

Narrative Reconstruction and Aesthetic Shifts in AIGC Generated Imagery: Pathways for Building an Autonomous Chinese Knowledge System of Audiovisual Arts in the Age of Artificial Intelligence

Jingjing Wang¹, Naixin Hou², Zhihao Luo³, Yu Zou⁴, Haowei Guo⁵, Chengming Liu⁶, Shenghao Zhang⁶, Bo Gao⁷, Yimin Wang^{4*}

1. School of Art, Anhui Polytechnic University, Wuhu, Anhui, 241000, China

2. Lanzhou University, Lanzhou, Gansu, 730000, China

3. Industrial and Commercial Bank of China Limited, Suzhou Yangtze River Delta Integration Demonstration Zone Sub-branch, Suzhou, Jiangsu, 215200, China

4. Faculty of Humanities and Arts, Macau University of Science and Technology (MUST), Macau, 999078, China

5. School of Animation and Digital Arts, Communication University of China (CUC), Beijing, 100024, China

6. Hong Kong Baptist University, Hong Kong, 999077, China

7. The City Vocational College of Jiangsu, Nanjing, Jiangsu, 210036, China

*Corresponding author: Yimin Wang, kenjiyiminwang@gmail.com/963417627@qq.com

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: Generative artificial intelligence, commonly referred to as AIGC, produces images and videos through text prompts and is transforming the production mechanisms, narrative structures, and aesthetic standards of audiovisual art. AIGC imagery relies on data training and probabilistic sampling to create visually convincing results. Its narratives are often constructed retrospectively on the basis of generated content, while authorship is shifting from direct control of the camera to the design and refinement of prompts. Drawing on observations from creative practice, this study examines recurring problems in AIGC imagery, including character inconsistency, the breakdown of spatial continuity, and the inaccurate representation of cultural symbols. It further argues that an autonomous Chinese knowledge system of audiovisual arts should be developed through coordinated efforts at the conceptual, methodological, and cultural levels. The study suggests that AIGC, as a generative medium, expands the possibilities of artistic expression while also creating risks of aesthetic homogenization and cultural misinterpretation. Developing an appropriate theoretical response to these changes is therefore essential to the innovation of audiovisual arts in the age of artificial intelligence.

Keywords: Illusion of Realism; Prompt Labor; New Popular Arts and Literature; Autonomous Knowledge System of Audiovisual Arts

Published: Aug 26, 2026

DOI: <https://doi.org/10.62177/chst.v3i3.1629>

1. Introduction

Traditionally, film theory assumes that cinematic images retain a material connection with the objects of representation. Since moving images in a movie are captured by a camera, the viewer assumes that everything shown on screen has existed before the lens ^[1]. In case of a fictional film, the plot of a story is made up; however, the actors, the scenes, the lighting, and the physical movement involved in film production are materially present during production. Material connection of cinematic images to the real object has thus enabled films not only to be a form of artistic expression but also a way for societies to remember and register history.

The advent of Generative Artificial Intelligence has shaken up the traditional understanding of the image. Unlike traditional photography, which involves capturing the light reflected from an object in a scene, generative AI models do not need an object to pass through a camera. Through the use of training data and probability, visual content is generated computationally. An AI model could thus construct an imaginary city, an unreal person, or an apparently historical event that has never happened, yet still produce images that look like professionally produced films. Cinematicity of these images is not the result of photography. These are the result of the system having learnt visual conventions of films through the processing of large amounts of data related to films ^[2].

The relation between image and reality consequently changes. AIGC does not document the existence of an event, but rather generates a visual simulation of reality. In addition, the place of image production is no longer the physical set, but rather the computational environment. To observe how the simulated realism works, practical tests have been conducted using Jimeng and Midjourney platforms. In several rounds of generation, images related to the topics of old Chinese industrial city, video hall in a small county town, and futuristic Chang'an Avenue have been generated. The systems succeeded in creating convincing lighting, colour, and atmosphere in a very short period of time. However, the generation of images has been inconsistent in terms of architectural space, written language, and recurring human characters.

It is not only about the substitution of one technology for another; the change brings the shift in the basis from which the knowledge about the process of audiovisual creation is gained. Traditional filmmaking is based on the experiential knowledge that comes from direct interaction with material processes. Directors work with actors and objects within a scene, cinematographers operate with cameras and light, performers build characters using their body, and editors manipulate duration and rhythm with the help of recordings. AIGC relies on an entirely different set of skills ^[3]. The result of the process of creation is determined by such factors as the formulation of prompts, the tendencies of the models used, training data, and selecting from multiple generated options. Therefore, the creator in the new process is no longer focused solely on representing the world with filming.

