

The Impact of AI Algorithm Optimization on Content Quality Enhancement in Short Video and Micro-Film Production

Shan Ge

Hangzhou Anheng Information Technology Co., Ltd. Zhejiang Hangzhou 310000

Abstract: *The rapid evolution of artificial intelligence (AI) is fundamentally reshaping the landscape of short-form video and micro-film production. This paper provides a systematic analysis of the pivotal role played by AI algorithm optimization in enhancing the overall quality of this dynamic media format. We specifically investigate its transformative impact across three core dimensions: content creation, post-production efficiency, and end-user engagement. By leveraging advanced deep learning architectures, including Generative Adversarial Networks (GANs) for visual synthesis, Convolutional Neural Networks (CNNs) for scene understanding and aesthetic enhancement, and Recurrent Neural Networks (RNNs) for narrative structure prediction, AI-driven tools are enabling unprecedented capabilities. These innovations facilitate not only a surge in video generation speed and a significant uplift in image resolution and stability but also the development of more coherent narrative logic and deeply immersive interactive experiences. However, the integration of AI also presents challenges, such as the risk of stylistic homogenization and emerging ethical concerns regarding deepfakes and data provenance. This study concludes that the strategic optimization of AI algorithms is a critical driver for quality advancement, pushing the boundaries of creative expression and setting new technical benchmarks for the industry. Future work should focus on developing more explainable and controllable AI systems that empower, rather than replace, human creativity.*

Keywords: Artificial Intelligence; Algorithm Optimization; Short Video Quality; Micro-film; Deep Learning; Generative Adversarial Networks; Content Creation.

1. INTRODUCTION

In recent years, with the continuous advancement of science and technology, artificial intelligence has experienced unprecedented "leapfrog" development. In the field of film and television design and production, AI has greatly promoted the development of short videos and micro-films. The impact of AI on short videos and micro-films is mainly concentrated in algorithms and their optimization, achieving significant improvements in both creative efficiency and artistic expression, while also bringing new challenges. Tu (2025) proposed ProtoMind, a modeling-driven approach for NAS and SIP message sequence analysis aimed at smart regression detection[1]. Addressing challenges in recommendation systems, Wang (2025) introduced a joint training method for propensity and prediction models using targeted learning, specifically for data missing not at random[2]. In healthcare, Wang, Y. (2025) developed a transformer-augmented survival analysis model for efficient adverse event forecasting in clinical trials[3]. Financial risk management is advanced by Wang, Z. et al. (2025), who conducted an empirical study on designing and optimizing an AI-enhanced intelligent financial risk control system for multinational supply chains[4]. Model optimization remains a key theme, with Wu et al. (2023) presenting Jump-GRS, a multi-phase structured pruning method for neural decoding[5]. For industrial applications, Xie and Liu (2025) designed InspectX, a system that optimizes industrial monitoring via OpenCV and WebSocket for real-time analysis[6]. In legal tech, Xie et al. (2024) advanced text classification by applying a Conv1D-based approach for multi-class classification of legal citation texts[7]. Generative models continue to find novel applications, as seen in Xu's (2025) work on CivicMorph, which uses generative modeling for public space form development[8]. Yang (2025) focused on intelligent consultation systems, introducing an identification method based on a Prompt-Biomrc model[9]. Zhang, Yuhan (2025) contributed to the ad tech space with AdOptimizer, a self-supervised framework for efficient ad delivery in low-resource markets, and also addressed LLM development tools with InfraMLForge for rapid development and scalable deployment[10,11]. Network analytics was tackled by Zhang, Yujun et al. (2025), who proposed MamNet, a novel hybrid model for time-series forecasting and frequency pattern analysis in network traffic[12]. In green finance, Zhang, Zongzhen et al. (2025) leveraged deep learning for carbon market price forecasting and risk evaluation under climate change[13]. For computer vision, Zheng et al. (2025) presented Diffmesh, a motion-aware diffusion framework for human mesh recovery from videos[14]. Agricultural technology was enhanced by Zhou (2025), who devised a swarm intelligence-based multi-UAV cooperative coverage and path planning system for precision pesticide spraying in

irregular farmlands[15]. System reliability is addressed by Zhu (2025) through REACTOR, a framework for reliability engineering with automated causal tracking and observability reasoning[16]. Finally, Zhuang (2025) explored the evolutionary logic and theoretical construction of real estate marketing strategies in the context of digital transformation[17].

