International Core Journal of Social Sciences
ISSN: 3080-8774 (Print), ISSN: 3080-8782 (Online) | Vol. 1, No. 1, 2026

Research on Ethical Risks and overnance Mechanisms in AI-Assisted News Production

Xinru Cao, Yi Zhang

School of Languages and Media, Anhui University of Finance & Economics, Bengbu Anhui 233030, China

Abstract: Generative artificial intelligence (AI) is deeply integrated into the entire process of news planning, gathering, editing, and distribution. The application of technologies such as automated writing, algorithmic recommendation, and deep synthesis continues to expand, improving news production efficiency while constantly challenging the core principles of journalistic professionalism. This paper, based on ethics and journalistic professionalism, systematically examines the ethical risks of AI-assisted news production through a combination of case analysis, literature review, and comparative research. It analyzes the technological, economic, and institutional motivations behind these risks and proposes a layered and collaborative governance solution based on domestic and international governance practices. The study finds that AI-assisted news production faces multiple risks, including the erosion of factual accuracy, the solidification of algorithmic bias, the ambiguity of responsible parties, and the polarization of the public communication sphere. The probabilistic generation logic of technology, the profit-driven nature of the traffic economy, and the lag in existing laws and regulations collectively constitute the root causes of these risks. Building a responsible AI news governance system requires coordinated efforts at three levels: macro-level institutional regulation, meso-level organizational self-discipline, and micro-level literacy cultivation. This will propel the news industry from anxieties about technological substitution to a virtuous cycle of human-machine collaborative evolution.

Keywords: Artificial intelligence; News production; Algorithmic ethics; Media governance; Journalistic professionalism.

1. Introduction

1.1. Research Background and Significance

The large-scale application of generative artificial intelligence (AI) is driving the transformation of AI from specialized technologies to general-purpose services, bringing profound changes to the production processes and dissemination logic of the global news industry. Major domestic and international media outlets such as the Associated Press, Reuters, Xinhua News Agency, and Peoples Daily have successively established intelligent editorial departments, integrating AI tools into their daily production systems. The efficiency improvements brought about by technology empowerment have become an industry consensus. As of August 2025, 538 generative AI services had been registered in my country, most of which involve the production, processing, and distribution of news information. CCTV Video used AI technology to complete the entire production process of the short drama "Chinese Mythology," and the Daily Economic News AI video platform enabled the rapid mass production of short news videos, fully validating the value of technology in reducing costs and increasing efficiency.

While technology is rapidly becoming widespread, the core values of journalistic professionalism face severe challenges. In 2025, multiple cases emerged in various regions involving the use of AI to generate false information about peoples livelihoods and fabricate sudden safety incidents. The rapid spread of this content caused public panic, and the operators were dealt with according to law. Data from Stanford Universitys 2023 Digital Tr ust Research shows that AI-generated false information is 3.7 times more difficult to identify and spreads 2.4 times faster than traditional rumors, indicating a continued erosion of social trust in the news industry. The ethical lapses in AI-assisted news production not only affect the survival and development of the news industry but also concern the health of the public information ecosystem and the stability of social order. Systematic research on this issue has significant theoretical and practical value.

1.2. Definition of Core Concepts

AI-assisted news production is based on the core principle of human-machine collaboration. Unlike models where AI autonomously replaces production, AI acts as a tool in all stages of news production, while humans retain the final decision-making power over content review and publication. The Financial Times establishment of AI editor positions and the New York Times company -wide AI tool training both demonstrate an industry practice oriented towards human-machine collaboration. Generative AI is essentially an extension of human news production capabilities, rather than a replacement for professionals.

The connotations of news ethics in the digital age continue to expand. Traditional news ethics focus on basic norms such as truthfulness, objectivity, and fairness. With the intervention of AI, issues such as algorithmic transparency, data bias, attribution of responsibility, and the human-machine boundary have become new ethical issues. Relevant research by the United Nations shows that AI systems in the industry commonly exhibit stereotypes based on gender and race. When such biases are embedded in news production, they directly distort news value judgments and violate the professional requirements of fairness and impartiality in journalism.

