Risk Traceability of False Content Generation Enabling by Machine Learning Technology in the Intelligent Media Era

Abstract: The rapid evolution of intelligent media has driven machine learning technologies to deeply integrate into content production. While lowering the barriers to content creation and enhancing dissemination efficiency, these technologies also provide technical support for the generation and spread of false content, giving rise to a series of challenges in social governance, information security, and ethical considerations. Set against the backdrop of the intelligent media era, this paper examines the core technical pathways and manifestations through which machine learning enables the creation of false content, analyzes the diverse risk profiles associated with such content generation, and focuses on addressing the technical bottlenecks and practical challenges in current risk tracing efforts. Ultimately, it establishes a tripartite risk tracing framework encompassing "technical tracing, subject tracing, and regulatory tracing," while proposing targeted optimization strategies to provide theoretical insights and practical guidance for regulating machine learning applications, curbing the spread of false content, and maintaining order within the intelligent media ecosystem.
Keywords: intelligent media; machine learning; fake content; risk tracing; technological governance

I. Introduction

1.1Research Background

With the groundbreaking advancements in artificial intelligence technology, human society has entered a new era of information dissemination dominated by intelligent media. Powered primarily by algorithms and leveraging technologies such as machine learning and deep learning, intelligent media enables automated content generation, precise content delivery, and personalized distribution, fundamentally transforming the content production and dissemination models of traditional media. Machine learning technology, with its powerful capabilities in data mining, feature extraction, and pattern generation, has broken down the professional barriers to content creation, allowing ordinary users to rapidly produce various types of content—including text, images, audio, and video—without requiring specialized skills, thereby significantly enriching the content offerings of intelligent media.

However, the double-edged nature of technology is particularly pronounced in the field of intelligent media. While machine learning has become widespread, it has also been exploited by malicious actors as a tool for generating and disseminating false content. From AI-generated fake news and comments to deepfake facial videos

and voice imitations, and further to algorithmically produced misleading marketing messages and malicious rumors, the efficiency, authenticity, and reach of false content have all seen qualitative improvements, characterized by "low cost, high realism, broad dissemination, and difficulty in detection." The World Economic Forum's 2026 Global Risks Report lists "misinformation and disinformation" alongside geopolitical conflicts and armed conflicts as one of the five major global risks, highlighting that the harm caused by false content has become a significant challenge for global governance.

The proliferation of false information not only misleads public perception and undermines social trust, but also risks triggering societal panic, infringing upon individual rights, disrupting market order, and even threatening national security. For instance, an operator of a WeChat public account fabricated the claim that "300,000 residents had evacuated Shanghai" using AI, sparking public panic and resulting in administrative detention; in a case solved by Wuhan police in Hubei Province, a suspect exploited AI face-swapping technology to bypass facial verification systems, altered corporate legal entity information, and induced followers to engage in fraudulent investment schemes, causing substantial financial losses. Against this backdrop, accurately identifying false content and tracing its origin and dissemination pathways has become crucial for addressing the challenges of combating misinformation in intelligent media. Research on risk tracing for false content generation empowered by machine learning technologies has emerged as a key focus in both academic studies and social governance.

1.2 Research Significance

Theoretical Significance: This paper systematically examines the technical logic and risk characteristics of machine learning technologies enabling the generation of false content, delves into the core challenges and implementation pathways for tracing the risks associated with such content, enriches research findings in fields such as intelligent media governance, machine learning ethics, and information security, establishes a tripartite risk tracing framework encompassing "technology-subject-regulation," provides a theoretical framework and conceptual guidance for subsequent research, and fills the gap in current studies on false content risk tracing by completing the logical of "technology empowerment – risk generation – traceability governance."

Practical Significance: Aligned with current practical needs for combating false content, this paper proposes targeted risk traceability optimization strategies. These strategies provide actionable guidance for regulatory authorities, media platforms, and technology R&D institutions, enabling them to enhance the efficiency and accuracy of false content tracing. This helps effectively curb the creation and dissemination of false content, safeguard the healthy and orderly development of the intelligent media ecosystem, protect public legitimate rights and interests as well as societal welfare, and mitigate various security risks associated with false content.

1.3 Research Methods and Innovations

Research Methodology: This study employs a literature review approach to systematically synthesize domestic and international publications on intelligent media, machine learning, and the generation and risk tracing of false content, thereby establishing the theoretical foundation for the research. It utilizes case analysis, incorporating recent landmark cases such as AI-driven fake news and deepfake fraud, to examine the risk characteristics and tracing challenges associated with false content generation. Additionally, through logical analysis, the study elucidates the technical mechanisms by which machine learning facilitates the creation of false content, constructs a risk tracing framework, and proposes optimization strategies.

