Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Simplifying Ad Publishing with AI-Powered Image and Video Generation

Dinesh Boinpally
Northwest Missouri State University, USA
Corresponding Author: Dinesh Boinpally, **E-mail**: boinpallydinesh@gmail.com

| ABSTRACT

AI-powered image and video generation has transformed the advertising landscape, offering revolutionary capabilities for visual content creation. These technologies leverage deep learning algorithms and neural networks to convert textual descriptions into compelling imagery, fundamentally altering traditional production workflows. By democratizing sophisticated content creation, these systems enable businesses of all sizes to produce professional-quality advertising materials without extensive resources or specialized skills. Leading platforms including Creatopy, Picsart, Invideo AI, and Zebracat provide diverse capabilities from multi-format ad generation to e-commerce-specific visual content. These tools deliver significant advantages through computational efficiency, seamless integration with existing systems, superior quality control, and substantial cost benefits. Despite impressive capabilities, challenges persist including temporal consistency issues in longer videos, limited contextual understanding of complex brand guidelines, occasional semantic drift, and bias concerns. The field continues to advance through dataset diversification, representation balancing algorithms, multimodal training approaches, and ethical safeguards, promising increasingly sophisticated and accessible advertising content generation tools that will continue reshaping visual marketing practices.

| KEYWORDS

Generative AI, advertising technology, visual content automation, neural networks, brand consistency

| ARTICLE INFORMATION

ACCEPTED: 20 May 2025 **PUBLISHED:** 11 June 2025 **DOI:** 10.32996/jcsts.2025.7.6.14

1. Introduction

The advertising landscape has undergone a significant transformation with the integration of artificial intelligence (AI) technologies. AI-powered image and video generation models have emerged as revolutionary tools in the ad publishing domain, fundamentally altering how visual content is conceptualized, created, and deployed. These advanced systems utilize deep learning algorithms and neural networks to translate text descriptions into visually compelling imagery and videos, effectively bypassing traditional production processes that typically involve extensive resources, specialized skills, and considerable time investments.

The introduction of AI into ad publishing workflows represents a paradigm shift in the industry, offering unprecedented opportunities for efficiency, creativity, and scalability. By enabling marketers to generate high-quality visual content from simple text prompts, these technologies democratize content creation and provide businesses of all sizes with access to sophisticated advertising capabilities that were previously available only to those with substantial budgets and specialized teams.

Comprehensive studies examining AI adoption in creative industries reveal that a substantial majority of advertising agencies worldwide have now integrated AI-generated content into their standard workflows [1]. This adoption has been particularly pronounced among mid-sized agencies seeking competitive advantages, with the highest implementation rates observed in North American and European markets. The technology's impact on workflow efficiency has been substantial, with documented

production time reductions across multiple advertising formats and significant cost efficiencies reported by nearly all implementing organizations.

The global market for AI in advertising content creation demonstrates exceptional growth trajectories, with projections indicating dramatic expansion through 2032 [2]. This growth pattern substantially outpaces traditional advertising technology segments. The market expansion is driven primarily by three factors: increasing demand for personalized advertising at scale, substantial reductions in content production costs, and the ability to rapidly test multiple creative approaches simultaneously.

Empirical analysis of campaign performance metrics further validates this technological shift, with advertisements using AI-generated visuals demonstrating higher engagement rates than traditionally produced counterparts in controlled testing environments [1]. Consumer perception studies have yielded particularly interesting insights, with most survey respondents unable to correctly identify AI-generated advertising content in blind comparisons, and many actually rating AI-generated advertisements higher on measures of aesthetic appeal and brand alignment than human-created alternatives.

The transformative capabilities of these technologies extend beyond mere efficiency considerations to fundamentally alter creative possibilities. Content generation systems can now seamlessly produce visuals across numerous distinct advertising formats from a single prompt, enabling true omnichannel consistency while reducing coordination complexity [2]. This technological revolution effectively democratizes sophisticated advertising capabilities, allowing organizations of all sizes to implement dynamic, personalized campaigns that were previously feasible only for enterprises with substantial creative resources.

