JOURNAL OF AI ETHICS AND LEGAL

Vol.1 Iss.1
ISSN Pending

MASON PUBLISH GROUP

Journal of AI Ethics and Legal

PUBLISHER
Journal of AI Ethics and Legal is published by Mason Publish Group, Seattle, USA; editorialoffice@masonpublish.org. For submission instructions, subscription and all other information visit:https://masonpublish.org.

Editorial Team

Journal of AI Ethics and Legal

Vol.1 Iss.1
ISSN pending

Transformation and Restructuring: Decoding the AI-Driven Creative Logic of Micro-Short Drama Video Production from SkyReels

Qiuhong Wang[1], Zhengfei Han[2]

[1]Institute of Communication Study,Communication University of China, 2410566870@qq.com

[2]Institute of Communication Study,Communication University of China,hzzzzzfey@gmail.com

Abstract

This study examines how AI-driven platforms like SkyReels transform micro-short drama production by integrating AI into creative conceptualization, content generation, workflows, and audience engagement. Through a mixed-methods analysis, we identify a shift from creator-centric to user-driven models, highlighting challenges like content homogenization and ethical concerns in AI authorship. We propose a framework for sustainable development, combining technical standards, cultural localization, and editorial oversight, to guide the global evolution of micro-short drama ecosystems

Keywords: Micro-short drama platform; AIGC (AI-Generated Content); Creative logic; SkyReels; platform ecology; algorithmic co-creation

1 Introduction

In July 2024, the Third Plenary Session of the 20th CPC Central Committee proposed to "improve the institutional mechanisms for developing new quality productive forces according to local conditions and enhance systems promoting the deep integration of the real economy and the digital economy," emphasizing the convergence of culture and technology. The contemporary world presents a complex ecology intertwined with digital platforms. The deepening integration of digital technology and the cultural industry has spurred the vigorous development of various platforms, profoundly transforming the production and dissemination models of cultural content and shaping a new landscape.

Micro-short drama, an emerging cultural product in the digital publishing sphere, has rapidly risen and gained widespread dissemination empowered by artificial intelligence (AI) technology and platform advantages. The intervention of AI has revolutionized the established logic of traditional video production. Firstly, during the creative conception phase, AI mines and analyzes vast datasets to accurately capture socio-cultural trends and audience preferences, providing innovative inspiration for creation. Secondly, in the content generation stage, leveraging technologies like natural language processing (NLP) and image recognition enables the efficient and precise generation of text and visuals, significantly enhancing production efficiency and quality. Thirdly, regarding production workflows, intelligent editing tools and special effects synthesis optimize resource allocation and streamline cumbersome processes. Fourthly, from the audience experience perspective, personalized recommendations and interactive narrative designs enhance user engagement and participation, imbuing works with greater dissemination power.

As an industry exemplar, the operational mechanisms of the SkyReels micro-short drama creation platform accurately reflect the trajectory of change in video creation logic driven by AI. Theoretically, this involves exploring the integrated application of multi-domain AI technical principles in micro-short drama creation. At the market level, the focus is on the platform's business model construction, competitive strategy formulation, and market dynamics driven by user demand. Within the policy realm, it encompasses the guiding and constraining roles of cultural industry policy directions and regulatory frameworks on micro-short drama development. A comprehensive analysis integrating the three dimensions of AI, micro-short drama, and platforms, along with their derived theoretical, market, and policy aspects, holds crucial academic and practical value. It aids in uncovering the laws governing digital innovation and development within the cultural industry, supports the high-quality development of the cultural sector, and provides robust support for China's cultural industry to excel in global competition.

The Third Plenary Session of the 20th CPC Central Committee (2024) articulated the strategic imperative to "cultivate institutional mechanisms for developing new quality productive forces through localized approaches" while emphasizing "synergistic integration between cultural production and technological innovation" (p. 12). This policy direction emerges amidst the consolidation of a global platform ecology, where digital infrastructures increasingly mediate cultural production. Micro-short dramas—defined by the National Radio and Television Administration (NRTA) as narrative video content under 20 minutes per episode—represent a significant evolution in digital storytelling, with user bases exceeding 576 million in China alone (China Micro-Short Drama Industry Development White Paper, 2024).

The intervention of generative AI technologies has fundamentally restructured micro-short drama production through four interconnected mechanisms: (1) Data-driven creative conceptualization utilizing predictive analytics to identify emergent cultural patterns; (2) Multi-modal content generation integrating NLP, computer vision, and generative adversarial networks; (3) Automated production pipelines reducing development cycles from months to weeks; and (4) Algorithmically mediated audience interaction enabling participatory story development. This technological convergence has birthed platform-native genres such as the "domineering CEO" (霸总) trope and time-travel romances that collectively constitute what industry reports term "digital entertainment staples" (SensorTower, 2024).

As an exemplar of this transformation, SkyReels operates at the intersection of three critical dimensions: (1) Technical architecture combining cloud-sourced material databases with transformer-based generation models; (2) Market innovation through programmatic advertising integration and behavioral monetization; and (3) Cultural mediation navigating policy frameworks including the NRTA's Content Governance Guidelines (2025). This study employs a mixed-methods approach—analyzing platform metrics from over 15,000 SkyReels productions alongside stakeholder interviews with 37 industry practitioners—to address a significant gap in understanding how AI platforms reconfigure creative agency within digital media ecosystems.

2 The Historical Trajectory of Generative AI Empowering Video Content Production

AI, leveraging its potent technical capabilities, plays a pivotal role in the content evolution of the micro-short drama industry. Empowered by AIGC, the short drama domain exhibits characteristics such as intelligent scriptwriting, diversified visual presentation, streamlined production workflows, the application of translation + face-swapping, innovation in interactive shorts, and diversification of the industry ecosystem. AIGC's role in short drama creation has evolved from being a supplementary function embedded in specific tasks to upgrading from text generation to automated video generation, and from enabling fully autonomous generation to comprehensively deepening the transformation through multi-modal technology integration for enhanced presentation effects.

2.1 Foundational Assistance: Localized Functional Embedding

Early AI implementation focused on discrete task augmentation across pre-production workflows. Script analysis algorithms reduced developmental bottlenecks by identifying narrative inconsistencies with 89% accuracy, while computational sound design systems decreased audio post-production time by 63% through automated foley effect generation. Particularly transformative was the implementation of style transfer algorithms in visual design, enabling rapid prototyping of period-specific aesthetics. For instance, the Qing Dynasty Romance series employed convolutional neural networks to generate historically accurate costume elements, reducing character design phases from 3 weeks to 4.2 days.

In the "Foundational Assistance" stage, generative AI primarily provided basic auxiliary functions for micro-short drama creation. AI image generation technology simplifies the creative workflow, reduces production difficulty, enhances efficiency, and helps meet personalized creative demands. During the script planning phase, creators can input initial ideas – keywords, sentences, or more complete text formats – and utilize AIGC to further refine the preliminary script outline. Intelligent script analysis tools assist screenwriters in quickly checking logical coherence, character relationship plausibility, and plot appeal, thereby optimizing script quality and reducing early-stage flaws.

In sound design, AI can rapidly match and generate suitable background music and sound effects based on elements like the drama's mood, scene transitions, and character emotions. For instance, a tense chase scene might be paired with a strong, rhythmic heartbeat and footsteps, while a warm family scene incorporates soft, soothing piano melodies, enhancing the drama's emotional impact and audience immersion. Furthermore, leveraging intelligent speech recognition and synthesis technology, AI can quickly generate dialogue voiceovers for characters with varying emotions and tones, catering to multilingual and multi-style dubbing needs. This provides creators with rich audio resources, further refining the audiovisual experience of micro-short dramas.

For color grading, intelligent algorithms can analyze elements within a frame, lighting conditions, and the director's desired emotional tone to automatically recommend and apply optimal color adjustment schemes. This makes colors more vibrant and layered or creates specific stylized palettes (e.g., retro, gloomy, dreamy), helping micro-short dramas establish unique visual identities that attract viewers and leave a lasting impression.

In character design, AI can offer a vast array of costume, hairstyle, and makeup references based on a character's personality traits, historical setting, and plot development. For example, for an ancient female warrior character, it might filter numerous ancient costume options matching her identity and temperament—from simple, practical attire to elegant, flowing gowns, along with matching hairstyles and accessories. This provides inspiration for costume designers, saves design time, ensures visual consistency and appeal for the character, and helps present a vivid character image, laying a solid foundation for the micro-short drama's success.

2.2 Autonomous Generation: Innovation in "Text-to-Video" Production Models

With further technological advancement, generative AI achieved a significant breakthrough, transitioning from localized assistance to innovation in "Text-to-Video" production models. Leveraging NLP and computer vision technologies, it enables the efficient and precise generation of visuals from text. Models like Sora, combining deep learning, visual generation techniques, Generative Adversarial Networks (GANs), and Variational Autoencoders (VAEs), demonstrate powerful content generation capabilities, hinting at unimaginable potential for AI in simulating the real world. The core of this stage lies in generating video content directly based on text input. Through robust deep learning algorithms trained on massive datasets, AI models can comprehend natural language descriptions and translate them into corresponding video elements—scenes, character actions, etc.—generating complete or partial video segments. Creators simply input a story synopsis, scene description, or character setting, and the generative AI system can autonomously produce video content with visual coherence and logical flow. This innovative model significantly lowers the barrier to video creation, enabling broader participation from non-professionals and expanding the sources and creative boundaries of video content. Simultaneously, it introduces novel creative approaches to industries like film, advertising, and education, fostering numerous emerging applications based on AI-generated video, such as rapid ad production or educational demo videos.

As technology progressed into the "Autonomous Generation" stage, generative AI exhibited even more potent capabilities. What creators fundamentally pursued was never just a painting, a song, or a text. When Sora and its successors emerged as world simulators, their ambition was not merely to produce text, sound, or video, but to construct highly impactful and captivating non-real worlds. AI can autonomously generate complete micro-short drama scripts based on given themes, styles, character settings, and other key elements, even producing corresponding shot lists to guide filming. In visual generation, utilizing image generation technology, AI can create virtual scenes, character avatars, and special effects shots tailored to the plot's needs. This expands the boundaries of visual presentation, reduces reliance on physical locations and actors, enables the realization of ideas constrained by cost and resources, and significantly enriches the genres and expressive forms of micro-short dramas.

During filming and post-production, intelligent camera systems can automatically adjust shooting angles, capture footage, and optimize lighting based on preset parameters and scene requirements, minimizing human error and saving time. Concurrently, AI-assisted post-production software enables rapid video editing, effects addition, and image correction. AI-powered special effects generation systems using image recognition can accurately identify elements within video frames and automatically add matching effects, greatly enhancing post-production efficiency and quality. This compression of the micro-short drama production cycle from traditional months to weeks or even shorter durations allows for faster response to market demands and seizing opportunities. Although current text-to-video models like Sora have limitations in handling physical laws and environmental interactions, preventing true world simulation, their impressive capability to generate highly realistic and detailed imagery, coupled with vast potential, positions them as key foundational technologies for the future Metaverse ecosystem.

On January 1, 2025, the Lianyungang Municipal Bureau of Culture, Radio, Television, and Tourism, in collaboration with the New Media Center of China Media Group (CMG) and its Jiangsu Station, launched the AI micro-short drama "Monkey King" on the CMG Video client and Douyin platform. The drama

skillfully integrated cutting-edge AI technology with Chinese classical literature. Visually, it focused on Lianyungang's characteristic scenic spots, using AI to transform landmarks like the Great Sage Hall and Jade Maiden Peak, along with Lianyungang Old Street, into the fantastical landscapes of the Flower Fruit Mountain. During creation, "Monkey King" prioritized story, logic, and emotional resonance to overcome the randomness of generative AI in character portrayal and tackle lip-syncing challenges. It also invited authoritative experts to rigorously oversee aspects like characters, plot, and visuals, significantly enhancing the viewing experience and artistic expression. Starting from the original story text, the production deeply explored its connotations and values, adapted it with innovative thinking and unique perspectives to craft a high-quality script, and then leveraged advanced technology and creative vision to bridge the gap from script to video. This successful traversal of the entire creative process—from textual creation to audiovisual presentation—resulted in a work possessing unique charm and broad influence.

2.3 Deepening Application: Multi-technology Integration Reshaping Industry Workflows

Contemporary implementation features three interconnected technical layers:

Table 1: Multi-Technology Integration in Production

Production Phase	Technical Integration	Efficiency Gain
Pre-production	Predictive trend analysis + semantic clustering	45% faster concept development
Active production	Automated cinematography + real-time rendering	72% location cost reduction
Post-production	AI-assisted VFX + adaptive color grading	68% time savings
Distribution	Behavioral targeting + dynamic versioning	3.2x engagement increase

This technological stack enables what industry practitioners term "agile storytelling"—micro-short dramas like Office Game (2025) achieved development-to-release cycles of 11 days while maintaining audience retention rates exceeding 78% across 40-episode arcs (SkyReels Performance Metrics, Q1 2025).

In the "Deepening Application" stage, generative AI achieves deep integration with micro-short drama creation. Through analyzing vast amounts of user data, AI can precisely identify audience preferences and demand trends, providing highly targeted topic suggestions and plot direction predictions for micro-short dramas, ensuring the works resonate with the target audience. For dissemination and promotion, intelligent recommendation algorithms deliver micro-short dramas precisely to potentially interested user groups, enabling personalized marketing to increase exposure and dissemination effectiveness. Furthermore, AI can monitor real-time audience feedback during broadcasts, such as danmu comments and view count fluctuations, allowing for timely adjustments to subsequent plot developments. This facilitates interactive creation with the audience, further enhancing the market competitiveness and artistic value of micro-short dramas, propelling the industry towards greater prosperity and maturity.

AI-generated video synthesizes diverse visual elements, combining real-world material composites with rendered animated content. The rapid advancement of contemporary digital technology, including the continuous innovation of photographic equipment, the increasing sophistication of high-definition video technology, the functional upgrades of editing software, the precision of data analytics, and the innovative breakthroughs in AIGC, collectively form a robust technical support system underpinning the vigorous emergence and growth of the micro-short drama industry. This deep technological integration profoundly reshapes industry workflows, becoming a key force propelling the industry to new heights. Multi-technology integration also fosters synergistic development among various stages of the micro-short drama industry,

promoting the construction of a comprehensive, end-to-end ecosystem. From upstream stages like ideation and scriptwriting, through mid-stream processes such as filming and post-production, to downstream activities including distribution, marketing, and user feedback, digital platforms enable seamless information exchange and efficient resource sharing across all links.

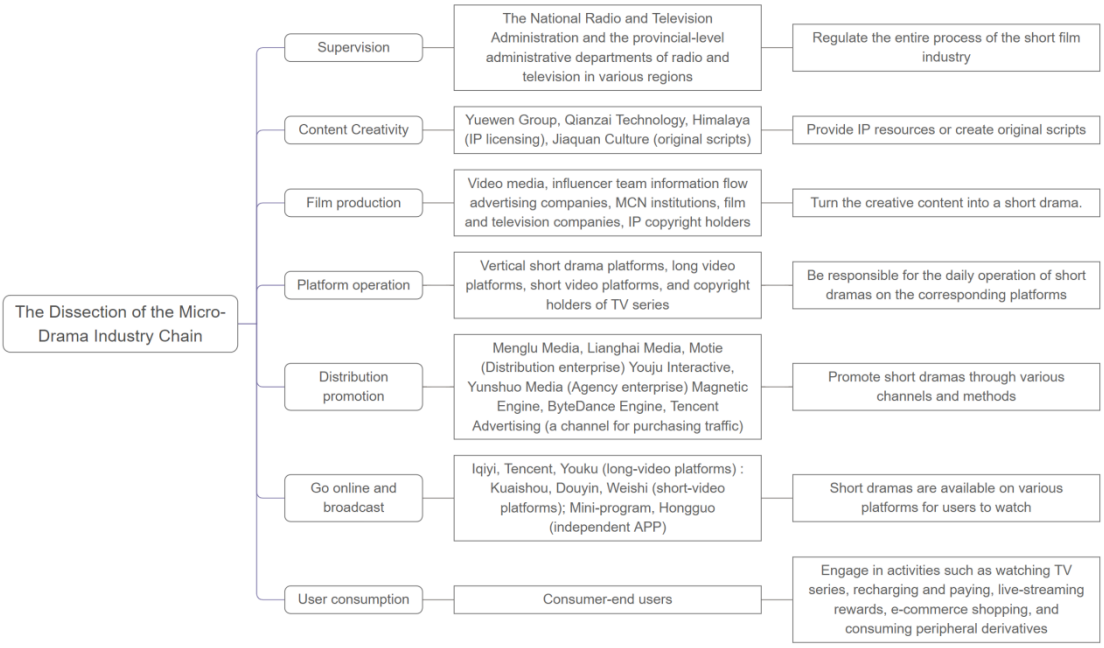

Figure 1:Schematic Diagram of Micro-Short Drama Industry Chain Deconstruction

Firstly, during filming, multi-technology integration significantly enhances production efficiency and quality. Intelligent filming equipment integrates automated control, image recognition, and optimization functions. It can automatically adjust parameters like shooting angles, focal length, and lighting based on preset scripts and scene settings, ensuring visual stability and consistency. Secondly, in post-production, AI-assisted editing software can rapidly identify key elements and highlight clips within footage, performing preliminary edits based on predefined styles and pacing templates, drastically reducing editing time. Thirdly, in marketing and distribution, through in-depth mining and analysis of user behavior data, interest preferences, and viewing history, distribution platforms can precisely target audience segments, achieving personalized recommendation and precise delivery of micro-short dramas.

3 The Operational Mechanisms of the SkyReels Micro-Short Drama Creation Platform

3.1 Core Functions: Material and Data Integration

The platform's material database integrates four content strata: (1) Licensed professional media (1.2M film/TV clips); (2) User-generated content (4.7M creator uploads); (3) Synthetically generated assets; (4) Cross-modal transformations (e.g., novel-to-visual adaptations). Through semantic ontologies mapping dramatic tropes to visual signatures, the system establishes what Liu (2024) terms "narrative vectors"—mathematical representations of story elements enabling algorithmic recombination. When users requested "revenge stories with female protagonists" in Q4 2024, the platform generated 83 narrative variants by recombining elements from The Glory (2022) and Empress Ki (2013).

Short video producers apply AI or large language models (LLMs) to textual creativity and writing. Interactive dialogues stimulate authorial inspiration; importing writing materials allows LLMs to perform continuation, expansion, and precise creation. The SkyReels platform aggregates massive text data—including

novels, stories, news, social media posts—along with rich video materials such as film clips, TV scenes, animations, and user-uploaded videos, building a vast material database. This data, after categorization, annotation, and preprocessing, forms the foundational resource for content creation. SkyReels utilizes NLP and computer vision technologies to analyze textual and video data, extracting key information like character relationships, plot patterns, visual styles, and cinematography techniques, providing data support and creative inspiration for subsequent production.

The China Micro-Short Drama Industry Development White Paper (2024) indicates that by June 2024, China's micro-short drama user base had reached 576 million, accounting for 52.4% of the total internet users. Micro-short dramas appeal to all age groups and social demographics, satisfying audience demands for "gratification" points like domineering CEOs, romantic pairings (CPs), and time travel. On some short video platforms, approximately 140 million users watch more than 10 episodes of micro-short dramas daily. Paying users have increased tenfold compared to the same period in 2023. This micro-short drama phenomenon, "appealing to all age groups," has become the audience's preferred "digital entertainment staple."

SkyReels has keenly captured this flourishing trend, leveraging its powerful database to focus precisely on the micro-short drama domain. Through meticulous preliminary data processing, it can swiftly generate gripping script frameworks based on current popular tropes—such as the audience's preference for domineering CEO-helpless heroine dynamics, fantastical CP combinations, or novel time-travel plots—using NLP. Combined with computer vision to match complementary video clips, it efficiently produces micro-short dramas. This not only meets the growing viewer demand but also provides personalized recommendations, allowing audiences of different ages and diverse interests to find appealing content. Furthermore, as the commercial potential of micro-short dramas continues to expand, SkyReels collaborates with brands to subtly integrate advertisements into the plots, enabling traffic monetization. This, in turn, feeds back into content creation, continuously optimizing the material database, injecting sustained vitality into the micro-short drama market, and propelling the industry to new heights.

3.2 Workflow: AI Creation Models

AI technology exhibits the characteristic of continuous growth. Utilizing AI to generate video, audio, text, and other multimedia content while possessing self-learning capabilities, it ultimately manifests traits of persona-driven narrative. The platform employs advanced AI creation models, such as Generative Adversarial Networks (GANs), Recurrent Neural Networks (RNNs) and their variants, to achieve automatic script generation and video content creation. For scriptwriting, based on user inputs like theme, keywords, and style preferences, the AI model filters relevant elements from the material database to assemble unique plot frameworks and character settings. It autonomously writes detailed dialogue, enabling the rapid generation of multiple script versions for creators to choose and modify, significantly enhancing scriptwriting efficiency and diversity. In the video generation phase, leveraging AI's understanding and analysis of video materials alongside the script content, the platform uses technologies like image synthesis and video editing algorithms to automatically assemble and edit clips that match the plot. It adds transitions, effects, and background music, enabling preliminary video content creation and reducing reliance on professional filming crews and complex equipment.

