Envisioning the Future of Communication

Vol 2, No 1 (2025)

Envisioning the Future of Communication - Conference Proceedings vol. 2

Artificial intelligence in journalism

Theodora Saridou

doi: 10.12681/efoc.7909

Artificial intelligence in journalism: Practices, challenges and regulatory framework

Theodora Saridou*

Abstract

In recent years, computational advancements combined with social and economic shifts have facilitated the integration of artificial intelligence (AI) technologies in a remarkably wide range of fields. The field of journalism is not unaffected by the explosive growth of AI tools and systems. News organizations worldwide are using AI techniques to carry out tasks throughout the news production process. Automated approaches are applied, for example, in journalistic research, editing or monitoring of audience participation. However, reshaping the way of working is challenging in the light of the risks involved and the serious issues that arise. This paper explores how media organizations utilize AI, especially for managing user-generated content, and delves into pressing issues such as data protection, fairness, and human rights. It also assesses the emerging regulatory landscape surrounding AI in journalism. Despite the benefits to news production, navigating complexities requires careful consideration of the associated challenges and ethical implications.

Keywords: Artificial intelligence, participatory journalism, moderation, ethics.

Introduction

The intersection of audience participation and journalism has long been a dynamic space, where the influx of user-generated content (UGC) has transformed traditional journalistic practices, such as gatekeeping and framing. Citizens can participate in journalism (Engelke, 2019; Krumsvik, 2018; Loosen et al., 2022) and contribute to the news production by professional journalists on a mainstream platform (Abott, 2017). At the same time, content generated by users outside the media organization's platform is acquired and used by professional media (Saridou & Veglis, 2021).

This evolution, while enriching the news landscape with diverse perspectives, also presents a host of complex legal and ethical dilemmas. As Saridou et al. (2019) noted, the coexistence of amateur and professional content often poses threats on accountability and authenticity, challenging the very essence of journalistic integrity. To mitigate and manage these issues, media organizations employ moderation techniques enabling them to organize, monitor, and authenticate content (Boberg et al., 2018; Wolfgang, 2016). However, manual moderation is a time-intensive duty for professionals, who must handle substantial content volumes amidst time

* Ph.D, Academic researcher, Aristotle University of Thessaloniki, Adjunct lecturer, University of Western Macedonia, saridout@jour.auth.gr.

constraints; a process which requires significant human and financial investment (Wang, 2020). Consequently, many media entities adopt automated methods and artificial intelligence (AI) techniques to figure this task out or integrate semi-automated approaches incorporating, for example, machine learning (ML) into non-automatic comment moderation processes (Risch & Krestel, 2018). Additionally, news media outlets have increasingly embraced ML applications in diverse capacities, from expediting journalistic research to aggregating and cross-referencing data (Underwood, 2019). While these automated systems demonstrate remarkable efficacy, they can pose notable implications for human rights, notably impacting freedom of expression and the right to non-discrimination. Additionally, they may, to some degree, encroach upon rights to privacy and data protection.

In order to better understand the role of AI in journalism, it is necessary to focus on ethical and legal considerations as well. The transparency of algorithms, the safeguarding of users' data privacy, and the accountability of automated decision-making processes are -among others- crucial factors which can ensure the protection of journalistic standards and professional integrity. In this vein, the exploration of AI impact on participatory journalism management transcends mere technological innovation. This paper thus embodies a critical examination of the evolving dynamics between media, technology, and society. It explores the evolving regulatory framework concerning AI, aiming to shed light both on possible risks and on ethical concerns.

Methodology

The study adopted a qualitative research design, utilizing a literature review process to explore both the use of AI in participatory journalism management and the ethical considerations deriving in this framework. The goal of this review was to synthesize existing knowledge, identify trends, gaps and challenges, and provide a comprehensive understanding of how AI technologies are being integrated into the management of participatory journalism. To specify the search, the paper focused on the most relevant studies that specifically discuss the application and impact of AI in participatory journalism management, on research that provides empirical evidence related to AI tools and techniques in journalism and on studies that explore the ethical, social, and managerial implications of AI in audience participation.