Research Innovations: First, an innovative research perspective that transcends single technological or governance approaches, adopting a holistic approach covering technology empowerment, risk generation, and traceability governance to establish a tripartite traceability system encompassing "technology traceability, subject traceability, and regulatory traceability," enabling multi-dimensional and comprehensive traceability governance. Second, content innovation that integrates the latest technological developments and case studies to systematically analyze emerging risks and traceability challenges posed by machine learning-enabled generation of false content, yielding more targeted and actionable strategies. Third, conceptual innovation that combines the "empowerment" and "governance" aspects of machine learning technology, addressing both risks arising from technological abuse and its application in risk traceability, thereby achieving synergistic advancement of technology and governance.

II. Definition of Relevant Concepts and Theoretical Foundations

2.1Definition of Core Concepts

2.1.1 Intelligent Media

Intelligent media refers to a media format that centers on artificial intelligence technologies (particularly machine learning and deep learning), integrating big data, the Internet of Things, cloud computing, and other technologies to achieve full automation and intelligence across the entire content lifecycle—from production and distribution to reception and feedback. Its core characteristics include algorithm-driven operations, data-driven support, personalized content delivery, and strong interactivity. This paradigm breaks away from the traditional one-way communication model of traditional media, fostering a multi-subject, multi-channel, and multi-format content dissemination ecosystem that encompasses various platforms such as short-video platforms, social media, and intelligent news apps.

2.1.2 Machine Learning Technology

Machine learning is the core branch of artificial intelligence, referring to the technology that enables computers to autonomously learn from large datasets, extract features, identify patterns, and make independent decisions and predictions through algorithms. Its core workflow follows the sequence of "data input – model training – feature extraction – result output." Common techniques include supervised learning, unsupervised learning, reinforcement learning, and deep learning. Among these, deep

learning (e.g., generative adversarial networks, Transformer models, diffusion models) serves as the core technology for generating fake content, capable of producing highly realistic text, images, audio, and video.

2.1.3 False Content

The false content referred to in this article denotes various types of information generated using machine learning technology that deviate from objective facts and are misleading or malicious, including false texts (news, comments, rumors, etc.), false images (forged photos, synthesized images, etc.), false audio (imitated speech, synthesized audio, etc.), and false videos (deepfake videos, synthesized scene videos, etc.). Compared to traditional false content, machine learning-generated false content is characterized by high generation efficiency, high realism, rapid dissemination speed, and strong concealment, making it difficult for the general public to identify.

2.1.4 Risk Traceability

False content risk traceability refers to the process of identifying the origin of false content (e.g., the creator, generation technology, or equipment) and its dissemination path (e.g., platforms, nodes, or channels) through technical means, institutional design, and stakeholder collaboration. This identifies the responsible entities involved in the creation and spread of false content, providing a basis for risk prevention, accountability, and governance optimization. Its core objective is to achieve "traceable origins, trackable paths, and accountable responsibilities."

2.2 Relevant Theoretical Foundations

2.2.1 Algorithm Governance Theory

Algorithm governance theory examines the role, risks, and governance approaches of algorithms in societal operations. Its core proposition is that algorithms, as a novel form of power, while empowering social development, may also give rise to risks such as algorithmic bias, algorithmic abuse, and information bubbles. Achieving the ethical operation of algorithms requires technical standards, institutional constraints, and stakeholder collaboration. This theory provides a theoretical foundation for analyzing the abuse risks and traceability governance of machine learning algorithms in generating false content, emphasizing the need for algorithmic regulation and technical optimization to mitigate the risks posed by algorithmic misuse.

2.2.2 Information Communication Theory

Information Communication Theory encompasses core elements such as the communication process, communicators, communication channels, and communication effects. Its central tenet posits that information dissemination is a dynamic process involving multiple actors and stages, with communication effects influenced by various factors including communicators, channels, and content. This theory provides a theoretical foundation for analyzing the generation and dissemination mechanisms of false content and tracing its transmission pathways, enabling a comprehensive mapping of the entire process from creation to spread and identifying key nodes in the traceability process.

2.2.3Technical Ethics Theory

This theory examines the relationship between technological advancement and human societal ethics, emphasizing that technological development must adhere to ethical principles, respect human rights, and uphold social fairness and justice. It provides theoretical support for analyzing the ethical risks posed by machine learning technologies in generating false content, guiding the risk tracing process to balance technological progress with ethical standards, thereby achieving coordinated advancement of technological governance and ethical constraints.

2.2.4Collaborative Governance Theory

The collaborative governance theory emphasizes that multiple stakeholders (government, enterprises, social organizations, the public, etc.) work together through division of labor and coordination to address complex social issues and maximize governance effectiveness. This theory provides theoretical guidance for establishing the "technology-subject-regulation" trinity risk traceability system in this paper, highlighting the need for coordinated participation by the government, platforms, technology R&D institutions, and the public in combating false content traceability to create a synergistic governance effort.