2. THE MEANING OF ARTIFICIAL INTELLIGENCE AND ITS DEVELOPMENT HISTORY

To understand the technologies related to artificial intelligence, one must first clarify what artificial intelligence is. Its full English name is Artificial Intelligence, abbreviated as **AI**. Artificial intelligence refers to technologies and methods that use computer programs or machines to simulate and realize human intelligence, enabling computers to possess perception, understanding, and judgment capabilities. It is an interdisciplinary field that integrates computer science, psychology, philosophy, and other disciplines. In layman's terms, artificial intelligence is a "higher-order" application of computer technology.

Artificial intelligence technology originated in the 1950s, known as the "embryonic stage" of AI. During this period, research focused mainly on establishing theoretical foundations and symbolic methods. In 1950, Alan Turing proposed the "Turing Test," providing a standard for evaluating whether a machine possesses intelligence. In 1956, the Dartmouth Conference was held, where John McCarthy and others coined the term "artificial intelligence" and initiated systematic research.

From the 1960s to the 1970s, artificial intelligence entered its "golden age." During this period, AI research achieved fruitful results, especially in expert systems, pattern recognition, and machine learning.

From the 1980s to the 1990s, AI research entered a "low-tide period." Due to limited computing power, immature algorithms, and a lack of data, AI research did not achieve widespread application.

From 2000 to the present, AI research has entered a "renaissance and industrialization period." With the popularization of the internet and the development of big data, cloud computing, and the Internet of Things, AI research has been provided with abundant data and powerful computing capabilities. Breakthroughs in machine learning and deep learning have enabled significant progress in areas such as speech recognition and image recognition. AI has also been widely applied across various fields, achieving industrialization.

3. AI ALGORITHMS AND THEIR OPTIMIZATION

3.1 Definition of Algorithms

In artificial intelligence, algorithms are the core driving force and the soul of AI, propelling continuous development and advancement. So, what is an AI algorithm? It is a mathematical model used to simulate human intelligent behavior. These algorithms process and analyze large amounts of data to help AI systems make decisions, learn, and solve problems. Algorithms are the core of AI technology, enabling machines to simulate human thought processes and perform complex tasks.

3.2 Classification of Algorithms

Artificial intelligence encompasses many algorithms; the main categories and their representative algorithms are as follows:

(1) Machine-learning algorithms: an important branch of AI that enables computers to extract patterns and regularities from observed and learned data in order to predict future events. Machine-learning algorithms are further divided into classification, clustering, and regression algorithms.

(2) Deep-learning algorithms: a subfield of machine learning that mimics the structure of the human brain's neural networks, using multi-layer neural networks to process large volumes of complex data. Deep-learning algorithms include Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Generative Adversarial Networks (GAN).

(3) Logical-reasoning algorithms: a rule- and symbolic-logic-based approach commonly used for knowledge representation and solving reasoning problems. Predicate logic is a typical representation method; it expresses knowledge by defining predicates and rules and derives conclusions through logical inference.

3.3 Algorithm Optimization

It is important to note that algorithms are not 100 % accurate; they have a certain level of precision. The higher the precision, the better the information provided matches user needs; conversely, the information diverges significantly from user requirements.

First, AI algorithms rely on large amounts of data for training and learning; if the data itself is biased or insufficient, the algorithm's accuracy will suffer. In addition, algorithm design and optimization also affect accuracy, and different algorithms suit different tasks and scenarios.