2. Characterization of Ethical Risks in AI-Assisted News Production

2.1. Content Production Side: The Dissolution of Authenticity

The probabilistic generation mechanism of large language models inherently possesses the flaw of "illusion." The models predict content output based on data associations rather than verifying facts, a characteristic that makes them a new threat to news authenticity. In 2024, the false information circulating on social media that "the mortality rate of Chinese people born in the 1980s is 5.2%" was found to be erroneous content generated by AI through illusion. An investigation by the French Press Union revealed that factual errors generated by AI when rewriting news are often overlooked by operators and even incorrectly cited in public knowledge entries[1].

The widespread adoption of deepfake technology further undermines the factual basis of news. In 2025, former New York Governor Andrew Cuomo used AI to create fabricated campaign footage, presenting fictional scenes and images in a way that appeared real, breaking the inherent connection between images and facts. Research from Columbia University indicates that when public figures actively use deepfakes for propaganda, societys tolerance for false content gradually increases, making the harm of maliciously fabricated content harder to detect. The case of French news websites using AI to mass-produce news further pushes the production of false content to an industrial scale, transforming the authenticity of news from an occasional error into a structural crisis[2].

2.2. Value Judgment Side: Algorithmic Solidification of Bias and Discrimination

Algorithm systems lack value neutrality, and social stereotypes carried by training data are inherited and amplified by the model, forming algorithmic bias. A 2025 study by the University of Zurich tested several mainstream large language models, and the results showed that the models significantly reduced their evaluation of texts labeled with Chinese authors, and the bias was more obvious when it came to sovereignty issues [3]. A study by UN Women also confirmed that AI systems generally have dual biases of gender and race. When such biases are embedded in news production, they can lead to the underestimation of the news value of specific groups and the reinforcement of regional and group stereotypes[4].

The traffic logic of algorithm-based recommendations further distorts the judgment of news value. Recommendation systems rely primarily on user preferences and interaction data, rather than the public importance of news content. In 2025, the Cyberspace Administration of China summoned some social media and news platforms, pointing out that platform algorithms favored content such as resentment towards the wealthy, gender antagonism, and regional discrimination, creating a closed loop for the spread of negative content, squeezing out rational and professional content, and resulting in a chaotic industry phenomenon where bad money drives out good. The existence of the algorithms black box makes t his process difficult to supervise externally, and the public nature of news is gradually being replaced by traffic logic[5].

2.3. Labor Ethics: Alienation and Responsibility Myths of Practitioners

The ambiguous definition of responsibility for AI-generated content creates a liability gap[6]. When AI-generated content infringes on rights or spreads false information, there is a lack of clear legal basis for assigning responsibility to developers, deployment platforms, and users. In a 2025 French AI news copyright infringement case, the court could only order the blocking of the infringing website; the actual operators and responsible parties were difficult to trace. The existing legal framework is inadequate to address the new liability issues brought about by AI[7].

The professional autonomy of practitioners is gradually weakening due to reliance on technology[8]. Journalists are shifting from content producers to AI content moderators and fact-checkers. Long-term reliance on AI for tasks such as data compilation and drafting will erode professional creative and in-depth thinking abilities. A Philippine news websites AI chatbot, due to a system update malfunction, continuously outputted outdated information. If practitioners do not conduct manual verification, erroneous content will directly enter the public dissemination process, amplifying the dissemination risks brought about by technology[9].

2.4. Public Communication Terminals: Information Cocoons and Public Opinion Polarization

Personalized algorithm recommendations continue to build a closed information environment, with users being precisely pushed content that matches their preferences, and information barriers between different groups are constantly deepening. In 2025, the Cyberspace Administration of China summoned mainstream platforms to discuss the information cocoon problem, demanding that they optimize recommendation models, strengthen human intervention, and address the fragmented dissemination caused by algorithms. Research by Professor Zhang Zhian of Fudan University points out that low-quality content generated by AI, with its production speed and traffic advantages, is squeezing out the space for professional media, continuously weakening the integrative function of news as a public instrument, and making it increasingly difficult to form consensus on public issues[10].

The emotional narrative characteristics of AI-generated content make it easy to incite public sentiment. The lax review standards for AI content by self-media and non-professional organizations allow extreme and emotional content to spread rapidly, further exacerbating public opinion polarization and undermining the rational atmosphere of public communication.