III. The Mechanisms and Manifestations of Machine Learning Technology Enabling False Content Generation in the Intelligent Media Era

3.1 Core Technology Path

The core methodology of machine learning-driven fake content generation follows a "data training – feature extraction – model generation – content output" workflow. By analyzing vast datasets of authentic content, the system learns its characteristics and patterns to produce highly realistic counterfeit material. Deep learning serves as the foundational technology, with key approaches including:

1. Generative Adversarial Networks (GANs): These deep learning architectures combine generators and discriminators through adversarial training to produce fake content. The generator creates the content, while the discriminator distinguishes it from real material. Through iterative optimization, the generator produces content so convincing it passes genuine content detection. GANs are pivotal for deepfake videos and synthetic images—enabling face replacement, scene synthesis, and generation of highly realistic fake videos that are nearly indistinguishable from originals. Advanced variants like Conditional GANs, DCGANs, and StyleGANs further enhance generation quality and control, allowing precise manipulation of facial orientation, expressions, and other details.

2. Transformer Model: The Transformer model, based on the self-attention mechanism, efficiently processes sequential data and serves as the core technology for generating fake text. Through training on extensive text datasets, it can simulate human linguistic logic and expression patterns, producing fluent, coherent, and

context-appropriate fabricated content—including fake news, fake reviews, and fake copy. In recent years, large language models built on the Transformer architecture (such as ChatGPT, DeepSeek, and Gemini) have significantly enhanced the quality of fake text generation, producing logically sound and emotionally rich content that can even mimic the speech styles of specific individuals with remarkable concealment.

3. Diffusion Model: Through a training process involving forward noise addition and backward denoising, the diffusion model progressively approaches the true content distribution, enabling the generation of highly faithful images, videos, and other content. Compared to GAN models, diffusion models excel in content diversity and realism, effectively addressing limitations of GANs such as blurred textures and geometric distortions. They are widely used in scenarios like generating fake images and high-fidelity avatars. Open-source frameworks like Stable Diffusion and DALL-E-2 further lower the barrier to generating fake content.

4. Speech Synthesis Technology: Machine learning-based speech synthesis technology (TTS), trained on human speech data, can simulate the vocal characteristics of specific individuals and generate highly realistic synthesized audio. This technology enables speech imitation and conversion—for example, mimicking the voices of celebrities or loved ones to produce fabricated audio for fraudulent activities or rumor dissemination. Furthermore, when integrated with video synthesis technology, speech synthesis achieves synchronization between speech and lip movements, significantly enhancing the realism of deepfake videos.

5. Transfer Learning Technology: Transfer learning enables the transfer of trained model parameters to new tasks, lowering the technical barriers and data requirements for generating fake content. Criminals can leverage transfer learning to rapidly produce counterfeit content using existing models without requiring substantial training data or specialized technical expertise, thereby exacerbating the proliferation of fake content. For example, trained face generation models can be swiftly applied to forge faces of various individuals through transfer learning, significantly improving generation efficiency.

3.2 Main Forms of Manifestation

Underpinned by the aforementioned technical approaches, the fabricated content generated through machine learning in the intelligent media era manifests in diverse forms—including text, images, audio, and video—with its application scenarios continuously expanding. These can be categorized as follows:

1. False textual content: The most prevalent form, encompassing fake news, fabricated reviews, malicious rumors, and deceptive copywriting. For instance, some independent media outlets leverage large language models to rapidly produce fabricated social or entertainment news, misleading public perception; criminals employ generated product reviews and user feedback for deceptive marketing schemes that disrupt market order; while certain individuals utilize AI to create inflammatory rumors that trigger societal panic. Statistics indicate that a specific MCN agency once generated 4,000 to 7,000 fake news articles daily via AI software, severely disrupting public order. A 2024 report from Tsinghua University's School of

Journalism and Communication revealed a 99.91% surge in AI-generated economic and corporate-related rumors over the past year.

2. False image content: Utilizing technologies such as GANs and diffusion models, counterfeit photos and synthetic images are generated, primarily employed for spreading rumors, defamation, and false advertising. Examples include fabricated negative photos of celebrities, synthetic images of accident scenes, and exaggerated product efficacy visuals. Additionally, AI face-swapping technology extends this category by replacing faces in images to create fake portraits, infringing on others' right to portrait rights. For instance, Dilraba was replaced with an AI-generated face in a short drama, sparking a portrait rights dispute; some criminals also forge inappropriate photos of individuals for extortion purposes.

3. False audio content: Using speech synthesis technology, fake audio recordings that mimic the voices of specific individuals are primarily used for fraud and rumor dissemination. For example, mimicking the voices of relatives, friends, or superiors to solicit money; imitating celebrities 'voices to publish false statements or claims that mislead the public; or even criminal organizations using such audio to forge instructions for financial scams. Additionally, cover works like "AI Sun Yanzi," which replicate celebrity voices, are suspected of infringing on others' voice rights and represent a derivative form of false audio content.