As for the regulatory framework, the paper delved into reports that specifically discuss the regulation, legislation and policy aspects of AI, and especially in official documents from recognized authorities such as government bodies, international organizations, and leading research institutions. Due to the constantly evolving landscape of AI regulations, attempts to explore the field demand continuous attention, since global policy shifts could reshape the framework and implications of AI governance.

Automation and artificial intelligence in participatory journalism management

To deter malicious activities by bots and prevent mass submissions, media websites often implement a mandatory fully automated verification process for user participation (Saridou & Veglis, 2016; Sivakorn et al., 2016). CAPTCHAs (Completely Automated Public Turing tests to tell Computers and Humans Apart) typically prompt users to identify and verify a distorted image containing alphanumeric characters. Google's reCAPTCHA service stands out as the predominant solution in this domain, widely embraced by numerous websites to fend off automated bot attacks (Sivakorn et al., 2016). Designed to streamline the experience for genuine users while complicating text recognition for automated systems, reCAPTCHA employs sophisticated risk assessment mechanisms. It dynamically adjusts the difficulty of challenges based on individual user interactions, sometimes necessitating the selection of specific images from a collection of similar ones (Sivakorn et al., 2016).

Furthermore, as moderators can face overwhelming numbers of comments, with some of them harassing or hateful, platforms typically provide tools such as word filters for creators to automate aspects of moderation (Shagun et al., 2022). Automated moderation uses predefined filters to detect and replace prohibited words, phrases or IP addresses before comments are published (Veglis, 2014). However, users easily bypass such filters by inserting symbols or asterisks. In these cases, natural language processing (NLP) proves to be more effective, as it provides the possibility to focus on each specific phenomenon separately through a series of computational techniques for automatic analysis and representation of the human language (Cambria & White, 2014). A typical field of its application is hate speech, since the classification of a message as malicious or not should not only be based on the verbal part; peripheral factors need to be taken into account, such as the wider discourse context in which the message is included, the images and the accompanying videos, the time of posting or even the identity of the author and the recipient. These detection systems can also be bypassed with the use of static or moving images, memes and graphics (Lamerichs et al., 2018; Matamoros-Fernández & Farkas, 2021; Perifanos & Goutsos, 2021).

Over time, a range of tools and interfaces have been developed to assist journalists with gathering information and identifying newsworthy items from UGC, from event and eyewitness detection to source curation and verification (Wang & Diakopoulos, 2021). Many news organizations have pioneered participatory journalism platforms allowing audience submissions. Recently, media outlets have increasingly leveraged AI technology, notably ML, across various phases of news production.

Specifically, the American newspaper New York Times employs advanced methods for automated content moderation, utilizing ML algorithms trained on extensive textual datasets (Binns et al., 2017). The organization utilizes Google's Perspective system, employing ML to automatically identify and flag toxic comments. Drawing from an extensive dataset of million comments, the system assesses every new comment on a toxicity scale ranging from zero to one hundred based on their resemblance to previously identified as toxic. As a result, comments

are directly approved and published or escalated to moderators for further review (Etim, 2017; Wang, 2020). Moreover, The Washington Post newspaper used the NLP technique of sentiment analysis to investigate the possible removal of critical details from public reports by the U.S. Agency for International Development's Inspector General. The comparison between draft reports and their final versions showed that more than 400 negative references were removed before publication (Higham and Rich 2014; Stray 2016). Aiming to validate sources and combat misinformation, the Pheme project uses ML techniques to help journalists observe the veracity of potential rumors propagating on social media, find reliable sources, and recognize how rumors surface and develop (Wang & Diakopoulos, 2021).

Challenges and ethical considerations for media organizations

Apart from the opportunities to enhance the media production process, such as using ML to sift through large numbers of documents, applying analytic tools for audience understanding or developing sociotechnical processes to support fact-checking (Diakopoulos, 2019; Trattner et al., 2021), the incorporation of AI tools in journalistic practice seems far from unhindered. The collection of personal data from millions of Facebook users without their consent by Cambridge Analytica company, and the debate over the role of Twitter bots in the 2016 US presidential election are two of the most well-known examples of AI misuse. Among others, issues of privacy, security and data protection, along with the constant need for transparency and justice call for thorough consideration. Traditional media organizations may face new threats since it is easy for both malicious editorial and non-editorial players to present misinformation and disinformation as news. In such a landscape of untrustworthy sources, responsible media entities should struggle to fortify their role as reliable sources of information (Trattner et al., 2021).