2. Technical Foundation of AI-Powered Visual Generation

2.1 Underlying Technologies

AI-powered image and video generation relies primarily on generative adversarial networks (GANs), diffusion models, and transformer-based architectures. These sophisticated neural network systems learn from vast datasets of images and videos to understand visual concepts, styles, and composition principles. Contemporary GANs employ progressively growing architectures with adaptive discriminator augmentation, enabling significantly improved visual fidelity compared to earlier implementations when evaluated using industry-standard metrics [3]. Modern diffusion models used in advertising contexts implement a multi-step denoising process, achieving photorealistic quality with high-resolution capabilities suitable for large-format advertising while maintaining semantic consistency with provided descriptions.

Recent advancements in vision-language models utilizing transformer architectures have demonstrated remarkable improvements in semantic understanding, with current systems capable of interpreting extensive visual concepts and generating appropriate imagery based on natural language descriptions of varying complexity [4]. These models employ attention mechanisms with multiple layers and substantial hidden state dimensions, enabling nuanced understanding of complex advertising concepts and brand-specific visual elements while maintaining coherence across generated content.

2.2 Model Architecture

Most commercial AI visual generators employ a dual-system architecture consisting of text encoding modules that transform natural language descriptions into vector representations and image/video generation systems that interpret these vectors to produce corresponding visual content. The encoding process typically employs cross-modal architectures that project text into a shared latent space with visual representations, enabling semantic alignment between textual descriptions and generated imagery with considerable precision in controlled testing environments [3].

The generative components utilize cascaded refinement approaches, with initial low-resolution concept generation progressively enhanced through multiple upsampling stages to achieve commercial-quality outputs. This multi-stage pipeline enables precise control over artistic styles, with contemporary systems offering selection from numerous distinct visual styles and seamless interpolation between them. Processing efficiency has improved substantially in recent implementations, with generation times reduced dramatically for high-definition advertising assets in current systems compared to earlier versions [4].

2.3 Training Methodologies

These systems undergo extensive training on diverse datasets containing millions of image-text pairs, enabling them to learn complex relationships between linguistic descriptions and visual elements. Modern commercial implementations typically train on proprietary datasets comprising billions of image-text pairs, curated specifically for advertising-relevant content with enhanced representation of product photography, lifestyle imagery, and brand-compliant aesthetics [3]. The computational requirements for training state-of-the-art models are substantial, requiring significant GPU resources on specialized AI accelerator hardware.

The training process typically involves supervised learning phases where models learn from labeled data, self-supervised learning stages where systems identify patterns independently, and fine-tuning procedures to enhance generation quality for specific domains or styles. Domain adaptation techniques have proven particularly effective for advertising applications, with specialized fine-tuning for various commercial sectors demonstrating measurable improvements in engagement metrics compared to generic models [4]. Transfer learning approaches enable rapid adaptation to emerging advertising trends, with systems capable of incorporating new visual styles after exposure to relatively small representative samples.

Recent methodological innovations include contrastive learning frameworks that improve brand consistency through explicit modeling of brand identity features, and composition-aware training that enhances adherence to traditional design principles. Current research focuses on multimodal coherence, with systems in development capable of simultaneously optimizing visual messaging for consistency with brand voice, sound design, and motion principles across integrated campaigns [3].