IV. Risk Profiling of Machine Learning-enabled False Content Generation in the Intelligent Media Era

Machine learning technology enables the generation of fake content, not only disrupting the information ecosystem of intelligent media but also triggering a range of multifaceted risks that affect personal rights, social order, national security, and other domains. These risks exhibit systemic, contagious, and covert characteristics, as detailed below:

4.1 Individual Level: Rights Infringement and Security Threats

4.1.1 Violation of Personal Privacy and Portrait Rights

The misuse of deepfake technology and voice synthesis technology enables the illegal collection and forgery of individuals 'biometric data—including facial images, voice recordings, and portraits—thereby infringing upon personal privacy and portrait rights. For instance, criminals may harvest others' facial data to create fabricated facial videos or images for defamation or extortion; AI face-swapping technology can replace individuals' faces without consent, violating their portrait rights. In a case solved by Wuhan police in Hubei Province, suspects illegally breached computer information systems using AI face-swapping technology to alter corporate legal entity information, which constitutes a fundamental violation of personal information and privacy.

4.1.2 Damage to Personal Reputation

The dissemination of false content can severely damage an individual's reputation. Particularly, false rumors and negative misinformation targeting public

figures or ordinary individuals can mislead public perception and lower their social standing. For example, fabricated negative news or inappropriate videos about public figures can seriously harm their reputation; false accusations or malicious defamation against ordinary individuals can affect their work and daily lives, and may even lead to psychological issues.

4.2 Social Level: Order Chaos and Trust Crisis

4.2.1 Disruption of social public order

The rapid dissemination of false information can trigger social panic and rumor spread, disrupting normal social public order. For example, false pandemic updates, disaster reports, or safety alerts may cause public panic, leading to irrational behaviors such as panic buying and evacuation, thereby undermining social stability; false information about social conflicts can escalate tensions and provoke group clashes. The false claim "300,000 residents evacuated from Shanghai," fabricated by an official WeChat account operator, once caused localized panic and disrupted normal social order.

4.2.2 Destruction of the Social Trust System

In the era of intelligent media, the authenticity of information serves as the foundation of social trust. The proliferation of false content leads to a significant decline in public trust in media and information, creating a "trust deficit." When "seeing is believing," the public struggles to distinguish truth from falsehood, developing skepticism toward various types of information. This subsequently affects social interactions, market transactions, and other societal activities, undermining the social trust system. Moreover, AI-generated false content blurs the line between authentic and fabricated information, impairing public cognition and plunging people into cognitive confusion.

V. The Dilemma in Tracing the Sources of False Content Risks in the Era of Smart Media

Currently, although countries and platforms are actively exploring methods for tracing the sources of false content, the challenge remains significant due to the rapid evolution of machine learning technologies, the complexity and concealment of false content, and the inadequacies in tracing systems. These challenges primarily manifest in three dimensions: technology, stakeholders, and regulation.

5.1 Technical Level: Lagging Traceability Technology and Insufficient Adaptability

The rapid iteration of machine learning has continuously advanced fake content generation techniques, enhancing their realism and concealment, while the development of fake content traceability technologies lags behind, failing to keep up with the pace of technological progress. For example, deepfake technology evolves extremely rapidly, with new milestone models emerging every two months on average,

whereas traceability technologies primarily target existing fake content generation methods and lack sufficient capability to track emerging types of fake content, making precise tracing difficult.

Current false content often exhibits multimodal integration, such as text + images or audio + video. Since different modalities employ distinct generation techniques, tracing methods also vary, significantly increasing the complexity of tracking multimodal false content. Existing tracing technologies primarily target false content of a single modality, lacking collaborative multimodal tracing approaches that enable full-process tracing. Although multimodal fusion detection technologies have advanced, their application in tracing remains underdeveloped.

During the generation of fraudulent content, criminals employ various anti-traceability measures—such as removing generation traces, modifying metadata, and using anonymous tools—which hinder traceability technologies from identifying the source of the false content. For example, publishing fraudulent content via virtual private networks (VPNs) or anonymous accounts conceals the actual IP address; modifying the metadata of the content and deleting critical information like the generation tool and time make it difficult for traceability systems to obtain reliable clues. Furthermore, adversarial attacks compromise the accuracy of traceability technologies, as criminals interfere with the model's judgment through techniques such as prompt injection and input perturbation.

5.2 At the Subject Level: Ambiguous Traceability Responsibilities and Insufficient Collaboration

5.2.1 Ambiguous accountability for traceability

The creation and dissemination of false content involve multiple entities, including content generators (criminals or malicious users), distribution platforms (smart media platforms), and technology providers (machine learning R&D institutions). However, the current lack of clear accountability mechanisms leads to mutual buck-passing among stakeholders, hindering coordinated traceability efforts. For instance, content generators are often anonymous, making tracing difficult; distribution platforms lack sufficient incentives for traceability and typically focus solely on removing false content while neglecting traceability efforts; technology providers fail to fulfill their regulatory responsibilities and inadequately curb the misuse of their technologies.

