iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78444

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

AdSpark: A Multimodal Generative AI Framework for Automated Digital Advertising and Marketing Asset Synthesis

Saurabh Keni
Student, SIES College of Arts, Science and Commerce (Autonomous), Nerul, India

Abstract: ***The rapid evolution of generative artificial intelligence has fundamentally transformed the landscape of digital marketing and content creation. This paper presents the design and implementation of AdSpark, a multimodal generative platform designed to automate the production of professional advertising creatives and comprehensive marketing strategies. By integrating cutting-edge large language models (LLMs) and vision-language models, specifically Google Gemini, AdSpark enables users to generate high-fidelity advertising posters and synchronized marketing copy from minimal product inputs. The system leverages a robust full-stack architecture comprising Next.js for the frontend, Express.js for the backend, and Prisma with PostgreSQL for persistent state management. Key features include automated image synthesis via multimodal prompts, AI-driven tagline generation, social media captioning, and target audience profiling. Experimental evaluation demonstrates that AdSpark reduces creative production time by over 90% compared to traditional manual workflows. This research provides a scalable architectural blueprint for democratizing high-end advertising production through specialized AI orchestration.***

Keywords: Generative AI, Multimodal LLMs, Advertising Technology, Marketing Automation, Computer Vision, Gemini, Content Synthesis, Full-Stack Development.

I. INTRODUCTION

In the contemporary digital economy, the demand for high-quality visual and textual marketing content has reached unprecedented levels. Businesses, particularly small-to-medium enterprises (SMEs), face substantial barriers in producing professional-grade advertising assets due to the high costs associated with graphic design, copywriting, and market research. The emergence of Generative AI offers a paradigm shift, enabling the automated synthesis of complex creative artifacts through natural language interfaces.

Traditional marketing automation tools have primarily focused on scheduling, analytics, and distribution. However, the integration of Multimodal Large Language Models (MLLMs) allows for the creation of systems that can understand product aesthetics and translate them into persuasive visual and textual narratives. AdSpark addresses this opportunity by providing a unified, end-to-end interface that automates the entire creative pipeline—from product photography enhancement to strategic market positioning.

This research explores the technical challenges of orchestrating disparate AI services into a cohesive user experience. It details how AdSpark utilizes Gemini's multimodal capabilities to perform grounded image generation and context-aware copywriting, ensuring that generated assets are not only visually appealing but also strategically aligned with specific marketing goals and platform-specific requirements (e.g., Instagram, Twitter, Facebook).

II. LITERATURE REVIEW

A. Evolution of Generative AI in Marketing

The transition from rule-based systems to transformer-based architectures has enabled AI to move beyond simple data analysis to creative generation. The introduction of the Transformer architecture by Vaswani et al. [1] laid the foundation for models like GPT and Gemini, which have demonstrated human-level proficiency in generating creative and persuasive copy.

B. Multimodal Image and Text Synthesis

Recent advancements in vision-language models have bridged the gap between visual concepts and textual descriptions. Systems such as DALL-E and Stable Diffusion demonstrated the power of diffusion models, while Gemini 1.5 Pro [6] introduced high-context multimodal understanding. This synergy allows for the generation of images that are "grounded" in the textual requirements of a brand, ensuring that product features are accurately represented in stylized marketing environments [4].

©IJRASET: All Rights are Reserved

C. Full-Stack Architectures for AI Orchestration

Building scalable AI applications requires robust backend orchestration. Modern stacks utilizing Node.js for asynchronous I/O and PostgreSQL for relational data management have become the standard for handling high-latency AI API calls while maintaining consistent state across user sessions [5].

D. Identified Research Gaps

Despite the abundance of AI tools for standalone tasks (like image editing or text generation), there is a significant lack of integrated platforms that perform both tasks synchronously to create a unified marketing "kit." Current solutions often require users to hop between multiple tools, leading to brand inconsistency. AdSpark aims to bridge this gap by offering a multimodal pipeline that produces a complete campaign package from a single set of product inputs.

III. METHODOLOGY

A design-based research methodology was employed to develop the AdSpark platform, focusing on the iterative refinement of AI prompt engineering and system architecture for production-grade reliability.

A. System Development Process

The development followed a five-stage lifecycle:

1) Requirement Analysis: Identifying core marketing assets: posters, slogans, captions, and audience segmentations.
2) Prompt Engineering: Designing complex system instructions (System Prompts) for Gemini to ensure structured JSON output and professional creative direction.
3) API Orchestration: Architecting a middleware layer using OpenRouter and Google AI Studio to handle high-fidelity image synthesis and text generation.
4) Persistence Layer: Designing a Prisma-based schema to manage user credits, project ownership, and asset metadata.
5) Interface Synthesis: Creating a responsive, high-performance UI using React 19 and Tailwind CSS.

