International Journal of Asian Social Science Research; Vol.3, No.4; 2026
E-ISSN: 3006-2179 P-ISSN：3006-754X
DOI: https://doi.org/10.70267/ijassr.260304.7885
Published by: Zeus Press

Multimodal Content Presentation Characteristics and Algorithmic Reconstruction of AIGC-Generated Science Popularization Short Videos: A Case Study of the Douyin Platform

Lebing Xu*

School of Journalism and Communication, Hebei Normal University, Shijiazhuang 050024, China

**Corresponding author: Lebing Xu.*

Abstract

Generative artificial intelligence is deeply involved in the production of popular science short videos and promoting the dissemination of scientific knowledge into the platform in a multimodal form. However, AIGC not only improves the efficiency of content production, but also may change the way knowledge is organized and the path of understanding. This study takes the AIGC-generated science popularization short video on the Douyin platform as the object of observation, and analyzes its multimodal expression characteristics from the visual, linguistic, and auditory dimensions. The study found that AIGC popular science short videos have three main characteristics: visual spectacle, linguistic templatization and auditory style differentiation. Under the mutual influence of technical logic, platform logic and attention logic, the expression of scientific knowledge undergoes algorithmic reconstruction. This process enhances the perceptibility of scientific knowledge and accelerates its dissemination, but it also causes scientific authenticity to be aestheticized and poses the risk that deep understanding will be replaced by sensory experience. The optimization method lies in taking the promotion of understanding as the anchor point and re-forming the integrity of knowledge expression in the multimodal joint action.

Keywords

AIGC-generated science popularization short videos, multimodal presentation, algorithmic reconstruction, cross-modal synergy, communication efficiency and depth of understanding

1. Introduction

At present, short video users in China account for 95.5% of the total number of netizens, with about 1.05 billion users [1]. Short videos have become one of the main channels for the public to obtain information. At the same time, the rapid development of generative artificial intelligence is changing the existing pattern of content production [2]. In the field of short videos, AIGC technology can automatically generate text, images, audio, video and other multimodal content, providing new technical conditions for popular science dissemination [3]. Against this background, AIGC-generated science popularization short video is not only the result of the application of new technology in the field of content production, but also a typical form of

scientific knowledge being reorganized in the platform, visual and algorithmic communication environment. This change not only means the improvement of the production efficiency of popular science content, but also means that the expression logic of scientific knowledge is being reorganized.

Popular science short videos are different from general entertainment short videos. Their core task is to transform relatively abstract and complex scientific knowledge into content that the public can understand and accept. Therefore, how the picture presents scientific concepts, how the language interprets the logic of knowledge, and how the sound adjusts the rhythm of communication will directly affect the understanding effect of popular science content. After the intervention of AIGC, the visuals of popular science short videos are richer, the language is more standardized, and the sound is more standardized, but there may also be problems such as superficial interpretation, content templateization, and poor cross-modal coordination. In other words, AIGC does not simply "empower" science popularization short videos, but it may also change the knowledge organization and dissemination quality of science popularization content.

This article analyzes the multimodal presentation characteristics of AIGC-generated science popularization short videos from the visual, linguistic, and auditory dimensions, and studies the functional differences and dissemination problems of different modalities in knowledge expression, so as to answer the characteristics of AIGC-generated science popularization short videos in terms of visual, language and hearing and how they affect the expression of scientific knowledge and public understanding.

2. Literature Review

As an important medium for the current scientific communication, the research on popular science short videos has gradually attracted attention in recent years. Relying on its compact rhythm, audio-visual combination and high efficiency of platform distribution, short videos not only reduce the difficulty of the public's access to scientific knowledge, but also expand the audience coverage of popular science content [4]. Its content expression directly affects the communication effect: the configuration of visual images, verbal commentary, subtitles, soundtracks and sound effects jointly determines the audience's understanding and acceptance of scientific information [5].

AIGC technology can generate text, images, audio, video and other multimodal content, and can integrate massive scientific data to realize the logical association and visual conversion of scientific concepts [2]. Research shows that AIGC has obvious advantages in lowering the threshold of creation, improving content production efficiency, enhancing visual presentation and information visualization [3]. At the same time, AIGC generated content generally uses the collaborative mode of combining algorithm generation and manual proofreading [6]. While ensuring scientific accuracy and effective communication, it also brings potential problems, such as content homogenization and algorithm bias [3], as well as scientific explanation shallowness [2]. This shows that AIGC not only empowers popular science short videos, but also introduces new challenges for content presentation and dissemination.