At the same time, as far as the development of new popular arts and literature is concerned, AIGC reinforces the role of platform-based production even more. The quick generation and reproduction of audiovisual content and their adaptation to familiar visual forms might contribute to greater involvement but at the same time make the style repetitive and emphasize visual spectacle rather than story or culture. Therefore, the discussions of artificial intelligence as a tool for empowerment present only a partial perspective on the consequences of its application. The other extreme position, which says that artificial intelligence would definitely replace the artist is also wrong.

What needs to be done in order to achieve more relevant conclusions, is to reconsider the basic problems of audiovisual art. The questions that have to be answered include how the sense of reality is formed when images are not based on photography anymore, how stories are affected by the generative processes, how creative authorship is to be understood when the models participate in creating images, and how the cultural meanings are to be chosen, transformed, or distorted by training data ^[4].

2. The Ontological Specificity of Generative Media: The Loss of Indexicality and the Production of Simulated Realism

Photographic indexicity of cinema is another defining feature. Indexicity was firstly introduced in the semiotic theory of Charles Sanders Peirce and then was elaborated in film ontology by such scholars as André Bazin and Siegfried Kracauer ^[9-10]. According to the traditional perspective, the idea of indexicity implies that a photographic image carries the physical trace of an object put before the camera. The light reflected by the existing body, place or event is physically incorporated into the image. Thus, the photographic image preserves the evidence of a real existence.

The film image provides not only visual information; it also creates an expectation that what is seen had a prior connection with reality ^[5]. This evidentiary power became the key aspect in the authority of documentary film, television news, and historical footage. Even in those film theories that contradict about whether the cinema has to reveal the reality or rebuild it in editing, there was one material premise at the start. Before arranging, interpreting and manipulating with the image, some object of the physical world has to be first photographed.

AIGC negates this assumption since its images can be created without an appearance of an original object before the lens. Generative models use patterns learned from huge datasets of images. They create certain expectations about what a face, an urban nightscape, an ancient palace or a cinematic closeup usually look like. Then the model generates the output by combining these patterns according to the semantic guidance of the user. Although the result can be similar to the photograph, the generation occurs not as a product of the physical process linking the photograph with the object.

What AIGC reproduces thus is not a trace of the world but a learned pattern of representation. It imitates how the reality used to be represented in photographs, films, ads, illustrations and online pictures. The object that it shows does not have to exist ^[6]. The realism of its images is created not through the physical contact with the reality but through the visual conventions.

This change is manifest in the viewing experience. A generative system can reproduce the skin texture, the directional lighting, the blur caused by the lens, the hazy atmosphere, the film grain and even the shades of color with a great precision. This instantly activates the viewer's awareness of how photographic and cinematic images look like. The generated picture may seem credible within fractions of seconds because it is consistent with the expectations about how the world photographed professionally should look like.

The viewer does not necessarily believe in the reality of the depicted event. What becomes believable is the image's conformity with the appearance of the photographic truth. One can define this process as an illusion of realism since the image becomes credible by reproducing the formal features that audiences have learned to associate with the camera-based representation ^[1]. The persuasive power of the image comes from its visual familiarity rather than from the real connection with reality.

This credibility may be rather fragile. The closer inspection may show the extra fingers, repeating textures, deformed lettering, inconsistency of facial features, objects not occupying the volume correctly and other irregularities. While such defects are considered temporary technical flaws to be eliminated in the future development of the technology, they may also be interpreted in a broader context. AIGC model does not reconstruct the reality according to the stable understanding of bodies, architecture, language, and physical laws. It calculates visual probabilities.

Thus, the AIGC is different from the camera in that it does not make a copy of the material world. It creates new images based on the patterns pre-existing in the visual culture. It imitates the representations of the reality rather than reality itself ^[7]. What it retains is the accumulated visual appearance of the world in its existing representations, not the material consistency of the world beyond.

This difference was revealed in practical experiments with Midjourney and Jimeng. For instance, one of the tasks was to generate a night street of an old industrial city in China. The generated images quickly acquired the cinematic mood. Wet streets were reflecting the neon lights, the pastel colors indicated the industrial decline, and the generated film grain gave the picture the corresponding visual atmosphere.

In no way should the issues above be attributed to poor software performance. They demonstrate the limitation of a system that has learned to recognize the statistical look of urban images without grasping their meaning in a human way ^[8]. It understands that there might be some connection between rain, neon light, industrial architecture, crowded streets, and cinematic framing, yet it lacks knowledge about the construction of streets, the positioning of bodies in them, and the functioning of written language within them.