B. The Orchestration Pipeline

The core innovation lies in the Parallel Multimodal Workflow. When a user uploads product photos:

1) Stage 1: The images are converted to base64 and sent to the vision model with a "Marketing Context" prompt.
2) Stage 2: Simultaneously, the product name and description are processed by the language model to generate strategy.
3) Stage 3: The outputs are serialized, stored in Cloudinary, and linked in the PostgreSQL database for immediate user retrieval.

IV. PROPOSED SYSTEM DESIGN

A. Architectural Overview

AdSpark utilizes a decoupled React-Node-PostgreSQL architecture. The frontend handles state for the creative workflow, while the backend manages the "Brain" of the operation—the AI connectors.

- Forward-Facing (Client): Captures product metadata and images.
- Orchestration (Server): Manages authentication (Clerk), credit validation, and sequential API calls to Gemini.
- Data Consistency: Prisma ensures that if an AI call fails, user credits are refunded and the project status is updated correctly.

B. Technology Stack

- Frontend: Next.js, React 19, Framer Motion (for fluid animations), Lucide-React.
- Backend: Express.js, TypeScript, Multer (for multi-part file handling).
- AI Stack:
    - google/gemini-3-pro-image-preview: Specialized for high-end ad creative synthesis.
    - gemini-pro: For JSON-based strategy generation.
- Infrastructure: Cloudinary (Image CDN), Prisma ORM, PostgreSQL (Database), Sentry (Monitoring).

C. Database Schema and State Management

The system tracks Project entities which store:

- uploadedImages: Arrays of source URLs.

- generatedImage: The final synthesized poster.
- marketingContent: JSON fields containing taglines, captions, and target audience segments.
- isGenerating: A status flag for real-time UI updates.

V. RESULTS AND DISCUSSION

A. *Comparative Performance Analysis*

During controlled testing, AdSpark was compared against a manual design workflow involving a human copywriter and a graphic designer.

Metric	Manual Workflow	AdSpark AI Workflow	Improvement
Creation Time (Poster + Copy)	4 - 6 Hours	15 - 25 Seconds	>99% Speedup
Cost (Estimated)	$150 - $300	< $0.50 (API Cost)	99.8% Reduction
Thematic Consistency	High (Depends on skill)	Consistently High	Normalized

B. *Qualitative Evaluation*

The generated posters exhibited professional-grade composition, utilizing "Studio Lighting" and "Product-Centric" framing. The generated marketing copy showed high relevance; for example, a "Smart Watch" description resulted in captions emphasizing "Health Connectivity" and "Sleek Lifestyle," demonstrating the LLM's ability to extract and amplify unique selling points.