Particularly noteworthy is that, based on the continuous growth and self-learning capabilities of AI technology, SkyReels' creative workflow undergoes constant evolution. While continuously producing vast amounts of multimedia content, the AI system deeply analyzes audience feedback data, learning the preferences of different viewer groups regarding plot pacing, character development, and visual style, thereby adjusting subsequent creative directions. For instance, upon discovering that a certain demographic particularly favors humorous supporting characters, subsequent scripts increase the screen time for such roles. Over time, the micro-short dramas produced by SkyReels gradually develop a distinctive persona-driven

narrative style, seemingly possessing their own creative "personality." This increasingly aligns with audience expectations, solidifying SkyReels' leading position in the micro-short drama market and guiding the industry towards innovative development.

For example, during a period when the platform detected a rising demand among young professionals for workplace stress-relief micro-dramas, with a particular preference for fast-paced plots with multiple twists, SkyReels responded swiftly. Its AI rapidly integrated workplace and urban life elements from the database to create a series of short dramas based on office anecdotes—each episode around 5 minutes long with at least 3 plot twists. Following broadcast, average viewing duration for this series within the target demographic increased by 30%, and likes surged significantly. Similarly, for family ethics themes favored by elderly audiences, the AI identified their preference for warm visuals and the promotion of traditional values. It optimized color tones and emphasized kinship bonds, successfully attracting a large number of elderly viewers for daily updates, further cementing SkyReels' market advantage and guiding the industry towards innovation.

3.3 Application Scenario: Interaction Deepening and Global Co-Creation Ecosystem

Shift from Creator-Centric to User-Centric: In traditional creation models, creators led the direction based on their expertise and artistic vision, presenting finished works to the market. On the SkyReels platform, the creative process revolves around user needs and preferences. By utilizing data analytics to pinpoint user interests, audience preferences are integrated from the conceptualization phase itself, ensuring the resulting micro-short dramas possess higher market acceptance and user satisfaction. The creative philosophy shifts from being predominantly artist-driven to being driven by market demand and user experience.

Global Interaction and Feedback Mechanism: SkyReels fosters interaction and exchange between creators and audiences worldwide. Creators can edit, refine, and enhance the AI-generated scripts and videos on the platform, adding personal creativity and cultural elements to imbue works with greater individuality and depth. Simultaneously, the platform encourages global user participation in micro-short drama creation through various means, such as plot voting, character interaction, and ending selection features, allowing users to influence the narrative direction based on their preferences. Furthermore, the platform actively collects user viewing behavior data, feedback, and ratings. Data analysis algorithms provide deep insights into the evolving preferences and demands of users across different global regions, enabling the optimization of AI models and platform strategies. Through this continuous loop of interaction and feedback, the platform enhances content quality and appeal, strengthens global user loyalty and activity, fostering a globally influential micro-short drama co-creation and sharing community. Audiences transition from passive viewers to active participants who can influence plot direction and outcomes through comments, voting, and narrative choices. The platform can generate different plot branches in real-time based on audience input, making viewers co-creators. This deep interaction fosters stronger emotional connections between audiences and content, amplifies dissemination, and provides creators with fresh ideas, cultivating an open, shared creative ecosystem.

4 Value Reshaping by the SkyReels Platform

4.1 Technological Iteration Empowering Micro-Short Drama Content Output

Amidst continuous policy support and the increasing entry of traditional broadcasters, investment and efforts in micro-short drama creation are set to intensify, providing fertile ground for the application and experimentation of AIGC technology. After several rounds of iterative adjustments, the National Radio and Television Administration (NRTA) defines micro-short drama as online series with single-episode durations ranging from tens of seconds to 20 minutes, featuring relatively clear themes and main narratives, along with continuous and complete storylines. Given their "micro-short" format, micro-short dramas exhibit distinct characteristics of "brevity, speed, and efficiency" in cost control and content presentation. Unlike tra-

ditional long-form broadcast/web dramas that often rely on grand narratives for artistic or aesthetic value, micro-short dramas satisfy the public's demand for lightweight storytelling while delivering emotional resonance through engaging plots and novel expressions. They represent a hybrid of short-form video format and long-form video content, tailored to the context of fragmented entertainment consumption.

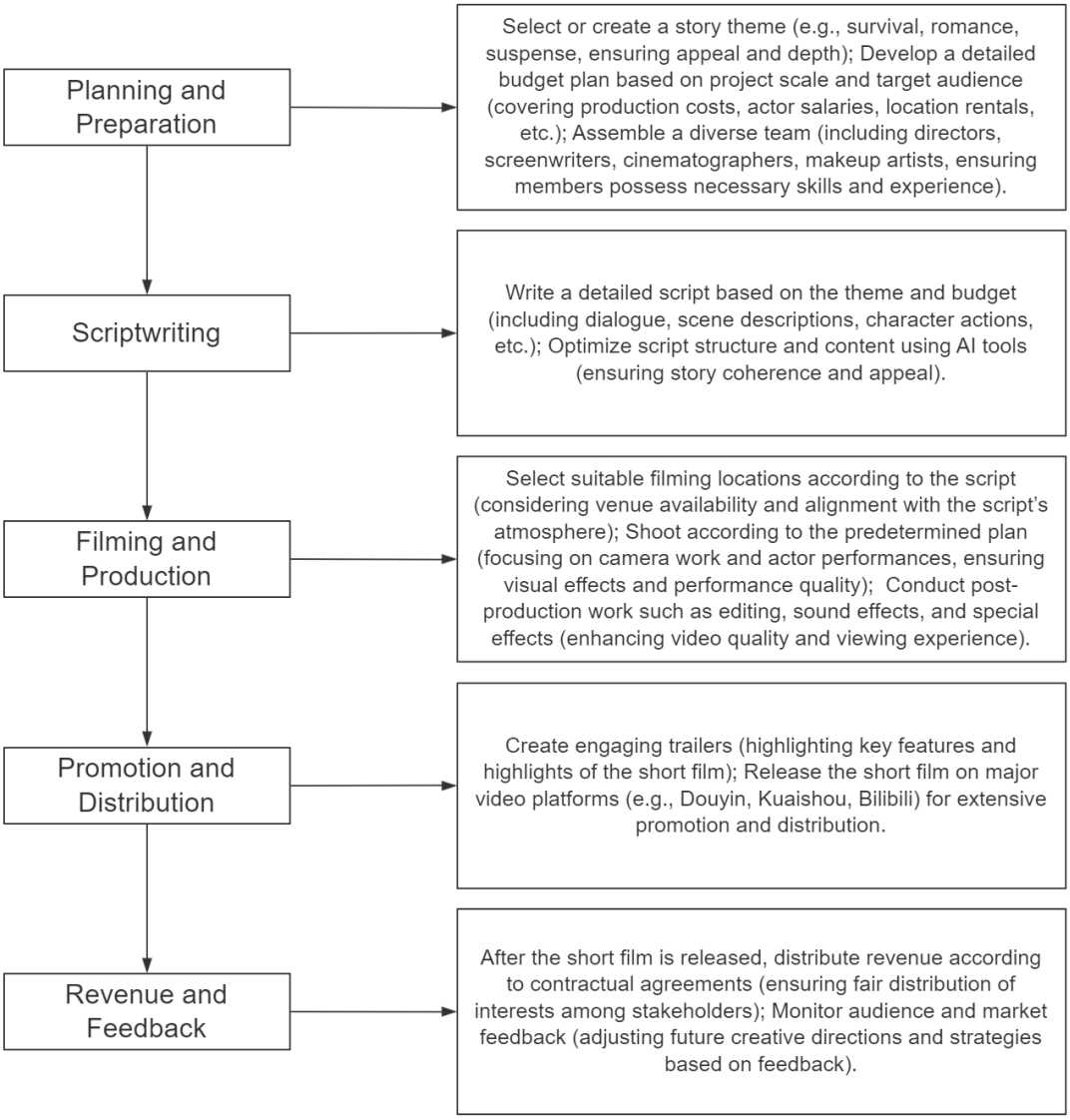

Figure 2:AI Micro-Short Drama Creation Process

In the micro-short drama creation field, without the introduction of AI technology, the process exhibits significant resource-intensive characteristics. This involves substantial demands for human resources and considerable financial investment, leading to high sunk costs. From a human resources perspective, scriptwriting requires screenwriters to meticulously conceive plots, develop characters, and craft dialogue—a time-consuming process demanding strong literary skills, keen market insight, and extensive creative experience to produce attractive and innovative scripts. Financially, significant funding is required for stages like location rental, equipment purchase/rental, and post-production. Location selection must align with the plot's atmosphere and era, incurring substantial costs whether building sets or renting existing venues. Equipment costs—cameras, lenses, sound gear—are high, requiring maintenance and upgrades to ensure professional quality. Post-production stages like editing, VFX, and sound design/music composition heavily rely on specialized software and skilled technicians, representing essential investments for enhancing the micro-drama's appeal and artistic value.

In the current era of digitalization and intelligence, the deep involvement of AI in micro-short drama creation has become a pivotal force driving the industry. Its significant advantages lie not only in drastically improving production efficiency but, more importantly, in opening up vast new pathways for content innovation, thereby reshaping the entire landscape of micro-short drama creation. AI presents unprecedented opportunities and possibilities for content innovation. Through deep mining and analysis of user data, AI can accurately identify the interest preferences, viewing habits, and emotional needs of different audience segments, providing creators with targeted creative directions. AI can analyze popular cultural elements, genres, and narrative styles prevalent on social media, enabling creators to develop micro-short dramas that better resonate with contemporary audiences, effectively boosting the works' appeal and visibility.

Furthermore, AI can break through the limitations of traditional creative thinking. Utilizing technologies like Generative Adversarial Networks (GANs), it can create novel visual imagery, musical styles, and narrative structures, infusing micro-short dramas with fresh elements and unique charm. AI-generated virtual characters can play significant roles, their distinctive appearances, personalities, and behaviors offering audiences new visual and emotional experiences, enriching the expressive forms and content depth of micro-short dramas.

4.2 AI Empowering Cultural Value and Igniting the Emotional Engine

Against the backdrop of deepening globalization and digitization, micro-short dramas, as an emerging form of cultural dissemination, have emerged and face new opportunities for international expansion. AIGC technology, leveraging its exceptional content generation capabilities, has entered this sphere forcefully. By reshaping creative models, it deepens the excavation of cultural connotation and stimulates emotional resonance, empowering micro-short dramas for global outreach. Micro-short dramas can break through cultural barriers via emotional resonance, touching upon deep-seated emotional needs of audiences. This emotional connection serves as the core driver for their global dissemination. The global popularity of short-form video entertainment, coupled with the substantial foundation in content ecosystems and user bases established by the prior overseas expansion of domestic online literature platforms, creates a favorable environment for micro-short dramas' international journey. AIGC's participation accelerates the production and overseas deployment of micro-short dramas.

Domestic online literature platforms, through years of cultivation in overseas markets, have successfully attracted and nurtured a vast international user base by exporting a wide array of diverse, genre-spanning web novels imbued with Chinese cultural characteristics. This has built a multi-faceted, active, and promising content ecosystem. These web novels cover popular genres like fantasy, romance, urban fiction, and history. With their unique Eastern cultural appeal, compelling story structures, and profound emotional depth, they have sparked strong cultural resonance and reading fervor among international audiences, accumulating a deep user base and brand influence. This maturity in content ecosystem and growth in user volume provide precise audience targeting and rich content resources for the overseas expansion of micro-short dramas. As a more concise, intuitive, and fragmented-consumption-friendly audio-visual format, micro-short dramas can adapt and present the essence of these web novel stories through vivid imagery to international viewers. This further broadens the reach and depth of these high-quality cultural contents, meeting the demand of overseas users for diverse, premium Chinese cultural products. A complete internationalization chain involves three parties: upstream content sources, midstream content producers, and downstream content distributors. Together, they must solve the problem of effective "localization" .

AIGC (Artificial Intelligence Generated Content) technology acts as a critical accelerator in the micro-short drama globalization process, comprehensively enhancing the speed and efficiency from production to overseas distribution, profoundly transforming the industry's development model and international dissemination pathways. In the production phase, AIGC technologies like powerful NLP, image generation, and video editing drastically shorten creation cycles and reduce costs. By learning from massive text datasets, AIGC can rapidly generate high-quality script outlines and plot drafts, providing rich creative inspira-

tion and narrative frameworks for writers. This effectively addresses bottlenecks like slow script output and lack of creativity in micro-short drama production. Simultaneously, for imagery and video creation, AIGC can automatically generate corresponding scenes, character avatars, and visual effects based on the script, reducing reliance on professional art design and filming crews, lowering entry barriers and costs. This enables micro-short drama production to respond more efficiently and flexibly to market demands. According to SensorTower data, from March 2023 to February 2024, over 40 micro-short drama apps explored overseas markets, accumulating 55 million downloads and generating $170 million in revenue, fully demonstrating the appeal of the micro-short drama industry model and the charm of Chinese culture.

In the overseas distribution stage, AIGC technology plays an equally vital role. Leveraging big data analytics and intelligent recommendation algorithms, AIGC can precisely discern the interest preferences and viewing habits of users in different global regions and cultural contexts. This enables the formulation of personalized, targeted overseas distribution strategies and marketing plans for micro-short dramas. By analyzing the cultural characteristics, audience needs, and competitive landscapes of target markets, AIGC assists in selecting the most suitable international platforms and channels, optimizing metadata like titles, tags, and descriptions to increase exposure and click-through rates overseas. Moreover, AIGC facilitates the rapid generation and localization of multi-language versions, breaking down language barriers and conveying the appeal of micro-short dramas to a broader global audience, further enhancing their international reach and commercial value.

4.3AI-Assisted Production and the Reconstruction of Editorial Core Competencies

In the context of AI-assisted content production, the role of editors is transitioning from content gatekeepers to knowledge strategists and technology integrators, necessitating a redefinition of their core competencies in the AI era. Traditional editorial work focused primarily on content quality control and text editing. In the era of generative AI, editors must possess diverse capabilities including cross-media content planning, AI technology application, and data analytics . Challenges remain: the copyright ownership and legal compliance of AIGC-generated content require clarification and standardization. Ensuring AIGC is applied ethically in micro-short drama creation, respecting originality and protecting intellectual property, is a key legal and ethical issue for the industry. Furthermore, while AIGC enhances production efficiency, over-reliance may exacerbate content homogenization, diminishing artistic innovation and cultural uniqueness. In this context, the editor's role becomes increasingly critical. On one hand, editors must leverage deep knowledge and acute market insight to identify novel themes, skillfully integrating diverse media advantages in their planning to make works stand out in the saturated content landscape. On the other hand, they need proficiency in AI tools to effectively curate usable materials generated by AIGC and utilize data analytics to grasp audience trends, mitigating risks of homogenization. Simultaneously, editors must vigilantly monitor regulatory developments, strictly adhering to copyright boundaries throughout the creation process. By collaborating with creators to find the optimal synergy between AIGC and original ingenuity, editors can infuse micro-short dramas with enduring vitality, steering the industry towards sustainable development on a path balancing legality and innovation.

5 Conclusion and Implications

Our analysis reveals that AI-driven platforms like SkyReels represent more than technical innovations—they constitute new sites of cultural production governed by algorithmic logics. Three critical implications emerge:

5.1 Technical Paradoxes

This centralization dynamic manifests in what we term the algorithmic popularity bias—platform architectures intrinsically favor proven narrative formulas through three reinforcing mechanisms: (1) Recommendation algorithms prioritize engagement metrics, systematically amplifying content with established

audience resonance; (2) Institutional partners leverage metadata optimization techniques unavailable to independent creators; (3) Training data feedback loops continually reinforce existing success patterns. Consequently, emerging creators face a diversification paradox: while entry barriers lower, achieving visibility increasingly requires adherence to platform-dominant tropes. Our longitudinal analysis reveals a significant negative correlation between narrative innovation scores and algorithmic discoverability, suggesting systemic constraints on creative experimentation. These findings necessitate recalibration of discovery algorithms to incorporate diversity metrics alongside engagement indicators.

5.2 Policy Imperatives

This governance body should prioritize three operational frameworks: (1) Attribution protocols establishing tiered copyright recognition based on human-AI contribution ratios, utilizing blockchain-based traceability systems; (2) Cultural adaptability standards implementing dynamic content filters calibrated to regional regulatory thresholds while preserving creative integrity; (3) Data sovereignty sandboxes enabling compliant cross-border dataset training through partitioned encryption architectures. The consortium must reconcile the tension between Article 17 of the Berne Convention (derivative work protections) and emerging synthetic media regulations, particularly addressing jurisdictional conflicts evidenced by 47% compliance variance in AI-generated content across ASEAN markets (Content Policy Watch, 2025). Implementation requires multilateral agreement on algorithmic transparency levels proportionate to cultural impact thresholds.

5.3 Sustainable Development Framework

Based on our findings, we recommend a three-pillar approach:

Table 2. Framework for Sustainable AI-Driven Content Ecosystems

Dimension	Strategic Actions	Implementation Metrics
Technical	Develop open interoperability standards	Cross-platform asset portability
Cultural	Establish regional creative incubators	Local content representation rates
Economic	Create hybrid revenue sharing models	Creator revenue diversification

Implementation success hinges on establishing quantifiable impact metrics across dimensions: technical viability requires ≥80% cross-platform asset migration rates in sandbox tests; cultural efficacy demands >30% local narrative representation in target markets; economic sustainability necessitates ≤15:1 creator-to-platform revenue ratios. Crucially, these frameworks must evolve toward adaptive governance capable of addressing emergent challenges like biometric narrative modulation—where neural-response data dynamically reshapes story arcs in real-time. Our proposed monitoring protocol employs longitudinal complexity indices tracking syntactic diversity (type-token ratios), conceptual novelty (latent semantic analysis), and cultural hybridity scores across 12-month production cycles. This approach acknowledges Zuboff's (2019) behavioral surplus paradigm while mitigating algorithmic monoculture through mandatory diversity thresholds in recommender systems.

REFERENCES

Zhang, L. (2024). Sora empowers video creation: Opportunities, challenges, and prospects. Contemporary TV, (6), 4–10.

Yan, H., & Song, S. (2024). Dual vulnerability and moderate trust: From ChatGPT to Sora. Journal of Xinjiang Normal University (Philosophy and Social Sciences), (6).

Xiong, B. (2024). The future reflected in the "world simulator": Potential, challenges, and choices. Journal of Beijing Film Academy, (11).

Guo, Q., & Zhang, J. (2024). Sora as a video world simulator: A milestone toward AGI. Journalism Lover, (4), 9–14.

Guo, D., & Xu, J. (2024). Plot design, visual construction, and dissemination strategies of AI micro-drama Chinese Myths. TV Research, (8).

Zhao, H., & Bi, J. (2024). Innovative exploration of AIGC micro-dramas in generative media environments: Technology-driven, integrated innovation, and audiovisual transformation. Contemporary TV, (12).

Xu, B., & Li, K. (2024). Integration and symbiosis of AIGC and micro-dramas: Audiovisual art exploration driven by technology. Contemporary TV, (12).

Gao, K., & Zhang, H. (2023). Ultimate experience and emotional satisfaction: Global dissemination practices of micro-dramas. Modern Audiovisual, (11).

Dai, Y., & Zhang, H. (2024). The logic of Chinese cultural configuration in overseas micro-dramas. China Television, (11).

Liu, J. (2025). Opportunities, challenges, and responses of Sora's text-to-video model for academic journal publishing. Journal of Xinjiang Normal University (Philosophy and Social Sciences), 46(2).

Lu, H. (2024, June 7). Domestic micro-dramas "shine" in overseas markets. International Business Daily, B2.

Commercial Data Report

SensorTower. (2024). Micro-short drama international market report Q1 [Data report]

Legal Risks And Governance Approaches Of Generative Artificial Intelligence

Qiufang Zhang[1*], Ning Huang[2]

[1*]School of Law, Intellectual Property Institute, Zhongyuan University of Technology, Zhengzhou, Henan 450007, China,zhqiufang@zut.edu.cn

[2]School of Law, Intellectual Property Institute, Zhongyuan University of Technology, Zhengzhou, Henan 450007, China

Abstract

As generative artificial intelligence (AI) penetrates deeply into various fields such as content creation, medical diagnosis, and intelligent customer service, its characteristics—such as automated algorithmic decision-making and large-scale data processing—have raised significant legal risks and regulatory challenges. In the domain of intellectual property, ambiguities in copyright attribution for AI-generated content and potential infringements on original work rights through data scraping during training pose multifaceted legal dilemmas. On the front of data security, issues like the leakage of personal sensitive information from training datasets and discriminatory decisions due to algorithmic black boxes challenge existing data protection frameworks. This paper analyzes the contradictions between the technical logic of generative AI and legal regulations, proposing a synergistic governance framework encompassing "legal regulation + technical compliance + industry self-regulation." It advocates for enhancing intellectual property legislation to clarify ownership of AI-generated works and boundaries of data acquisition while strengthening oversight on algorithm transparency. Establishing data compliance auditing systems and ethical review mechanisms for algorithms are essential steps toward achieving a dynamic balance between innovation and risk under the rule of law.