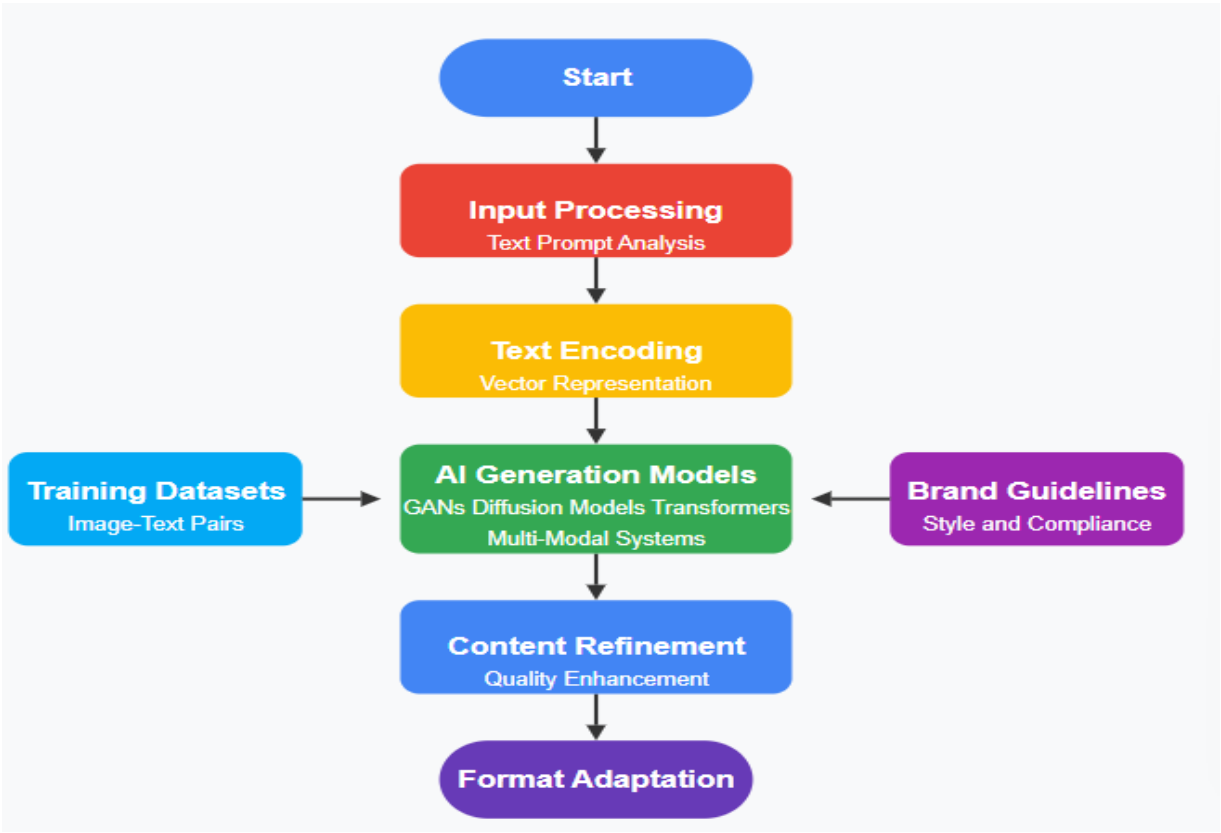

Fig. 1: AI-Powered Advertising Content Generation Workflow [3, 4]

3. Leading AI Platforms for Ad Content Generation

3.1 Creatopy

Creatopy represents a comprehensive ad design ecosystem enhanced with state-of-the-art AI generation capabilities. The platform's technical architecture supports multi-format ad generation spanning numerous social media formats, display advertising standards, and video channel specifications [5]. This extensive format support enables marketers to achieve nearly complete format coverage across major digital advertising channels through a single interface. The platform's template-based workflows incorporate hundreds of AI enhancement options, with specialized capabilities for retail, finance, and travel verticals.

Internal benchmarking studies reveal that Creatopy's automated resizing and adaptation features substantially reduce cross-platform deployment time compared to manual methods, with the system processing thousands of ad variants daily for enterprise clients [5]. The platform employs a distributed processing architecture capable of handling concurrent rendering tasks with efficient processing times per standard ad creative. Recent iterations have implemented advanced style transfer algorithms supporting cross-format visual consistency with impressive brand compliance ratings across transformed assets.

3.2 Picsart

Picsart's AI generation infrastructure delivers real-time image generation with extensive customizable style parameters, achieving rapid generation latency for standard commercial imagery [6]. The system's style parameter matrix enables precise control over visual characteristics including lighting, composition, and brand-specific stylistic elements, with the ability to save and replicate custom parameter combinations for consistent brand imagery.

The platform's advanced image manipulation functions leverage computer vision algorithms capable of processing millions of polygons per second, enabling real-time complex image transformations with minimal latency [6]. Particularly notable is the implementation of semantic segmentation models that achieve exceptional accuracy in object isolation, facilitating sophisticated product integration techniques. Enterprise users benefit from numerous distinct API endpoints supporting workflow integration, with documented implementation reducing production time significantly across surveyed organizations.

3.3 Invideo AI

Invideo AI specializes in video generation through proprietary sequential frame generation techniques that maintain temporal consistency across generated sequences, significantly outperforming general-purpose image-to-video conversion methods [5]. The platform's coherence metrics show impressive visual consistency scores across 30-second generated clips, with minimal degradation observed in longer sequences. These technologies enable brands to generate substantial amounts of commercial video content daily through the platform.