In a global survey conducted by the London School of Economics and Political Science in collaboration with the Google News Initiative on the use of AI in 71 news organizations from 32 countries (Beckett, 2019), journalists were asked about the challenges they faced in their daily work routine due to the integration of AI. Participants referred to practical issues, such as the lack of resources and the unwillingness to invest the existing resources in this direction. The cultural resistance including the fear of losing jobs or changing work habits, and a general hostility to new technology, the lack of knowledge about AI along with a lack of strategic managerial insight were also highlighted as key issues. Notwithstanding the above, journalists also mentioned the difficulty of setting the right priorities.

Survey participants were also asked in which areas AI brings or could bring a difference in relation to the ethics and culture of their media organizations. They initially highlighted the financial part, expressing the concern whether making cuts from AI-generated savings could lower editorial standards or whether reinvestments could instead be used to improve journalism quality and effectiveness. In addition, journalists referred to the risk of algorithmic biases, when the misuse of data or the use of bad data could lead to inaccuracies and discrimination against

specific social groups or views (Beckett, 2019). In a similar vein, the intensification of information bubbles and "echo chambers" arises as a widely expressed concern for media organizations and poses serious threats to freedom of speech, democratic values, and the quality of public debate (Ge et al., 2020). On the other hand, journalists underlined positive aspects where AI could improve the way editorial decisions are made, and enhance capabilities of human intelligence, such as critical thinking and creativity (Beckett, 2019).

More specifically, while ML has frequently been proposed as a wholesale solution to proactively remove undesirable posts, it requires gathering enough training data on rule violations, may not be adaptive to new kinds of infractions, may further complicate fairness and justice issues, and may have decisions that are not comprehensible (Gillespie, 2018; Gorwa, 2020; Jiang, 2019). Recent research suggests that algorithms designed by software engineers to detect hate speech show biases regarding, for example, gender (Shagun et al., 2022). Moreover, the computational and economic power required to train large language models as well as the access to large datasets, which are the basis of current NLP systems, lies in the hands of private companies such as OpenAI, Meta or Google (Prem & Krenn, 2023).

Regulatory framework

During the very recent years, a large set of declarations and guidelines for the ethical use and development of AI has started blooming. These declarations lead to different similar approaches for introducing sets of principles as a departure point for discussing the responsible development of AI (Díaz-Rodríguez et al., 2023). In November 2021 UNESCO published the Recommendation on the Ethics of Artificial Intelligence, which has been adopted by 193 member states and is built on four pillars: Respecting human rights and dignity, living in just, peaceful, and interconnected societies, ensuring diversity and inclusion and developing the natural environment and ecosystem. The Recommendation focuses on eleven thematic areas, including data, gender, education, communication and the economy. Values and principles are designed to be respected by all actors involved in the AI system life cycle, being amenable to change through amendments to existing and new legislation and business guidelines (Díaz-Rodríguez et al., 2023).

Also in 2021 the European Commission proposed the first EU regulatory framework for AI, known as Artificial Intelligence Act (AI Act – AIA). In June 2023 Members of European Parliament (MEPs) adopted Parliament's negotiating position on the AI Act. Tripartite negotiations between the Parliament, the Commission and the Council followed and led to a final agreement in December 2023. The proposed Act takes the form of a regulation, which means that once adopted the rules and provisions are directly applicable within all 27 European Member States (Helberger & Diakopoulos, 2022). The new rules focus on the specific use of AI systems, defining a classification with different requirements and obligations tailored to a "risk-based approach", where the obligations for an AI system are proportionate to the level of risk that it poses (Díaz-Rodríguez et al., 2023).

In Greece, general issues of AI are regulated by law 4961/2022, which obliges public and private bodies to a series of actions when using AI applications. Some of them are the algorithmic impact assessment, the obligation of transparency and of keeping records for employed AI systems, as well as the ethical use of data according to specific rules.