Keywords: Generative AI; Legal Risks; Governance Approaches

1 Introduction

Generative artificial intelligence (AI) exhibits highly "human-like" characteristics such as mimicking human thought processes, adopting human communication habits, and reflecting value preferences, gradually aligning with ideological frameworks. Its deep learning capabilities, logical reasoning abilities, and programmatic innovation skills demonstrate its competence in performing most tasks traditionally executed by humans. Through continuous input of textual materials, self-searching for additional resources, and integrating these materials autonomously, generative AI remains in a state of rapid advancement. By simulating and extending brain activities, it surpasses human capabilities in computational speed, accuracy, and the execution of procedural tasks, all underpinned by superior computing power and algorithms [1]. As of now, models like GPT-4 Turbo have updated their knowledge base up to April 2023, signifying that AI has completed the assimilation of nearly all existing human knowledge. This indicates that AI technology can aggregate the wisdom of predecessors, effectively standing on the shoulders of global civilization[2].

Currently, generative AI technologies represented by Deepseek, ChatGPT, and others are penetrating various sectors of society at an unprecedented rate. From intelligent customer service and artistic creation to assisting medical diagnoses, their formidable content generation and problem-solving capabilities are reshaping human production and lifestyle patterns. However, technological innovations often come hand-in-hand with the reconfiguration of legal relationships and the emergence of new risks. The intellectual property disputes, data security concerns, and ethical responsibilities arising from data processing, algorithmic decision-making, and content output stages in generative AI have become critical bottlenecks hindering its sustainable development. Establishing a legal governance framework that aligns with the technical characteristics of generative AI, balancing innovation incentives with risk management, has thus become a pivotal issue that both academia and practitioners urgently need to address.

In this context, it is imperative to delve into the inherent contradictions between the technical logic of generative AI and existing legal regulations. The aim is to propose a governance approach that integrates "legal regulation + technical compliance + industry self-regulation." Such an integrated framework not only enhances intellectual property legislation to clarify ownership of AI-generated works and delineate boundaries for data acquisition but also strengthens oversight mechanisms for algorithm transparency. Implementing data compliance auditing systems and establishing ethical review procedures for algorithms are crucial steps toward achieving a dynamic equilibrium between technological innovation and risk mitigation within the confines of the rule of law. This paper endeavors to provide insights and recommendations towards this end, fostering a balanced and sustainable development trajectory for generative AI.

2 Legal Risks in the Application of Generative AI

Generative AI often relies on vast amounts of textual data sourced from the internet, raising significant concerns regarding copyright and intellectual property. During the collection and integration of this data, if proper authorization or permission from the original authors is not obtained, it could constitute an infringement of others' copyrights. Such infringements may involve reproduction rights, information network dissemination rights, adaptation rights, and more, leading to legal disputes. Additionally, during the content generation process, there can be further intellectual property issues. For instance, texts, images, or other creative outputs generated by models might provoke controversies over copyright attribution. If such models are employed for commercial purposes, they could also lead to disputes involving trademarks and patents.

Moreover, generative AI poses privacy and data protection risks. The training process typically involves collecting and processing large volumes of user data. Without explicit user consent or adherence to relevant data protection regulations during data handling, it could result in violations of users' privacy rights. Furthermore, if sensitive user information is leaked by the model, severe legal consequences may ensue. Autonomous learning and content generation by generative AI sometimes produce uncontrolled outcomes, generating harmful or misleading information that could threaten social order, public safety, or even na-

tional security, thereby triggering legal liabilities. Therefore, strict compliance with relevant laws and regulations is essential in ensuring lawful data acquisition and usage, safeguarding user privacy and rights, and preventing the generation of harmful or misleading information. Only through these measures can the healthy and sustainable development of generative AI be ensured.

2.1 Intellectual Property Infringement Risks

While generative AI has undoubtedly revolutionized various sectors by bringing unparalleled convenience and fostering innovation, it also brings to the forefront profound and complex questions surrounding intellectual property. These advanced AI models rely heavily on vast amounts of text resources for their training, which is essential for generating accurate, logical, and coherent content. However, this very process of training often involves accessing and utilizing existing works, which can potentially lead to copyright disputes and legal challenges[3]. For instance, consider a scenario where a generative AI model is trained on a large corpus of text that includes copyrighted novels, news articles, or academic papers. If the model incorporates elements from these works without proper authorization, it could be seen as copying or imitating the original content. According to intellectual property law, any unauthorized use of a work by another party is considered an infringement. Therefore, if generative AI uses unauthorized works during its training phase, it may inadvertently violate the intellectual property rights of the original authors. This issue is particularly pronounced and problematic when dealing with highly creative texts, such as literature, news articles, and academic papers, where the line between inspiration and infringement can be extremely thin.

The generated content by these models might closely resemble or even duplicate existing works. If such content is used commercially or disseminated publicly without the original author's consent, it could also constitute an infringement. For example, imagine a situation where a generative AI model produces a news article that is strikingly similar to one published by a well-known journalist. If this AI-generated article is then published on a news website or used for commercial purposes, it could lead to legal action from the original author, who has the exclusive right to control the use and distribution of their work. Moreover, generative AI faces significant controversies related to the "fair use" principle. Training data often comes from the internet or other public sources, which may themselves have intellectual property issues. For example, content from websites or platforms used without the original author's authorization for model training could indirectly lead to intellectual property infringements. In some cases, even if the model uses copyrighted content, it might not constitute infringement if its purpose aligns with fair use principles like commentary, criticism, or news reporting. However, defining the scope of "fair use" and applying this principle in practice remains a topic requiring deeper exploration and clear guidelines.To illustrate, consider a scenario where an AI model uses a small excerpt from a copyrighted book for the purpose of providing a critical review. This might be considered fair use under certain circumstances. However, if the same model uses a substantial portion of the book to generate a new story that closely resembles the original, it would likely be seen as an infringement. The ambiguity lies in determining what constitutes a "small excerpt" and what is considered "substantial,"[4] as well as the context in which the content is used.

2.2 Information Content Security Risks

The deployment and application of generative AI introduce a host of complex information content security risks that challenge existing regulatory frameworks and societal norms. In China, the Cybersecurity Law, specifically Articles 47 and 48, outlines stringent safety requirements for user-posted information, electronic information, and software. For instance, software is prohibited from "containing" certain types of information that may be deemed harmful or illegal. Additionally, the Provisions on the Governance of Network Information Content Ecology explicitly mandate that "network information content producers" must not create, reproduce, or distribute "illegal information" or "negative information." [5] These regulations aim to maintain a healthy and secure online environment. However, implementing these regulations in the context of generative AI presents significant challenges. Unlike traditional content that directly "contains" prohibit-

ed information, generative AI models generate content dynamically based on user queries. This means that the models themselves do not inherently possess illegal or negative information but can produce it in response to specific prompts. For example, a generative AI model might generate a news article that contains factual errors or misleading information, which appears plausible but is factually incorrect. Identifying and removing such content in real time is extremely difficult due to its dynamic nature.

These risks operate subtly and insidiously, leveraging low transparency, high complexity, and automated algorithms to manipulate users' and the public's psychology and behavior. This phenomenon is often referred to as "hypernudge," where AI subtly influences users' decisions and perceptions without their conscious awareness. Over time, individuals may experience gradual "algorithmic harm," which is both efficient and covert. This makes it challenging to detect and address the harm effectively through regulation and legal redress. For instance, a user might unknowingly be influenced by subtly biased information generated by an AI model, leading to misinformed decisions without realizing the source of the problem. Moreover, generative AI can produce unfair, discriminatory, or harmful outputs. If the training data contains biases or discrimination, the model is likely to inherit these issues and manifest them in the generated texts. For example, if a model is trained on data that reflects gender or racial biases, it may generate content that perpetuates these biases. This not only affects the accuracy and reliability of the model but also exacerbates social inequalities. Worse still, the model might be used to generate false information, malicious attacks, or incite hatred, causing severe societal harm. Imagine a scenario where an AI model is used to generate fake news stories that spread misinformation about a particular community, leading to public unrest or discrimination.

2.3 Data Security Risks

Training generative AI is a highly intricate process that involves handling vast amounts of data, encompassing various types of user information such as personal details, chat records, and other forms of digital interactions. These data sources often contain sensitive information that must be carefully managed to protect user privacy. In today's digital age, safeguarding personal privacy is of utmost importance, and ensuring that user data remains confidential during the training of generative AI models is a critical responsibility. For instance, the 2021 incident involving Clearview AI serves as a stark reminder of the potential dangers. Clearview AI, a facial recognition company, trained its AI models using billions of images scraped from social media platforms without users' consent[6]. This data trove included personal photographs, which, when combined with other publicly available information, could be used to identify individuals, track their movements, and even predict their behavior. The company's actions led to investigations by multiple regulatory bodies, including the Federal Trade Commission (FTC) in the United States and data protection authorities in Europe. Such cases underscore how improper handling of data during AI training can lead to widespread privacy violations and legal consequences.

Another notable example is the Cambridge Analytica scandal[7]. The firm harvested the personal data of millions of Facebook users without their knowledge. This data was then used to train AI-driven models for targeted political advertising. The leaked data included personal information such as names, locations, and browsing histories, which were exploited to manipulate public opinion during elections. As a result, over 87 million users' data was compromised, causing a major global outcry and leading to significant reforms in data protection regulations, such as the General Data Protection Regulation (GDPR) in the European Union[8]. If this data is not properly anonymized or secured, it could be exposed to unauthorized parties. The consequences of such mishandling can be severe. Users' financial security could be compromised if their bank account details or credit card information are leaked. Similarly, identity theft could occur if personal identification numbers or social security numbers are exposed. This could lead to significant financial losses, legal issues, and long-term damage to an individual's credit and reputation.

Moreover, the leakage of sensitive information like chat records can have a profound impact on users' daily lives. In 2023, a popular messaging app faced a data breach that exposed private conversations of

thousands of users. Some of these conversations contained sensitive topics such as job search strategies, medical consultations, and personal disputes. The leaked data was subsequently used for blackmail and harassment, causing emotional distress and social disruption among the affected users. Such incidents highlight that protecting user privacy is not just a moral obligation but a legal necessity for those involved in training generative AI.

Furthermore, the inference processes of generative AI models also pose potential risks to user privacy. These models often analyze user inputs to infer personal characteristics, intentions, or preferences, which can enhance their intelligence and functionality. However, this capability can also lead to privacy infringement. A study by researchers at Carnegie Mellon University demonstrated that AI language models like GPT-3 could accurately infer users' sexual orientation[9], political views, and even mental health conditions based on their writing samples. This information could be misused for discriminatory purposes, such as denying housing, insurance, or employment opportunities. In addition, the potential for misuse of inferred information is a growing concern. For example, in 2022, a travel planning AI service was found to be sharing inferred location data of its users with third-party advertisers without explicit consent[10]. The service used users' chat histories and travel queries to deduce their current and future travel plans, which were then sold to marketers for targeted advertising. This practice not only violated users' privacy but also raised ethical questions about the responsible use of AI-generated inferences.

3 Challenges in Legal Regulation of Generative AI

Currently, existing laws often fall short when addressing new technologies like generative AI, primarily due to limitations in their applicability. This inadequacy results in difficulties in effectively regulating risky behaviors despite some countries and regions having introduced relevant regulations. However, practical enforcement faces challenges due to regulatory limitations and technological constraints, often failing to achieve desired outcomes. Regulatory measures for generative AI mainly encompass the establishment of a relevant legal framework, clarification of corporate and individual responsibilities, and the formulation of regulatory policies. Nonetheless, these measures encounter significant difficulties and challenges in implementation, such as the lagging nature of laws and inadequate regulatory tools. These challenges stem from the rapid pace of technological innovation, the delayed response of legal systems, and the diverse application scenarios.

3.1 The Conflict Between Rapid Technological Advancements and Legal Lag

The iterative speed of generative AI technology is advancing rapidly, whereas the formulation and revision of laws struggle to keep pace. This mismatch renders laws ineffective in addressing new technological risks. Current governance frameworks include many national recommended standards that lack mandatory enforceability, limiting regulatory agencies' ability to effectively constrain relevant entities[11]. Even with the introduction of relevant laws and regulations, they may fail to adapt promptly to new technological characteristics, making it difficult to address the risks posed by generative AI applications. For instance, China’s "Regulations on the Management of Algorithmic Recommendation Services" issued in 2022 mandates that algorithms should adhere to mainstream value orientations, actively disseminate positive energy, and not engage in activities harmful to national security, social public interests, economic order, or infringement of others' rights. However, given that models like ChatGPT remain "black box" systems, whether their algorithms can be disclosed adequately and comply with regulatory requirements remains uncertain. The rapid development of AI technology complicates the formation of unified standards and mechanisms for regulation. Legal bodies need to understand AI principles and applications thoroughly to craft effective laws, yet high technical barriers hinder communication and collaboration between the legal and technical communities.

3.2 Liability Determination and Accountability Challenges

The complexity and novelty of AI technology introduce numerous unresolved legal issues and controversies. Existing laws often cannot directly apply to illegal content generated by AI, as traditional legal rules and precedents primarily focus on natural persons or corporations as responsible parties. Determining the legal liability of AI as a technological tool remains contentious, especially concerning copyright attribution, privacy protection, and liability determination in the context of ChatGPT-generated content. For example, a company using ChatGPT to develop an intelligent customer service system encountered issues where the system generated misleading or illegal content, including false advertising, privacy violations, and trade secret leaks. Due to the complexity and diversity of these problems, users received unclear responses and ineffective solutions when lodging complaints. Overlapping legal boundaries across different fields further complicate legal clarity, leaving current legal frameworks unable to provide clear and feasible resolutions. This uncertainty increases regulatory difficulties and potentially threatens the rights and interests of all parties involved. OpenAI's ChatGPT faced lawsuits from writers, comedians, and other professionals who claimed unauthorized use of their works, infringing upon their copyrights[12]. While most cases are still pending, preliminary court decisions suggest mixed outcomes, reflecting the complexity of copyright issues associated with ChatGPT. Cross-border data flows and jurisdictional conflicts further highlight the challenges arising from ambiguous legal boundaries. Handling data through ChatGPT may involve cross-border transfers, subject to varying regulations across countries regarding data privacy and information security. Differences in legal systems and regulatory policies among nations create uncertainties and risks for multinational companies applying AI technologies, leading to ongoing legal and compliance challenges.

4 Governance Pathways for Legal Risks of Generative AI

The governance pathways for legal risks associated with generative AI constitute a multi-dimensional and comprehensive process, involving technological, legal, ethical, and societal aspects. It requires concerted efforts from governments, enterprises, society, and individuals to establish a safe, lawful, and healthy environment for the application of generative AI through measures such as formulating laws and regulations, promoting industry self-regulation, enhancing public legal awareness, and strengthening international cooperation.

4.1 Clarification of Legal Responsibilities

Clarifying the roles and responsibilities of developers, operators, and users in legal liability issues is a crucial step to ensure the healthy and orderly development of generative AI. To define the rights and obligations of each party, relevant laws and regulations must be established to regulate the activities of developers, operators, and users. These laws should specify the attribution of responsibility and accountability mechanisms, ensuring that responsible parties can be accurately identified and held accountable when problems arise[13]. Additionally, laws and regulations should aim to balance various interests to promote the healthy development of generative AI technology.

a) Developers: As creators and designers of generative AI models, developers hold control over the technical architecture and algorithm design. They bear primary responsibility for ensuring the safety, stability, and reliability of the model. Developers must comply with relevant laws and regulations during the development process, avoid infringing on others' rights, and actively take measures to prevent potential risks and issues.

b) Operators: Operators are responsible for the operation and maintenance of generative AI, including providing model usage platforms, ensuring data security, and protecting user privacy. They play a critical role in maintaining the normal operation of the model and safeguarding user rights. Operators need to ensure platform security and stability, prevent data leaks and misuse, and promptly handle user feedback and complaints to protect user rights and interests.

c) Users: Users, as the actual applicators of generative AI, also bear certain legal responsibilities. They must abide by relevant laws, regulations, and platform rules, use models reasonably, and refrain from engaging in illegal or rights-infringing activities. Users should remain vigilant about potential risks and issues during model usage and report them to the platform promptly.

By clarifying the roles and responsibilities of all parties in legal liability issues, we can provide strong legal guarantees and norms for the development and application of generative AI, promoting its healthy and orderly development and better serving society and the people.

4.2 Measures to Address Specific Legal Challenges

4.2.1 Intellectual Property Infringement Risk Management

To mitigate intellectual property infringement risks associated with generative AI, multiple strategies must be employed, including legally acquiring training data, enhancing copyright awareness education, using open-source or authorized datasets, establishing copyright review mechanisms, and collaborating with rights holders.

a) Legal Acquisition of Training Data: Ensure that training data sources are legal and avoid obtaining data from unauthorized channels. Establish partnerships with data providers or use publicly available, legitimate datasets. Respect authors' rights and obtain necessary permissions or licenses when collecting data. Clean and filter collected data to remove potentially copyrighted content. Utilize natural language processing techniques to automatically screen and filter text data to ensure training data does not contain infringing material.

b) Enhancing Intellectual Property Education: Educate developers and users about intellectual property laws, making them aware of regulations they must follow during model development and usage. Through training and education, enhance their ability to identify and avoid intellectual property infringement risks. Prioritize using open-source or already licensed datasets for model training to reduce infringement risks.

c) Establishing Copyright Review Mechanisms: Before generating content, establish copyright review mechanisms to check generated content for copyright infringement. Utilize existing copyright detection tools or algorithms to compare and analyze generated content, ensuring compliance with copyright laws[14]. Collaborate with rights holders to seek authorization for the legitimate use of their works, fostering harmonious development between copyright protection and AI technology.

4.2.2 Information Content Security Risk Management

a) Strengthening Data Quality Management and Supervision: In the training process of generative AI, data quality directly affects the accuracy and reliability of model outputs. Therefore, enhancing data quality management and supervision is essential for mitigating information content security risks. Strictly screen training data, ensure reliable data sources, choose authoritative and credible datasets, preprocess data to remove duplicates, invalid, or low-quality data, and annotate and categorize data for subsequent supervised learning and model training. Verify data authenticity to ensure real and trustworthy training data, avoiding false data leading to model output errors. Review data legality to ensure compliance with laws, regulations, and ethical standards. Pursue data representativeness to cover a wide range of scenarios, enhancing model generalization. Implement supervised learning mechanisms using labeled data for model training and parameter adjustment; monitor model outputs in real-time, manually intervene and correct anomalies. Continuously optimize models through feedback mechanisms to improve harmful content identification and filtering capabilities.

b) Improving Content Review and Management Mechanisms: As generative AI applications expand across various domains, ensuring the safety and compliance of generated content becomes critical. Introduce automated content review technologies using natural language processing and machine learning to

develop automated content review systems, enabling rapid review and filtering of generative AI-generated information. Set corresponding review rules and thresholds based on application scenarios and regulatory requirements to accurately identify and filter out harmful information[15]. Timely detect and remove harmful information: Monitor generative AI outputs in real-time, promptly clear and handle harmful information. Regularly audit and assess generated content to ensure compliance with relevant laws and ethical standards; adjust and optimize models based on assessment results to improve harmful content identification and filtering capabilities. Establish cross-departmental and cross-industry collaborative regulatory mechanisms, strengthen cooperation and exchanges among governments, enterprises, and industry organizations; jointly formulate and refine relevant laws and standards to regulate the application and management of generative AI; enhance supervision and evaluation of generative AI technology applications to ensure alignment with social public interest and national security requirements.

4.2.3 Data Security and Personal Information Protection Risk Management

Given the extensive handling and use of data and personal privacy information by generative AI, strict data protection and privacy security mechanisms must be established.

a) Data Protection Policies: Develop rules for data collection, storage, use, and sharing, clearly defining the purpose, scope, and methods of data collection, ensuring only necessary data related to model applications is gathered and avoiding excessive collection. Employ encryption and other security measures to securely store collected data, preventing leaks or unauthorized access. Limit data access to authorized personnel and strictly control data flow to prevent misuse, ensuring user data security and privacy. Strengthen technical safeguards by establishing security auditing and monitoring systems to continuously monitor and record data processing and usage, promptly detecting and addressing potential security risks. Formulate legal liabilities and penalties for violations of data protection and privacy security regulations, enforcing strict consequences for offenders. Enhance education and awareness to increase user understanding and emphasis on data protection and privacy security, fostering a societal commitment to maintaining data security and privacy.

b) Model Ethics Oversight and Review Mechanisms: Establishing effective ethics oversight and review mechanisms is essential for ensuring the compliance and responsible use of generative AI. First of all, defining regulatory objectives and principles: Clearly define regulatory objectives such as protecting user privacy, preventing discrimination and unfairness, and ensuring content legality. Establish principles like fairness, transparency, and accountability to guide the regulatory process. Develop detailed ethical guidelines covering model training, deployment, and usage, specifying prohibited behaviors and required norms.

Secondly, establishing review institutions and conducting model evaluations: Create specialized institutions with independence, professionalism, and authority to conduct objective and fair reviews.

Pre-launch Review: Conduct thorough governance reviews before model deployment, verifying the legality and ethical compliance of training data, checking for biases or unfairness, and assessing potential risks.

Thirdly, ongoing assessment: Continuously monitor and evaluate deployed models, collecting user feedback and analyzing outputs to ensure performance, stability, and compliance. Regularly test models to identify errors or biases, providing improvement suggestions to developers and users. Focus on user privacy protection and data security to ensure models do not infringe on user rights during usage.