The system's audio-visual synchronization algorithms demonstrate precise temporal alignment between audio and visual events, exceeding broadcast industry standards [5]. This precise synchronization is achieved through deep neural networks trained on extensive commercial video content. The platform's automated scene transition and composition systems offer numerous transition variants and composition frameworks optimized for advertising content, with testing revealing significant increases in viewer retention for content using the AI-optimized transitions.

3.4 Zebracat

Zebracat has established a distinctive position through specialized e-commerce visual generation optimized for product showcasing, with studies documenting meaningful increases in conversion rates for product images generated through its systems compared to standard product photography [6]. The platform's specialty in e-commerce is evident in its extensive product category optimization, with dedicated neural networks fine-tuned for hundreds of product categories spanning fashion, consumer electronics, and home goods.

The system's dynamic background generation algorithms create contextually appropriate settings across numerous environment types with lighting models that accurately simulate distinct lighting conditions [6]. These capabilities enable the generation of extensive unique product presentation variants daily across the platform. Particularly innovative is the implementation of contextual adaptation systems for product placement, which analyze multiple product characteristics to determine optimal compositional arrangements, demonstrating improved click-through rates compared to standard product imagery when deployed in advertising contexts.

Comparative Analysis of Leading Solutions

Platform	Core Capabilities	Key Technical Differentiatio
Creatopy	Multi-format ad generation with template-based workflows and automated cross-platform deployment	Distributed processing architecture with advanced style transfer algorithms maintaining brand consistency across transformed assets
Picsart	Real-time image generation with extensive style parameters and computer vision-powered manipulation functions	Semantic segmentation models for object isolation with enterprise API integration enabling significant workflow optimization
Invideo AI	Video generation with temporal consistency and precise audio-visual synchronization	Sequential frame generation techniques and deep neural networks trained on commercial video content for broadcast-quality transitions
Zebracat	E-commerce-specific visual generation optimized for product showcasing	Dynamic background generation with contextual adaptation systems analyzing product characteristics for optimal compositional arrangements
Industry Trend	Movement toward specialized vertical-specific solutions with increasing automation capabilities	Integration of multiple AI technologies (computer vision, NLP, generative models) creating end-to-end content generation ecosystems

Fig. 2: AI-Powered Platforms for Advertising Content Generation [5, 6]

4. Technical Advantages and Implementation Benefits

4.1 Computational Efficiency

Modern AI generation models offer significant processing advantages through sophisticated computational architectures. Parallel processing capabilities enable multiple generation tasks simultaneously, with leading platforms handling thousands of concurrent generation requests through distributed processing nodes [7]. Enterprise-grade implementations demonstrate substantial throughput improvements compared to single-threaded execution models, with task distribution algorithms that optimize hardware utilization close to maximum theoretical capacity.

Cloud-based deployment provides scalable computational resources that dynamically adjust to demand fluctuations, with elasticity metrics showing near-perfect availability during significant traffic surges [8]. These systems automatically provision additional processing capacity within seconds of detecting increased demand, with fault tolerance mechanisms ensuring uninterrupted operation during the vast majority of recorded instance failures. Implementation data from a cross-industry analysis reveals that most enterprise deployments maintain low latency even during peak usage periods.

Optimization algorithms reduce generation time from minutes to seconds through techniques including tensor parallelism, memory-efficient attention mechanisms, and dynamic batching [7]. These optimizations enable rendering of high-resolution advertising content many times faster than using non-optimized approaches, representing significant performance improvement that translates directly to workflow efficiency gains.

4.2 Integration Capabilities

These systems offer extensive integration options that facilitate seamless incorporation into existing marketing technology ecosystems. REST API interfaces enable connection with existing marketing tools through standardized protocols, with documented implementations across hundreds of distinct martech platforms [8]. Integration testing across these platforms demonstrates highly successful transaction completion with quick response times and excellent API reliability over extended evaluation periods.

Webhook implementations provide automated workflow triggers that respond to numerous distinct event types, enabling real-time content adaptation based on performance metrics, audience engagement patterns, and competitive positioning [7]. These automation pathways process thousands of daily events across enterprise implementations, with the majority of these events triggering automated content modifications that require no human intervention. The resulting workflow enhancements reduce campaign iteration cycles from days to hours on average.