Second, complexity of understanding and interpretation. It may fail to fully grasp the context or nuances of a problem, especially when handling issues involving multiple meanings, metaphors, or specific cultural backgrounds.

Third, impact of computational resources and environment. The operating environment of an algorithm also influences its accuracy—for example, network latency and device performance.

So how can algorithmic precision be improved? This requires algorithm optimization.

Common optimization methods are as follows:

Select appropriate data structures. Choosing the right data structure can markedly improve algorithmic execution efficiency.

Reduce algorithmic time and space complexity. By analyzing an algorithm's time and space complexity, employ more efficient algorithms or improve existing ones to cut execution time and memory usage.

Algorithm parallelization. Run algorithms in parallel using multiple threads or processes to boost execution efficiency.

Optimize compiler options and configuration. Enable compiler optimization flags, increase heap size, reduce context switching, etc., to enhance program performance and efficiency.

Code-level optimization. Improve execution efficiency by reducing lines of code, minimizing function calls, using efficient data types, and similar techniques.

4. THE IMPACT OF OPTIMIZATION ALGORITHMS ON SHORT VIDEOS AND MICRO-FILMS

With the rapid development and widespread adoption of artificial intelligence, it is now possible to rely on large AI models to create short videos and micro-films. Simply input the production requirements into the model, and it will automatically generate the desired short video or micro-film according to the user's specifications.

Currently, several domestic large models have made significant progress in short-video production. For example, Vidu by Shengshu Technology offers text-to-video and image-to-video capabilities; Zhipu AI released the video-generation model "Qingying"; SenseTime introduced the controllable character-video model Vimi; Alibaba DAMO Academy's "Xun Guang" platform and Kuaishou's "Keling AI" have also achieved breakthroughs in short-video creation. In addition, open-source projects like MoneyPrinterTurbo can automatically generate video scripts, footage, subtitles, and background music from a single topic or keyword, ultimately producing high-definition short videos and micro-films.

Abroad, numerous companies have developed large models for designing and producing short videos and micro-films. OpenAI's Sora, for instance, can generate high-quality short videos from text prompts, featuring multiple characters, complex scenes, and precise themes with photorealistic visuals. Stability AI, Luma AI,

Runway, and HaiperAI have also released their own video-generation models. Although these models face challenges such as cultural and contextual differences, their continuous improvement never stops. Meta's Movie Gen model can create vivid video and audio clips from user instructions, leveraging deep-learning techniques and multimodal design to transform textual descriptions into dynamic visuals while offering a rich selection of video styles.

Thus, both domestically and internationally, great importance is attached to developing AI large models for short-video and micro-film production. The reason is that these models can greatly enhance production efficiency, reduce costs, and even spark unexpected creative "inspiration."

However, it should be noted that not all AI-generated short videos or micro-films meet user expectations; often, the output deviates from the user's needs, sometimes significantly. I believe this is mainly due to two factors. First, when users provide input conditions to the model, the instructions may be inaccurate, incomplete, or too brief, leading to substantial deviation. If the input is comprehensive and precise yet the generated content still diverges markedly from the user's requirements, the issue likely lies with the model itself.

This requires first understanding the definition of a large model: an AI large model is an artificial-intelligence model with an enormous number of parameters, trained through deep-learning algorithms and artificial neural networks. These models take vast multimedia data resources as input and complete large-scale training via complex mathematical operations and optimization algorithms. From this definition we can see that an AI large model is jointly constituted by algorithms and data; the algorithm is the model's "brain," and through learning and training it can extract useful information and patterns from massive datasets. Moreover, AI large models are built on deep-learning algorithms and consist of multiple neural networks, each containing many neurons. Thus, if the algorithm is sub-optimal, it will directly affect the accuracy of the information (including text, audio, video, and images) generated by the large model.

So how can we optimize the algorithms in AI large models?