5.2.2 Insufficient motivation for platform traceability

As the primary vehicle for disseminating false content, intelligent media platforms bear significant responsibility for tracing such content. However, their profit models conflict with this traceability obligation. The platform's core revenue stems from content traffic, and false content often achieves high dissemination rates, generating substantial traffic and revenue for the platform. This creates insufficient motivation for platforms to trace false content; they may even deliberately conceal relevant information, hindering traceability efforts. Furthermore, traceability initiatives increase operational costs, further diminishing platforms' incentive to

engage in traceability activities.

5.2.3 Lack of a collaborative traceability mechanism involving multiple stakeholders

Tracing the origin of false content requires coordinated efforts from multiple parties, including the government, platforms, technology research and development institutions, and the public. However, current mechanisms for communication and collaboration among these stakeholders are ineffective, resulting in a fragmented approach. There is insufficient information sharing between government regulators and platforms, inadequate technical coordination between platforms and R&D institutions, and limited public access to traceability channels and capabilities. These factors lead to low efficiency in traceability efforts and hinder comprehensive chain tracing of false content. For example, regulatory traceability technologies cannot be effectively integrated with platform content data, and the traceability technologies developed by R&D institutions struggle to be rapidly implemented in practical platform operations.

5.3 Regulatory Level: Incomplete Legal Framework and Insufficient Regulatory Capacity

5.3.1 Incomplete relevant laws and regulations

Currently, the legal framework governing how machine learning technologies enable the generation and tracing of false content remains inadequate, lacking clear legal provisions and penalty standards. For instance, the legal responsibilities of content generators, distribution platforms, and technology providers remain undefined; there are no specialized laws regulating emerging forms of false content such as deepfake content; and issues related to data collection and privacy protection during content tracing lack explicit legal guidance, undermining the legal foundation for effective tracing efforts and hindering the realization of "legally binding and enforceable regulations." Although China has issued documents like the "Measures for Identifying Artificial Intelligence-Generated Synthetic Content" and the "Regulations on Deep Synthesis in Internet Information Services," these legal frameworks still require further refinement to meet the governance demands posed by modern false content.

5.3.2 Insufficient regulatory technology and capacity

Government regulatory agencies possess relatively outdated regulatory technologies and lack advanced tools for tracing fake content, making it difficult to combat the massive volume and complexity of such content. Additionally, regulatory personnel lack the necessary expertise, particularly in machine learning and fake content generation technologies, hindering effective traceability efforts. Furthermore, the cross-border dissemination of fake content complicates regulation, as varying national standards and technological capabilities, coupled with the absence of an effective international coordination mechanism, impede the effective tracing of transnational fake content.

5.3.3 Outdated Regulatory Approach

Current false content governance primarily adopts a "post-event supervision" model, where identification, removal, and traceability measures are implemented only after false content has been disseminated. This approach fails to achieve "prevention before occurrence and control during occurrence," resulting in traceability efforts being initiated only after harm has been caused, thereby significantly diminishing their effectiveness. Moreover, the regulatory framework lacks flexibility, making it ill-suited to address the rapid evolution of false content generation and dissemination in the intelligent media era, and unable to promptly respond to emerging risks posed by new types of false content.

VI. Optimized Path for Tracing the Origin of False Content Risks in the Intelligent Media Era

To address the current challenges in tracing the sources of false content, and in alignment with the evolving trends in machine learning technology and the practical needs of intelligent media governance, this paper establishes a tripartite framework for false content risk tracing — "technical tracing, subject tracing, and regulatory tracing." Through technical optimization, coordinated efforts among stakeholders, and enhanced regulatory mechanisms, the framework enables the "traceability of origins, tracking of dissemination paths, and accountability determination" for false content. The specific optimization pathways are as follows:

6.1Technical Traceability: Upgrading Traceability Technology to Overcome Technical Bottlenecks

Increase R&D investment in fake content traceability technologies, encourage research institutions and enterprises to conduct traceability studies, and prioritize the development of technologies targeting deepfake and multi-modal fake content. For example, develop deep learning-based multi-modal collaborative traceability algorithms to enable full-process tracing of fake content across text, images, audio, and video; create blockchain-based traceability technologies that leverage blockchain's immutable and traceable characteristics to document the entire generation and dissemination chain of fake content for precise tracing. Additionally, drawing inspiration from the "Measures for Identifying Artificially Generated and Synthetic Content," integrate traceability technologies with identification technologies—using both implicit and explicit identifiers to provide actionable clues for tracing.

To combat traceability challenges posed by fraudulent content, corresponding countermeasures have been developed. These include employing metadata recovery techniques to restore deleted or modified metadata and identify the sources of fabricated content; utilizing IP tracking technology to bypass anonymity tools and pinpoint the actual IP addresses of content creators; and developing adversarial traceability techniques to mitigate cybercriminal attacks and strengthen the robustness

of traceability systems. For instance, the RECCE ("Reconstruction–Classification") learning framework proposed by Yang Xiaokang's team at Shanghai Jiao Tong University enables precise identification of forged content sections and localization of traceability clues, making it highly applicable for widespread implementation.