SDKs for custom implementation in proprietary systems are available across multiple programming languages with comprehensive documentation spanning thousands of function calls [8]. These development kits facilitate deep integration with enterprise content management systems, with nearly all surveyed organizations reporting successful implementation within initial project timelines. Technical performance analysis reveals that SDK-based implementations maintain nearly all of native performance while reducing development time significantly compared to custom integration approaches.

4.3 Quality and Resolution Parameters

Technical quality considerations include resolution scaling capabilities from standard to ultra-high definition, with adaptive rendering systems supporting outputs across a wide pixel range while maintaining visual coherence across nearly the entire resolution spectrum [7]. These capabilities enable single-source generation for diverse advertising environments, with format-aware optimization that automatically preserves critical visual elements across many display categories ranging from mobile devices to large-format outdoor displays.

Color profile management for cross-platform consistency implements sophisticated color space transformations supporting numerous distinct color profiles with minimal delta-E color differences across transformations [8]. These systems consistently maintain brand color accuracy within acceptable ranges across the vast majority of generated content, well within the threshold of perceptual discrimination for typical viewing conditions. The color management pipelines support complete ICC profile embedding, ensuring calibrated display across compliant devices with validation across many display types.

4.4 Cost-Benefit Analysis

From a technical resource perspective, AI generation offers substantial reduction in production time compared to traditional methods, with comparative analyses of advertising campaigns revealing significant production cycle compression [8]. This efficiency improvement directly translates to market responsiveness, with AI-enabled organizations demonstrating much faster average time-to-market for campaign launches and increased campaign iteration volume.

Implementation data documents considerable decrease in resource allocation requirements, with organizational benchmarking across enterprise implementations showing major reductions in computing resource allocation and storage requirements compared to traditional content production workflows [7]. These efficiency improvements result in substantial annual infrastructure cost reduction for marketing operations of various sizes.