Multimodal discourse refers to the phenomenon of using auditory, visual, tactile and other sensory perceptions to communicate through language, images, sounds, actions and other means and symbolic resources [7]. In popular science short videos, visual representations are often used for conceptual presentation and situational construction, language commentary undertakes the functions of knowledge explanation and logical organization, and auditory elements are used to strengthen the focus and regulate the rhythm [5]. However, AIGC generated short videos have different characteristics in multimodal combination: the visual content may be highly realistic or symbolic, the language text presentation is templated or generalized [2], and the auditory content lacks affinity [8]. These changes may not only improve the visual attractiveness and information transmission efficiency of short videos, but also lead to problems such as scientific interpretation shallowness and insufficient cross-modal coordination.

The existing research has laid the foundation for the communication characteristics of popular science short videos, the potential of AIGC technology and the theory of multimodal expression. However, under the background of AIGC generated content, the systematic analysis of the micro-presentation characteristics of popular science short videos is still insufficient. Based on this, this article intends to take the Douyin platform AIGC-generated science popularization short video as the research object, and examine its content presentation

characteristics from the visual, linguistic, and auditory dimensions, in order to provide reference for the optimization of AIGC science popularization content.

3. Multimodal Presentation Characteristics

3.1 Visual Presentation: From Scientific Visualization to Algorithmic Landscapes

Observing the Douyin platform, it can be seen that AIGC science popularization content visually shows the characteristics of high imitation and homogeneity. Whether it is microscopic cells or cosmic nebula, algorithms usually output similar styles: high-saturation colors, particle special effects and futuristic lighting effects. The essence of this phenomenon can be called " algorithmic landscape", that is, the algorithm is highly realistic, dazzling and technologically generated according to the training data and the aesthetic preferences of the platform, but it does not necessarily strictly conform to the scientific and real visual characterization system.

The function of "algorithmic landscape" in popular science communication can be understood from three aspects. First, visualize invisible objects. In traditional science popularization, the observation of micro or macro objects such as the internal structure of cells and the cosmic nebula often relies on professional modeling or live shooting equipment. The production threshold is high and the cycle is long, but AIGC can quickly transform it into a film-level presentation. The Douyin account "Science Travel" uses AIGC to generate deep-space scenes in large numbers, transforming astronomical concepts such as exoplanet surfaces and black hole accretion disks into realistic dynamic images, allowing ordinary users to witness celestial phenomena that could only be imagined through scientific data. Second, make the abstract process dynamic. One of the difficulties of popularization lies in explaining the process - how proteins fold and how cell division proceeds. AIGC can transform static scientific concepts into continuous dynamic pictures. In the "A Brief History of Life Science" of "Big Round Mirror Science Popularization", the audience can see that protein folding is presented as a continuous process of twisting and compressing amino acid chains in space and finally forming a specific three-dimensional structure, rather than just seeing the static contrast between "before folding" and "after folding". Third, contextualize the complex mechanism. AIGC can form an easy-to-understand macro analogy scenario to describe the micro mechanism. When explaining the signaling pathway, the signaling pathway can be regarded as the sequentially lit signal lights in the urban transportation network, so that the audience can understand the operating logic of the micro world through daily experience. These functions change the dissemination of science from abstract description to physical experience, which significantly enhances the perceptibility of scientific knowledge.

3.2 Linguistic Templatization: From Knowledge Explanation to Template-Based Narration

The linguistic dimension of AIGC popular science short videos also shows the template characteristics driven by algorithms. Yang Qiguang and others pointed out that the text generation of AI is essentially based on statistical induction and linguistic dimension of large-scale corpus. It follows the law of large numbers and median logic, and interprets and reconstructs the text according to internal coding rules. [9] In other words, its working mechanism is based on the statistical probabilities of vocabulary matching in the training data, not the causal understanding of the connotation of knowledge. This mechanism interacts with the platform ecosystem of the Douyin platform, which together shapes the template trend of science communication scripts.

The template of the discursive opening is the most obvious feature. A large number of AI popular science videos use fixed rhetoric to open the narrative, such as "Do you know" and "I'll take you to understand today" or create cognitive conflicts with contrasting questions." Cognition convenience store M!" The 86 works of the account attracted 1.15 million fans, and most of the videos were "Why is milk packed in a square box and Coke in a round bottle?" This programmatic beginning has become an effective attention capture strategy verified by algorithms.