From the definition of a large model we know that, both domestically and internationally, current large-model products from related companies all adopt deep-learning algorithms (primarily neural-network algorithms). Yet the accuracy and effectiveness of each company's product differ—why is that? The reason is that each company's large-model product may not use a single algorithm but rather a combination of multiple algorithms, and the algorithm-combination techniques of each company may also differ. This is the fundamental reason why the large models developed by various companies for short videos and micro-films differ in accuracy, quality, and effect.

At present, no company—domestic or foreign—has fully open-sourced its large-model code, especially the algorithm implementation and optimization parts, so outsiders cannot gain a detailed understanding of each company's technical implementation system and can only speculate on the general technical direction. Therefore, when optimizing algorithms for short-video and micro-film generation, we must rely on two factors: first, each company's own optimization of the algorithms within its large model and its updates to algorithm-combination techniques; second, optimization of general-purpose algorithms—for example, we should currently focus on optimizing the three algorithms of convolutional neural networks (CNN), recurrent neural networks (RNN), and generative adversarial networks (GAN). Optimization methods can include modifying the underlying mathematical models, reducing time and space complexity, and so on. Finally, we can also consider constructing entirely new mathematical models whose execution efficiency is higher and whose time and space complexity are lower, thereby achieving the goal of algorithm optimization.

5. SPECIFIC APPLICATION SCENARIOS OF AI TECHNOLOGY

5.1 Intelligent Generation and Assisted Creation

Scripts and storyboards: NLP algorithms (e.g., GPT-4, Claude) can generate script outlines or storyboard suggestions based on keywords, shortening pre-production planning time; AIGC tools (e.g., Runway ML) can automatically generate storyboard images from textual descriptions.

Auto-editing: Computer-vision-based algorithms (e.g., Adobe Sensei) identify highlight moments in footage—such as peak expressions or actions—for one-click rough cuts; AI can also auto-match shot transitions to music beats (e.g., CapCut's smart beat-sync feature).

Visual enhancement: Super-resolution algorithms (e.g., ESRGAN) upscale low-resolution clips; style-transfer techniques (e.g., Stable Diffusion) quickly apply cinematic, cyberpunk, or other artistic looks.

5.2 Personalization and Interactivity

Dynamic adaptation: Algorithms automatically adjust video versions (e.g., Douyin's A/B testing for multiple thumbnails) based on user preferences such as watch time and engagement data.

Interactive storytelling: In micro-films, AI can analyze real-time audience emotional feedback (via camera or click behavior) to dynamically shift story branches, akin to Netflix's "Black Mirror: Bandersnatch."

6. EFFICIENCY AND ARTISTIC EXPRESSION GAINS

6.1 Leap in Efficiency

Shorter production cycles: Post-production for micro-films that once took weeks can be compressed to days through AI matting (e.g., RemBG) and automatic color grading (DaVinci Resolve AI).

Cost reduction: AI virtual actors (e.g., Synthesia) replace some live-action shoots, cutting location and talent expenses; AI voice-over (e.g., ElevenLabs) supports multilingual generation.

6.2 Expanding Artistic Boundaries

Breaking physical limits: AI-generated scenes (e.g., MidJourney + Unreal Engine) build fantastical worlds with less reliance on live filming.

Data-driven aesthetics: Algorithms analyze audiovisual patterns of viral content—such as shot duration and color saturation—to help creators refine artistic expression (e.g., Bilibili creators using ChatGPT to optimize titles and tags).

7. CONCLUSION

In summary, optimizing AI algorithms will directly improve the precision of large models in generating short-form videos and micro-films, bringing significant benefits to users. Continuous refinement of AI algorithms is reshaping the production paradigm of short-form and micro-film content, propelling the industry toward greater efficiency, intelligence, and personalization. Future breakthroughs in cross-modal learning and real-time rendering will unlock even more creative possibilities for the content industry.