Develop unified technical standards for tracing false content, specifying technical indicators, application specifications, and data formats to achieve interoperability among different traceability technologies and platforms, thereby enhancing traceability efficiency. For example, establish multi-modal traceability standards and data sharing protocols to standardize traceability practices; create a technical evaluation framework to assess the performance and accuracy of traceability technologies, facilitating their optimization and upgrading. Simultaneously, align with international traceability standards to strengthen the global competitiveness of China's traceability technologies.

6.2 Subject Origin Tracing: Clarify the Responsible Party and Strengthen Collaborative Cooperation

Define the traceability responsibilities of content generators, distribution platforms, and technology providers, establishing a responsibility framework where "the generator bears responsibility, the distributor assumes accountability, and the technology provider oversees compliance." For instance, content generators must be accountable for their creations and cooperate with traceability efforts; distribution platforms must fulfill their primary responsibility by implementing robust traceability mechanisms to promptly track the origin and dissemination paths of false content; technology providers must integrate traceability features during development, strengthen oversight against misuse of technology, and support traceability initiatives. Additionally, specify the legal liabilities of each entity and impose penalties on those failing to fulfill their traceability obligations.

Encourage intelligent media platforms to establish robust mechanisms for tracing false content, incorporating traceability efforts into platform performance evaluations to incentivize improved compliance. For instance, platforms should maintain traceability databases recording the creation time, responsible entities, and dissemination paths of false content; they are encouraged to collaborate with research institutions and regulatory bodies to adopt advanced tracing technologies. Concurrently, optimize platform revenue models to prioritize content quality over "traffic supremacy," prompting proactive traceability initiatives. Platforms demonstrating effective traceability efforts that successfully curb false content dissemination should receive policy support and incentives.

Develop a coordinated traceability framework involving the government, platforms, technology research and development institutions, and the public, enhancing communication and collaboration among all parties to create a unified traceability effort. For example, government regulatory authorities should lead the establishment of a traceability information sharing platform to facilitate data exchange between platforms and R&D institutions; R&D institutions provide technical support

to platforms and regulators to assist in traceability efforts; the public can contribute clues about false content to regulators and platforms through dedicated traceability feedback channels, thereby participating in the traceability process. Additionally, a collaborative traceability working mechanism should be established, with regular coordination meetings held to address challenges encountered during traceability operations.

6.3 Regulatory Traceability: Improving the Legal Framework and Enhancing Regulatory Capabilities

Accelerate the development of legal frameworks governing the use of machine learning technologies for generating and tracing false content, clarifying legal bases, liability assignments, and penalty standards for such activities. For example, enact a dedicated "False Content Governance Law" specifying regulations for the creation, dissemination, and tracing of false content; enhance regulatory frameworks for deepfake and AI-generated content to define the legal responsibilities of technology providers, platforms, and content creators; address data collection and privacy protection issues during traceability processes to ensure lawful implementation. Additionally, refine the implementation rules of the "Measures for Identifying Artificially Generated and Synthetic Content" to facilitate deeper integration of identification and traceability technologies, thereby strengthening source tracing capabilities.

Increase technological investment in government regulatory agencies, introduce advanced fake content traceability technologies and equipment, and establish an intelligent regulatory platform to enable real-time monitoring, identification, and tracing of fake content. For example, develop a big data and artificial intelligence-based fake content traceability platform that integrates content data and traceability information from various sources to achieve comprehensive oversight throughout the entire lifecycle of fake content; strengthen professional training for regulators to enhance their understanding of machine learning and fake content generation techniques, thereby improving traceability capabilities; simultaneously, enhance international regulatory cooperation by establishing a cross-border fake content traceability coordination mechanism to address challenges posed by transnational fake content.

Shift from the traditional "post-event supervision" approach to establishing a full-process regulatory framework of "pre-event prevention – in-process control – post-event traceability." For instance, strengthen pre-event oversight of machine learning applications to standardize technology development and usage while preventing misuse; implement real-time monitoring mechanisms for false content during operations to promptly identify and curb its rapid dissemination; conduct precise post-event traceability to hold responsible entities accountable, forming a closed-loop supervision system of "prevention – control – traceability – accountability." Simultaneously, adopt a combination of "flexible supervision and stringent penalties" to both encourage proactive compliance with traceability

obligations and impose severe penalties for violations.

VII. Conclusion and Outlook

7.1 Research Conclusion

In the era of intelligent media, the rapid advancement of machine learning technology has provided robust technical support for the generation and dissemination of false content. Such content is characterized by "low cost, high realism, widespread circulation, and difficulty in identification," posing multifaceted risks to individual rights, social order, and national security. The core technical approaches enabling the creation of false content include Generative Adversarial Networks (GANs), Transformer models, diffusion models, and speech synthesis technologies, manifesting in various forms such as fabricated texts, images, audio, and videos. Currently, tracing the origins of false content faces challenges including technological backwardness, ambiguous primary responsibilities, and inadequate regulatory frameworks, significantly undermining the effectiveness of countermeasures against false content.