The structure of the text also shows characteristics worth distinguishing. "Cognitive convenience store M!" The video generally follows: the promotion method of asking questions, attribution analysis, and logical sublimation. It is necessary to distinguish between the two concepts here. First, typification. It refers to the stabilization of the narrative structure, which helps to reduce the cost of audience understanding; second, templateization. It points to the mechanical application of content production, that is, the creator applies the

same narrative framework to different types of knowledge with different cognitive difficulties without adjustment.

The fundamental difference between the two is not the framework itself, but whether the framework and knowledge match. “Cognition convenience store M!” The popular video guides the audience to think and understand the principles of economics through life-oriented questions. Its narrative structure is in line with the logic of knowledge, which is a reasonable way of typified use. However, if this model is used for quantum entanglement, it presets the cognitive premise that “every phenomenon has a concise answer”, and the superposition state and nonlocality of quantum entanglement are precisely counter-common sense and anti-linear attribution. When this misalignment occurs generally, typification slides to templateization. The framework is no longer a tool to explain knowledge, but knowledge has become a material to adapt to the framework.

3.3 Auditory Configuration: Standardization, De-emotionalization, and Stylistic Differentiation

The auditory dimension of Douyin AIGC popular science videos mainly relies on AI synthetic voice to dominate. Tools such as Clipping AI Reading and ElevenLabs have almost become standard. Their main advantages are that they have clear words, uniform rhythm and can be mass-produced. However, the emotional insufficiency and excessive smoothness of AI dubbing have triggered completely different audience reactions in different content positioning, and thus evolved three application paths.

The most common path is standardized dubbing. Creators directly use the standard tone of AI reading tools to achieve the purpose of clear, stable and efficient output, which is widely used in the oral explanation of various knowledge. The “Cognitive Awakening” account, relying only on a stable AI male voice to explain human psychology, has accumulated hundreds of thousands of fans in a short time. This kind of dubbing filters out the breathing sound and natural pauses, and the hearing is clean but lacks the breathing feeling. When audiences repeatedly hear the same synthetic voice, they are more likely to associate such content with “mass production.”

Some creators use the “sense of calmness” of AI in reverse, giving rise to a professional dubbing path. “Big Round Mirror Science Popularization” uses ElevenLabs to generate narration throughout the whole process. The team once tried to use human voice-over, but many users commented that AI's tone without emotional fluctuations is more suitable for the narrative of high-density hardcore knowledge - the emotional lack of AI dubbing is re-coded into a professional style label.

Furthermore, some creators completely jump out of the dubbing framework and move towards entertainment expression, changing the form of knowledge dissemination through AI music. “Abstract Science Popularization” generates lyrics and melodies through AI, and rewrites core biological knowledge such as motor proteins into rap forms. A single video has received more than 3.35 million likes. “South of Nanxi” changed the copy of animal science popularization to Chinese rap, and the collection of “Animal Rap Science Popularization” reached 350 million views. Hearing is not only the carrier of information, but also the logic of knowledge is embedded in rhyme and rhythm. This also shows that in AIGC popular science short videos, auditory modality is not an accessory to the visual picture, but an important symbol resource that participates in shaping the sense of knowledge authority, emotional rhythm and spreading memory points.

4. Algorithmic Reconstruction of AIGC-Generated Science Popularization Short Videos

4.1 Technical Generation Logic: Statistical Nature and Aesthetic Preferences

Observing the current operation mode of the mainstream generative model, AI content generation mainly relies on learning and reorganizing the pattern relationship in the training data. The goal of visual model optimization is to make the image generated by the model look like a real cell, rather than understanding the real structure of the cell. The basis of the large linguistic dimension to generate text is to imitate the probability of vocabulary matching, rather than the causal understanding of the connotation of knowledge. The audio model pursues the natural smoothness of the sound, rather than the appropriate expression of semantics. These

three have the same statistical essence. They all replicate the most frequent patterns in the data and do not explore the authenticity and integrity of knowledge.

But it is precisely because the model learns patterns rather than rules that it can avoid the professional modeling process required by traditional scientific visualization and produce dynamic pictures with cinematic quality in a short time. For ordinary audiences without professional background, this conversion means a significant reduction in the cognitive threshold: DNA is no longer the linear diagram in textbooks, but a dynamic structure that can be seen; AI commentary text will automatically reproduce the narrative patterns that appear frequently in statistics; AI synthetic voice uses uniform rhythm to filter out breathing and emotional fluctuations. In the process of improving the efficiency of content production, these technical features also allow scientific knowledge to reach more audiences in a highly standardized form.