Visible artificiality is not always detrimental to the aesthetic value of an AIGC image. In experimental, surrealistic or futuristic artworks, spatial inconsistency and bodily distortion can add to the intended style of the piece. Unnatural movement, weird faces, and strange environment may be used for creating a mood of dreaming, strangeness, or even fear.

Furthermore, some viewers can treat any image created by a generative model as something that has marks of machine

creation. This is why they will be looking for some technical inconsistencies in fingers, eyes, signs, shadows, and architecture. Pleasure of viewing an AIGC image might lie in recognizing its production. A new pattern of perception, thus, emerges. Conviction and suspicion coexist here. Image is credible on the whole, while being dubious in details.

This opens additional expressive opportunities for audiovisual art, yet brings about serious trouble as soon as generative realism is applied to journalism, historical reconstruction, and public communications. An image that has no physical connection with an event can be endowed with the visual authority that was characteristic of photography all the time ^[9]. It can be used as an evidence if the context in which it appears is not provided.

Also, the process of creation of such images makes one develop a new type of relation between visual material and narration. Many traditional filmmakers would start with an image, atmosphere, memory, and dream before proceeding to the development of a story. Akira Kurosawa and Andrei Tarkovsky are examples of directors who treated visual imagination as a base out of which they could build up the whole narrative structure. Image-led creation is, thus, not exclusive to generative technology.

The difference is in origin and amount of images used ^[10]. In a traditional film, visual imagination is shaped by personal memory, artistic preparation, drawing, writing, observation, and material realization. With AIGC technology, there will be many possible images generated randomly by a machine before a storyteller realizes what story these images are parts of. Unexpected images might suggest characters, settings, and conflicts that were not conceived before. Then, the story would be developed under the pressure of image.

This kind of creative process became especially apparent in recent Chinese AI concept films published online since 2024. Many concept videos inspired by the story of Journey to the West have paid little attention to dramatic construction. Instead, they are based on the accumulation of visually saturated images of Sun Wukong, magical landscape, ruined city, mechanical temple, and futuristic Chang'an.

These images often look like music videos, opening sequences, or lengthy trailers. Their appeal mainly stems from their spectacularity, mood, rhythm, and emotional quality. Rather than following a narrative line of action and consequence, viewers are presented with a sequence of strong impressions. It relies on visual accretion rather than traditional narrative development.

This implies that the social power associated with photography is simultaneously broadened and fundamentally transformed by the emergence of AIGC. Consequently, any consideration of the impact of synthetic images must go further than an analysis of style. It must consider not only the ethical deployment of the images but also the context within which the images circulate and how that context might be governed.

3. The Reorganization of Narrative: From Planned Structure to Retrospective Construction

Conventional film and television production usually starts with the creation of a narrative structure. Characters, relationships, goals, temporal order, turning points, and conflicts are described in a screenplay. The job of the director, cinematographer, production designer, actors, and editors is to interpret this structure using audiovisual means ^[11]. Their decisions may differ from the script, yet usually take place in the context of an established dramatic structure.

This preference for the anticipation is tightly connected to the organization of the film industry. Finances, scheduling, casting, locations management, set construction, costumes preparation, and technical coordination usually demand some reasonable understanding of the film content. Narrative structure, in such a way, is not only an artistic choice. It is also a practical tool that allows for the collaboration.

Generative image creation, however, often takes the opposite route. A creator starts with a phrase, mood, color palette, location, or partially formed visual idea. The model generates several possible scenes. Some of them are rejected, and others are selected because of their peculiar faces, forms, gestures, or atmosphere. Only after choosing these images, the creator tries to think about what sort of story can possibly link them.

Thus, the process of creating a narrative structure becomes retrospective. Not the creation of the images to meet the demands of an already established narrative but the development of a story to meet the existing images ^[12]. The generated material

becomes a source of narrative suggestions. Then writing, editing, narration, sound, and music are used to present these images as parts of the sequence of events.

It does not mean, however, that the generative images are unstructured. The more important shift is the moment and way when the structure is built. The power of creating a narrative goes away from the screenplay and to the complex combination of generation, comparison, selection, sequencing, and re-interpretation. The story becomes apparent in the course of processing visual fragments instead of specifying them beforehand.

One could call it the visual notebook narration. The creator gathers visual fragments as he or she gathers observations, sketches, memories, or field notes. Relationships are created only afterwards with the help of ordering and interpretation. The result is not the creation of a traditional dramatic world but a sequence of visual meanings becoming understandable in terms of rhythm, juxtaposition, voice-over, and emotional association.