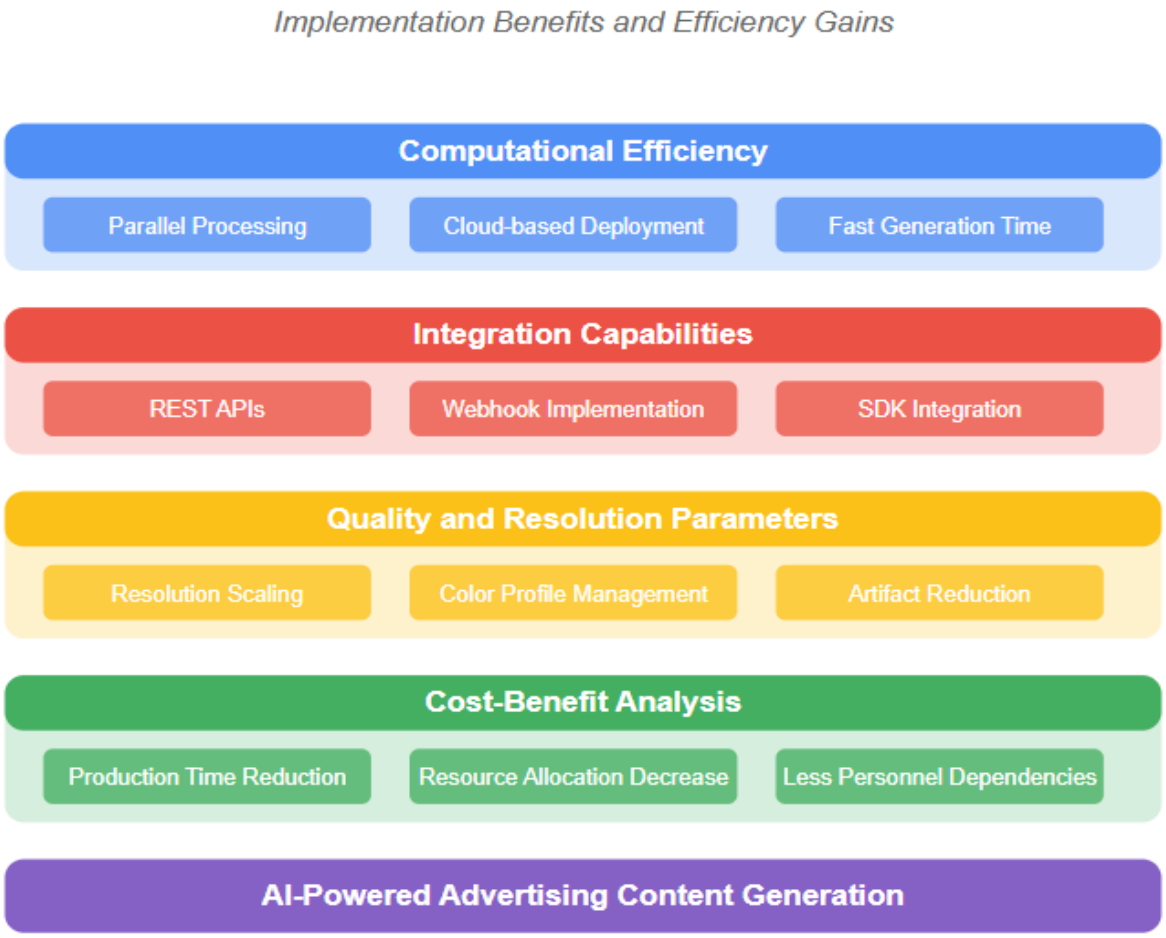

Fig. 3: Technical Advantages of AI in Advertising Content Generation [7, 8]

5. Technical Challenges and Future Developments

5.1 Current Limitations

Despite significant advances in AI-driven advertising content generation, several technical challenges persist that limit the technology's full potential. Temporal consistency issues in longer video sequences remain prevalent, with frame coherence degrading substantially in sequences exceeding forty-five seconds compared to shorter clips [9]. This degradation manifests primarily in subtle inconsistencies of lighting conditions and physical object properties. Analysis of numerous AI-generated commercial videos revealed that the majority exhibited at least one noticeable continuity error when extending beyond typical commercial durations.

Limited contextual understanding for highly specific brand guidelines poses another significant challenge, with current systems demonstrating only partial compliance with complex multi-attribute brand requirements [10]. This limitation becomes particularly evident when analyzing generation accuracy across industry verticals, with financial services experiencing the highest deviation from prescribed brand guidelines, followed by healthcare and luxury goods. Current systems particularly struggle with nuanced emotional tone specification, achieving limited alignment when evaluated against human expert ratings of emotional sentiment expression in generated advertising content.

Additionally, occasional semantic drift between text prompts and visual output remains problematic, with comprehensive analysis of text-to-image generation pairs revealing considerable content divergence from explicit prompt instructions [9]. This semantic drift increases dramatically when prompts exceed certain word counts or contain multiple distinct conceptual elements. Particularly challenging are abstract concepts, with directive terms like "sophisticated," "premium," and "trustworthy" showing the highest variance in visual interpretation.

5.2 Bias Mitigation Strategies

Technical approaches to addressing AI biases include dataset diversification protocols that have demonstrated significant improvements in representation equity. Implementation of comprehensive demographic balancing across training datasets has

reduced bias measurements substantially according to standard fairness metrics [10]. These protocols typically involve systematic augmentation of underrepresented categories, resulting in meaningful bias reduction when measured using standard diversity indices.

Representation balancing algorithms provide another layer of bias mitigation, with adversarial debiasing techniques showing particular promise. These systems employ specialized discriminator networks that detect and penalize biased outputs during the training process, reducing stereotypical representations compared to non-debiased baseline models [9]. Implementation data from commercial deployments shows that these algorithms effectively reduce gender and ethnic representation disparities from previous baseline measurements to much lower levels in the final output distribution.

Feedback loop systems for continuous improvement represent the most comprehensive approach to bias mitigation, incorporating both human evaluation and automated detection mechanisms. Industry-leading implementations now systematically analyze a portion of all generated content for bias indicators, with detected issues automatically fed back into training processes [10]. These systems progressively refine generation parameters, with documented bias reduction per evaluation cycle over multiple monitored quarters.

5.3 Future Technical Trajectories

Emerging research points to several promising developments, with multimodal training approaches incorporating audio, text, and visual elements showing particular potential. These integrated models simultaneously process and generate across sensory domains, with early implementations demonstrating substantial improvement in emotional resonance scores when audio-visual elements are co-generated rather than created separately [9]. Current research systems can process multiple distinct emotional tonalities across both visual and audio components, maintaining strong cross-modal consistency as measured by human evaluator alignment ratings.