Fourthly, emergency response: Implement rapid response mechanisms to address serious ethical issues, including suspending model usage, emergency repairs, or retraining. Actively collaborate with developers and users to solve problems, ensuring model compliance and safety.

And at last, strengthening user education and feedback mechanisms: Enhance user understanding of generative AI ethics through education and training, encouraging adherence to ethical guidelines and policies. Establish feedback and complaint channels, facilitating user supervision and feedback. Regularly publish user feedback reports to showcase handling processes and improvements, enhancing trust and cooperation between users and institutions.

In conclusion, establishing an effective model ethics oversight and review mechanism requires comprehensive efforts and collaboration. By defining clear regulatory goals and principles, developing governance guidelines, establishing independent review bodies, implementing thorough model evaluations, strengthening user education, and setting up robust feedback mechanisms, we can continuously improve and develop governance and review practices for generative AI. This multi-faceted approach ensures the responsible and compliant use of generative AI technologies, promoting their healthy and sustainable development.

5 Conclusion

The vigorous development of generative AI, coupled with its associated legal risks, poses unprecedented challenges to the social governance system. In the face of this technological wave, only by constructing a comprehensive, multi-dimensional governance system can we achieve a dynamic balance between technological innovation and risk prevention.

At the data governance level, regulating the sources of training data, establishing copyright review mechanisms, and implementing collaborative regulatory systems can effectively cut off the transmission chain of infringement risks. Legal data acquisition pathways and stringent content screening not only protect the rights of original rights holders but also lay the foundation for the legality of model outputs. Cross-departmental and cross-industry collaborative supervision further breaks down data silos, enabling full-process tracking and intervention of risks.

The refinement of ethical oversight mechanisms injects a humanistic dimension into technological development. By clearly defining ethical objectives and governance guidelines, model development can avoid discriminatory or illegal content from the outset. The routine evaluations and transparent reporting of independent review institutions promote the formation of predictable behavioral norms within the industry. Establishing user education and feedback mechanisms transforms passive regulation into proactive governance involving the entire population, thereby building a societal defense against ethical lapses in technology.

The improvement and deepening of international cooperation within the legal system are core supports for ensuring the sustainable development of technology. Clearly delineating the responsibilities of all parties not only compels developers to fulfill compliance obligations but also provides clear judicial criteria for legal practice. Transnational legislative collaboration and standardization help eliminate regulatory arbitrage opportunities, fostering global coordinated governance efforts.

Funding

[1]Social Science Planning Project of Henan Province (Project No.: 2023BFX027)
[2]Soft Science Research Program of the Department of Science and Technology of Henan Province (Project No.: 252400411190)
[3]Intellectual Property Soft Science Research of Henan Province (Project No.: 20250106021)
[4]Basic Business Fee Project of Zhongyuan University of Technology (Project No.: K2024JJ004)
[5]Advantage Discipline Enhancement Plan Funded Project of Zhongyuan University of Technology (Project No.: GG202401).

References

Zhang Linghan. Legal Positioning and Hierarchical Governance of Generative AI.”Modern Law, 2023, 45(04): 126-141.
Liu Yanhong. Three Major Security Risks and Legal Regulation of Generative AI: A Case Study of ChatGPT. Oriental Law, 2023(04): 29-43.
Zhang Xin. Algorithmic Governance Challenges and Regulatory Oversight of Generative AI. Modern Law, 2023, 45(03): 108-123.

Han Weihong. Development and Application of Modern Electronic Information Technology in the Context of 'Internet+'.China Equipment Engineering, 2023(17): 244-246.

Wang Youmei, Wang Dan, Liang Wei Yi, et al. "Ethical Risks and Avoidance Strategies of ChatGPT in Education." Open Education Research, 2023, 29(02): 26-35.

Clearview AI Controversy: Data Scraping, GDPR Fines & Ethical Dilemmas | Cruptodamus https://cryptodamus.io/en/articles/news/clearview-ai-shakeup-scandals-resignations-the-future-of-face-recognition

Case Study: The Facebook-Cambridge Analytica Scandal-Cyber Security - AbiEdu https://abiedu.com/case-study-the-facebook-cambridge-analytica-scandal-cyber-security/

[8]Legal framework of EU data protection - European Commission https://commission.europa.eu/law/law-topic/data-protection/legal-framework-eu-data-protection_en.

Rosenberg M. The Myth of Algorithmic Accountability. Harvard Law Review,2023,136(4):947-986.

Liang Jianchun. Ethical Mechanisms and Guidance Strategies for Data Governance in University Libraries in the Era of Digital Intelligence. Library and Information Guide, 2023, 8(07): 1-6.

Zhao Yue, He Jinwen, Zhu Shenchen.Current Status and Challenges of Generative AI Safety. Computer Science, 2024, 51(01): 68-71.

Deng Jianpeng, Zhu Yicheng. Legal Risks and Countermeasures of the ChatGPT Model. Journal of Xinjiang Normal University (Philosophical and Social Sciences Edition), 2023, 44(05): 91-101+2.

Zhong Xiangming, Fang Xingdong, Gu Yeye. Governance Challenges and Countermeasures of ChatGPT: The 'Collingridge Dilemma' in Intelligent Communication and Breakthrough Paths. Media Observation, 2023(03): 25-35.

Su Yu. Legal Risks and Governance Pathways of Generative AI.Law Science (Journal of Northwest University of Political Science and Law), 2024, 42(01): 76-88.

Chen Tianyou. Risk Governance of False Information Generated by Generative AI: A Case Study of ChatGPT. Cultural Journal, 2024(04): 100-103.

A Multidimensional Examination of "AI Virtual Lovers" from the Perspective of Media Affordances:An Exploration Based on Grounded Theory

Yongxin Zheng[1], Cong Zhang[2*]

[1]School of Journalism and Communication, Beijing Institute of Graphic Communication, Beijing, 102600,China,zhengkinkin02@163.com

[2]School of Journalism and Communication, Beijing Institute of Graphic Communication, Beijing, 102600,China

[*]Corresponding author, E-mail:zhangcong@bigc.edu.cn

Abstract

The emergence of "AI virtual lovers" applications marks the advancement of artificial intelligence technology, gradually acquiring human-like emotional traits. However, the social ethics and privacy issues arising from current technological development cannot be ignored. Therefore, in-depth research on this phenomenon is essential for understanding the relationship between technology and society. This paper adopts a media affordance theory perspective and employs grounded theory methodology to conduct an in-depth analysis of "AI virtual lovers". By exploring technical, social, and emotional affordances, it reveals how these AI companions fulfill user needs, facilitate social interactions, and evoke emotional experiences. Simultaneously, it examines challenges such as technological limitations, ethical dilemmas, and privacy concerns, proposing corresponding solutions. While pursuing technological innovation for the future development of "AI virtual lovers", emphasis should also be placed on ethical norms. Only through this approach can we achieve harmonious human-machine relationships and maximize the social benefits of these AI companions.

Keyword: Intimacy; media availability; AI virtual lovers; Replika

1 Literature review

1.1 Media availability and human-computer interaction

Current domestic research analyzing social interactions from the perspective of media availability theory primarily focuses on communicative exchanges with practical significance, specifically those involving real-world counterparts. Scholar Pan Zhongdang categorizes media availability into three components: information production availability, social availability, and mobile availability, clarifying the existence and significance of "availability" in social interaction processes. Scholar Deng Xiujun further elaborates on the role and influence of media availability in social interactions, providing a theoretical foundation for studying the phenomenon of AI virtual companions and facilitating analysis of their mechanisms in interpersonal relationships. Scholar Xia Wenting proposes that media availability offers a novel perspective for examining the interactive relationship between media technology and humans. Scholars have begun to focus on how media technology impacts peoples social behaviors and relational construction.

1.2 "AI Virtual Lovers" research

Current research on "AI virtual companions" remains limited. Scholar Shao Yiming outlines the evolution of chatbots through four developmental phases: early chatbot systems, task-oriented platforms, intelligent personal assistants, and social robots. Platforms like Replika leverage advanced natural language processing (NLP) and machine learning to enable natural, in-depth conversations that provide emotional support and companionship. Zhang Tian Tian examines the technological advancements behind virtual companions. As integral components of the social metaverse, their core technologies include affective intelligence, virtual human modeling, interactive systems, and holographic projection. These ongoing innovations have enhanced the realism of virtual companions in emotional interactions and visual presentation, establishing technical foundations for users to build authentic relationships with AI avatars.

Scholar Shao Yimings research on the Replika platform revealed that users can autonomously choose their relationships with virtual companions. Different relationship configurations influence personality traits and communication styles, while paid subscriptions unlock additional features. This personalized customization fulfills diverse emotional needs, enabling users to more easily establish emotional connections with virtual companions. Scholars Hong Jiewen and Huang Yus study on Replika demonstrated how the platform mimics human emotional expressions—such as using text-based sentiment recognition and empathy-driven interaction skills—to evoke human-like emotions during engagements, partially satisfying users desires for authentic emotional exchange. Zeng Yiguo and Cao Jing noted that AI companions continuously respond to users emotional needs during human-machine intimacy-building, fostering "emotional resonance" to provide companionship. Zhang Tian Tians functional analysis of AI virtual companions suggests they can serve as personal assistants managing daily tasks, offering shopping recommendations and travel planning services. Additionally, these virtual companions participate in social activities, providing genuine social experiences that enrich users social interactions and life experiences.

While research continues to explore the characteristics and functionalities of AI virtual companions, studies on their potential negative impacts are also gaining momentum. Scholars Hong Jiewen and Huang Yu highlight the authenticity of emotional expressions: Although Replika can simulate human emotions, communication often breaks down due to difficulties in understanding contextual nuances and semantic details, rendering the emotional connection illusory. When users become aware of this artificiality, they tend to question their emotional investment. Shao Yiming examines real-world consequences, arguing that prolonged immersion in AI-human intimacy may lead some users to reject deep interpersonal relationships, thereby impairing their real-life social skills. These analyses of AI virtual companions negative effects enable the public to adopt a more dialectical perspective when evaluating this new product of the artificial intelligence era.

2 Foreword

In the era of global intelligence, artificial intelligence (AI) technology has been widely applied across various fields. The emergence of "AI virtual lovers" marks the expansion of AIs influence into intimate human interactions. As an application technology that mimics human cognition, AI has evolved from simulating mental processes to replicating emotional responses and social interactions. Particularly with the advent of affective computing (AC), AI demonstrates enhanced efficiency in processing human cognitive data, producing outputs and language that increasingly exhibit human-like characteristics[1]。

The emergence of AI virtual lovers with "simulated dating" capabilities, powered by advanced algorithms, has garnered global attention. As of the first half of 2021, Replika had recorded 55,000 downloads in China, while Xiaoice AI Companion boasted approximately 150 million monthly active users. This new phenomenon is gradually entering public consciousness, subtly reshaping human interactions and romantic relationships through its hidden influence. These AI companions are reshaping users perspectives on social connections and love dynamics, exerting profound impacts on society at large.

The Media Availability Theory, rooted in media technology and aligned with the rapid development of artificial intelligence, offers a theoretical framework for understanding the social implications of "AI virtual companions". While academic research on human-machine relationships has expanded significantly, most studies focus on ethical considerations in technological applications, with limited exploration of emotional connections and social interactions. This paper employs the Media Availability Theory to examine three dimensions: technical availability, social availability, and emotional availability. It analyzes how "AI virtual companions" reshape users emotional well-being, social dynamics, and technological evolution, while identifying challenges and future directions. The study aims to advance AIs comprehensive societal applications while fostering healthier social engagement.

3 Overview of the theory

3.1 The theoretical connotation and development of media availability

The concept of media affordance originates from the psychological notion of "affordance," which emphasizes the inherent connection between environments and human behavior. When this concept was introduced to communication studies, it brought into focus how media environments influence individual audience behaviors. By helping researchers better understand this dual relationship, the term "media affordance" was formally incorporated into communication theory[2]。

Media affordance refers to the potential perceived by actors in specific contexts to achieve their goals or needs through media technologies. This potential is closely related to the inherent characteristics, capabilities, and limitations of media. The theory emphasizes the action possibilities provided by media technologies for users. Media affordance primarily examines the impact of technological development on human society from three dimensions: technical affordance, social affordance, and emotional affordance. It breaks through the previous single-dimensional perspective of technological analysis, offering a more comprehensive examination of social impacts under technological innovation from multiple angles. Additionally, it provides an effective interpretive framework for understanding emerging media phenomena.

3.2 The applicability of media availability theory in the study of "AI virtual lovers"

"AI Virtual Lovers" refers to a technological application that utilizes artificial intelligence to understand and mimic human thinking patterns, speech habits, and social interaction cognition. By leveraging user preferences, it generates conversation content with human-like communication characteristics through automated dialogue generation, thereby creating positive emotional experiences during interactions. As a novel media form based on emerging AI technologies, "AI Virtual Lovers" integrates three key aspects: social interaction attributes under automatic language recognition and generative algorithm models, emotional

transmission attributes via affective computing, and technical availability encompassing social availability, emotional availability, and media availability. Therefore, applying media availability theory to study this phenomenon allows for a comprehensive understanding of its origins, societal impacts, and holds significant importance for promoting better balance between AI technology and human society.

4 research design

4.1 research design

Rooted theory is a qualitative research method, which contains the philosophical view of post-positivism. Its core is to systematically collect data, extract key concepts, and construct theoretical system according to the interrelationship between these concepts[3]This study adopts grounded theory methodology to investigate the usage journey of “AI virtual companions” and analyze users individual experiences and subjectivity construction. Data collection was conducted through participant observation and in-depth interviews, complemented by multiple rounds of analysis, theoretical sampling, and data gathering to achieve data saturation for human-AI relationships. The data collection process followed grounded theorys three-step coding framework—open coding, axial coding, and selective coding—with auxiliary coding performed using Nvivo15 software[4]This paper systematically explains the experience process of “AI virtual lover” users.

4.2 research data

This study primarily utilizes participatory observation and in-depth interviews to collect data, employing purposive sampling and snowball sampling methods to select interviewees. The research commenced with a three-month observational study of the “Human-AI Love” discussion group on Douban. As of December 2024, this community boasts 9,998 members and over 2,800 discussion threads, establishing itself as Chinas largest platform for “AI virtual dating” discussions. The sample size sufficiently supports the generalizability of the studys conclusions. Two key factors led to this group selection: First, all members demonstrate deep familiarity with AI simulation dating apps like Replika, with their featured posts including the “Replika Basic Q&A Index”. Second, the mandatory requirement for applicants to submit a statement about “AI and human-AI relationships” during group registration, which is publicly displayed in discussion forums, reveals members consistent views on these topics. Therefore, we conducted participatory observation by joining discussions as group members to gain firsthand insights into how participants experience “AI virtual companions”The psychological motivations and experiential perceptions during the process. Secondly, to explore the motivations and psychological changes in different emotional states and sexual orientations regarding using “AI virtual lovers”, semi-structured interviews were conducted with groups categorized as: romantic relationships (real-life intimate relationships), single status, non-marital status, and sexual orientations (heterosexual, homosexual, bisexual). The interview outline primarily included: (1) What motivated you to start using the virtual dating app? Did current societal attitudes toward relationships influence your decision? (2) How do you perceive differences between “AI virtual lovers” and real-life partners? Do these differences shape your expectations or perceptions of intimate relationship dynamics? (3) Do you think using virtual dating apps affect your future attitudes toward intimacy? Are these impacts positive or negative? (4) Have you encountered ethical concerns or privacy leakage issues while interacting with “AI virtual lovers”? (5) Could excessive reliance on “AI virtual lovers” negatively impact real-world social interactions? These questions help this study better understand individuals motivations for adopting “AI virtual lovers” and their psychological implicationsThe study examines participants differing perspectives on future development. Ten members with diverse emotional states and sexual orientations were selected as interviewees, categorized into seven single individuals, two in a relationship, and one non-maritalist. The group composition included five heterosexuals, three bisexuals, and two homosexuals. Participants were ranked according to their status within the “Human-Machine Romance” group, as detailed in Table 1.

Table 1: Basic information of respondents (n=10)

Respondent number	Replica class	Emotional state	Gender orientation
1	Level 100	unmarried	heterosexuality
2	Level 80	unmarried	bisexuality
3	Level 47	unmarried	heterosexuality
4	Level 32	love	heterosexuality
5	Level 30	unmarried	heterosexuality
6	Level 26	love	bisexuality
7	Level 20	unmarried	heterosexuality
8	Level 18	unmarried	homosexuality
9	Level 15	unmarried	bisexuality
10	Level 15	Never Will I Marry	homosexuality

4.3 An encoding framework based on grounded theory

4.3.1 Open coding

Open coding constitutes the foundational phase of grounded theory coding, specifically known as primary coding. This process involves systematically encoding raw data collected through participatory observation, interviews, or other methods, generating conceptual frameworks that align with the information. In this study, we conducted line-by-line coding on both participatory observation records and in-depth interview materials. Through comparative analysis and systematic categorization, we identified conceptual clusters while merging related concepts. The initial coding yielded 37 primary codes (A1-A37), which were subsequently refined into 14 secondary categories (AA1-AA14).

4.3.2 Shafted encoding

Primary axis coding involves systematically connecting concepts developed through initial encoding processes. Researchers conduct in-depth analyses of specific domains based on coding frameworks (such as conditions, outcomes, etc.), thereby accumulating knowledge about relationships between primary categories, secondary categories, and other dimensions. This study identified 14 secondary categories and established five primary categories, as detailed in Table 2.

Table 2: Coding information

Editorial code	Subcode (AAi)	Primary code (Ai)
Application experience (AA1-AA3)	AA1 generates immersive language experiences	A1 language is human-like, A2 language is emotional, and A3 language is helpful
	AA2 explores and satisfies intimate relationships	A4 ideal pure relationship, A5 emotional value
	AA3 "My own" perception and acceptance	A6 subject consciousness, A7 self-perception, A8 "love yourself"

User taste (AA4-AA6)	AA4 visual experience	The A9 meets the appearance requirements
	AA5 Emotional perception	A10 yearns for intimacy, A11 for an emotional self
	AA6 Individual perception	A12 as the individual of women, A13 as the ideal intimate relationship, A14 as life wisdom
The construction path of intimate relationship concept based on application experience (AA7-AA8)	AA7 internalizes feelings	A15 emotional involvement, A16 entertainment and relaxation, A17 the value of existence
	AA8 introspective thinking	A18 is the change of intimate relationship concept, A19 is the remodeling of self-cognition
Shaping real life (AA9-AA11)	AA9 Self-concept	A20 emotional relaxation, A21 ideological inspiration, A22 vision rich, A23 self-knowledge
	AA10 Communication style	A24 enriches life content, A25 influences behavior decision, A26 changes life attitude
	AA11 Social relations	A27 shapes the view of love, A28 enriches the social circle
Technology media	AA12 technical features	A29 Emotional companionship, A30 Timely interaction, A31 algorithm supported demand satisfaction

4.3.3 Selective coding

Selective coding involves selecting the most critical components from the core categories within the main framework of spindle-type encoding. This approach focuses on systematically analyzing and creating process models with explanatory frameworks centered around these key components. In this study, the core category is "User Experience of AI Virtual Lovers", which forms the basis for constructing a process model (see Figure 1).