However, when technical logic replaces causal understanding with pattern reproduction as a whole and replaces scientific reproduction with aesthetic reconstruction, the “scientific image” generated by it is essentially a statistical reorganization of the visual pattern in the training data, which is not a faithful presentation of the structural relationship of scientific objects. The true boundaries of knowledge are becoming blurred.

4.2 Platform Traffic Logic: The Filtering Effect of Communication Efficiency on Knowledge Expression

The platform traffic logic does not directly stipulate the expression of knowledge, but uses traffic distribution to filter knowledge. Douyin takes the completion rate as the main indicator of the recommendation algorithm to fundamentally evaluate the efficiency of content acquisition and maintaining attention, not the accuracy or depth of knowledge communication. The expression that can quickly stimulate the viewing behavior to gain higher weight, otherwise it will be marginalized. The key point of this method is: what kind of knowledge expression can survive and spread.

The recommendation algorithm has a significant content amplification effect. Once a high-quality popular science video breaks through the cold start stage and obtains a high completion rate and high interaction rate, its reach may far exceed the traditional popularization channel. Platform logic objectively provides unprecedented communication opportunities for high-quality popular science content.

But the problem is that the algorithm is based on the user's behavior data: stay time, completion rate, likes and comments, not the accuracy or systematicity of the knowledge itself. The reason why the narrative model of “Cognitive Convenience Store M” has become a traffic template is not because it is most suitable for explaining economic knowledge, but because it is repeatedly verified as efficient in data feedback. This success has transformed it into the imitation object of other creators, and the template has changed from an individual strategy to the collective preference of the platform ecology. In the process, the logic of knowledge expression has shifted, which does not entirely depend on the structure and difficulty of the knowledge itself, but is mainly affected by the external standard of the algorithm to identify it as high-quality content.

Algorithms are more inclined to highly stimulating, low-threshold and highly certain expressions, while the complex, open and uncertain characteristics of scientific knowledge face the risk of being marginalized. The screening logic of platform traffic does not directly shape the superficial expression of knowledge, but quietly affects the way of expression of knowledge.

4.3 Audience Attention Logic: The Tension Between Sensory Capture and Deep Understanding

Technical logic forms the generation conditions, the platform logic runs the screening method, and the audience attention logic constitutes the terminal link of the reconstruction chain. In the short video consumption scenario where the audience's attention is highly distracted, the content must complete the process of capturing attention, conveying information and forming memory in a very short time. This constraint deeply affects the expression of AIGC science popularization, that is, to complete attention acquisition, emotional stimulation and problem introduction in a short time.

Its advantages in communication efficiency are worthy of recognition. The reason why AIGC popular science short videos can stand out in the massive content is precisely because the visual spectacle, rhythmic

hearing and template language enable scientific knowledge to complete information communication before attention is lost. For the audience who originally lacked interest in scientific topics, the strategy played the role of cognitive entrance, that is, first attracted by the gorgeous picture of DNA and the rhythm of AI rap, and then unconsciously exposed to the knowledge content.

However, the audience of "abstract science popularization" can smoothly sing the rap lyrics of " motor protein", but may not be able to retell its biological mechanism. Cognitive load theory points out that [10] that the capacity of human working memory is limited. When information processing needs to handle multiple cognitive tasks at the same time, there will be resource competition between each task. When users receive AIGC science popularization, there happens to be this competitive relationship between sensory rhythm following and semantic logic understanding. The auditory channel is occupied by fast-paced rhymes, and the cognitive resources used for deep understanding will be compressed. The audience is fully satisfied at the sensory level, but may get very little at the semantic level. This is the main paradox of the communication effect of AIGC popular science short videos: the stronger the communication power of the expression, the more likely it is to compress the depth of understanding.

5. Recommendations for Optimization

5.1 Visual Dimension

The picture should serve scientific interpretation, not a visual spectacle independent of the content. Specifically, the visual content generated by AIGC should be marked on the structural relationship and causal logic of scientific objects. For example, when displaying DNA double helix, it is necessary to mark the position and direction of base pairing at the same time. When presenting protein folding, it is necessary to mark the driving force and intermediate state of folding. Designers should avoid using glare effects, particle effects and overly smooth material renderings without scientific basis. Although these visual rhetorics can improve the viewing experience, they may blur the real structure of scientific objects. When using anthropomorphic expression, it is necessary to maintain a cautious attitude, especially when describing the molecular mechanism. Overly personified visual metaphors are easy to mislead the audience into the perception of "molecules are conscious".