3. The Genetic Logic and Underlying Motivations of Ethical Risk

3.1. Technical Logic: The Uncertain Nature of Generative AI

The probabilistic generation logic of large language models fundamentally conflicts with the core requirement of news seeking certainty in facts. Models rely on lexical association predictions for output, failing to eliminate the illusion problem at its root . Deep synthesis technology enables the generation of all elements of audio and video, overturning the traditional empirical logic of "a picture is worth a thousand

words," blurring the lines between true and false content, and exponentially increasing the difficulty of fact-checking. The uncertainty and uncontrollability of technology present entirely new challenges to the verification and validation of news facts.

3.2. Economic Logic: The Coercion of Attention Economy Driven by Traffic

Platform capitalisms revenue mechanism reshapes the value orientation of news production, directly linking content views and interaction rates to revenue. This drives numerous self-media outlets and MCN agencies to use AI to mass-produce sensational content to gain traffic. A 2025 Xinhua News Agency investigation revealed that an MCN agency in Jiangxi operated hundreds of self-media accounts, using AI to generate thousands of pieces of false information daily to attract traffic, earning over ten thousand yuan per day. AI-driven rumor-mongering has formed a complete black market profit chain, with ethical oversight completely abandoned under the drive for profit.

3.3. Institutional Logic: The legal definition of human-machine relationships is lagging behind.

The current legal system fails to clearly define the legal status of artificial intelligence (AI), preventing AI from independently assuming legal responsibility. The allocation of liability for issues such as infringement and misinformation lacks clear rules. Core issues such as the ownership of copyright for AI-generated content and the accountability for factual errors are not clearly defined by law. Even in French AI news infringement cases, judgments still rely on traditional intellectual property rules, without establishing specific regulations for AI-generated content. This lag in law and regulation has created a regulatory vacuum characterized by ethical lapses. The high cost of rights protection and the low cost of violating the law exacerbate industry chaos.

4. Practical Observations and Lessons Learned from Domestic and International AI News Governance

4.1. Hard constraints: regulation and legislation

The European Union has established a world-leading AI regulatory system. Its Artificial Intelligence Act, which came into effect in 2024, classifies news recommendation systems as high-risk AI systems, setting stringent requirements such as transparency and human oversight. In 2025, the EU proposed amendments to its Digital Comprehensive Act, postponing the implementation of strict regulations for high-risk AI systems. This has sparked questions from NGOs regarding the protection of digital rights, highlighting the challenge of balancing regulation and innovation as a common issue in global AI governance.

my country has adopted a regulatory approach of developing and governing simultaneously. In 2023, the "Interim Measures for the Administration of Generative Artificial Intelligence Services" established a filing system, requiring AI services with public opinion attributes to disclose filing information. In 2025, the Cyberspace Administration of China launched the "Qinglang" special campaign, focusing on rectifying the problem of AI-generated false information and gradually improving the supporting rules for technical supervision and content governance.

4.2. Soft Constraints: Industry Self-Regulation and Media Practice

Major international media outlets have established guidelines for the use of AI. The Associated Press, Reuters, and other organizations have clearly defined core principles such as requiring human review of AI-generated content, prohibiting AI from independently publishing news, and disclosing the extent of AI involvement. The AI chatbot on the Philippine website Rappler limits its information sources to its own reports and verified data, mitigating the risk of AI illusions through information fencing, thus forming a valuable industry practice.

Domestic media outlets adhere to a parallel approach of technological empowerment and professional oversight. Applications such as the Peoples Dailys "Creative Brain AI+" platform, Xinhua News Agencys digital reporters, and CCTVs AI -powered micro-dramas all retain the core element of manual review, upholding journalistic professionalism in their technological applications. The gradual improvement of industry self-regulation and institutional norms has become a crucial support for compensating for regulatory lags and preventing ethical risks.

5. Building a Responsible AI-Powered News Governance Mechanism

5.1. Macro-institutional Level: Agile Governance and Legal Coverage

The pace of AI technology iteration far exceeds the traditional legislative cycle, making static regulatory rules ill-suited to the needs of industry development. Establishing a regulatory sandbox mechanism is therefore inevitable, allowing for technological experiments within a controllable scope and dynamic adjustments to regulatory rules to balance innovation and risk. my countrys "Measures for the Identification of Artificial Intelligence-Generated and Synthetic Content," to be implemented in 2025, requires AI content to carry non-editable metadata identifiers to enhance content traceability. Subsequent improvements to supporting technical standards and oversight mechanisms are needed to ensure the rules are effectively implemented. Algorithm recommendations must be required to disclose core logical parameters to break down the algorithms black box and safeguard public communication interests.