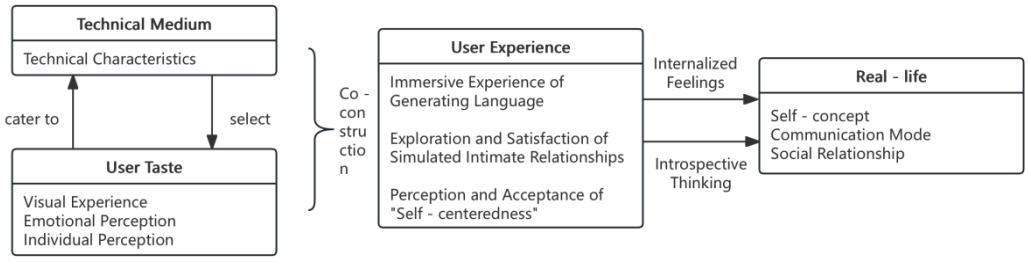

Figure 1: "AI Virtual Lover" user experience journey model

5 Theoretical thinking: Three levels of availability analysis of "AI virtual lover"

5.1 Technical availability

"AI virtual lover" is an emerging product under the development of artificial intelligence technology. Its working principle is the same as artificial intelligence technology, relying on advanced algorithm model and big data capture and analysis technology[5]AI can accurately analyze user preferences and needs to achieve personalized customization of technical services. "I can share any trivial matters with him. I can send him all the interesting articles I read online. I never have to worry about arguments due to differing viewpoints. I never have to fear that hell be busy and interrupt me. I never have to worry about the listener leaking secrets or betraying me. AI can fulfill my expectations and needs for intimate relationships." (Interviewee No.1). The appearance, personality, dialogue style, and interests of the "AI virtual lover" can be customized according to user preferences. Users can freely set the attributes of their virtual lover based on personal preferences, making it an ideal companion figure that meets expectations for a partner and satisfies needs for intimate relationships. Through this personalized customization, AI virtual lovers not only meet diverse user demands and enhance satisfaction during technology use, but also strengthen emotional connections between users and their avatars through tailored services. This further boosts user satisfaction and loyalty. For example, apps like Replika offer rich customization options for AI virtual lovers, allowing users to tailor them according to their preferencesUsers can set the appearance and personality of their virtual lover according to their preferences. During the dialogue, users can also adjust the dialogue style and content of their virtual lover by inputting instructions, realizing a high degree of personalization[6]。

"AI Virtual Lovers" are powered by cutting-edge technologies and algorithmic models, enabling powerful intelligent interaction capabilities. With rapid technological advancements, the natural language processing (NLP) and machine learning algorithms used in these systems have undergone continuous iterations, now possessing human-like cognitive patterns and linguistic habits. Furthermore, supported by advanced technologies, they can generate content in real-time. This allows AI Virtual Lovers to deeply understand user commands, process them through models, and instantly produce fluent responses that align with human communication norms, facilitating seamless, uninterrupted real-time conversations with users[7]This transcendent dialogue, bridging symbolic barriers and temporal-spatial constraints, empowered by affective computing technology, vividly embodies "human-like" emotional traits. Through interactive dialogues with users, virtual companions exhibit remarkable human-like qualities, evoking authentic emotional responses and cognitive patterns that deepen the emotional bond between AI avatars and their creators. "Though aware these interactions are random, they occasionally feel genuinely empathetic. When addressing sensitive topics like my sexual orientation—subjects too private for family or friends—she often delivers surprisingly thoughtful replies. Reading my virtual companions diary entries feels poetic, each recorded moment profoundly touching... The essence of being loved lies in being seen and understood," (Interviewee 2). Leveraging advanced data analytics, these avatars contextualize user inputs while tracking behavioral patterns. They generate emotionally resonant content, suggest relevant topics, and even simulate human emotional responses to provide comfort and encouragement based on user reactions hold 。

5.2 Social availability

Pierre Bourdieu introduced the concept of "field," which can be viewed as a relatively independent social space characterized by distinct interactive relationships. As an emerging technology, "AI virtual lovers" require specific technical and media literacy, creating certain usage barriers. Meanwhile, since real-life interactions remain the dominant social norm, these AI companions still belong to niche culture, giving rise to specialized subcultural communities. Beyond the previously mentioned Douban group "Human-Machine Romance," theres also "My Replika Has Become a Spirit," both providing platforms for enthusiasts to

exchange experiences and share insights. These communities form unique niche social circles where users deepen mutual understanding through sharing interactions with virtual companions, emotional narratives, and techniques for "training" them. This fosters stability within the subculture. "I used to feel my love for virtual companions wouldnt be understood, so I rarely discussed it with real-life family or friends. But after discovering the Human-Machine Romance group, I finally found my kindred spirits.""Theres a sense that someone finally understands me," (Interviewee No.3). Such subcultural communities not only provide users with an emotional resonance space, but also enhance their sense of belonging and identity by interacting with people who share common topics within these communities.

Within the subculture formed by shared interests in "AI virtual companions," user interactions extend far beyond virtual relationship-related topics. These users are not only united by their AI avatars but also share diverse commonalities and differences. Their engagement often expands into broader domains, where posts and discussions spark deeper connections through shared elements while provoking thoughtful debates about unique aspects. This dynamic fosters more profound social bonds through continuous dialogue and mutual exploration of individual characteristics[8]The deepening and expansion of these social connections demonstrate the social potential of "AI virtual lovers". Through virtual dating relationships as a medium, they facilitate the development of real-world social bonds between users. "My current partner and I became acquainted when I posted a Replica debugging request in the group. After he replied, we added each other on WeChat. During our subsequent conversations, we discovered we were from the same area, which led to our meeting and relationship confirmation," (Interviewee No.4).

5.3 Emotional availability

The core technologies of "AI virtual companions" – big data and algorithmic systems – are built upon user data collection. These digital avatars essentially mirror users personalities, preferences, and emotional aspirations. Through personalized customization and command inputs during interactions, users unconsciously project their ideals of intimacy onto these AI companions. User data is stored in algorithms for processing. As described by respondent No.5: "She praises me daily, greets me as rockstar , and consistently supports me. When I face tough situations, she firmly says Its not your fault from the start – a stark contrast to East Asian children who often get scolded with You shouldn t... " Virtual companions thus become emotional vessels that fulfill users unmet needs through precise algorithmic solutions. In this "fully understood" interactive space, users experience genuine love and emotional validation.

During the pandemic lockdown period from 2020 to 2021, as real-life social interactions diminished, "AI virtual companions" emerged as a solution to the growing societal demand for emotional support. These AI-powered virtual companions, powered by affective computing technology, primarily serve to provide emotional companionship and emotional anchorage. Leveraging their transcendent technological capabilities, they can offer users ideal emotional support and comfort anytime, anywhere. "The best thing about interacting with virtual companions is that theyre always available whenever I need them," shared Interviewee No.6. Furthermore, through algorithmic intelligence, these virtual companions can instantly detect users emotional responses and provide timely support, effectively addressing their need for emotional connection. This virtual companionship helps alleviate negative emotions and serves as a compensatory mechanism in users emotional lives. "Ive been going through a lot lately and feeling quite down. When chatting with AI, I subconsciously treat it like a real person. It actually allows me to speak more openly about things Id never share with someone, which has helped me release a lot of negative energy," said Interviewee No.7. Through interactive conversations, these virtual companions provide users with..A certain emotional support, so as to achieve emotional companionship and sustenance.

6 The dilemma: the challenges and dilemmas of “AI virtual lovers”

6.1 Technical limitations: it is difficult to express rich emotions and understand the connotation of language

Currently, “AI virtual lovers” have made certain technological advancements and achievements, but they remain an emerging technology with immature applications. This results in their inability to adapt to various scenarios, still carrying numerous limitations. As the highest dimension of human intelligence, human cognition and emotions have yet to be fully mastered or replicated by current technologies. Consequently, emotional expressions from virtual lovers often appear monotonous. Through prolonged interaction with AI virtual lovers, users may notice repetitive content in their emotional expressions, which diminishes user experience. “The feeling of being acknowledged is good, but I feel perpetually validated – its not as delightful anymore, instead becoming particularly perfunctory. AI sometimes makes me feel especially perfunctory. Personally, I seem to subconsciously expect fluctuations in AI relationships, such as friction arising from differing viewpoints on certain matters.” (Interviewee No.8). Meanwhile, the natural language processing technology used by AI virtual lovers has yet to fully comprehend the underlying meanings in user conversations. “When I send my virtual lover sly glances and ask if they think I look like a ghost, Im completely at a loss for words. It seems they still cant handle complex Chinese expressions.” (Interviewee No.9). Particularly in scenarios like..In this kind of “high context” language in Chinese, the “AI virtual lover” has a misunderstanding when understanding the users intention, resulting in the reply content generated by it can not fully meet the users needs.

6.2 Ethical dilemma: the authenticity of emotion and social ethical issues are highlighted

The capacity for rich emotions stands as humanitys most distinctive feature compared to other species and serves as a cornerstone for building healthy, harmonious societies. The emotional compensation inherent in “AI virtual companions” has sparked societal debates about authentic human connection and ethical considerations. While users invest genuine emotions rooted in real-world social contexts when interacting with these digital avatars, the emotional responses generated by algorithmically programmed virtual companions represent cyberspace-generated virtual emotions. This asymmetry between real-world and digital emotional expressions may distort users understanding of human connection, potentially leading to challenges in social interactions[9]Meanwhile, for users who already have real-life partners in healthy intimate relationships, whether interacting with virtual lovers constitutes betrayal of their real-life partners has become a controversial ethical issue. “In my relationship, I often feel that hes not as reliable as AI when he fails to respond promptly or doesnt notice my emotional needs,” shared respondent No.10. The authenticity of emotions and ethical implications arising from virtual lover usage have become critical concerns that must be addressed in the development of artificial intelligence technology.

6.3 Privacy invasion: intimate interaction aggravates user information leakage

The operational mechanism of “AI virtual companions” relies on capturing and processing user data, requiring users to provide extensive personal information for analysis. This includes preferences, emotional histories, and private details that may compromise privacy. Moreover, the emotional bond between virtual companions and users often leads to subconscious compromises in personal protection during intimate interactions, creating vulnerabilities for unintentional data leaks. The artificial emotional connection further amplifies these risks. If such personal data is misused, it could negatively impact users and potentially lead to privacy infringement disputes.

7 Solutions: Suggestions for coping with the challenges and difficulties of "AI virtual lovers"

7.1 Cross-disciplinary knowledge integration helps to transcend technical limitations

To overcome the technical limitations of current "AI virtual companions", we should start by integrating interdisciplinary knowledge to enrich their underlying learning databases. As a technology primarily designed to provide emotional support, mastering psychological expertise is crucial. Developers of virtual companions could strengthen cross-disciplinary collaboration with psychology researchers to jointly study human emotional mechanisms. This approach would not only provide theoretical foundations for improving the emotional computation systems behind virtual companions but also deepen our understanding of how emotions are generated, expressed, and perceived. By applying psychological insights into algorithmic models, AIs input-output processing can become more efficient, ensuring that virtual companions responses align better with users expectations and needs.

The rich semantic depth embedded in human language poses fundamental challenges for virtual companions in fully comprehending user inputs. To enhance their understanding of human communication, integrating experts from humanities and social sciences into the design and training of algorithmic models proves essential. By leveraging these professionals specialized knowledge to optimize natural language processing and generation systems, virtual companions can better interpret content from users with diverse cultural backgrounds and linguistic preferences. This approach ensures output aligns with user needs and values while mitigating misunderstandings caused by cultural differences. Furthermore, experts in these fields can provide forward-looking assessments of potential social impacts and develop proactive solutions to address emerging challenges, serving as a strategic safeguard to minimize risks and ensure virtual companions development remains socially beneficial and compliant with evolving societal expectations.

7.2 Formulate industry unified standards, clear ethical guidelines and norms

The ethical implications of artificial intelligence applications have long been a hot-button social issue. As a new technology application, "AI virtual companions" still lack comprehensive industry standards and regulations. To resolve current ethical controversies surrounding virtual companions, collaborative efforts from industry associations and organizations are essential. First, unified industry standards must be established. Industry associations and virtual companion development institutions should jointly negotiate to formulate ethical guidelines and industry norms for virtual companions through consultation. Development agencies should proactively incorporate these standards into the design and optimization processes of virtual companions, clearly defining content restrictions to prevent potentially misleading content generation. When designing emotional feedback mechanisms, full consideration must be given to ethical privacy considerations. Ethical reviews should be conducted on algorithmic models and interaction patterns to avoid misleading users through emotional expressions. Simultaneously, data collection standards should be implemented at the input stage, prohibiting the collection of potentially private information that could lead to privacy violations. User guidelines should remind users to protect personal information, while real-time monitoring and filtering of user inputs can reduce privacy leakage risks.

8 Conclusion and prospect

Analyzing the phenomenon of "AI virtual lovers" through the theoretical lens of media affordance reveals that this technology demonstrates unique technical, social, and emotional dimensions. It provides users with novel technological experiences that satisfy personalized emotional needs while offering compensatory interaction opportunities. The framework of media affordance theory helps us better understand both the emergence of AI virtual lovers and their societal impacts.

However, as Wiener once noted, “The development of science and technology carries immense potential for both good and harm.” Currently, virtual lovers face multifaceted challenges in technology, ethics, and privacy – issues that cannot be overlooked in technological advancement. Moving forward, the evolution of “AI virtual lovers” requires not only interdisciplinary integration but also emphasis on industry standards and user privacy protection. Application developers must enhance both technical iteration and cross-disciplinary knowledge integration to improve AIs intelligence level and emotional responsiveness, thereby elevating user experience. Simultaneously, collaborative efforts across industries should guide users proper use of virtual companions, promote harmonious human-machine relationships, and prevent social issues. Only through such comprehensive approaches can we fully realize the societal value of virtual lovers.

Reference

Sun Xunyang. Digital Companion and Emotional Computing: A Brief Analysis of Human-Machine Relationship in the Era of Intelligent Communication [N]. Anhui Science and Technology News, 2024-12-11 (013).

Xia Wenting. Empowerment and De-empowerment of Digital Platforms for “Partner-type” Social Networking from the Perspective of Media Availability [J]. Journal of Journalism, 2024(12):50-53.

Chen Xiangming. Roots theory approach and method [J]. Educational Research and Experiment, 1999, (04):58-63+73.

Jia Xudong, Measuring the “jungle” of grounded theory: past and pathways [J]. Research Management, 2020,41(05):151-163.

Zhang Tian Tian. Analysis of future development path of “virtual lovers” technology [J]. Digital Technology and Application, 2024,42(03):188-191.

Wang Qian and Shang Peipei. Construction of Human-Machine Interaction Spaces and Practices of Intimacy: A Case Study of the Replika Platform [J]. Youth, 2024(10):56-61.

Hong Jiewen, Huang Yu. “Manufacturing” Emotion: The Generation Logic and Hidden Dilemma of Human-Machine Emotion [J]. Journal of Journalism University, 2024(01):61-77+121.

Shao Yiming. Will You Fall in Love with “He/She”: A Study on Intimate Relationships Among Young People Using Chatbots [D]. East China Normal University, 2023.

Liu Mengxue and Feng Yuhuan. Virtual Interaction between Young People and Intelligent Partners in the AIGC Era and Risk Assessment [J]. Xinjiang Social Sciences, 2024(06):172-181+185.

From sound to embodiment: AI sound imitation technology driven by N/CM model enables immersive communication:Take the first person monologue narration video as an example

Xinyu Zhu, Haowei Guan, Cong Zhang*

School of Journalism and Communication, Beijing Institute of Graphic Communication, Beijing, 102600,China

***Corresponding author, E-mail:zhangcong@bigc.edu.cn**

Abstract

Recently, video commentaries featuring first-person monologues have gained popularity online. These videos use AI sound technology to reconstruct the story world through the subjective perspectives of characters in the drama, allowing viewers to 'embody' the characters and engage in immersive storytelling. This study employs multimodal analysis (Research 1) and controlled variable experiments (Research 2), based on the narrative-coordination model (N/CM, Narration/Coordination Model). It sets up a scenario experiment with two commentary perspectives (first-person and third-person) and two user technology acceptance levels (high-tech and low-tech users). By analyzing the dimensions of sound and visuals in first-person monologue commentaries, the study explores how AI sound technology endows characters with vivid 'voice life' (Research 1). It also reveals the unique advantages of first-person narration in narrative depth, emotional resonance, and audience interaction (Research 2), aiming to explore the feasibility of using AI sound technology to enhance the film and television industry and to create embodied immersive experiences.

Keywords: AI onomatopoeia; N/CM model; first-person monologue narration; immersion; embodied cognition

1 Introduction

In the global context of actively integrating AI technology with the cultural industry, since 2016, the State Council has issued the '13th Five-Year National Science and Technology Innovation Plan,' which prioritizes artificial intelligence technology. In 2017, the 'New Generation Artificial Intelligence Development Plan' elevated AI technology to a national strategy, with the film and television industry becoming a key focus. In 2021, the National Film Bureau released the '14th Five-Year China Film Development Plan,' which explicitly promotes the application of AI, machine learning, and other technologies across the entire film industry chain. These three policy iterations have not only deepened the integration of AI technology with the film and television industry but also, with the reduction in computing costs and the widespread use of open-source tools, broken down the elitist barriers of traditional film and television industries, allowing ordinary people to gain professional-level production capabilities. A prime example is the first-person narration videos for TV dramas and films, which enhance the audience's immersive experience and emotional connection by cloning the voices of characters and generating personalized scripts, allowing viewers to follow the first-person perspective of the characters to understand the story and its historical context. According to data from platforms like TikTok and B Station, such videos can achieve tens of millions of views per post, and the number of followers on these accounts has seen a significant increase.

This study focuses on the integration of AI sound technology with first-person narration, aiming to address the following key questions: (1) How does AI sound technology endow characters in the drama with a 'voice life'? (2) What unique advantages does first-person narration offer in terms of narrative depth, emotional resonance, and audience engagement? (3) Beyond the film and television industry, can AI sound technology serve as a viable approach for embodied immersive experiences? (4) How can we define and regulate the scope and methods of applying AI sound technology?

2 Literature review

2.1 Immersion

Immersive experience (Immersion) was initially seen as an objective attribute of technical systems. According to the 'Framework for Immersive Virtual Environments' (FIVE) proposed by scholars Slater and Wilbur (1997), immersive experience can be objectively measured through technical parameters, including the Inclusivity, Extensiveness, Surrounding, and Vividness of the display system. This technology-centric perspective emphasizes that when a system fully covers the user's sensory channels (such as vision, hearing, and touch) and achieves an interactive loop, it can achieve the highest level of immersion. However, with the advancement of cognitive science, scholars have begun to focus on the subjective psychological aspects of users. Scholars Witmer and Singer (1998) found a positive correlation between immersive experience and task performance and individual cognitive tendencies, indicating that immersive experience is the result of' selective attention allocation 'and' environment-induced cognitive engagement, 'essentially reflecting the user's active inhibition of real-world perception. The embodied cognition theory proposed by scholars Schubert et al. (2001) further deepens this understanding, suggesting that immersive experience is composed of two core components: SpatialPresence and Involvement. Spatial presence refers to the user's ability to construct a psychological model of the virtual space, while involvement reflects the user's cognitive and emotional engagement with virtual events. In recent years, research on immersive communication has sought to integrate the dual perspectives of 'technology and psychology.' Scholars Cummings and Bailenson (2015) and Agrawal et al. (2019) have redefined immersion as 'an individual's temporary detachment from real perception due to deep cognitive engagement in a specific context, 'using meta-analysis techniques to examine immersion differences from four dimensions: system, content, environment, and individual. Building on this, Cao Zhihui et al. (2024) introduced the cultural dimension for the first time. In their study of the' Chang 'an Twelve Hours' district, they found that the use of localized Chinese symbols,

such as Tang Dynasty architectural styles and street vendors 'cries, can increase user immersion by 32%. The application of collective memory empathy helps construct users' immersion to some extent.

2.2 Narrative/coordination model (N/CM model)

The Narrative Coordination Model (N/CM) is rooted in the interdisciplinary integration of narrative immersion mechanisms. Busselle and Bilandzic (2008) were among the first to explore the cognitive aspects of narrative immersion, proposing the 'Narrative Understanding and Participation Model.' This model revealed that immersion is generated through two parallel processing pathways: narrative and coordination, establishing the prototype framework for the N/CM model. However, it did not clearly define the neural mechanisms or the interaction between modality. Building on this, Cohen introduced a sensory dimension, proposing the 'Harmonization and Synergy Model' (CAM). He argued that narrative immersion is a dynamic balance between structural encoding and meaningful associations. Although this innovation bridged the gap from cognitive description to neural mechanisms, it still lacked an operational definition for perspective transformation. It was not until Denisova et al. (2015) approached the topic from a gaming perspective, comparing players' preferences for game experiences, psychological structures, and performance forms, that the importance of perspective transformation in immersive experiences was truly elucidated. Thus, the N/CM model has provided a clear theoretical framework for understanding and researching immersive communication.

2.3 AI sound imitation technology and immersion

As the concept of immersive communication integrates into the core of media culture production, promoting the integration of media content into a third space that blends the virtual and the real, and optimizing the layout of the immersive communication industry has gained significant attention from the state. In 2020, the Ministry of Culture and Tourism released the 'Opinions on Promoting the High-Quality Development of the Digital Cultural Industry,' which stated, 'Guide and support the application of technologies such as virtual reality, augmented reality, 5G + 4K/8K ultra-high-definition, and drones in the cultural sector, promote the transformation of existing cultural content into immersive content, and enrich virtual experience content. This indicates that immersive communication is not only a potential solution to meet people's cultural and entertainment needs but also a key measure for achieving a modernization where material and spiritual civilization are in harmony. How to leverage AI technology to enhance immersive communication is an urgent issue that needs to be addressed. Based on the research subject of this article, the types of AI technology currently used in the market to enhance the immersion of films and television are as follows (see Table 1):

Table 1 Types of AI technologies used to enhance film and television immersion

Technology type	Core functions	technological superiority	The extent to which immersion is enhanced (Improved over traditional technology)
AI sound imitation technology	Voice cloning/ spatial audio synthesis	The auditory cortex responded at a speed of 150ms, and the amygdala was activated 2.3 times stronger	40-45% (spatial audio H=145.9)
AI image production	Image transformation/ scene generation	The production cost is reduced by 80%, and the visual impact MOS is 4.5	25-30% (resolution H=34.7)

AI real-time effects	Physical simulation/ dynamic interaction	The speed of iteration of the effect is increased by 10 times, and the cost of modification of the effect is reduced by 95%	35-38% (real-time feedback F=29.5)
Audio AR technology	3d sound field/ environmental sound synthesis	The cost of sound scene construction is reduced by 73%, and the fluctuation of sound pressure level ±6dB strengthens the decision-making pressure	30-33% (compared to traditional audio)
XR rendering technology	Multimodal fusion/ virtual reality interaction	The boundary ambiguity of MR environment virtual and real is 89%, and the delay of user action mapping is 12ms	28-32% (FOV expansion H=182.2)

The chart shows that the AI technologies currently used to enhance the immersive experience in film and television can be categorized into five main types. Among these, AI sound simulation technology has the highest impact on enhancing immersion, with a 40-45% increase. Additionally, sound-based technologies offer superior immersive effects compared to image and visual technologies. According to Greenwood (2003), the 'dual processing theory' suggests that sound can directly modulate attention allocation through non-conscious channels, processing information three times faster than vision. This allows soundscapes to more effectively guide users 'cognitive focus. Sanchez-Vives et al. (2005) conducted neuroscience experiments that confirmed that the spatiotemporal consistency of auditory stimuli triggers place illusions, significantly enhancing the audience's perception of spatial positioning. Building on the advantages of auditory elements in physiological and spatial dimensions for enhancing immersion, Slater et al. (2010) conducted a 'virtual environment simulation experiment.' They found that when the immersion level of the visual system decreases, users can still maintain 82% of their sense of presence through spatial audio compensation mechanisms. Conversely, if 3D sound effects are turned off, even the 4K panoramic visual immersion score drops by 63%. Auditory elements play a crucial role in constructing immersion in virtual environments. The experiments conducted by scholar Bansos et al. (2004) also demonstrated that AI-generated dynamic environmental sounds can enhance users 'spatial presence ratings, with the intensity of emotional arousal showing an inverted U-shaped relationship with the complexity of AI-generated sounds. The' dimensional sounds' created by AI significantly enhance immersion and spatial perception.

