Research on the Mechanism of Emotional Distortion and Credibility Restoration in Multimodal News Driven by Generative AI

Yantong Lin[1,a], Yulin Zhao[2,b,*]

[1]Faculty of Humanities and Arts, Macau University of Science and Technology, Guangzhou, China

[2]School of Journalism and Communication, Northwest University of Political and Law, Xi'an, China

[a]739116995@qq.com, [b]18192052826@163.com

[*]Corresponding author

Abstract: With the rise of generative artificial intelligence, multimodal news production has entered a new stage of automated and intelligent expression. However, emotional distortion problems caused by excessive algorithmic generation and style transfer threaten news credibility and public trust. This study examines the mechanism of emotional expression distortion in AI-driven multimodal news and explores credibility restoration strategies. By analyzing the causes of distortion from content, technology, and communication perspectives, and constructing governance and repair pathways, this research aims to provide theoretical support and practical reference for improving the quality and trustworthiness of news communication in the intelligent era.

1. Introduction

1.1. Research Background

Generative AI technologies, including large language models such as GPT, diffusion-based image generators, and multimodal AI systems, are increasingly integrated into news production and journalism practices. These tools enable the automatic generation of textual content, images, audio, and video, greatly expanding narrative possibilities and enriching communication channels. For instance, AI can quickly produce news summaries, generate illustrative visuals, or create interactive multimedia stories, enhancing both efficiency and audience engagement. However, the use of generative AI also introduces significant challenges. The emotional content conveyed through AI-generated news—whether in language tone, visual representation, or audio expression—may be unintentionally exaggerated, misrepresented, or distorted, leading to inconsistencies between the intended message and audience perception. Such distortions pose risks to journalistic accuracy, potentially spreading biased interpretations or creating misunderstandings among the public. Moreover, the opacity of AI generation processes can make it difficult for audiences to discern whether content has been algorithmically produced or human-authored, raising concerns about accountability and transparency in news dissemination.

1.2. Research Significance

Examining the mechanisms behind emotional distortion in AI-driven multimodal news and exploring strategies for credibility repair is of both theoretical and practical importance. Theoretically, it contributes to research on algorithmic journalism, media psychology, and the ethics of AI-mediated communication. Practically, understanding these mechanisms helps journalists, media organizations, and policymakers develop effective strategies to safeguard the authenticity of news, uphold ethical standards, and maintain public trust. In an era of rapidly evolving generative technologies, such research is essential for ensuring that AI enhances rather than undermines the credibility and social responsibility of news communication.

2. Theoretical Foundation of Generative AI and Multimodal News

2.1. Characteristics of Generative AI in News Production

Generative AI has become a transformative tool in modern news production, supporting automatic text composition, image generation, video synthesis, and audio production [1]. These technologies enable news organizations to produce content at unprecedented speed, reducing reliance on manual labor while maintaining consistency across multiple formats. One of the most prominent characteristics is efficiency: AI can rapidly generate large volumes of content, summarize lengthy reports, or create visualizations without extensive human intervention. Scalability is another key feature, as generative AI systems can simultaneously serve diverse platforms, audiences, and media channels, ensuring broad distribution of news products. Beyond efficiency and scalability, generative AI also enhances creativity in news storytelling. By synthesizing multimodal elements, AI can generate illustrative graphics, simulate realistic scenarios, or create immersive audio-visual narratives, enriching narrative styles and enhancing audience engagement. However, this creative capacity is a double-edged sword; while it offers novel storytelling methods, it also introduces challenges in maintaining factual accuracy, emotional appropriateness, and ethical standards in news content.

2.2. Concept and Features of Multimodal News

Multimodal news represents an evolution from traditional text-based reporting to integrated communication that combines text, images, video, and audio to convey information [2]. By leveraging multiple sensory channels, multimodal news enhances comprehension, engagement, and memorability. Its key characteristics include high immersion, enabling audiences to experience events more vividly through synchronized visual, auditory, and textual cues. Strong interactivity is another hallmark: multimedia news can incorporate clickable elements, interactive visualizations, and AI-driven chat interfaces, allowing audiences to explore content according to their preferences and actively participate in information interpretation. Furthermore, multimodal news demonstrates emotional resonance: by integrating expressive visual imagery, tone-modulated audio, and narrative text, it can evoke feelings that reinforce the informational message and shape audience perception. These features collectively create a rich, dynamic, and emotionally engaging communication experience, but they also amplify the potential for emotional distortion when AI-generated elements are misaligned or exaggerated.