5.2. Meso-level Organization: Ethical Embedding of Human-Machine Collaboration

Human involvement is a fundamental principle of AI-assisted news production. While AI can perform auxiliary tasks, the final decision-making power regarding content review and publication must remain with humans . News organizations should establish AI ethics assessment mechanisms to conduct pre-launch testing on AI tools training data, probability of bias, and illusion rate, embedding ethical requirements into the entire technology application process. The practice of Zhejiang Provincial Communication Brain Company demonstrates that combining AI capabilities with content risk control and operational management can achieve a balance between technological empowerment and

risk prevention.

5.3. Micro-level literacy: Two-way empowerment between practitioners and the public

Journalists need a comprehensive upgrade in their AI literacy, not only mastering the use of AI tools but also possessing the ability to identify AI errors and verify facts. The New York Times has conducted company-wide AI training, focusing on risk identification and professional oversight of AI content, aligning with the industrys new demands on professional competence. Public media literacy needs to improve simultaneously, moving beyond the traditional notion of "a picture is worth a thousand words" in an era of widespread AI-generated content, understanding the logic behind AI content generation, and developing rational information consumption and discernment abilities.

6. Conclusion

The integration of artificial intelligence and journalism has entered an irreversible stage of development, and the core issue of industry development has shifted from whether to use AI to how to use it responsibly. This paper systematically reviews the ethical risks of AI-assisted news production, analyzes the risk drivers from three levels: technology, economy, and system, and proposes a governance path that coordinates macro, meso, and micro levels based on domestic and international practices, providing a reference for the reconstruction of journalistic professionalism in the context of human-machine collaboration. This study has certain limitations; the depth of case analysis is limited by publicly available information, and there is still room for expansion in the scope of interviews and the research on the underlying technological logic.

In the future, human-machine collaboration will become the norm in news production. Media organizations need to find a balance between technological application and professional integrity, platform companies need to shoulder social responsibility beyond traffic-driven logic, and the public needs to cultivate media literacy adapted to the intelligent era. Only through the concerted efforts of multiple stakeholders can we safeguard the authenticity, objectivity, and public nature of news in the AI era and promote the healthy development of the news industry and the public information ecosystem.

Acknowledgements

This work is supported by Anhui University of Finance & Economics 2025 Undergraduate Research innovation fund project fund, Project number: XSKY25058ZD.

References

[1] Yang Zhijie. French website shut down for using AI to generate infringing news? [J/OL]. North American Intellectual Property News, 2025-12-30.

[2] Xinhua News Agency. Xinhua Viewpoint | The Business Behind Sensationalism - Unveiling the Profit Chain of AI Rumor-Making [N/OL]. China.com.cn, 2025-09-17.

[3] European Times. Swiss Media: Artificial Intelligence Shows Bias Towards Text from China [N/OL]. European Times, 2025-11-14.

[4] Weng Xiuqi. The Era of Self-Deception: When Politicians Begin to "Deeply Fake" Themselves [N/OL]. Mirror News, 2025-10-29.

[5] Hunan Daily. Three-point update | Toutiao and Weibo were summoned for talks, what do you think? [N/OL]. New Hunan, 2025-09-24.

[6] European Times. EU's relaxation of AI regulations and privacy rules is seen as a "major setback" in digital protection [N/OL]. European Times, 2025-11-20.

[7] Xinhua News Agency. 538 Generative Artificial Intelligence Services Complete Filing [N/OL]. CCTV.com, 2025-09-11.

[8] Rowan Philip. How Newsrooms Are Using AI Chatbots to Leverage Their Own Reporting — and Build Trust[J/OL]. Global Investigative Journalism Network, 2025-08-07. https://gijn.org/stories/newsrooms-using-ai-chatbots-leverage-reporting/

[9] Zhang Yang, Wang Yang, Cao Ning. Exploration and Challenges of Human-Machine Collaborative Communication in News Media in the AIGC Era [J]. Youth Journalist, 2025(07):4-7.

[10] Zhang Zhian, Liu Jie. Journalistic Professionalism in Platform Society: Challenges and Reconstruction [J]. International Journalism Review, 2023, 4.