REFERENCES

- [1] Tu, Tongwei. "ProtoMind: Modeling Driven NAS and SIP Message Sequence Modeling for Smart Regression Detection." (2025).
- [2] Wang, Hao. "Joint Training of Propensity Model and Prediction Model via Targeted Learning for Recommendation on Data Missing Not at Random." AAAI 2025 Workshop on Artificial Intelligence with Causal Techniques. 2025.
- [3] Wang, Y. (2025). Efficient Adverse Event Forecasting in Clinical Trials via Transformer-Augmented Survival Analysis.
- [4] Wang, Z., Chew, J. J., Wei, X., Hu, K., Yi, S., & Yi, S. (2025). An Empirical Study on the Design and Optimization of an AI-Enhanced Intelligent Financial Risk Control System in the Context of Multinational Supply Chains. *Journal of Theory and Practice in Economics and Management*, 2(2), 49–62. Retrieved from <https://woodyinternational.com/index.php/jtpem/article/view/208>
- [5] Wang, Zhiyuan, et al. "An Empirical Study on the Design and Optimization of an AI-Enhanced Intelligent Financial Risk Control System in the Context of Multinational Supply Chains." (2025).
- [6] Wu, Xiaomin, et al. "Jump-GRS: a multi-phase approach to structured pruning of neural networks for neural decoding." *Journal of neural engineering* 20.4 (2023): 046020.

- [7] Xie, Minhui, and Boyan Liu. "InspectX: Optimizing Industrial Monitoring Systems via OpenCV and WebSocket for Real-Time Analysis." (2025).
- [8] Xie, Y., Li, Z., Yin, Y., Wei, Z., Xu, G., & Luo, Y. (2024). Advancing Legal Citation Text Classification A Conv1D-Based Approach for Multi-Class Classification. *Journal of Theory and Practice of Engineering Science*, 4(02), 15–22. [https://doi.org/10.53469/jtpes.2024.04\(02\).03](https://doi.org/10.53469/jtpes.2024.04(02).03)
- [9] Xu, Haoran. "CivicMorph: Generative Modeling for Public Space Form Development." (2025).
- [10] Yang, J. (2025, July). Identification Based on Prompt-Biomrc Model and Its Application in Intelligent Consultation. In *Innovative Computing 2025, Volume 1: International Conference on Innovative Computing* (Vol. 1440, p. 149). Springer Nature.
- [11] Zhang, Yuhan. "AdOptimizer: A Self-Supervised Framework for Efficient Ad Delivery in Low-Resource Markets." (2025).
- [12] Zhang, Yuhan. "InfraMLForge: Developer Tooling for Rapid LLM Development and Scalable Deployment." (2025).
- [13] Zhang, Yujun, et al. "MamNet: A Novel Hybrid Model for Time-Series Forecasting and Frequency Pattern Analysis in Network Traffic." *arXiv preprint arXiv:2507.00304* (2025).
- [14] Zhang, Zongzhen, Qianwei Li, and Runlong Li. "Leveraging Deep Learning for Carbon Market Price Forecasting and Risk Evaluation in Green Finance Under Climate Change." *Journal of Organizational and End User Computing (JOEUC)* 37.1 (2025): 1-27.
- [15] Zheng, Ce, et al. "Diffmesh: A motion-aware diffusion framework for human mesh recovery from videos." 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE, 2025.
- [16] Zhou, Dianyi. "Swarm Intelligence-Based Multi-UAV CooperativeCoverage and Path Planning for Precision PesticideSpraying in Irregular Farmlands." (2025).
- [17] Zhu, Bingxin. "REACTOR: Reliability Engineering with Automated Causal Tracking and Observability Reasoning." (2025).
- [18] Zhuang, R. (2025). Evolutionary Logic and Theoretical Construction of Real Estate Marketing Strategies under Digital Transformation. *Economics and Management Innovation*, 2(2), 117-124.