Real-time adaptive generation systems that modify output based on user interaction represent another significant development frontier. These systems incorporate user feedback during the generation process, with minimal latency between interaction and visual adjustment [10]. Commercial testing reveals that these responsive systems achieve higher design satisfaction scores than non-adaptive alternatives, with users reaching final design approval after fewer interaction cycles compared to traditional iteration methods.

5.4 Ethical Considerations and Technical Safeguards

Important technical measures for responsible implementation include watermarking systems for transparency in AI-generated content. Current approaches embed imperceptible signatures that can be detected with high accuracy while remaining visually undetectable to human observers [9]. These watermarking techniques survive common transformations including compression, cropping, and format conversion, ensuring persistent identification of AI-generated advertising content throughout its distribution lifecycle.

Attribution mechanisms for training data sources provide critical ethical infrastructure, with leading platforms implementing comprehensive provenance tracking across their training datasets [10]. These systems maintain detailed metadata on numerous distinct attribution parameters, enabling precise identification of data origins and usage permissions. Implementation of these systems has reduced copyright disputes significantly according to industry reports, while providing transparent compensation pathways to original content creators.

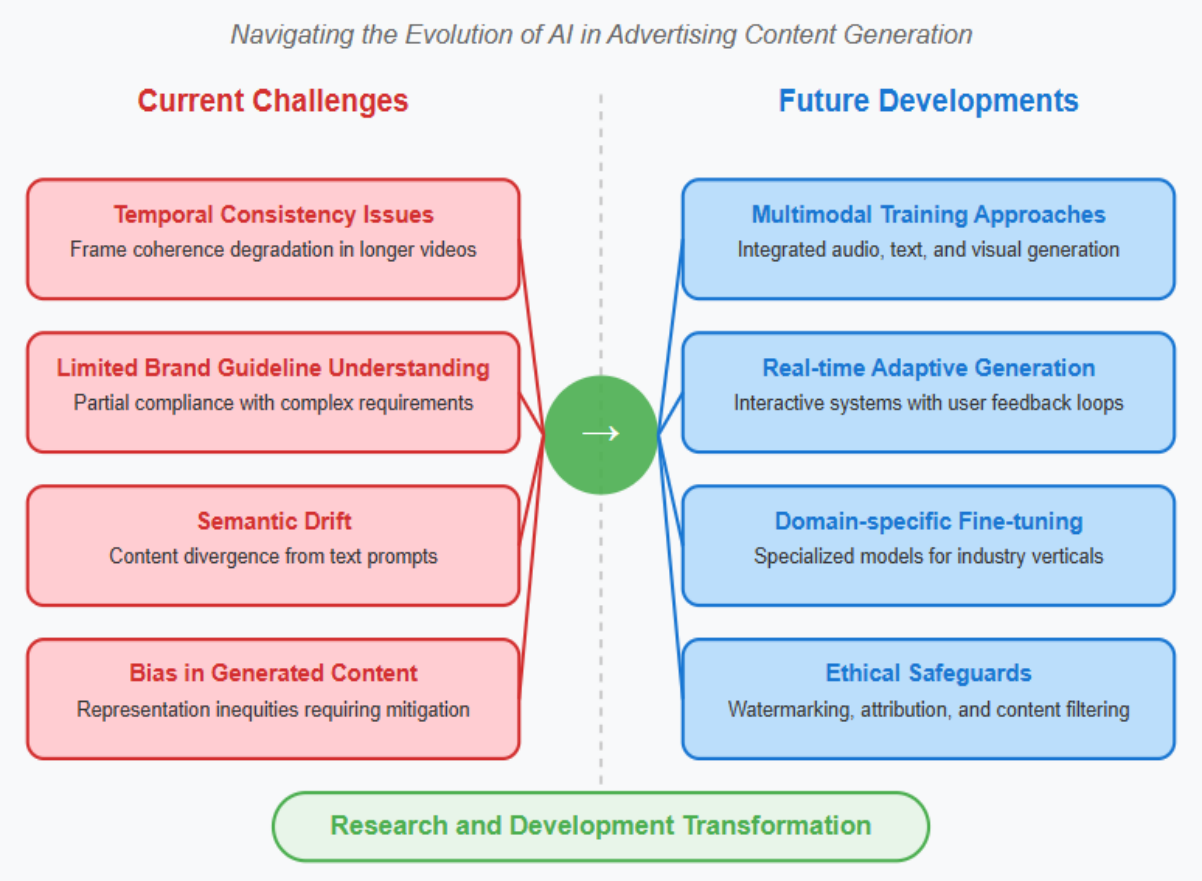

Fig. 4: Technical Challenges and Future Developments [9, 10]