To address the challenges in tracing the origins of false content, this paper establishes a tripartite risk tracing system comprising "technical tracing, entity tracing, and regulatory tracing," and proposes targeted optimization strategies: at the technical level, accelerate the development of tracing technologies, overcome technical bottlenecks, promote the lightweight and widespread adoption of tracing technologies, and establish a technical standards framework; at the entity level, clarify the responsible entities for tracing, strengthen platform accountability, build a collaborative tracing mechanism involving multiple stakeholders, and enhance public tracing capabilities; at the regulatory level, improve relevant laws and regulations, enhance regulatory technologies and capabilities, innovate regulatory models, and establish a data-sharing mechanism for tracing information. Through these multidimensional and comprehensive optimizations, precise tracing of false content can be achieved, providing support for the governance of intelligent media ecosystems.

7.2 Future Outlook

Future research can focus on the following areas: First, conduct in-depth studies on how novel machine learning technologies enable the generation of false content, with particular emphasis on exploring traceability methods for emerging technologies such as diffusion models and large language models to enhance the adaptability of traceability systems. Second, refine the implementation details of traceability strategies through concrete case studies, verifying their effectiveness and practicality—for example, by optimizing traceability technologies and collaborative mechanisms based on real-world practices in platform-based false content governance. Third, strengthen research on cross-border false content traceability and explore international collaborative traceability frameworks to address risks associated with transnational false content dissemination. Fourth, prioritize ethical guidelines for

machine learning technologies by integrating ethical constraints into traceability systems, achieving synergistic progress between technological solutions and ethical governance. Additionally, with the ongoing implementation of policies such as the "Measures for Identifying Artificially Generated and Synthetic Content," further research should focus on deeply integrating identification and traceability technologies to improve efficiency and accuracy, providing robust support for combating false content in the era of intelligent media.

References

[1]Saridou T ,Dimoulas A C .Artificial Intelligence in Participatory Environments: Technologies, Ethics, and Literacy Aspects[J].Societies,2026,16(4):127-127.

[2]Parashar J ,Hung T B ,Upreti K , et al.An enhanced hybrid framework for IoT healthcare security using blockchain-driven multimedia data analysis and cybersecurity techniques[J].Array,2026,30100759-100759.

[3]Jin Z ,Sun X ,Xi X , et al.Simulon: An AI - Assisted, PyTorch - Native Framework of Molecular Dynamics and Modeling[J].Journal of Computational Chemistry,2026,47(8):e70364-e70364.

[4]A.O. A ,K.T. L ,O.F.W. O .Movie Recommendation system with sentiment analysis using deep learning algorithms[J].Egyptian Informatics Journal,2026,33100905-100905.

[5]Singh K A .AI-Driven Multi-objective Optimization Framework for Enhancing Cost, Quality, and Manufacturability in Product Design[J].Journal of Engineering Research and Reports,2026,28(1):271-283.

[6]Yang R ,Liu T .Research on the Teaching Reform of Advertising Creativity Courses Based on the OBE Concept[J].Journal of Education and Educational Research,2025,16(3):173-181.

[7]Moumen N ,Bahi H ,Makhoul N , et al.AI for Motorized Travel Time Index Prediction: Enhancing Spatio-Temporal Urban Mobility Performance in Smart Cities[J].Urban Science,2025,9(12):499-499.

[8]Ying X ,Zhang J ,Wang S , et al.The Value Foundation and Development Path of China's Digital Law under the Guidance of Xi Jinping's Thought on the Rule of Law[C]//University of West London.Proceedings of the 4th International Conference on International Law and Legal Policy （part1）.School of Marine Law and Humanities,Dalian Ocean University;,2025:139-145.

[9]Zheng Y ,Chen M .The Return of Life in Primary School Education in the Intelligent Era:Reflection and Practice Based on Educational Philosophy[C]//University of West London.Proceedings of the 4th International Conference on International Law and Legal Policy （part1）.Huaiyin Normal University;Jiangnan University;,2025:146-154.

[10]Li C .Determination of Copyright Infringement Liability of Online Game Live-Streaming Platforms[C]//University of West London.Proceedings of the 4th International Conference on International Law and Legal Policy （part2）.Zhengzhou Foreign Language New Fengyang school;,2025:364-373.

[11]Liang R .Legal Protection of Patients' Right to Know in Artificial Intelligence Diagnosis and Treatment[C]//University of West London.Proceedings of the 4th International Conference on International Law and Legal Policy （part2）.College of Liberal Arts and Law, Beijing University of Chemical Technology;,2025:457-462.