2.3. Potential Risks of Generative AI in Multimodal Expression

Despite its advantages, the use of generative AI in multimodal news introduces several risks. One major concern is mismatched emotion–content relations, where the emotional tone of images, audio, or text does not correspond with factual information, potentially misleading audiences. Stylistic exaggeration is another risk, as AI may overemphasize dramatic visuals or sensational language to capture attention, undermining journalistic neutrality [3]. Moreover, AI often exhibits lack of contextual appropriateness, struggling to interpret the nuances of cultural, social, or situational contexts, which may lead to inappropriate or insensitive emotional expressions. These risks collectively threaten the credibility, fairness, and ethical integrity of news reporting, highlighting the necessity of understanding the mechanisms behind emotional distortion.

3. Mechanisms of Emotional Distortion in AI-Generated Multimodal News

3.1. Content-Level Distortion Mechanism

At the content level, AI systems often overemphasize emotionally charged keywords or visual exaggerations. For example, when summarizing a political event, AI-generated headlines may disproportionately highlight conflict or controversy, creating an emotional tone inconsistent with the neutral facts presented in the body text. Similarly, image-generating algorithms may amplify facial expressions or dramatize scenes beyond reality, resulting in an emotional mismatch that can mislead audiences and bias their perceptions.

3.2. Technical-Level Distortion Mechanism

Technical mechanisms also contribute to distortion. Generative AI relies heavily on training data, and biases present in these datasets—such as the overrepresentation of sensational events—can skew outputs toward extreme emotional expressions [4]. Additionally, algorithmic preferences in model architecture and weighting may amplify emotional polarity, while insufficient adaptation to specific contexts reduces sensitivity to nuanced reporting. These technical limitations exacerbate the gap between intended and perceived emotional content, producing outputs that may overstate risks or intensify audience reactions.

3.3. Communicative-Level Distortion Mechanism

From a communicative perspective, the architecture and operation of algorithm-driven recommendation systems play a pivotal role in shaping the emotional impact of AI-generated news. Social media and news platforms typically optimize for engagement metrics such as clicks, likes, shares, and comments, which incentivizes the promotion of content that evokes strong emotional responses. As a result, AI models tend to prioritize news stories with heightened emotional appeal—such as sensational headlines, dramatic images, or emotionally charged video clips—over neutral or balanced reporting. This selective amplification creates a feedback loop: emotionally extreme content attracts more user interactions, which in turn signals the algorithm to further prioritize similar content, gradually skewing the visible information landscape [5]. Consequently, audiences are more frequently exposed to emotionally intense narratives, potentially reinforcing existing biases, heightening anxiety or outrage, and distorting their perception of events. This mechanism not only undermines the objective presentation of facts but also challenges the principles of journalistic impartiality, as algorithms inadvertently become active participants in shaping emotional narratives, amplifying sensationalism, and marginalizing measured or nuanced reporting. Over time, this systemic bias can contribute to the formation of echo chambers and polarization, further complicating the communicative environment of AI-mediated news.

3.4. Ethical and Cognitive Distortion Mechanism

In addition to technical and communicative factors, ethical and cognitive mechanisms significantly contribute to emotional distortion in AI-generated multimodal news. Unlike human journalists, AI lacks the capacity for moral reasoning, ethical reflection, or evaluation of potential social consequences. It cannot autonomously determine whether the amplification of particular emotional cues may mislead, provoke unnecessary fear, or trigger societal harm [6]. Simultaneously, audiences interact with AI-generated content through cognitive frameworks shaped by prior beliefs, emotional susceptibility, and interpretive heuristics. This human cognitive bias may lead readers to attribute intent, exaggeration, or emotional manipulation to AI-generated content, even when no conscious intention exists. The interplay between AI's ethical indifference and human interpretive biases can significantly amplify emotional distortion, resulting in the reinforcement of misinformation, heightened public anxiety, and erosion of trust in media institutions. Recognizing these ethical and cognitive mechanisms is therefore essential: it highlights the need for interdisciplinary strategies that combine technical safeguards, editorial oversight, audience education, and transparent communication practices. Only by addressing both the algorithmic drivers and the cognitive reception of emotional content can news organizations mitigate distortion, foster rational engagement, and restore credibility in AI-driven multimodal journalism.

4. Strategies for News Credibility Restoration

4.1. Ensuring Accuracy and Transparency of Content

Ensuring the accuracy and transparency of content is the cornerstone for maintaining credibility in AI-generated multimodal news [7]. With generative AI capable of producing text, images, audio, and video simultaneously, the risk of introducing false or exaggerated emotional elements increases. To mitigate this risk, it is essential to establish robust content verification and fact-checking

mechanisms that integrate AI-assisted detection with human judgment. AI systems can automatically flag inconsistencies between the narrative and underlying data, detect unusual sentiment intensities, or identify visual exaggerations in generated images or video clips. For instance, natural language processing models can cross-reference reported events against trusted databases, while computer vision algorithms can assess whether visual depictions correspond accurately to real-world occurrences. Beyond technical verification, transparency in reporting processes is equally critical [8]. News organizations should disclose which elements of a story are AI-generated, the sources of the training data, and the methods used to synthesize multimodal content. Such transparency not only improves accountability but also educates audiences about the AI's role in shaping narrative forms. Moreover, combining automated checks with editorial oversight ensures that factual integrity is preserved, emotional expression aligns with context, and audience trust is reinforced.