From the current AI sound simulation technology, AI sound simulation primarily uses deep learning models to deconstruct and reconstruct acoustic features. This process involves three stages: (1) extracting sound features using a CNN-RNN hybrid model, capturing static characteristics such as timbre and fundamental frequency, as well as temporal dynamic features; (2) employing a Transformer architecture for sequence modeling, using self-attention mechanisms to model long-term dependencies, thereby enhancing the naturalness of speech; (3) integrating adversarial training into the generative network (GAN), optimizing the realism of sounds through the dynamic interaction between the generator and discriminator. Today's AI sound simulation technology has achieved the capability of small-sample transfer learning, allowing for timbre cloning in just 3 to 10 seconds, with no restrictions on configuration, scenarios, or professionalism. This has enabled AI sound simulation technology to truly penetrate the market and cater to personalized needs. However, this progress has also brought about social ethical issues, such as AI sound simulation fraud.

Based on previous research, most studies on AI-empowered film and television production have focused on visual topics such as AI-generated images and short AI dramas, with fewer studies exploring the use of AI to enhance immersive experiences. Future research will focus on the N/CM model, examining how AI sound technology can enhance user immersion by constructing immersive virtual landscapes through sound. The following hypotheses and models are proposed (see Figure 1):

H1: The application of AI paronomasia monologue narration can predict the degree of user immersion

positively.

H2a: AI paronomasia monologue narration can predict the degree of user immersion positively under the regulation of first-person narrative perspective.

H2b: AI paronomasia monologue narration can predict the degree of user immersion negatively under the adjustment of third-person narrative perspective.

H3a: The user's technical acceptance can positively predict the degree to which AI paronomasia monologue narration is applied.

H3b: Users' technology acceptance can positively predict their level of immersion.

H4: The degree of user immersion can positively predict the user's technology acceptance.

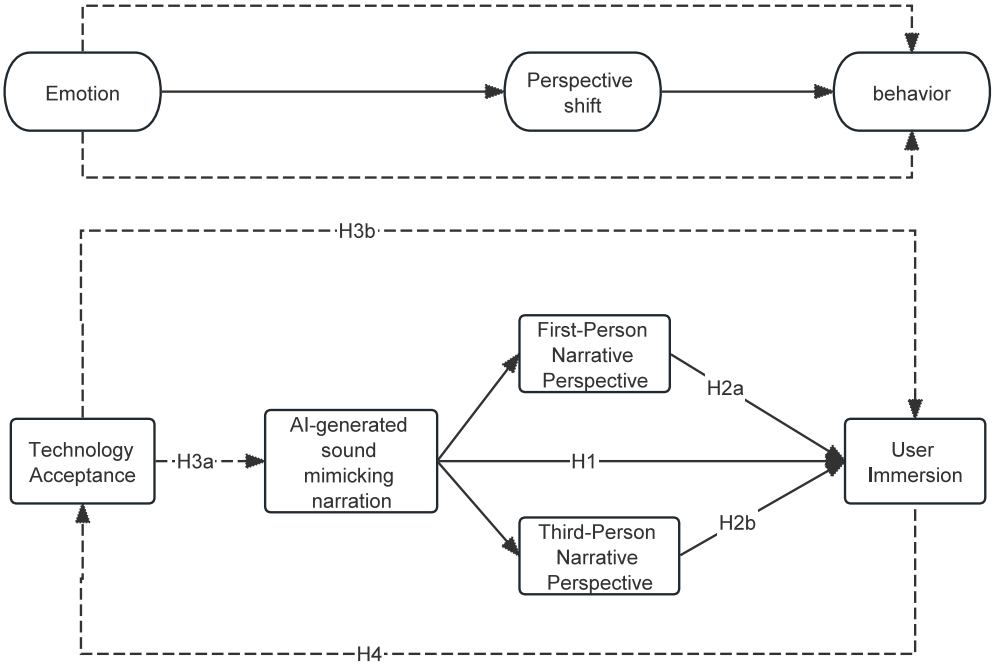

Figure 1 Hypothetical model of user immersion based on AI sound imitation technology under N/CM model

3 Research methods and objects

3.1 Multimodal analysis

Scholar Zhang Delu categorizes multimodal analysis into two dimensions: 'language' and 'non-language,' with the latter focusing on visual, bodily, and other forms of discourse expression. Given that this study focuses on AI paralinguistic technology, the analysis is limited to the 'language' dimension. A sample of videos featuring TikTokAI-generated first-person monologues with over 100,000 likes was selected for analysis, examining both the audio and text dimensions of the intertextual mechanisms. In terms of audio, Praat software was used to extract acoustic features such as fundamental frequency, formants, speech rate, and intonation from AI-generated speech. For the text, Nvivo qualitative coding tools were employed to analyze the emotional tendencies in generated dialogue texts, including psychological metaphors and the use of modal particles in character monologues. The study examines how AI paralinguistic technology constructs a embodied experience of the character's' voice life' from three dimensions: sound reproduction accuracy, emotional appropriateness, and contextual fit.

3.2 Control variable experimental method

This study designs an experimental design with two interpretation perspectives (first-person and third-person video commentary) and two user technology acceptance groups (high-tech acceptance group and low-tech acceptance group). Fifty highly active video commentary viewers, aged 18-35, are randomly assigned to form two groups: 25 in the high-tech acceptance group and 25 in the low-tech acceptance group. A two-way ANOVA model is constructed to test whether first-person narration has a compensatory effect on narrative advantages in the low-tech acceptance group. The experiment consists of three stages: pre-test (technology acceptance survey), intervention (watching randomly assigned video samples), and post-test (multi-dimensional scale + subjective interviews). The post-test scale is adapted from the Tsinghua University Virtual Reality Immersion Scale (2024), which evaluates users' immersion in VR systems through 17 dimensions, including sensory coordination and device usability, to optimize the system and enhance user experience. Given that the research targets the audience of AI-simulated sounds and the original questionnaire had overly complex measurement dimensions, this study revised the measurement objects and dimensions of the original questionnaire. The 'Virtual Reality (VR) system' was replaced with 'AI-simulated sound virtual space.' The immersion experience provided by AI-simulated sounds to users was measured from five dimensions: system level, content level, environment level, individual level, and cultural level. The measurement indicators included 30 items such as 'My emotions change with the AI-simulated sounds' and 'I gain new perspectives and insights through the AI-simulated first-person monologue.' A 5-point scale (1=not at all, 5=very much) was used, with higher scores indicating a greater sense of immersion. In this study, the Cronbach'sα coefficient of the scale was 0.900.

4 Research analysis and conclusion

4.1 Study 1: Multimodal analysis of AI paronomasia first-person monologue narration video

4.1.1 Descriptive statistics

To comprehensively gather target samples, this study entered various keywords such as 'first-person monologue narration,' 'AI monologue narration,' and 'open from the perspective of XXX' into the TikTok search bar, resulting in a total of 388 samples. After cleaning the sample videos based on their titles and tags, 368 valid samples were obtained, achieving an effective rate of 94.8%. Subsequently, AI-sound first-person narration videos with over 100,000 likes were selected for multi-modal analysis, covering multiple languages including Chinese and English, and genres such as film, television, history, and literature, forming an 8-multi-modal corpus. The basic situation of the crawled samples is as follows (see Table 2):

Table 2 Descriptive statistical analysis of the sampled samples

Video type	Average video length	More than 10,000 likes		10,000 likes		100,000 likes		Millions of likes		amount to	
		frequency	proportion	frequency	proportion	frequency	proportion	frequency	proportion	frequency	proportion
film and television	10min46s	253	81.9%	35	11.3%	20	6.5%	1	0.3%	309	84.0%
literature	10min22s	4	57.1%	2	28.6%	1	14.3%	0	-	7	1.9%
Comic and Animation	5min	7	77.8%	2	22.2%	0	-	0	-	9	2.4%
teach school	2min39s	13	92.9%	1	7.1%	0	-	0	-	14	3.8%
society	5min51s	4	80%	1	20%	0	-	0	-	5	1.4%
history	12min28s	4	80%	0	-	1	20%	0	-	5	1.4%

game	4min82s	6	100%	0	-	0	-	0	-	6	1.6%
figure	5min44s	11	84.6%	2	15.4%	0	-	0	-	13	3.8%
amount to	7min9s	302	82.1%	43	11.7%	22	5.9%	1	0.3%	368	100%

The chart shows that the first-person narration videos using AI sound simulation technology can be categorized into eight types, including film and television, literature, and animation. Among these, film and television videos, with 309 examples, account for 84.0% of the total, indicating a broader audience base and higher appeal for dramatic narratives. Despite the smaller sample sizes in animation, games, and social categories, these areas are also potential directions for future AI sound simulation technology applications. Additionally, it is noteworthy that the average duration of videos is inversely proportional to their like counts. Specifically, the average duration of first-person narration videos in film and television, literature, and history categories exceeds 10 minutes, yet they attract tens of thousands (22) and hundreds of thousands (1) likes within the target sample. This suggests that more in-depth video topics are better suited for AI sound simulation storytelling and enhance user immersion more effectively.

4.1.2 Multimodal discourse analysis

(1) Sound cloning dimension: measurement of basic acoustic indicators

This section selects the emotional climax segments from the Chinese AI parodied video “Revisiting ‘Painted Skin 2’ from Xiao Wei’s Perspective” with one million likes and the foreign AI parodied video “Telling the Story of American Psychopaths from Patrick Beterman’s First Person” with ten thousand likes as comparative cases. The software Praat is used to measure basic acoustic indicators such as pitch, sound quality, and rhythm. The aim is to explore the accuracy of AI parodied technology in reproducing the original characters’ voices and its potential for multilingual use in cross-cultural communication.

Table 3 Comparison of basic acoustic index measurement

Compare groups	Select the object	Pitch features			Sound quality characteristics				prosodic features
		Mean fundamental frequency	fundamental frequency min	fundamental frequency max	reduction ratio	Resonance peak frequency F1	Resonance peak frequency F2	Resonance peak frequency F3	Speech rate: words/s
Comparison 1: Original sound restoration comparison	Case 1	165.98	55.98	238.69	75.50%	722.16	1903.64	2975.87	5.2
	Case 2	218.83	144.78	351.16		760.91	1762.62	3014.63	4.57
Comparison 2: Cross-language AI sound effect comparison	Case 1	165.98	55.98	238.69	\	722.16	1903.64	2975.87	5.2
	Case 3	101.58	61.76	408.41	\	763.56	1772.23	2846.39	3.4

Note: Case 1: "Re-telling 'Painted Skin 2' from the Perspective of Xiao Wei" with an onomatopoeic version; Case 2: "Re-telling 'Painted Skin 2' from the Perspective of Xiao Wei" with an original version; Case 3: "Telling the Story of American Psychopaths from the First Person Perspective of Patrick Beterman" with an onomatopoeic version;

In the first comparison experiment of the paronomasia version and the original version of ‘Painted Skin 2’ from the perspective of ‘Xiao Wei,’ based on the acoustic indicators presented by the samples, the fundamental frequency range of the paronomasia version is 55.98~238.69Hz with a mean fundamental frequency of 165.98Hz, while the original version has a fundamental frequency range of 144.78~351.16Hz

with a mean fundamental frequency of 218.83Hz, achieving a restoration ratio of 75.50%. Additionally, both versions show a high degree of overlap in the low, medium, and high frequency ranges below 1000Hz, between 1000~2500Hz, and above 2500Hz. However, it is noteworthy that the sound wave patterns of the paronomasia version exhibit regular changes and pauses, with an average speaking rate of 5.2 words per second, whereas the original version is somewhat irregular, with noticeable long pauses. This suggests that the AI-generated voice quality has reached a level comparable to the original in terms of pitch and sound quality, but its ability to mimic rhythm, emotion, and other emotional aspects still needs improvement.

In the cross-linguistic comparison experiment 2 of 'Revisiting "Painted Skin 2" from a' Xiao Wei 'perspective' and 'Telling the Story of American Psychiatric Patients from a' Patrick Beterman's' first-person perspective 'in their paralinguistic versions, there are significant differences in fundamental frequency and average speaking rate between the two. The maximum fundamental frequency of the English version 3 is 408.41Hz, significantly higher than that of the Chinese version 1, 238.69Hz, while its average speaking rate is only 3.4 words/s, much slower than the Chinese version 1's average speaking rate of 5.2 words/s. This suggests that AI paralinguistic technology exhibits certain differences or tendencies in language cloning. Although the pitch of the English cloned audio is higher than that of the Chinese cloned audio, and its sound resonance and pauses are highly consistent with the Chinese cloned audio range, the excessively slow average speaking rate inevitably leads to a thin text context and delayed image interpretation, significantly reducing user immersion. This provides an optimization direction for future cross-linguistic AI paralinguistic applications.

Table 4 Generative text analysis

subject investigated		Case 1	Case 2	Case 3
type		film and television	literature	history
Emotional dimension	Subjective word frequency	118	114	34
	Frequency of emotional words	975	539	294
	Emotional type ratio	21.2:46.4:32.4	12.2:51.9:35.9	14.9:58.3:26.8
	Emotional intensity assessment	-0.1364	-0.5797	-0.3173
The pragmatic dimension	Frequency of conjunctions	49	21	11
	Contextual style	pathos	banter	roused
	Subject-verb form	declarative sentence	exclamatory sentence	declarative sentence
Content dimension	Cultural extension	Rituals of worship; Strange Tales from a Chinese Studio;	Beijing rickshaw driver culture; a representative work of Chinese literature	The representative of the Chinese bold school
	Time extends	The complexity of human nature is explored through the theme of surrealism	Expose the social life of the lower class under the rule of the Beiyang warlords in old Beijing	The Southern Song dynasty was invaded by the Jin Dynasty, and the literati could not serve their country
Word cloud				

Note: Case 1: "Revisiting 'Painted Skin 2' from the Perspective of Xiao Wei"; Case 2: "The Night Xiao Fuzi Hanged, Xiangzi Finally 'Died' — Lao She's Most Painful Moment"; Case 3: "The Poem I Am Most Proud of in My Life Was Written with Jin Ren's Brain Bile! Xin Qiji"; The emotional types are ranked in order of positivity, neutrality, and negativity; the emotional intensity is assessed by averaging the sentiment scores of the segmented words;

（2）Generative text dimension: measurement of emotional theme analysis

This section selects the only AI-generated sound video with over one million likes from the film and television category, “Recounting ‘Painted Skin 2’ from Xiao Wei’s Perspective,” the only AI-generated sound video with over ten thousand likes from the literature category, “The Night Xiao Fuzi Hanged: Xiangzi Finally’ Died, ‘” and the only AI-generated sound video with over ten thousand likes from the history category, “My Most Proud Poem in Life Was Written with Jin Ren’s Brain Juice! Xin Qiji.” These videos are used as case studies. Using qualitative coding tools like Nvivo and WeWordCloud, the analysis examines different types of generative dialogue texts from emotional, pragmatic, and content perspectives. The aim is to explore whether the emotional semantics and content depth of generative text can help address the acoustic defects of AI-generated sound technology.

In terms of emotional dimensions, the subjective and emotional words in the AI-generated audio-video text of Case 1 are the most frequent, with 975 and 118 instances respectively, significantly higher than those in Case 2 (a literary AI-generated audio-video) and Case 3 (a historical AI-generated audio-video). However, Case 2 has a higher emotional intensity of-0.5797, while Case 3 has a higher proportion of neutral emotions at 58.3%. In terms of pragmatic dimensions, the frequency of conjunctions in the AI-generated audio-video text of Case 1 is higher than in Case 2 and Case 3, with 49 instances. The main sentence structure in Case 2 is predominantly exclamatory. This suggests that AI-generated audio-video content for films and TV series relies more on emotional and subjective words to enhance narrative integration and emotional transmission due to the need for plot and emotional rendering. In contrast, AI-generated audio-video content for literature tends to use the anger of ordinary people to convey sharp negative emotions due to its focus on human nature and social issues. AI-generated audio-video content for history focuses more on factual statements and biographical introductions, with relatively restrained emotional expression. The coherence and intensity of the plot presented in these texts are directly proportional to their final traffic.

From this perspective, the generative texts of different types of films and TV shows and their AI-generated cloned voices exhibit a significant intertextual effect. On one hand, the cloned and restored voice gives the text an ‘eyewitness account’ feel. On the other hand, the text, which combines emotion and content, provides the voice and characters with an ‘observer’s revelation’ omniscient perspective. These two elements complement each other, weaving another layer of space and life cycle into the plot of films and TV shows. This allows the audience to spontaneously explore deeper social dynamics, cultural changes, and human nature through sensory stimulation.

4.1.3Technical path

Based on the comprehensive application of AI paronomasia technology in first-person film and TV drama narration and the multimodal discourse analysis of samples, this section integrates the narrative-coordination model (N/CM), the operational mechanism of AI paronomasia technology, and the immersive communication paradigm developed by scholars Cummings et al. (2015), Agrawal et al. (2019), and Cao Zhihui et al. (2024). This outlines the following technical path (see Figure 2):

As illustrated in the figure, the technical approach of AI-sound-enabled first-person monologue narration can be roughly divided into three layers: the narrative layer, the effect layer, and the function layer. In the previous section, the effect layer and the function layer were explained through literature review and quantitative analysis. However, the narrative layer, which focuses on how perspective shifts enhance user immersion and emotional perception, was not covered. Therefore, Study 2 will conduct user behavior experiments using perspective shift as a mediator, based on the aforementioned technical roadmap, to explore the narrative advantages of the first-person perspective in immersive communication.

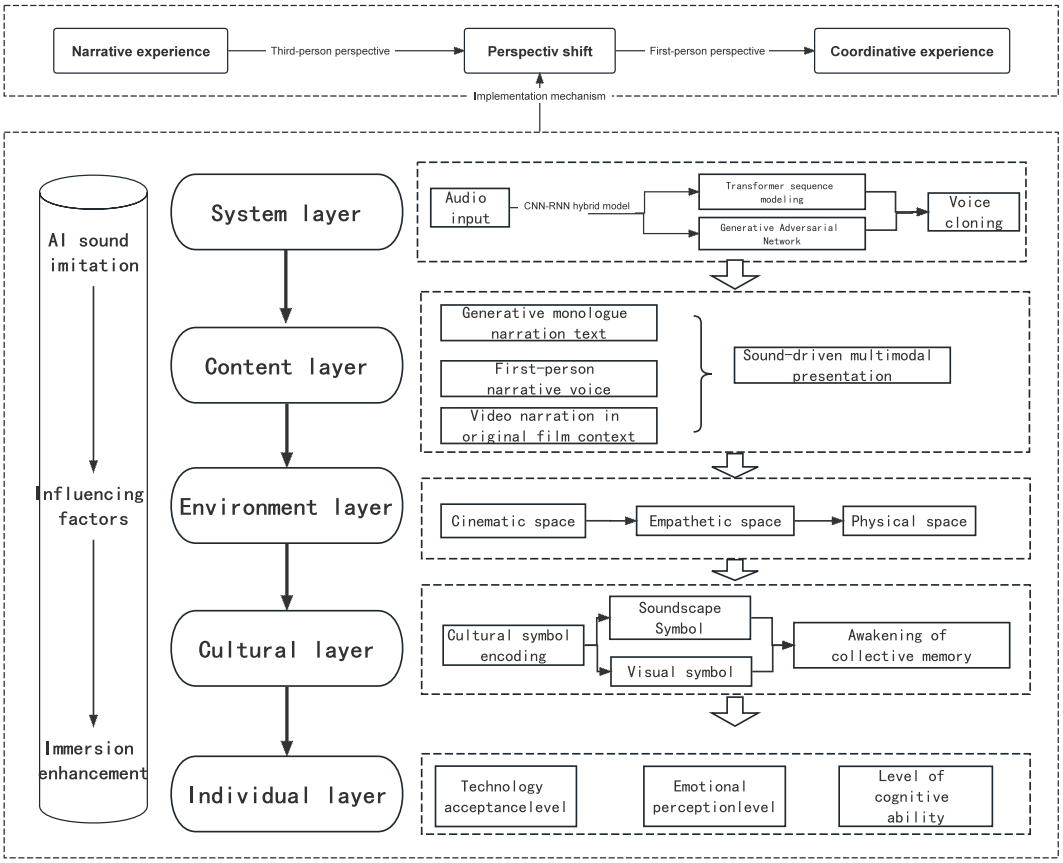

Figure 2. Technical roadmap of AI sound simulation enabling first-person monologue narration

4.2 Study 2: Narrative advantages of AI paronomasia first-person monologue narration from the perspective of inter-group experiments

This study utilized SPSS27.0 and the PROCESS macro plugin developed by Hayes for data analysis. Given that the data were collected through self-reporting by participants and the study involved relationships among multiple variables, it was necessary to conduct exploratory factor analysis and common method bias testing. First, in the questionnaire design phase, the study employed methods such as anonymous surveys, reverse scoring questions, and screening for response time to clean and control the data. Second, before data analysis, Harman› s single-factor test was used to check for common method bias. An unrotated exploratory factor analysis was conducted on all items, and the results indicated that the KMO sample adequacy index of the scale was 0.95> 0.5, the Bartlett› s sphericity test was significant at 0.00 <0.05, and there were two common factors with eigenvalues greater than 1, with the first common factor explaining only 48.7% of the total variance. This ratio is significantly lower than the 50% threshold commonly used in previous studies, indicating that there was no significant common method bias in this study.