Conclusions

In today's media landscape AI tools and techniques are often utilized by news organizations to facilitate the laborious duty of managing UGC. While NLP and ML algorithms contribute to content moderation or source verification, considerations regarding AI bias, privacy concerns, and the potential for censorship necessitate careful oversight and transparency. Beyond the legal part, concerns arise regarding the broader ethical issues of who has access to the databases and the large data centers required, who owns the very large computing infrastructures needed and, finally, how can the beneficial use of AI be ensured.

Against this background, addressing the manifold emerging issues requires a multidisciplinary approach, necessitating collaboration among developers, AI system designers, producers, providers, legislators, and journalists. This convergence entails partnerships between news organizations, academic institutions, and technology companies. Additionally, the importance of education and ongoing training, alongside the adoption of shared ethical guidelines between journalists and technologists concerning data usage should not be neglected. Above all, we have to keep in mind that technophobia cannot be the solution. Valuable advancements will persist, underscoring the importance of providing equitable access to tools for everyone and strengthening digital literacy initiatives. By effectively addressing the emerging challenges, we can leverage AI's transformative capacity to enhance journalistic methodologies while safeguarding the essential values of integrity and democracy.

References

Abbott, J. Y. (2017). 'Tensions in the scholarship on participatory journalism and citizen journalism'. *Annals of the International Communication Association*, 41(3-4): 278–297. https://doi.org/10.1080/23808985.2017.1350927.

Beckett, C. (2019). *New powers, new responsibilities. A global survey of journalism and artificial intelligence*. https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/

Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2017). Like trainer, like bot? Inheritance of bias in algorithmic content moderation. In G. Ciampaglia, A. Mashhadi, & T. Yasseri (Eds.), *Social Informatics*. *Lecture Notes in Computer Science*. Springer, Cham. https://doi.org/10.1007/978-3-319-67256-4_32.

Boberg, S., Schatto-Eckrodt, T., Frischlich, L., & Quandt, T. (2018). 'The moral gatekeeper? Moderation and deletion of user-generated content in a leading news forum'. *Media and Communication*, 6(4): 58-69. https://doi.org/10.17645/mac.v6i4.1493.

Cambria, E. & White, B. (2014). 'Jumping NLP Curves: A review of natural language processing research'. *IEEE Computational Intelligence Magazine*, 9(2): 48-57. https://doi.org/10.1109/MCI.2014.2307227.

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press, Cambridge. https://doi.org/10.4159/9780674239302.

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). 'Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation'. *Information Fusion*, 99. https://doi.org/10.1016/j.inffus.2023.101896.

Engelke, K. M. (2019). 'Online participatory journalism: A systematic literature review'. *Media and Communication*, 7(4): 31-44. https://doi.org/10.17645/mac.v7i4.2250.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Gorwa, R., Binns, R., & Katzenbach, C. (2020). 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance'. *Big Data & Society,* 7(1).

Helberger, N., & Diakopoulos, N. (2023). 'The European AI Act and how it matters for research into AI in media and journalism'. *Digital Journalism*, 11(9): 1751–1760. https://doi.org/10.1080/21670811.2022.2082505.

Higham, S. & Rich, S. (2014, October 22). *Whistleblowers say USAID's IG removed critical details from public reports*. The Washington Post. https://www.washingtonpost.com/investigations/whistleblowers-say-usaids-ig-removed-critical-details-from-public-reports/2014/10/22/68fbc1a0-4031-11e4-b03f-de718edeb92f_story.html

Jiang, J., Kiene, C., Middler, S., Brubaker, J.R., & Fiesler, C. (2019). *Moderation challenges in voice-based online communities on discord*. In Proceedings of the ACM on Human-Computer Interaction, 3(55): 1–23. https://doi.org/10.1145/3359157.

Krumsvik, A. (2018). 'Redefining user involvement in digital news media'. *Journalism Practice*, 12(1): 19–31. https://doi.org/10.1080/17512786.2017.1279025.

Lamerichs, N., Nguyen, D., Carmen, M., Melguizo, P., Radojevic, R., & Lange-Böhmer, A. (2018). 'Elite male bodies: The circulation of alt-right memes and the framing of politicians on social media'. *Participations*, 15(1): 180–206.

