies involving technologists, ethicists, and policymakers are essential to address the complex interplay of automation, accountability, and media credibility in the digital age.

Literature

Ahmad, R., Nguyen, Q.H., Choi, Y., & Lee, K. (2024) 'AI threats to politics, elections, and democracy: A blockchain perspective', Information, 2(4), pp. 386–400. Available at: https://www.mdpi.com/2813-5288/2/4/20 (Accessed: 11 March 2025).

Bontridder, N. and Poullet, Y. (2021) 'The role of artificial intelligence in disinformation', Data & Policy, 3. Available at: https://www.cambridge.org/ (Accessed: 8 February 2025).

Buchanan, B., Lohn, A.J., Sedova, K. and Musser, M., 2021. Truth, lies, and automation: How language models could change disinformation. Center for Security and Emerging Technology. Available at: https://cset.georgetown.edu/publication/truth-lies-and-automation/ [Accessed: 6 March 2025].

Gerbaudo, P. (2023) TikTok and the algorithmic transformation of social media publics: From social networks to social interest clusters. New Media & Society. Available at: https://doi.org/10.1177/14614448241304106 (Accessed: 3 March 2025).

Kertysova, K. (2018) Artificial Intelligence and Disinformation: How AI Changes the Disinformation Landscape. Security and Human Rights Monitor. Available at: https://www.shrmonitor.org/assets/uploads/2019/11/SHRM-Kertysova.pdf (Accessed: 9 February 2025).

Lange, B., & Lechterman, T. (2021). Combating disinformation with AI: Epistemic and ethical challenges. 2021 IEEE International Symposium on Technology and Society (ISTAS). Available at https://philpapers.org/archive/LANCDW-2.pdf (Accessed: 22 February 2025).

Menz, B., Modi, N., Sorich, M., & Hopkins, A. (2024). Health disinformation use case highlighting the urgent need for artificial intelligence vigilance. JAMA Internal Medicine, 184 (1), 92. https://doi.org/10.1001/jamainternmed.2023.5947 (Accessed: 17 March 2025)

Ndungu, M. W. (2023). Integrating basic artificial intelligence literacy into media and information literacy programmes. Journal of Information Literacy, 17(2). Available at: https://journals.cilip.org.uk/jil/article/view/690 (Accessed: 1 April 2025).

Northwestern Now (2023) Social media algorithms exploit how humans learn from their peers. Available at: https://news.northwestern.edu/stories/2023/08/social-media-algorithms-exploit-how-humans-learn-from-their-peers/ (Accessed: 4 March 2025).

Paik, J. (2023) Journalism ethics for the algorithmic era. Digital Journalism, 11(7), pp.893–913. Available at: https://doi.org/10.1080/21670811.2023.2200195 (Accessed: 18 March 2025).

Santos, F. (2023) 'Artificial intelligence in automated detection of disinformation: A thematic analysis', Journalism and Media, 4(2), pp. 679–687. Available at: https://doi.org/10.3390/journalmedia4020043 (Accessed: 1 March 2025).

Shaik, S. (2023). A review and analysis on fake news detection based on artificial intelligence and data science. Journal of Propulsion Technology. Available at: https://doi.org/10.52783/tjjpt.v44.i3.2107 (Accessed: 1 April 2025).

Shoaib, M. R., Wang, Z., Taleby Ahvanooey, M., & Zhao, J. (2023). Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models. arXiv.Org, abs/2311.17394. https://doi.org/10.48550/arxiv.2311.17394 (Accessed 15 March 2025)

Tuneva, M. (2024) Од глобални искуства до локални стандарди: Етички предизвици и решенија во новинарството [From Global Experiences to Local Standards: Ethical Challenges and Solutions in Journalism]. Скопје: Македонски институт за медиуми (МИМ). Available at: https://mim.org.mk/ (Accessed: 11 February 2025).

Twomey, J., Holmes, D., Hughes, N. and Redmond, S. (2023) Do deepfake videos undermine our epistemic trust? A thematic analysis of Twitter conversations during the Russo-Ukrainian war. PLOS ONE, 18(10), e0291668. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0291668 (Accessed: 19 March 2025).

Vinay, R., Spitale, G., Biller-Andorno, N. and Germani, F. (2024) Emotional manipulation through prompt engineering amplifies disinformation generation in AI large language models. arXiv preprint arXiv:2403.03550. Available at: https://arxiv.org/abs/2403.03550 (Accessed: 11 March 2025).

Wardle, C. and Derakhshan, H. (2017) Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe report DGI(2017)09. [online] Strasbourg: Council of Europe. Available at: https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html (Accessed: 1 March 2025).

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review. Available at: https://timreview.ca/article/1282 (Accessed: 17 March 2025).

About the author:

Marina Tuneva holds a PhD in Cultural Studies, with a dissertation titled "The Role of Communication Strategies in Multicultural Societies: A Case Study of the Ohrid Framework Agreement." She earned her Master's degree at the Institute for Security, Defense, and Peace of the Faculty of Philosophy in Skopje, and completed postgraduate studies in Diplomacy at the University of Malta. Tuneva graduated in Interdisciplinary Journalism Studies at Ss. Cyril and Methodius University in Skopje. In 2021, she obtained a certificate for the successful completion of the Distinguished Humphrey Fellowship Program in Media and Information from the Bureau of Educational and Cultural Affairs of the U.S. Department of State.

As an expert in public relations, diversity reporting, intercultural communications, and journalism ethics, she is actively engaged in numerous consulting and research projects in these areas, both in the country and internationally.

She is the author of several publications, including: "International Actors, the Conflict in Macedonia in 2001, and Peace Journalism," "Public Relations and Interculturalism: Communication and Information on the Ohrid Framework Agreement," "Guidelines for Inclusive Reporting on the Pandemic," "Media, Citizens, and Intercultural Communications," "Handbook for Reporting on Diversity," "Introduction to Public Relations," and others.