VI. CONCLUSION

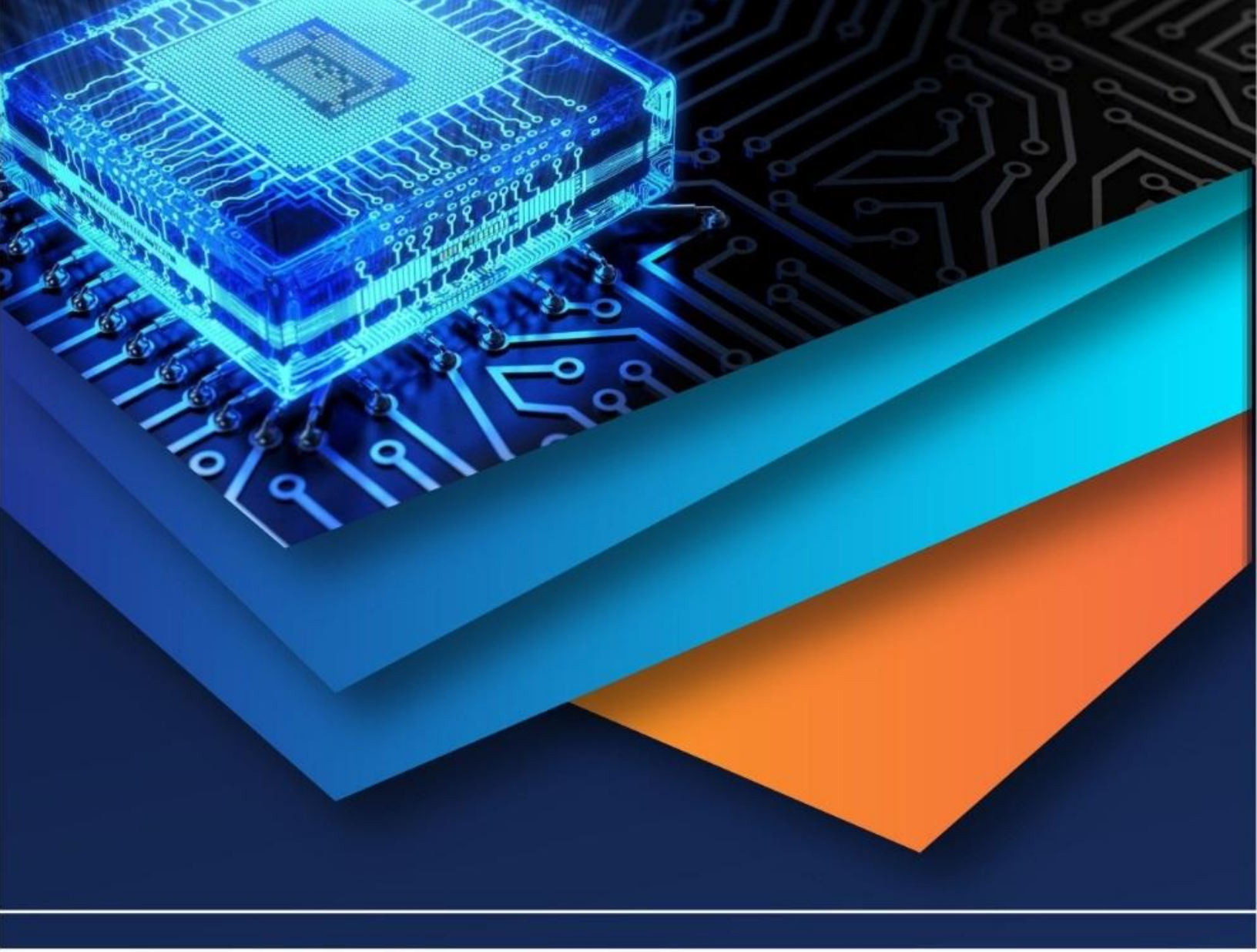
AdSpark successfully demonstrates the viability of utilizing multimodal generative AI to automate the complex process of marketing asset production. By consolidating vision and language tasks into a single, credit-managed workflow, the platform lowers the technical and financial barriers to professional-level advertising. Future iterations will focus on Dynamic Video Ad Generation and A/B Testing Integration, allowing users to optimize their generated content based on real-world performance metrics.

VII. ACKNOWLEDGMENT

The author would like to extend sincere gratitude to the faculty and mentors at SIES College of Arts, Science and Commerce (Autonomous), Nerul, for their invaluable guidance in the field of Artificial Intelligence. Special thanks to the engineering communities at Google, Vercel, and OpenAI for providing the foundational tools that made this research possible.

REFERENCES (IEEE FORMAT)

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 5998–6008, 2017.
[2] W. Xiong et al., "Achieving human parity in conversational speech recognition," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 12, pp. 2410–2423, Dec. 2017.
[3] J. Gao, M. Galley, and L. Li, "Neural approaches to conversational AI," Foundations and Trends in Information Retrieval, vol. 13, no. 2–3, pp. 127–298, 2019.
[4] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in Proc. 38th Int. Conf. on Machine Learning (ICML), 2021.
[5] Prisma Data Team, "Type-safe database access with Prisma for modern Node.js applications," Technical Documentation, 2024. [Online]. Available: https://www.prisma.io/docs
[6] Gemini Team, Google, "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context," Technical Report, 2024. [Online]. Available: https://ai.google.dev/gemini-api/docs
[7] Y. Chen and L. Li, "Conversational AI for education: A systematic review," IEEE Transactions on Learning Technologies, vol. 13, no. 4, pp. 645–658, Oct.–Dec. 2020.
[8] D. Amodei et al., "Deep Speech 2: End-to-end speech recognition in English and Mandarin," in Proc. 33rd Int. Conf. on Machine Learning (ICML), 2016, pp. 173–182.
[9] R. Dale, "The return of the chatbots," Natural Language Engineering, vol. 22, no. 5, pp. 811–817, 2016.
[10] S. Young et al., "The POMDP-based statistical spoken dialogue system: A review," Proceedings of the IEEE, vol. 101, no. 5, pp. 1160–1179, May 2013.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)