These shortcomings become particularly apparent in short AI films. Classical dramatic tension usually rests on the interconnection between desire, decision-making, action, resistance, and consequences. The character wants something, faces obstacles, makes decisions, and influences the situation with his or her actions. Generative films, on the contrary, progress through the replacement of one spectacular scene with another. Images may become darker, larger, stranger, or more emotional, but the relations between events stay obscure.

The viewers feel melancholy, fear, wonder, isolation, or historical greatness without understanding why the character acts and what are the consequences of his or her actions. Such films remind of remembered dreams^[13]. They include spectacular images and emotional continuity, but no dramatic causality. The atmosphere links scenes better than plot.

Such development raises the general theoretical question about the definition of narrative in generative audiovisual culture. Certain AIGC products may be more similar to illustrated sequence, comic, music video, concept album, or moving image installation than to a traditional film. Their key organizing principle is not the event developing in terms of action but the mood that becomes varied through the series of related images.

For such cases, it would be appropriate to use the framework that treats certain AIGC shorts as the instruments of emotional narrative. Their main purpose is not the explanation of what happened but the creation of an emotional environment. Images, music, voice-over, and editing are used to lead the viewer through a sequence of emotional states. The success of the work lies in the ability to preserve the experience, even if not building the whole dramatic plot.

Retrospective construction, however, may become a productive creative technique. Generated images may become a raw visual material that is developed into a more consistent story. The key thing is the ability of the creator to go beyond the fascination with the output and create relationships between characters, spaces, actions, and consequences. Writing, editing, performance design, and sound become necessary, as the model does not create the narrative structure on its own.

AIGC, therefore, may be seen as the instrument of expanding the possibilities of visual imagination, not the independent narrative system. The generative system can suggest the setting a creator never thought about, generate variations of a character, or show some unexpected combination of historical and futuristic imagery. But it cannot define why these things are important for the human experience. The narrative meaning still depends on interpretation, selection, and judgement.

The weakness of many AIGC narratives also shows the problem in audiovisual education that existed for a long time. Technical skills are more appreciated than the dramatic construction. The creators spend much time working with the camera, software, lighting, effects, and editing, giving less attention to the character motivations, causal progression, and thematic organization. As the generative systems start to make the generation of spectacular images easier, this disproportion becomes more obvious.

Therefore, the importance of the narrative education may become even higher than before. The creators, who do not know the basics of the construction, produce spectacular sequences. The ones who understand the process of dramatic progression can use generated images as material for the new types of stories. The availability of the technology does not reduce the necessity of the narrative knowledge. It increases the gap between producing images and creating a meaningful audiovisual work.

4. Changes in Authorship and Creative Labour: Prompt Practice, Selection, and Distributed Production

AIGC is transforming the processes of creation of audiovisual imagery and the notions of authorship, creative labour, and aesthetic value ^[15]. In the conventional production of films, the division of labour is structured through identifiable roles. Though film is collaborative, its contributions are usually traceable through a certain identifiable production chain.

This process becomes unstable in the case of generative systems. Once a prompt is provided, the model generates various details that the creator has not necessarily mentioned: the expression on the face, lighting, costume, texture, body pose and spatial arrangement. Thus, the idea that the user of AIGC only pushes the button needs to be reconsidered. Such an assessment ignores the fact that labour does not disappear in the process and is shifted into the prompt construction, testing, selection, corrections and narrative rearrangements.

Writing of prompts requires the translator of the atmosphere, historical era, distance of the shot, cultural context and tone of emotions into words that the model will be able to understand ^[14]. Prompt writing is not just the process of writing more words. Too many instructions will generate semantic conflicts, whereas too little information – visual drifting. The creator has to learn to balance between the need of control and uncertainty, learn the tendencies of a certain model and refine its prompts by comparing them.

As well as prompting, selection is a key element of the process. Generative systems offer several options, and the final product depends on which one the creator chooses, rejects, modifies and connects with others. Though the image looks attractive, it might still be rejected because it contradicts the period, culture, character or narrative. Thus, authorship becomes more visible through judgements, not the construction itself. The creator becomes the director, editor, curator and interpreter of the possibilities of the generated images.

Thus, authorship becomes distributed. The final image results from the collaboration of the intention of the creator, the preferences of the model, its training data and the settings of the platform. Therefore, criticism has to assess not only the creator, but also the model, dataset, interface and the platform. The allocation of responsibility becomes difficult in cases of reproduction of stereotypes, resemblance to living performers and circulation of misleading visual evidence. The theory of AIGC authorship should therefore discuss such aspects as consent, attribution, data provenance, representation and shared responsibility.