4.2. Improving Algorithm Interpretability and Auditability

Generative AI models, particularly large language and multimodal models, often operate as "black boxes," where the internal decision-making process is opaque to both journalists and audiences [9]. This lack of interpretability can obscure how emotional cues are generated, making it difficult to identify bias, exaggeration, or misalignment between emotion and factual content. To address this, enhancing algorithm interpretability and auditability is critical. Explainable AI techniques, such as attention visualization, feature attribution, and decision-tracing tools, can clarify how specific textual or visual elements are produced. These tools allow editors and researchers to understand which inputs led to particular emotional outputs, enabling targeted corrections and improvements. Additionally, regular algorithm audits should be institutionalized, focusing on both technical performance and ethical compliance. Audits can identify biases embedded in training data, evaluate the balance of emotional expression, and ensure that model outputs comply with journalistic standards. By establishing transparent reporting of algorithmic mechanisms and audit results, news organizations can reduce "black box" concerns, strengthen accountability, and provide evidence that emotional generation is governed responsibly. This dual approach—enhancing interpretability while maintaining rigorous audit practices—serves as a foundation for credible AI-assisted journalism.

4.3. Strengthening Human-Machine Collaborative Editing

While generative AI offers efficiency and creative capabilities, relying solely on automated outputs risks emotional misrepresentation and contextual inaccuracy. A human-in-the-loop editorial model is therefore essential. In this approach, journalists and editors supervise AI-generated content at multiple stages, from initial generation to final publication [10]. Humans can correct excessive emotional exaggeration, adjust tone to match the factual narrative, and ensure that visuals, audio, and text are contextually aligned. For example, AI may generate a visually striking image to illustrate a news event, but an editor can assess whether the depiction exaggerates the emotional intensity of the story, potentially misleading the audience. Human oversight also provides ethical guidance, preventing content that may be culturally insensitive, sensationalist, or socially destabilizing. Collaborative workflows between humans and machines create a feedback loop: AI learns from editorial corrections, improving the quality of subsequent outputs, while humans benefit from AI's speed and analytical capabilities. Ultimately, this model balances efficiency, creativity, and ethical responsibility, ensuring that multimodal news communicates emotions appropriately without compromising factual accuracy or credibility.

4.4. Enhancing Media Literacy and Public Participation

In addition to editorial and algorithmic governance, the audience plays a pivotal role in maintaining the credibility of AI-generated news. Improving media literacy enables readers to identify AI-generated emotional content, critically assess tone and intent, and discern potential exaggerations or distortions [11]. Educational initiatives, such as interactive tutorials, explanatory notes within news apps, and public workshops, can empower audiences to engage with AI news

more critically. Furthermore, public participation mechanisms—including interactive feedback channels, comment moderation systems, and crowdsourced verification tools—allow audiences to provide real-time corrections and observations. For example, viewers can flag emotionally misleading visual content or provide context that AI may overlook, creating a dynamic dialogue between producers and consumers. Encouraging such participation not only improves content accuracy but also reinforces audience trust, fostering a collaborative governance model where human oversight extends beyond journalists to include informed citizens. Through media literacy and participatory feedback, society collectively contributes to minimizing emotional distortion in AI-generated news.

5. Evaluation and Future Prospects of Credibility Restoration

5.1. Evaluation Indicators

The effectiveness of credibility restoration in AI-generated multimodal news can be evaluated through several key indicators. Emotional accuracy measures the alignment between emotional expression in multimedia content and the factual narrative, ensuring consistency and appropriateness. Content-objectivity consistency evaluates whether the overall story maintains factual integrity while conveying emotional cues. The public trust index reflects audience confidence in news organizations and their use of AI, capturing attitudinal and perceptual dimensions of credibility. Finally, communication effectiveness assesses whether audiences not only receive accurate and emotionally coherent information but also engage with it meaningfully, such as through sharing, discussion, or informed decision-making. Together, these indicators provide a multidimensional assessment framework, encompassing technical accuracy, emotional integrity, audience perception, and social impact.