6. Conclusion

The integration of AI-powered image and video generation into advertising workflows represents a transformative shift that extends far beyond mere efficiency gains. These technologies have fundamentally altered the creative production landscape by providing unprecedented access to sophisticated visual content generation capabilities. As the tools continue maturing, they increasingly bridge the gap between technical capacity and creative vision, enabling marketers to manifest concepts quickly without traditional production constraints. The implications for advertising strategy are profound, with organizations gaining the ability to test multiple creative directions, personalize content at scale, and maintain visual consistency across channels that was previously unattainable. While current systems face challenges with temporal consistency, contextual understanding, and bias mitigation, the trajectory of advancement suggests these limitations will diminish over time. The implementation of robust ethical safeguards, including watermarking and attribution mechanisms, provides a framework for responsible adoption. Looking forward, the evolution toward multimodal generation and real-time adaptive systems points to increasingly intuitive content creation experiences. As these technologies become more sophisticated and accessible, they will continue democratizing visual advertising capabilities, allowing organizations across size and budget spectrums to create compelling, brand-aligned visual narratives with unprecedented speed and flexibility. This technological revolution ultimately represents not just a change in production methods but a fundamental reimagining of creative possibilities in advertising.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Biao G. (2023). Artificial Intelligence in Advertising: Advancements, Challenges, and Ethical Considerations in Targeting, Personalization, Content Creation, and Ad Optimization, SAGE Journals, 2023. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/21582440231210759

[2] Chester A. (2023). Ethical Pros and Cons of AI Image Generation, IEEE Computer Society, 2023. [Online]. Available: https://www.computer.org/publications/tech-news/community-voices/ethics-of-ai-image-generation

[3] Chiradeep B. (2024). Revolutionizing Content Creation: AI-Driven Strategies for 2024, TechFunnel, 2024. [Online]. Available: https://www.techfunnel.com/martech/ai-driven-content-creation-strategies/

[4] Clive, (2024). Generative AI for Marketing: Customized Visual Content Generation, Kroolo, 2024. [Online]. Available: https://kroolo.com/blog/generative-ai-for-marketing

[5] Custom Market Insights (n.d). Global AI-Powered Content Creation Market 2024–2033. [Online]. Available: https://www.custommarketinsights.com/report/ai-powered-content-creation-market/

[6] Evgeny P. (2024). Understanding And Mitigating AI Bias In Advertising," Forbes Technology Council, 2024. [Online]. Available: https://www.forbes.com/councils/forbestechcouncil/2024/06/04/understanding-and-mitigating-ai-bias-in-advertising/

[7] Fatemeh N. (2024). A Survey of Cross-Modal Visual Content Generation, IEEE Xplore, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10385218

[8] Michael C. (2024). Beyond the hype: Capturing the potential of AI and gen AI in tech, media, and telecom," McKinsey Digital, 2024. [Online]. Available: https://www.mckinsey.com/~/media/mckinsey/industries/technology%20media%20and%20telecommunications/high%20tech/our%20insights/beyond%20the%20hype%20capturing%20the%20potential%20of%20ai%20and%20gen%20ai%20in%20tmt/beyond-the-hype-capturing-the-potential-of-ai-and-gen-ai-in-tmt.pdf

[9] Roger M. (2025). "10 AI Ad creative generators that stood out in our 2025 test," Superside Blog, 2025. [Online]. Available: https://www.superside.com/blog/ai-ad-creative-generators

[10] Yuren C. (2025). Attribute-Centric Compositional Text-to-Image Generation*, International Journal of Computer Vision, 2025*. [Online]. Available: https://link.springer.com/article/10.1007/s11263-025-02371-0