The same structure also determines visual style. AIGC is often praised for making cinematic imagery accessible, however, its visual style is determined mostly by the dominant patterns in the training data. Familiar lighting, colouring, depth, and texture have become the recurrent signs of sophisticated images. As such features are generated and rewarded by the recommendation systems, the cycle of generation, model output, audience preference and imitation becomes self-reinforcing. Thus, AIGC functions not only as the productive tool but also as the mechanism which creates aesthetic standards ^[3].

Thus, visual convergence becomes the common outcome. The works about mythology, urban memory, science fiction or historical fantasy might differ in their topics, however, use the same grammar of images. Visual polish becomes a replacement for the narrative elaboration, social experience and conceptual depth. The representation of China becomes especially vulnerable in such a framework. The models could reduce Chinese aesthetics to lanterns, dragons, curved roofs, ink painting, calligraphy and imperial architecture, but not see the industrial cities, regional variety, domestic space, youth culture and changes in society.

The main problem therefore is not only whether AIGC generates new images, but who defines the notions of beauty, authenticity and cultural recognizability. The globally trained models could steer the Chinese creators towards the familiar international conventions and narrow down the range of experiences considered aesthetically valuable. However, machine instability could become the tool of creativity as well. Distorted space, shifting faces, repetitive textures, unstable body could serve for the creation of surrealism, fragmented memory or historical displacement. AIGC should therefore be evaluated not by the ability to imitate the conventional cinema, but by the development of its visual and narrative logic.

The development of such process calls for the development of the independent Chinese knowledge system in the field of generative audiovisual art. Such concepts as indexicality, realism, montage, mise-en-scène and authorship have to be rethought within the framework of the prompt, latent space, probability, training data and platform parameters. The education of creative professionals has to be extended beyond the camera operation and the physical staging to semantic organization,

model literacy, comparative selection, continuity management and retrospective narrative construction^[11]. The traditional artistic education is crucial in this framework, however, it should be supplemented by the knowledge about the biases, synthetic evidence, stereotyping and uncertain responsibility.

The sustainable system has to be developed at the conceptual, methodological and cultural levels. It has to consist of the theories able to explain the generative images, methods of evaluation of prompts and output, and culturally relevant databases with the variety of regional, historical and social experiences of Chinese people. The Chinese language narrative materials and interdisciplinary systems of cultural correction are also necessary. The goal is not to isolate the Chinese audiovisual art from the global exchange or impose the national style. It is to guarantee that the creators could work with generative technologies without losing the cultural autonomy.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Baudrillard, J. (1994). *Simulacra and simulation*. University of Michigan Press. ISBN: 9780472065219.
- [2] Manovich, L. (2001). *The language of new media*. MIT Press. ISBN: 9780262632553.
- [3] Bordwell, D. (1985). *Narration in the fiction film*. University of Wisconsin Press. ISBN: 9780299157704.
- [4] Benjamin, W. (2008). *The work of art in the age of mechanical reproduction*. Penguin. ISBN: 9780141194272.
- [5] Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. NYU Press. ISBN: 9780814742952.
- [6] Mitchell, W. J. T. (1992). *The reconfigured eye: Visual truth in the postphotographic era*. MIT Press. ISBN: 9780262511580.
- [7] Cao, Y., Li, S., Liu, Y., et al. (2023). A comprehensive survey of AIgenerated content (AIGC): A history of generative AI from GAN to ChatGPT. *arXiv*. <https://doi.org/10.48550/arXiv.2303.04226>
- [8] Wu, J., Gan, W., Chen, Z., et al. (2023). AIgenerated content (AIGC): A survey. *arXiv*. <https://doi.org/10.48550/arXiv.2304.06632>
- [9] Bazin, A. (1967). *What is cinema?* University of California Press.
- [10] Peirce, C. S. (1931). *Collected papers of Charles Sanders Peirce*. Harvard University Press.
- [11] Rao, S. (2023). The media turn in new popular literature and art. *Contemporary Cinema*.
- [12] Peng, L. (2024). Generative AI and the restructuring of communication structures. *Modern Communication (Journal of Communication University of China)*.
- [13] Zhou, X. (2024). Artificial intelligence and new transformations in China's film and television industry. *Film Art*.
- [14] Liang, W., Shen, H., & Hou, N. (2026). Performative algorithmic deviance: Rationalization and social negotiation of traffic violations among Chinese food delivery riders. *Deviant Behavior*, 1–17.
- [15] Liang, W., Shen, H., & Hou, N. (2026). Learning to report with AI: Journalism students' professional identity and boundary work in the age of generative tools. *Journalism*. <https://doi.org/10.1177/14648849261458974>