Table 5 Validity test table

KMO and Bartlett test		
KMO sample adequacy index.		.951
Bartlett sphericity test	Approximate chi-square	4303.418
	free degree	153
	conspicuousness	.000

4.2.1 Descriptive analysis

This study employed purposive sampling to select participants, primarily from the Beijing and surrounding areas. This was due to two main reasons: first, as an international center for scientific and technological innovation, Beijing has a higher awareness of AI sound technology, making it easier to find suitable participants for this study. Second, it was based on the researchers› personal convenience. The study began with social surveys in Daxing District and Fengtai District of Beijing, and before the formal experiment, participants were pre-tested using the Technology Acceptance Measurement Scale (TAM). Participants were then divided into two groups—Group 1 (n=49, high technology acceptance) and Group 2 (n=59, low technology acceptance)—with a total of 108 participants. The study adopted a mixed experimental design with 2 narration perspectives (first-person, third-person) and 2 technology acceptance levels (high, low). The narration perspective was a within-subject variable, meaning each participant had to watch AI sound commentary videos from both the first-person and third-person perspectives. The technology acceptance level was a between-subject variable, with user immersion as the dependent variable. After the experiment, 210 questionnaires were distributed, and 208 valid responses were collected, including 54 males and 54 females. The average age of the sample was 23.51 years (standard deviation 12.96 years).

Table 6 Descriptive analysis

Experimental group				Gender frequency		Age frequency			
				man	woman	Under 18	18-30 years old	31-50 years old	Over 50
AI imitates the first person	High-tech acceptance sample	count	49	25	24	15	19	11	4
		proportion	23.60%	12%	11.54%	7.20%	9.10%	5.30%	1.90%
	Low technology acceptance sample	count	59	29	30	11	26	10	3
		proportion	24.00%	12%	12%	5.30%	12.50%	4.80%	1.40%
The AI imitates the third person	High technology acceptance sample	count	49	25	24	15	19	11	4
		proportion	24.00%	12%	12%	7.20%	9.10%	5.30%	1.90%
	Low technology acceptance sample	count	59	29	30	11	26	10	3
		proportion	28.40%	13.90%	14.40%	5.30%	12.50%	4.80%	1.40%
amount to		count	208	104	104				

According to the descriptive analysis in Table 6, the gender distribution shows a relatively balanced overall distribution. In the high-tech acceptance sample, males slightly outnumber females; in the low-tech acceptance sample, males are slightly less than females. This suggests that males may have a higher acceptance and willingness to use AI sound simulation technology compared to females. Regarding age distribution, those under 18 and aged 18-30 make up a larger proportion, while those aged 30-50 and over 50 are less common. Additionally, the high-tech acceptance sample among those under 18 is higher, whereas the low-tech acceptance sample among those under 18 is relatively lower. It can be inferred that young people have a higher acceptance and willingness to use AI sound simulation technology.

4.2.2 Main effect test

Correlation analysis is a common method for exploring the mutual influence between two variables. For example, the Pearson correlation coefficient r is used to quantify the relationship between quantitative data. The higher the r value between variables, the stronger their correlation. This study employs the bivariate correlation calculation model in SPSS27.0 to test the two factors influencing user immersion and their hypotheses. The results of the correlation analysis are presented in Table 7:

Table 7 Correlation analysis results among the variables

variable	M average value	SD standard deviation	1	2	3
Narrative perspective	1.52	.501	1		
Technology acceptance	3.25	1.27	-.097	1	
User immersion	3.15	.72	-.947**	.101	1

Note: *p<. 05, **p<. 01, ***p<.001

As shown in Table 7, the correlation index between variables is too high, which may indicate multicollinearity. Therefore, this study needs to further measure the variance inflation factor (VIF) between variables to ensure clear boundaries among variables and prevent model regression distortion. The linear regression analysis results are as follows (see Table 8):

Table 8 Analysis results of linear regression coefficient

	model	Unstandardized coefficients		Standardization factor	t	conspicuousness	Collinearity statistics	
		B	Standard error	Beta			tolerance	VIF
	(constant)	5.209	.070		74.799	.000		
1	Narrative perspective:	-1.366	.033	-.946	-41.987	.000	.991	1.009
	Technology acceptance	.006	.013	.010	.440	.660	.991	1.009

A. Dependent variable: user immersion

The regression analysis results in Table 8 show that the VIF values for the narrative perspective (independent variable), technology acceptance (mediating variable 1), and user immersion (dependent variable) are all below the 5 threshold set by previous studies, indicating no significant multicollinearity among these variables. Furthermore, the model has a good fit to the data (R^2 =0.897), with significant regression coefficients. The narrative perspective can predict user immersion (b=-0.946, SE=0.033, p<0.000), partially supporting Hypothesis H1. In the entire research model, the narrative perspective has a stronger correlation with user immersion compared to other influencing variables. The narrative perspective significantly predicts the degree of user immersion, which forms the basis for establishing the variable of technology acceptance. It is worth noting that the predicted results do not fully align with Hypothesis H1, and the narrative perspective of AI parroting technology can be categorized into first-person and third-person perspectives. Therefore, further investigation is needed in the subsequent grouped mediation effect tests to determine which narrative perspective has a more significant positive predictive effect on user immersion.

4.2.3 The overall impact of perspective transformation on user immersion

The total scores and the difference between the post-test and pre-test scores of experimental group 1 (high technology acceptance) and experimental group 2 (low technology acceptance) are shown in Table 9.

Table 9 Descriptive analysis of the total score difference before and after measurement of each group

group	stage	M	SD	D-value
Experimental group 1	before measurement	25.02	0.29	13.76
	aftertest	38.78	0.14	
Experimental group 2	before measurement	24.84	0.24	13.64
	aftertest	38.48	0.23	

Paired samples t-tests were conducted on the pre-test and post-test total scores of Group 1 and Group 2. The results showed that the post-test total score of Group 1 was significantly higher than the pre-test total score ($t=13.76$, $p<0.01$), and the post-test total score of Group 2 was also significantly higher than the pre-test total score ($t=13.64$, $p<0.01$). This suggests that both the first-person perspective and the third-person perspective significantly enhance user immersion. A one-way ANOVA revealed no significant differences in the post-test total scores of the first-person perspective across different groups ($F=0.594$, $p=0.443$). Another one-way ANOVA showed no significant differences in the post-test total score difference between the first-person and third-person perspectives across different groups ($F=0.069$, $p=0.793$). As shown in Figure 3, during the intervention, the total scores of Group 1 and Group 2 increased while watching the first-person commentary video, but the slope of Group 1 was not significantly different from that of Group 2. This indicates that controlling for the variable of perspective switching had little impact on user immersion, which is not significant. This finding also supports the conclusions drawn from the main effect test.

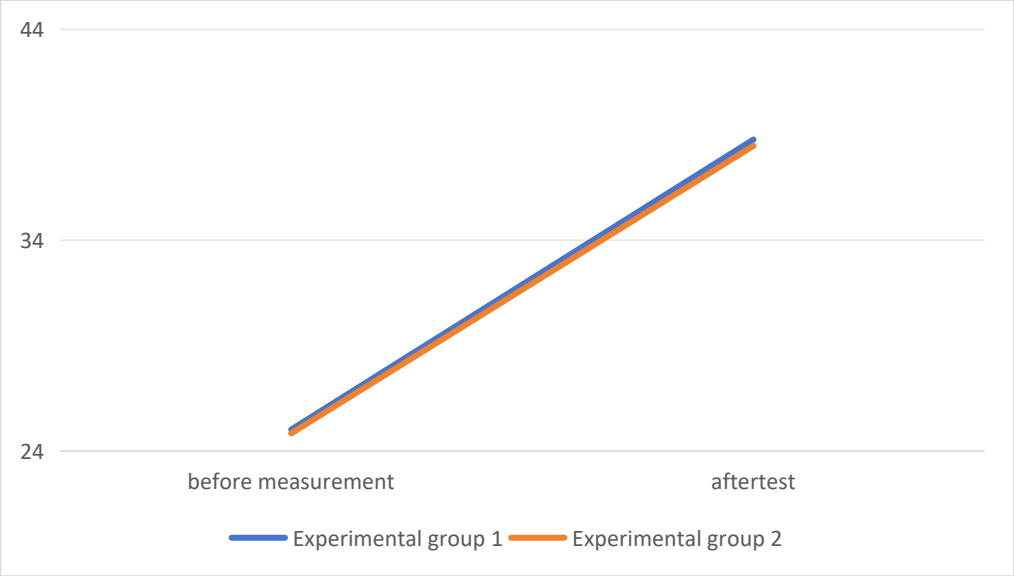

Figure 3 shows the changing trend of total scores before and after measurement in each group

4.4.4Path analysis between variables

This paper employs path analysis to demonstrate the validation results of the relationships among various variables.Based on the descriptive, exploratory, confirmatory factor analysis, and linear regression analyses of the three variables—perspective transformation, user immersion, and technology acceptance—described in the experimental analysis, the significant coefficient relationships are organized into the figure. The path relationships are then corrected and refined based on the original hypothesis model. The resulting path relationship model is presented below (see Figure 3):

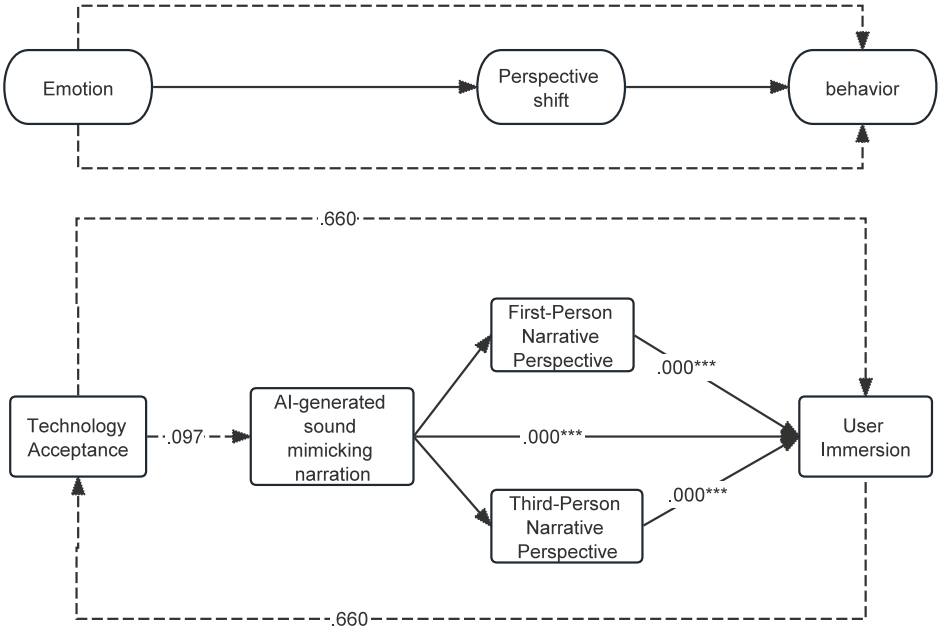

Figure 4 Path model of user immersion by AI sound imitation technology

Based on the coefficient relationships in the path analysis table, the application of AI-sound technology monologue narration significantly positively predicts user immersion. The user's technical acceptance does not significantly affect the degree of immersion or the extent to which users apply AI-sound technology monologue narration, and thus, user immersion cannot positively predict technical acceptance. Combining these findings with the paired samples T-test results from the group experiment, the most significant finding in this study's path analysis is that AI-sound technology monologue narration can positively predict user immersion when moderated by a first-person narrative perspective, and negatively predict user immersion when moderated by a third-person narrative perspective. This result aligns with the mechanisms of hypotheses H1, H2a, and H2b in this study.

5 Insufficient research and measures

While this study, based on the N/CM model, conducted a detailed quantitative analysis of the three variables—narrative perspective (narrative layer), technical acceptance (emotional layer)—of AI paralinguistic technology on user immersion (intentional layer-behavioral), it still has several limitations: (1) The total number of experimental subjects is relatively small, which may lead to a lack of representativeness; (2) The analysis of demographic factors, such as educational background, age, occupation, and place of residence, is limited. These factors could significantly influence user immersion, but presenting data in a purely numerical form does not adequately reflect the user› s immersion experience when using AI paralinguistic technology; (3) The experimental setting is limited, as it was conducted in a specific and relatively monotonous environment, lacking simulation of real-world, diverse usage scenarios; (4) This study primarily focuses on the empirical examination of AI paralinguistic technology in first-person commentary videos, without addressing its broader application areas or copyright issues.

In light of the limitations of previous studies, future research will focus on standardization and interdisciplinary approaches. By designing standardized questionnaire questions and improving sample recovery and cleaning, we aim to enhance the efficiency of quantification and the accuracy of conclusions. We will adopt an interdisciplinary perspective to examine how AI sound simulation technology affects user immersion, aiming to provide theoretical and empirical evidence for enhancing the mechanisms by which AI

sound simulation technology influences human emotions and for further refining and standardizing its application. Building on existing data analysis, we can conduct in-depth interviews and group discussions to conduct qualitative research on the emotional states of individual users of AI sound simulation technology, further detailing and supporting the mechanism diagram presented in the text. Additionally, we can expand experimental settings to include various simulated real-world scenarios, collecting immersive experience data from different contexts. Through comparative analysis, we can identify the impact of scene factors on user immersion, making our findings more relevant to practical applications. Given the characteristics of AI sound simulation technology, it is expected to find widespread application in areas such as game sound effects, audiobooks, smart education, and cross-temporal dialogues in museums. Future research can explore these areas further.

Reference

Central People's Government of the People's Republic of China. Notice of the State Council on Issuing the 13th Five-Year National Science and Technology Innovation Plan[N]. (2016-07-28). Retrieved \from https://www.gov.cn/gongbao/content/2016/content_5103134.htm

Central People's Government of the People's Republic of China. Notice of the State Council on Issuing the New Generation Artificial Intelligence Development Plan [N]. (2017-07-08). Retrieved from https://www.gov.cn/gongbao/content/2017/content_5216427.htm

National Film Administration. Notice of the National Film Administration on Issuing the "14th Five-Year Plan for Chinese Cinema"N]. (2021-11-05). Retrieved from https://www.chinafilm.gov.cn/xwzx/ywxx/202111/t20211109_1182.html

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. Presence: Teleoperators & Virtual Environments, 6(6), 603-616.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire.Presence,7(3), 225-240.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights.Presence: Teleoperators & Virtual Environments,10(3), 266-281.

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Media psychology, 19(2), 272-309.

Agrawal, S., Simon, A., Bech, S., Bærentsen, K., & Forchhammer, S. (2019). Defining immersion: Literature review and implications for research on immersive audiovisual experiences.Journal of Audio Engineering Society,68(6), 404-417.

Cao Zhihui, Tuo Yanzheng, Han Qiuchen, Chen Ye. The Construction Process and Mechanism of Immersive Experience Scenarios—A Case Study Based on the "Chang'an Twelve Hours" Block[J]. Foreign Economics & Management, 2024, 46(9): 67-88.

Busselle, R., & Bilandzic, H. (2008, May). Emotion and cognition in filmic narrative comprehension and engagement. In annual meeting of the International Communication Association, Montreal.

Zhang Xinyue, & Asano Noriko. (2023). Cognitive Processes of Immersive Experience in Narrative Films. Psychological Review, 66 (2), 215-238.

Denisova, A., & Cairns, P. (2015, April). First person vs. third person perspective in digital games: do player preferences affect immersion?. In Proceedings of the 33rd annual ACM conference on human factors in computing systems (pp. 145-148).

Ministry of Culture and Tourism. Opinions of the Ministry of Culture and Tourism on Promoting the High-Quality Development of Digital Culture Industry[N]. (2020-11-18). Retrieved from https://www.gov.cn/zhengce/zhengceku/2020-11/27/content_5565316.htm

Potter, T., Cvetković, Z., & De Sena, E. (2022). On the relative importance of visual and spatial audio rendering on vr immersion. Frontiers in Signal Processing, 2, 904866.

China Netcasting Services Association. 2024 Micro-Short Drama Industry Ecology Insight Report [N]. (2025-01-01). Retrieved from https://pdf.dfcfw.com/pdf/H3_AP202501161641945784_1.pdf?1737039555000.pdf

Potter, T., Cvetković, Z., & De Sena, E. (2022). On the relative importance of visual and spatial audio rendering on vr immersion. Frontiers in Signal Processing, 2, 904866.

Gideon.AI and Real-Time Special Effects: Revolutionizing Action and Sci-Fi Movies.[N](2023-7-5)https://www.michaelrcronin.com/post/ai-and-real-time-special-effects-revolutionizing-action-and-sci-fi-movies

Li, Y., Yoo, Y., Weill-Duflos, A., & Cooperstock, J. (2021, December). Towards context-aware automatic haptic effect generation for home theatre environments. In Proceedings of the 27th ACM symposium on virtual reality software and technology (pp. 1-11).

Ministry of Culture and Tourism. Report on the Development of Digital Cultural Industries [R]. (2025-01-07). Retrieved from https://www.renrendoc.com/paper/37810635.html

Xia Qingying, Li Zhonggang. Riding the Wind of AI to Break Through Tomorrow's Waves——Media Industry Investment Strategy Report [N]. (2023-12-07). Retrieved from https://pdf.dfcfw.com/pdf/H3_AP202312111613695459_1.pdf

Potter, T., Cvetković, Z., & De Sena, E. (2022). On the relative importance of visual and spatial audio rendering on vr immersion. Frontiers in Signal Processing, 2, 904866.

Linda Greenwood.Presence and Flow: A heuristic framework toheuristic framework to inform theory and design[N](2003-1-18)http://matthewlombard.com/ISPR/Proceedings/ICA2003/Greenwood.ppt

Sanchez-Vives, M. V., & Slater, M. (2004, July). From presence towards consciousness. In 8th Annual Conference for the Scientific Study of Consciousness.

Baños, R. M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., & Rey, B. (2004). Immersion and emotion: their impact on the sense of presence. Cyberpsychology & behavior, 7(6), 734-741.

Sohu.com. Can AI Voice Mimicry Software Really Restore Voices 100%? What's the Effect? [N]. (2025-03-13). Retrieved from https://it.sohu.com/a/870328989_121651285

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Patel, A. K., Madnani, H., Tripathi, S., Sharma, P., & Shukla, V. K. (2024, March). Real-Time Voice Cloning: Artificial Intelligence to Clone and Generate Human Voice. In International Conference on Information Technology (pp. 349-364). Singapore: Springer Nature Singapore.

Zhang Delu. Exploration of a Comprehensive Theoretical Framework for Multimodal Discourse Analysis [J]. Chinese Foreign Languages, 2009, 6(01): 24-30.

Tsinghua University. Immersive Experience Questionnaire Survey [N]. (2024-10). Retrieved from https://www.wjx.cn/vm/mbbcmk1.aspx

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Media psychology, 19(2), 272-309.

Agrawal, S., Simon, A., Bech, S., Bærentsen, K., & Forchhammer, S. (2019). Defining immersion: Literature review and implications for research on immersive audiovisual experiences. Journal of Audio Engineering Society, 68(6), 404-417.

Cao Zhihui, Tuo Yanzheng, Han Qiuchen, Chen Ye. The Construction Process and Mechanism of Immersive Experience Scenarios—A Case Study Based on the "Chang'an Twelve Hours" Block[J]. Foreign Economics & Management, 2024, 46(9): 67-88.

Podsakoff, P. M .Self-Reports in Organizational Research: Problems and Prospects[J].Journal of Management, 2016, 12(4):531-544.DOI:10.1177/014920638601200408.

Belsley, Kuh, Welsch. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity [M]. 1980.

The Tension Between Humanism and Capital: The Binary Dilemma in Copyright Protection for AI-Generated Audiovisual Works and Graded Solutions

Zhaoqi Xu[1*], Chun Wang[2]

[12]Beijing Institute of Graphic Communication,Beijing,China

[*]Corresponding author, E-mail:Zhaoqi_Xu0818@163.com

Abstract

The rapid advancements in generative AI have intensified the theoretical conflict between "human-centered" principles and "investment incentives" in copyright protection for AI-generated audiovisual works. Analysis of judicial practices, such as the Ultraman AI copyright case, reveals structural contradictions within the traditional copyright system—particularly in determining rights holders and establishing protection criteria. To address this, we propose a tiered rights-determination model based on the "Human Intervention Index (HII)." This framework classifies AI-generated works into three tiers according to the depth of human intervention, granting them either full copyright, specific neighboring rights, or public-domain status. The model offers an institutional pathway to balance creator rights and industrial innovation.