[12]Yu J .Cross-Cultural UX Design in Media Convergence:Challenges and Strategies for Global Platforms[C]//Sam Houston State University.Proceedings of the 4th International Conference on

Interdisciplinary Humanities and Communication Studies （part1） .University of Arts London:,2025:10-14.
[13]Qu Y ,Zhang J .Social Changes and the Evolution of Crime Patterns:Taking China's "Digital Transformation" as an Example[C]//Sam Houston State University.Proceedings of the 4th International Conference on Interdisciplinary Humanities and Communication Studies （part1） .School of Ocean Law and Humanities,Dalian Ocean University:,2025:164-171.
[14]An Y .Exploring the Impact of Artificial Intelligence Development on the Teaching Profession[C]//Sam Houston State University.Proceedings of the 4th International Conference on Interdisciplinary Humanities and Communication Studies （part1） .St.John Bosco High School:,2025:281-287.
[15]Zhang H .The Value Logic,Challenge and Development Path of Traditional Handicraft Empowered by Artificial Intelligence Technology[C]//Sam Houston State University.Proceedings of the 4th International Conference on Interdisciplinary Humanities and Communication Studies （part1） .School of Journalism & Communication,JiNan University:,2025:401-407.
[16]Chin M ,Wang X ,Zhong Y .AI Versus Authenticity: Consumer Sentiments Toward AI-Generated Advertisements on Social Media and Strategic Implications for Brands[C]//Sam Houston State University.Proceedings of the 4th International Conference on Interdisciplinary Humanities and Communication Studies （part2） .SJI International;School of Literature and Arts, South West University of Technology;Nanjing Forestry University:,2025:307-319.
[17]Cao J .Craftsmanship and Alienation: Reconstructing Labor Subjectivity from a Marxist Humanist Perspective in the Age of AI[C]//Sam Houston State University.Proceedings of the 4th International Conference on Interdisciplinary Humanities and Communication Studies （part3） .The Keystone Academy:,2025:546-551.
[18]Zhou Z .Intelligent Innovation in the Cultural and Creative Industries in the Era of Artificial Intelligence[C]//Sam Houston State University.Proceedings of the 4th International Conference on Interdisciplinary Humanities and Communication Studies （part3） .Art College of Yunnan Normal University:,2025:729-735.
[19]Xie Y .From Algorithmic Gazing to Cognitive Closure: The Invisible Discipline of the Information Cocoon on Human Cognitive Constructs[C]//Sam Houston State University.Proceedings of the 4th International Conference on Interdisciplinary Humanities and Communication Studies （part3） .Faculty of Humanities and Arts, Macau University of Science and Technology:,2025:198-205.
[20]Geng C .An intelligent framework combining deep learning and fuzzy logic for accurate remote language translation[J].Scientific Reports,2025,15(1):38736-38736.
[21]Nurlan Z ,Gabdullayev D ,Mukhametzhanova B , et al.Incident-aware smart prioritization framework for penetration testing and prevention of URL-based cybersecurity attacks in industry 4.0 IoT networks[J].Scientific Reports,2025,15(1):37792-37792.
[22]Liu X ,Meng Y .An Analysis on AI Hallucination from the Perspective of Media Archaeology[C]//Beijing Normal University.Proceedings of the 4th International Conference on Art, Design and Social Sciences.School of Arts and Communication,Beijing Normal University;School of Communication,Fujian Normal University:,2025:153-166.
[23]Yang Y .Research on the Challenges and Coping Strategies of News Authenticity in the Age of Social Media[C]//Beijing Normal University.Proceedings of the 4th International Conference on Art, Design and Social Sciences.Nanjing University of the Arts:,2025:190-198.

[24]Ád á m Francuz,Tam á s B á nyai.Intelligent Control Approaches for Warehouse Performance Optimisation in Industry 4.0 Using Machine Learning[J].Future Internet,2025,17(10):468-468.
[25]Qaisar S ,Rafique A ,Iqbal K , et al.Unlocking China's digital public diplomacy on X: A machine learning approach to big data analysis in Iran and Saudi Arabia.[J].Acta psychologica,2025,260105696.
[26]Fu D ,Ding L ,Zhang B , et al.Machine learning-assisted biomass colorimetric dipsticks for portable clinical nephropathy diagnosis by using smart mobile terminal[J].Chemical Engineering Journal,2025,522168059-168059.
[27]Zhang J ,Zhang D ,Huang X .Survey on Data-driven Intelligent Computing and Its Applications[J].Frontiers in Computing and Intelligent Systems,2025,13(2):106-111.
[28]Wang X .A Multimodal Analysis of Human Imitation of AI and Human – Machine Relations: A Case Study of 50 TikTok Videos[J].Frontiers in Computing and Intelligent Systems,2025,13(2):65-69.
[29]Luo L .A System Framework for Optimizing the Localization Media Strategy of Multinational Corporations Based on Multi-Source Data Fusion[J].Journal of Innovation and Development,2025,12(2):158-167.
[30]Li J .Automatic Evaluation System for News Communication Effectiveness Based on Topic Features and Deep Learning[J].International Journal of High Speed Electronics and Systems,2025,(prepublish):DOI:10.1142/S0129156425408010.