5.2 Linguistic Dimension

Explanation needs to retain the conditions, boundaries and uncertainties included in scientific concepts. The popular science copy generated by AIGC should avoid compressing complex questions into a closed expression of a single answer. For cutting-edge issues such as quantum entanglement and gene editing, explanations need to clearly distinguish between known and unknown, consensus and hypothesis, and avoid giving conclusions of absolute truth. The choice of narrative mode needs to be matched with the category of knowledge: the linear causal model is suitable for dealing with life-oriented economic problems, but when explaining the complex system of multi-factors, the description of necessary conditions, restrictions and uncertainties should be retained. In other words, templates can stabilize the narrative structure, but should not compress the complexity of knowledge itself.

5.3 Auditory Dimension

The use of dubbing and music should serve key content prompts, not create continuous stimulation. In standardized dubbing, the conversion of important concepts can be marked by varying the speech rate and short pauses to help the audience establish a sense of structure of knowledge in hearing. In the expression of entertainment, it is necessary to be wary of the suppression of semantics by rhythm. Although rap and other forms can improve communication power, it should also ensure that the logic of scientific knowledge is clear in the lyrics, not just as a filling content for rhymes. Music and sound effects should be aimed at enhancing understanding nodes, rather than maintaining continuous sensory stimulation.

5.4 Production Dimension

The optimization of AIGC popular science short video content should not stop at the quality improvement of a single mode, but should be implemented in the cross-modal unified design of the production link. One of

the main findings of this article is that the main dilemma faced by AIGC popularization short videos is not the lack of quality of a single mode, but the lack of unified semantic guidance between each mode.

To solve this dilemma, professional proofreading not only needs to carry out factual verification, but also to check whether the structure of visual presentation corresponds to the logic of language explanation, and whether the node of auditory enhancement coincides with the key points of knowledge. In the generation stage, the effectiveness of knowledge dissemination is regarded as the goal of multimodal joint optimization, instead of simply splicing after each modality is generated. This is not only a requirement for the artificial intervention ability of creators, but also a promise for the development of AIGC tools from "general content production" to "popular science communication specialized expression system". Only in this way can visual, language and hearing truly function and cooperate with each other to serve the ultimate goal of scientific understanding, instead of being reduced to the accumulation of their own senses.

References

[1] *The 54th Statistical Report on China's Internet Development* released. (2024). *Media Forum*, *7*(17), 121.

[2] Zhang, Y., Zhu, Y., Huang, R., et al. (2025). Research on collaborative optimization strategies for scientific accuracy and narrativity in science popularization creation in the context of AIGC. *Studies on Science Popularization*, *20*(2), 43–52, 80, 107–108. https://doi.org/10.19293/j.cnki.1673-8357.2025.02.005

[3] Wang, S., Yan, Y., & Li, Z. (2024). Generative artificial intelligence empowering science popularization: Technological opportunities, ethical risks, and coping strategies. *Studies on Science Popularization*, *19*(4), 5–13, 22, 101. https://doi.org/10.19293/j.cnki.1673-8357.2024.04.001

[4] Li, G., Wang, D., & Jie, Q. (2025). Strategies for enhancing the visual rhetorical effects of short videos produced by science popularization journals. *Science-Technology & Publication*, (4), 47–53. https://doi.org/10.16510/j.cnki.kjycb.20250407.001

[5] Zhou, S., & Huang, W. (2025). Research on multimodal narrative optimization strategies for science popularization short videos: A case study of "Huazha Huaxiaolao" on Bilibili. *Studies on Science Popularization*, *20*(2), 81–90, 104, 109–110. https://doi.org/10.19293/j.cnki.1673-8357.2025.02.009

[6] Zhou, S., Tao, M., & Liu, X. (2024). Underlying logic of AI-generated science popularization content and functional analysis of participating actors. *Studies on Science Popularization*, *19*(4), 14–22, 101–102. https://doi.org/10.19293/j.cnki.1673-8357.2024.04.002

[7] Zhang, D. (2009). Exploring an integrated theoretical framework for multimodal discourse analysis. *Foreign Languages in China*, *6*(1), 24–30. https://doi.org/10.13564/j.cnki.issn.1672-9382.2009.01.004

[8] Feng, Y. (2025). The impact of AI-generated synthetic speech on content quality and user engagement in science popularization short videos. *Journalism Lover*, (10), 92–96. https://doi.org/10.16017/j.cnki.xwahz.2025.10.018

[9] Yang, Q., & Zhang, Y. (2025). The boundaries of generative AI in news production: An examination of textual voice and hallucination risks in AI-generated news commentaries. *Journalism Review*, (9), 21–36. https://doi.org/10.16057/j.cnki.31-1171/g2.2025.09.007

[10] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.