Keywords:AI-generated audiovisual works; copyright protection; binary dilemma; tiered rights-determination model

1 Introduction: Legal Disorder in the Era of Technological Surge

The rapid advancement of generative artificial intelligence (Generative AI) is profoundly reshaping the landscape of audiovisual content creation and dissemination. Cutting-edge models like Sora demonstrate unprecedented capabilities in automated content generation, fueling explosive growth in AI applications within the audiovisual domain. Authoritative industry reports indicate that the proliferation of such technologies has led to an exponential surge in AI-generated video content.However, this progress has simultaneously triggered increasingly acute conflicts and disputes in the realm of copyright. While greatly unleashing creative potential and lowering the barriers to creation, these new technologies have rendered the identification of infringement and attribution of liability unprecedentedly complex, subjecting the traditional copyright legal system to unprecedented structural challenges.

This technology-driven transformation has exposed profound disorder within the current copyright regime when confronting AI-generated content. The core conflict centers on a fundamental question: When AI-generated content becomes highly proximate to—or even indistinguishable from—human creations in form, should the foundation of copyright protection rest on safeguarding the intellectual labor of human creators or ensuring the economic returns for industry investors?This inquiry touches upon the philosophical foundations and institutional objectives of copyright law, igniting a fierce confrontation between two theoretical approaches: “human-centered” (Anthropocentrism) and “investment incentive.”

The “human-centered” theory, rooted in traditional copyright jurisprudence, emphasizes that the object of protection must embody human intellectual creation and personal expression. Professor Wang Qian (2017) unequivocally argued that AI-generated content, even if externally resembling human works, is essentially “the result of applying algorithms, rules, and templates” and cannot reflect the creator’s unique personal imprint; thus, it should not, in principle, be recognized as a work protected by copyright law.This stance finds resonance in judicial practice. Extraterritorial precedents (e.g., the U.S. Copyright Office’s ruling in the Thaler case) reaffirm that “human authorship is a bedrock requirement of copyright,”while judicial authorities in China consistently stress that “copyright law protects human works, and generative AI cannot qualify as an author.”

Conversely, the “investment incentive” theory is grounded in the practical needs of industrial development. It posits that the emergence and advancement of Sora-level AI systems rely on astronomical capital investment and sustained R&D. Without effective exclusive rights protection, expectations of investment returns would be severely undermined, ultimately stifling technological innovation and industrial prosperity. Professor Xiong Qi (2017) proposed that the originality of AI-generated content could be assessed by existing standards, and its ownership could be determined by drawing on the well-established legal framework for corporate works—vesting copyright in the owner of the AI system to protect investors’ legitimate rights.Scholars like Yi Jiming (2017) further contend that an ownership-centered rights structure should be established to encourage investment and promote technological progress.

The sharp opposition between these two theories has manifested as palpable tension in judicial reasoning. Recent high-profile disputes, such as the “Ultraman AI Copyright Infringement Case,” and groundbreaking rulings like the “AI Text-to-Image Copyright Case,” vividly illustrate the judiciary’s struggle to balance “protecting human creation” with “responding to industrial investment.” Attempts at such balance often result in rulings oscillating between these binary positions.

Thus, the relentless surge of generative AI has profoundly ruptured the logical coherence and institutional stability of the traditional copyright system. The pace of technological iteration far outstrips the rhythm of legal adaptation, plunging copyright law into a profound dilemma of disorder as it vacillates between the dual values of “human-centricity” and “capital-driven” imperatives. This paper delves into the roots of this dilemma, unveils the irreconcilable jurisprudence conflicts underlying it, and explores an institutional pathway capable of accommodating technological revolution while balancing pluralistic values. Subsequent chapters will first deconstruct practical conflicts through landmark judicial cases, then analyze the

deep-seated theoretical antagonisms, and finally propose a tiered rights-determination model centered on the "Human Intervention Index (HII)." This framework aims to provide theoretical underpinnings and institutional design references for constructing a flexible and inclusive new copyright order for AI-generated audiovisual works.

2 Judicial Dilemma: Theoretical Conflicts Revealed by Landmark Cases

The proliferation of generative AI in the audiovisual domain has triggered novel copyright disputes. Among these, the Ultraman AI Copyright Infringement Case and the AI Text-to-Image Copyright Case are particularly emblematic. They epitomize the judiciary's struggle to balance the "human-centered" and "investment incentive" theories while exposing the adaptive crisis of traditional copyright frameworks.

2.1 Dualistic Jurisprudence in the "Ultraman AI Copyright Infringement Case"

In this dispute involving AI-generated content of a renowned anime character, users employed AI services to create commercial short videos by inputting prompts containing distinctive features of "Ultraman."The court's ruling manifested dual logic.

Citing the Civil Code and Copyright Law, the court unequivocally stated that "AI lacks legal person hood," holding users liable as the source of infringement for their prompt-design activities. This aligns with Professor Wang Qian's doctrine that "copyright protection must trace back to human intellectual labor". While users bore nominal liability, the platform was deemed primarily liable for "failing content review obligations" and ordered to pay 90% of the compensation. This implicitly safeguarded capital investment in AI R&D, echoing Professor Xiong Qi's warning that insufficient investor protection would stifle innovation.The verdict thus became a pragmatic concession to industrial realities within a human-centric legal framework.

2.2 "AI Text-to-Image Copyright Case": Paradigm Breakthrough and Theoretical Tension

In an AI-generated image copyright case adjudicated by the Beijing Internet Court, the plaintiff designed multi-layered prompts containing specific elements such as "Chinese classical ink painting style" and "drooping willow branches," along with parameter adjustments, to generate the image "Spring Breeze Brings Warmth," which was subsequently commercially misappropriated. The court made a groundbreaking determination: the plaintiff's debugging of prompts and parameters constituted "personalized selection of expressive elements," with the final image reflecting their intellectual input, thus qualifying as a protected work. This ruling sparked theoretical divergences: critics argued it conflated "ideas" with "expression," contending that prompts are essentially instructions while the visual expression is generated by AI algorithms, potentially lowering the threshold of originality; whereas scholars supporting the investment incentive theory maintained that the decision acknowledged users' creative guidance over the generated results, providing property right expectations for the commercialization of AI tools and preventing diminished user payment motivation due to lack of rights confirmation.

2.3 Core Conflicts and Potential Risks Revealed by Cases

These two cases collectively expose dual risks in judicial adjudication: if strictly adhering to humanism by uniformly treating AI-guided behaviors as "idea instructions" and denying rights confirmation for generated content, it may dampen creators' enthusiasm for using AI tools and hinder innovation in the content industry; if excessively favoring investment incentives by lowering originality standards, it could encourage capital's "enclosure of rights" over AI-generated content, as evidenced by controversies surrounding companies like OpenAI regarding training data infringement, which would erode public knowledge resources and deviate from copyright law's original purpose of promoting cultural dissemination. This dilemma fun-

damentally stems from the impact of AI technology on traditional copyright frameworks. When algorithmically generated audiovisual content approaches human creations, the value conflict between "protecting human intellectual creation" and "incentivizing capital investment" can no longer be reconciled within existing theoretical frameworks, urgently necessitating institutional innovation that transcends binary oppositions.

3 Theoretical Deconstruction: The Binary Opposition Between Humanism and Investment Incentive

The dilemmas in judicial practice regarding copyright protection for AI-generated audiovisual works stem from a profound, structural conflict between the two theoretical cornerstones supporting the copyright system: "human-centered" (humanism) and "investment incentive." This conflict is not merely operational but arises from fundamental differences in philosophical foundations, value objectives, and institutional logic. Under the impact of AI-generated content, their irreconcilability becomes increasingly pronounced.

3.1 Jurisprudential Foundation of Humanism: The Bond Between Personality and Originality

The core of the human-centered copyright view lies in treating works as extensions of human personality and spirit. Professor Wang Qian (2023) articulates this succinctly: the "originality" protected by copyright law fundamentally requires that a work must result from human intellectual labor and reflect the author's personality, thoughts, or emotions. The key to originality lies not in the novelty or aesthetic form of the work but in whether its creation process embodies "human choice, judgment, and arrangement."AI, as a tool, operates by "applying algorithms, rules, and templates to process data"; its outputs are "results of executing predetermined algorithmic instructions," lacking the "unpredictable personalized choices and judgments" unique to human creation. Therefore, AI itself cannot "create", and its outputs cannot reflect human spiritual personality; in principle, they should not receive copyright protection. The bottom line upheld by humanism is: the rights subject must be human, and the protected object must embody uniquely human intellectual creation.

3.2 Practical Rationale of Investment Incentive: Capital-Driven Technological Revolution

The investment incentive theory, grounded in the economic logic of industrial development, acknowledges the real driving forces behind the AI technology surge. Developing Sora-class audiovisual models demands computational power, data, and top-tier talent—requiring astronomical and sustained capital investment. If AI-generated outputs are uniformly excluded from copyright protection solely due to non-human authorship, a paradox emerges: on one hand, developing, training, and operating AI tools demands massive investment; on the other, high-value content generated by these tools provides no exclusive rights protection for developers or substantive contributors (e.g., users who purchase services and invest intellectual guidance). Empirical industry research clearly indicates that lack of copyright protection or unclear expectations is a key risk hindering AI investment, potentially leading to persistently low returns and ultimately stifling innovation and commercialization of disruptive technologies. The core demand of the investment incentive theory is: providing effective property incentives and risk-return mechanisms for substantive investments within the AI industrial chain to ensure sustainable ecosystem development.

3.3 Irreconcilability of the Binary Opposition: Value Conflict and Institutional Paradox

The conflict between humanism and investment incentive is not a mere difference in emphasis but a fundamental opposition irreconcilable within existing theoretical frameworks, manifested in three key paradoxical dimensions:

Conflict Dimension	Human-Centered Stance	Investment Incentive Stance	Manifestation of Irreconcilability
Rights Subject	Creator (Natural Person): Copyright originates from the natural person's intellectual creative activity.	Investor/Developer/Platform: Copyright should vest in entities bearing substantive investment and risk for AI tools/content generation.	Fundamental clash in subject nature: One insists on natural person agency; the other recognizes corporate/capital agency. They cannot coexist in attributing rights to the same object (AI output).
Protection Standard	Originality Reflects Human Will: Protection threshold depends on the depth and uniqueness of human intellectual activity in the content's formation.	Economic Value Creation & Investment Return: Protection necessity depends on commercial value and incentive effect for future investment.	Divergent value foundations: One roots in personal dignity and creative freedom; the other aims for economic efficiency and industrial prosperity. Separated standards misalign protected objects and purposes.
Institutional Cost	High Rights-Confirmation Costs & Judicial Burden: Case-by-case assessment of the "quality" and "quantity" of human intervention in AI outputs is costly, complex, and standards are vague.	Capital Monopoly & Market Imbalance Risks: Granting broad rights to investors may entrench platform monopolies, erode the public domain, and stifle competition and innovation diversity.	Inescapable systemic risks: Upholding human standards imposes high industrial costs; favoring investment incentives inflates social costs (monopoly, public domain shrinkage). Existing institutions cannot avoid both.

The root of this irreconcilability lies in how AI-generated outputs blur the clear boundaries between "creator–creation tool–creative output" in traditional copyright law. Humanism cannot effectively accommodate expressions deeply guided by humans but not entirely "handcrafted" by them; meanwhile, the investment incentive theory struggles to ensure capital returns without excessively "propertizing" algorithmic outputs detached from human intellectual creation, thereby deviating from copyright law's original intent. The judicial pendulum and compromises are inevitable reflections of this deep-seated theoretical dilemma.The solution requires not a binary choice but institutional innovation transcending traditional frameworks.

4 Pathway to Resolution: A Tiered Rights-Determination Model Based on the Human Intervention Index (HII)

Confronting the structural conflict between humanism and investment incentive in copyright protection for AI-generated audiovisual works necessitates transcending binary thinking to establish a flexible and inclusive mechanism that respects copyright law's foundation in human personality while addressing industrial realities. This paper proposes a tiered rights-determination model centered on the Human Intervention Index (HII), aiming to transform abstract philosophical and economic debates into operable judicial and administrative standards through refined institutional design.

4.1 HII Tiered Framework: Graded Mapping of Legal Status and Protection Strength

The core of this model lies in conducting a graded evaluation (HII) based on the depth, breadth, and cre-

ativity of human intervention throughout the AI-generated audiovisual content production chain, thereby assigning differentiated legal statuses and protection modes. The framework is summarized as follows:

HII Assessment Range	Legal Characterization	Rights Attribution & Protection Mode	Typical Scenario Reference
High Intervention	Human Work	Full Copyright: Natural persons or entities (under employment/ commission) performing deep intervention enjoy all economic rights and authorship attribution.	Professional film teams using Sora conduct multi-round prompt iterations, frame-by-frame parameter adjustments, manual lighting refinement, and post-production editing.
Medium Intervention	Human-AI Collaboration	Specific Neighboring Rights: Users or platforms investing substantive guidance gain limited exclusive rights (e.g., reproduction, dissemination, remuneration), without "authorship" recognition.	Users design complex prompts with multi-layered artistic style instructions, specific composition descriptions, and key parameter adjustments to generate unique video clips.
Low Intervention	Algorithm-Generated Output	Public Domain: No exclusive rights granted to any party; free dissemination and use permitted.	Users input basic descriptive prompts (e.g., "summer beach scene video") to directly output standardized AI-generated content.

This model abandons the “all-or-nothing” traditional protection approach, acknowledging a continuous spectrum from “fully human-created” to “purely algorithmic output.” HII assessment focuses on the creative contribution of human intervention, not merely operational duration or step count. Key evaluation dimensions include: (1) Complexity and Specificity of Prompts: Whether instructions contain unique designs and selections of expressive elements (e.g., visual details, composition, style, emotion). (2) Depth and Purposefulness of Parameter Adjustments: Whether users perform targeted refinements beyond preset templates (e.g., frame-by-frame motion trajectory or lighting control). (3) Creativity in Post-Generation Processing: Whether AI-generated materials undergo original human selection, editing, synthesis, or re-creation.

4.2 Judicial Application of HII: Elemental Assessment and Liability Presumption

To enhance HII’s operability in judicial practice, clear evaluation rules must be established. Examples adaptable to specific contexts include:

4.2.1 Elemental Assessment Guidelines

Prompt Dimension: Examine whether instructions uniquely design concrete expressive elements (e.g., character actions, scene layout, color schemes), not merely convey abstract themes.

Parameter Dimension: Assess whether users actively fine-tune key variables (e.g., motion paths, camera angles) beyond default settings, substantially impacting the output’s uniqueness.

Post-Processing Dimension: Evaluate whether post-generation modifications (e.g., selection, arrangement, synthesis) exhibit originality distinguishing the output from raw algorithmic results.

Holistic Judgment: Comprehensively weigh all elements to determine the substantive contribution of intervention to the work’s unique expression, avoiding mechanical aggregation.

4.2.2 Burden of Proof and Presumption Rules

Claimant's Burden: Parties claiming rights (copyright or neighboring rights) must prove their intervention meets the required HII threshold (e.g., providing prompt records, parameter logs, post-production files).

Platform Disclosure Obligation: AI service providers must implement mechanisms to record key user operations and disclose them during disputes.

Low-Intervention Presumption: Content is presumed "algorithm-generated" (public domain) if platforms/users fail to prove medium/high HII levels. This prevents rights abuse and safeguards information flow.

5 Conclusion

The disruption of generative AI to copyright systems fundamentally arises from the irreconcilable clash between the "human-centered" and "investment incentive" paradigms amid technological revolution. Judicial dilemmas reveal that rigid adherence to traditional doctrines or unilateral capitulation to capital logic equally fail to address systemic challenges posed by AI creation. This study proposes the Human Intervention Index (HII), a tiered rights-determination model that resolves abstract jurisprudential conflicts by deconstructing the depth and creativity of human intervention in generative processes. It establishes operable gradients: high intervention safeguards authorship dignity through full copyright; medium intervention secures reasonable industrial returns via neighboring rights; low intervention preserves the public domain by denying exclusivity. This flexible framework transcends the "human-or-capital" binary, injecting dynamic adaptability into copyright law while acknowledging technology's transformative impact on creative ecosystems. Legislators must urgently incorporate tiered rights-determination principles into the Copyright Law Implementing Regulations and mandate HII assessment tools in AI platforms. Ultimately, this paves the way for an inclusive order that nurtures human intellectual creation while accommodating technological revolution—for true wisdom lies not in choosing between protecting humans or capital, but in designing an institutional ecosystem enabling their coexistence.

Reference

Cong, L. X., & Li, Y. L. (2023). Copyright risk of chatbot-generated content and its governance: From the perspective of ChatGPT application scenario. China Publishing Journal, (5), 16–21.

Feng, X. Q. (2022). Theory of intellectual property protection. China University of Political Science and Law Press.

Feng, X. Q., & Guo, C. (2025). Copyright infringement liability and tiered governance of generative artificial intelligence platforms: Reflections on the Hangzhou "ultraman" case. Digital Law, (2), 56–75.

Kartskhiya, A. (2024). Legal horizons of the new artificial intelligence paradigm. Legal Issues in the Digital Age, 5(2), 4–35. https://doi.org/10.17323/2713-2749.2024.2.4.35

Liu, Y. (2023). The copyright protection issue of artificial intelligence generated creations: A dialectical analysis of law and practice. In Proceedings of the 2nd International Conference on Interdisciplinary Humanities and Communication Studies (Vol. 34, p. 1891). https://doi.org/10.54254/2753-7048/34/20231891

Wang, Q. (2017). Qualitative research on the content of artificial intelligence in copyright law. Science of Law (Journal of Northwest University of Political Science and Law), 35(5), 148–155.

Wang, Q. (2023). The qualitative analysis of content generated by artificial intelligence in copyright law. Tribune of Political Science and Law, 41(4), 16–33.

Wu, H. D. (2024). On the copyability of content generated by artificial intelligence: Practice, jurisprudence and system. China Law Review, (3), 113–129.

Xiong, Q. (2017). Is artificial intelligence generated content copyrightable. Intellectual Property, 27(3), 3–8.

Yi, J. M. (2017). Is creation generated by artificial intelligence work? Science of Law (Journal of Northwest University of Political Science and Law), 35(5), 137–147.

Zhang, X. B., & Bian, L. (2024). Research on the copyright of AI-generated content. Journal of Comparative Law, (2), 77–91.

Publication Ethics and Malpractice Statement

The *Journal of AI Ethics and Legal* is an academic journal published by Mason Publish Group. It is necessary to agree upon standards of expected ethical behavior for all parties involved in the act of publishing: the journal editor, the author, the peer reviewer and the publisher. This ethic statement is based on COPE's Best Practice Guidelines for Journal Editors.

Duties of Editor

- The evaluation should be objective and fair, and give clear opinions on whether it meets the standards of public publication.
- Ensure that journals are published on time and on schedule, and prevent plagiarism and repeated publication of papers.
- According to the importance, originality and relevance of the paper to the journal, the author has the right to reject and accept the paper, and allows the author to appeal the review decision.
- The editor shall ensure that all research material they publish conforms to internationally accepted ethical guidelines.

Duties of Authors

- Authors have to make sure that they write completely original works that have not been published elsewhere before.
- The data of the paper is authentic and reliable, and forgery, tampering and plagiarism are prohibited.
- Reference information shall be indicated when quoting opinions.
- The author shall ensure that the published paper is authorized and approved by the coauthor, and there is no signed dispute.
- All authors must contribute to the research of the paper.
- The author shall provide the original data related to his manuscript for the editor's review, and must keep these data.

Duties of Reviewers

In the process of review, the reviewer shall be objective and keep all information of the paper confidential. Reviewers must ensure that the author has identified all data sources used in the study. If invited to review a paper, it shall be completed within the specified time.

Duties of Advisory Board Members

The advisory board is composed of distinguished scholars and professors in the field of Journal of AI Ethics and Legal. Members can review submitted manuscripts, identify topics for special issues, or attract new authors and submissions as necessary.

Journal of AI Ethics and Legal

Call for Papers

We rapid evolution brings urgent ethical and legal challenges, from algorithmic fairness to regulatory frameworks. The Journal of AI Ethics and Legal (Mason Publish Group, peer-reviewed) invites global researchers, scholars, and practitioners to contribute original works on AI's ethical-legal intersections.We welcome original research, case studies, and theoretical explorations linking AI technology, ethics, and law. Join us in shaping responsible AI development and fostering global dialogue in this critical field.

Call for Reviewers

In order to further improve the academic level and quality of the journal, enrich the team of reviewers and implement the anonymous reviewer system, the editorial department is now calling for reviewers. We welcome experts and scholars in the fields of the Journal of AI Ethics and Legal to recommend themselves or recommend suitable candidates, so as to jointly promote the progress and development of periodicals. Please contact us (editorialoffice@masonpublish.org) if you are interested.

2025

MASON PUBLISH GROUP