5.2. Evaluation Methods

To operationalize these indicators, a combination of qualitative and quantitative methods is recommended. Content analysis can systematically examine textual, visual, and audio elements to identify mismatches between emotional tone and factual content. Machine–human comparative experiments involve comparing AI-generated content with human-edited versions to evaluate the extent of distortion and the effectiveness of corrections. Audience surveys and focus groups provide insights into public perception, trust, and engagement, offering feedback on whether emotional cues are interpreted accurately. Additionally, credibility assessment models, which may integrate sentiment analysis, engagement metrics, and fact-checking outcomes, allow for large-scale, data-driven evaluation of news content. Using multiple methods in concert ensures that evaluations are comprehensive, reliable, and actionable.

5.3. Future Prospects

Looking forward, news production will increasingly rely on deeper human-machine collaboration, where AI assists with generation, analysis, and preliminary editing, while humans supervise, correct, and contextualize outputs. Emotional control, authenticity assurance, and credibility restoration will remain central tasks in AI governance for journalism. Future developments may include adaptive AI models capable of learning from editorial feedback and audience reactions, systems that automatically detect potential emotional distortion, and interactive platforms that integrate public verification in real time. Ethical frameworks and standards will continue to evolve, guiding both algorithm design and human editorial practices. As generative AI matures, the goal will be to establish a sustainable ecosystem in which efficiency, creativity, emotional resonance, and factual accuracy coexist, ensuring that AI-enhanced news not only informs but also builds and preserves public trust.

6. Conclusion

This study investigates the emotional distortion mechanism in AI-driven multimodal news and

proposes strategies for credibility restoration. Emotional distortion mainly stems from mismatched content-emotion alignment, algorithmic bias, communicative amplification, and ethical deficits. Credibility restoration requires multi-pronged efforts, including content verification, algorithm transparency, human-machine collaboration, and public literacy enhancement. Ultimately, balancing technological efficiency with news ethics is the key to ensuring trustworthy, authentic, and sustainable news communication in the age of generative AI.

References

[1] Motevalian P, Lakshmipathy U, Correia A, et al. Breaking the Bottleneck in Lentiviral Production: Achieving Ultra-Low Infectivity-to-Particle Ratios and Cost Efficiency for Transformative Gene Therapies[J]. Cytotherapy, 2025, 27(5):S219.DOI:10.1016/j.jcyt.2025.03.446.

[2] Qiu Y , Liu H , Zhao M .A Review of Brain–Computer Interface-Based Language Decoding: From Signal Interpretation to Intelligent Communication[J].APPLIED SCIENCES-BASEL, 2025, 15(1):392.DOI:10.3390/app15010392.

[3] Gabora L, Ganesh K, Bashmakova I. Ignoring the role of reiterative processing and worldview transformation leads to exaggeration of the role of curiosity in creativity[J]. Behavioral and Brain Sciences, 2024, 47. DOI:10.1017/S0140525X23003527.

[4] Baiyan L , Xinyuan C , Jiangang X ,et al.Theoretical and technical investigation of automatic vertical drilling tools[J].Journal of Wuhan University of Science and Technology(Natural Science Edition), 2008.

[5] Ciriello R , Hannon O , Chen A Y ,et al.Ethical Tensions in Human-AI Companionship: A Dialectical Inquiry into Replika[J].Proceedings of the Annual Hawaii International Conference on System Sciences, 2024.DOI:10.24251/hicss.2024.058.

[6] Kuzmenkov V A , Soina I Y .Doublethink and Anomie: ethical and political context[J].Cuestiones Políticas, 2021, 39(69).DOI:10.46398/cuestpol.3969.15.

[7] Chen A , Evans R , Zeng R .Editorial: Coping with an AI-saturated world: psychological dynamics and outcomes of AI-mediated communication[J].Frontiers in Psychology, 2024, 15(15):1479981.DOI:10.3389/fpsyg.2024.1479981.

[8] Fox C .Beyond the "Tyranny of the Real": Revisiting Burke's Pentad as Research Method for Professional Communication[J].Technical communication quarterly, 2002, 11(4):p.365-388. DOI:10.1207/s15427625tcq1104_1.

[9] Fichtner U A , Knaus J , Graf E ,et al. Exploring the potential of large language models for integration into an academic statistical consulting service–the EXPOLS study protocol[J].PLoS ONE, 2024, 19(12).DOI:10.1371/journal.pone.0308375.

[10] Hernawan A H , Emilzoli M , Rullyana G ,et al.Enhancing Student Collaboration and Participation through Google Workspace in Higher Education[J].IJOEM Indonesian Journal of E-learning and Multimedia, 2025, 4(1):30-42.DOI:10.58723/ijoem.v4i1.359.

[11] ERSTAD, OLA, AMDAM, et al. FROM PROTECTION TO PUBLIC PARTICIPATION.[J]. Javnost the Public, 2013.