Loosen, W., Ahva, L., Reimer, J., Solbach, P., Deuze, M., & Matzat, L. (2022). 'X Journalism'. Exploring journalism's diverse meanings through the names we give it'. *Journalism*, 23(1): 39–58. https://doi.org/10.1177/1464884920950090.

Matamoros-Fernández, A. & Farkas, J. (2021). 'Racism, hate speech, and social media: A systematic review and critique'. *Television & New Media*, 22(2): 205–224. https://doi.org/10.1177/1527476420982230.

Perifanos, K. & Goutsos, D. (2021). 'Multimodal hate speech detection in Greek social media'. *Multimodal Technologies and Interaction*, 5(7): 34. https://doi.org/10.3390/mti5070034.

Prem, E. & Krenn, B. (2024). On algorithmic content moderation. In: Werthner, H., et al. *Introduction to digital humanism*. Springer, Cham. https://doi.org/10.1007/978-3-031-45304-5_30.

Risch, J. & Krestel, R. (2018, August 25). Delete or not delete? Semi-automatic comment moderation for the newsroom. In R. Kumar, A. Ojha, M. Zampieri & S. Malmasi (Eds.), *Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying* (pp. 166–176), Santa Fe, USA.

Saridou, T. & Veglis, A. (2016). 'Participatory journalism policies in newspapers' websites in Greece'. *Journal of Greek Media and Culture*, 2(1): 85-101. https://doi.org/10.1386/jgmc.2.1.85_1.

Saridou, T. & Veglis, A. (2021). Exploring participatory journalism through the integration of user-generated content in media organizations. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology*, 5th edition, (pp. 1152-1163). IGI Global. https://doi.org/10.4018/978-1-7998-3479-3.

Saridou, T., Panagiotidis, K., Tsipas, N., & Veglis, A. (2019). 'Towards a semantic-oriented model of participatory journalism management: Perceptions of user-generated content'. *Journal of Education, Innovation and Communication*, 1: 27-37. https://doi.org/10.34097/jeicom_SI_Dec2019-2.

Shagun, J., Chen, Q., Knauss, D., & Zhang, A. (2022). *Designing word filter tools for creator-led comment moderation.* In CHI Conference on Human Factors in Computing Systems (CHI '22), April 29 - May 5, 2022, New Orleans, LA, USA. https://doi.org/10.1145/3491102.3517505.

Sivakorn, S., Polakis, J., & Keromytis, A.D. (2016, March 29 – April 1). *I'm not a human: Breaking the Google reCAPTCHA*. In Black Hat Asia 2016, Singapore. https://www.blackhat.com/asia-16/briefings.html.

Stray, J. (2016, September 30 – October 1). *What do journalists do with documents? Field notes for natural language processing researchers*. [Conference presentation]. Computation+Journalism Symposium. Palo Alto, CA: Stanford University.

Trattner, C., Jannach, D., Motta, E. Costera Meijer, I., Diakopoulos, N., Elahi, M., Opdahl, A., Tessem, B., Borch, N., Fjeld, M., Øvrelid, L., De Smedt, K., & Moe, H. (2022). 'Responsible media technology and AI: challenges and research directions'. *AI Ethics 2*, 585–594. https://doi.org/10.1007/s43681-021-00126-4.

Underwood, C. (2019). *Automated journalism - AI applications at New York Times, Reuters, and other media giants*. New York Times. https://emerj.com/ai-sector-overviews/automated-journalism-applications.

Wang, S. (2020). 'The influence of anonymity and incivility on perceptions of user comments on news websites'. *Mass Communication and Society*, 23(6): 912–936. https://doi.org/10.1080/15205436.2020.1784950.

Wang, Y. & Diakopoulos, N. (2021). Journalistic source discovery: Supporting the identification of news sources in user generated content. In CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan. https://doi.org/10.1145/3411764.3445266.

Wolfgang, D. (2016). 'Pursuing the ideal'. *Digital Journalism*, 4(6): 764-783, https://doi.org/10.1080/21670811.2015.1090882.