Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Ad Genie: A Multimodal Generative AI Framework for Automated Marketing Campaign Creation Using Product Images, Textual Prompts, and Web Intelligence

Ms. Gouthami
Project Guide, Assistant Professor
Dept. of Computer Science and Engineering
Keshav Memorial Institute of Technology
Hyderabad, Telangana, India

K. Saanvi
Author, Student (UG Scholar)
Dept. of Computer Science and Engineering
Keshav Memorial Institute of Technology
Hyderabad, Telangana, India

M. Divya Bharathi
Co-Author, Student (UG Scholar)
Dept. of Computer Science and Engineering
Keshav Memorial Institute of Technology
Hyderabad, Telangana, India

Arekanti Mercy
Co-Author, Student (UG Scholar)
Dept. of Computer Science and Engineering
Keshav Memorial Institute of Technology
Hyderabad, Telangana, India

Chavan Supriya
Co-Author, Student (UG Scholar)
Dept. of Computer Science and Engineering
Keshav Memorial Institute of Technology
Hyderabad, Telangana, India

Nenavath Sreelatha
Co-Author, Student (UG Scholar)
Dept. of Computer Science and Engineering
Keshav Memorial Institute of Technology
Hyderabad, Telangana, India

Abstract

Digital marketing requires product understanding, customer insight, competitive awareness, creative writing, and platform-specific communication. For small businesses, independent sellers, student entrepreneurs, freelancers, and influencers, producing effective campaigns is difficult because it demands both creativity and continuous market research. Existing AI copywriting tools can generate promotional text, but many depend mainly on text prompts, produce generic outputs, and do not fully incorporate visual product cues or real-time market context. This paper presents *Ad Genie*, a multimodal generative AI framework for automated marketing campaign creation. The proposed system accepts a product image and a campaign or product description as input, extracts visual and semantic features using vision-language models, generates search queries for market intelligence, retrieves trend and review-oriented information from online sources, and produces structured campaign assets using a large language model. The generated outputs include social media posts, a blog concept, a short promotional video script, target audience persona, market trends, sentiment summary, and structured intermediate results. The prototype demonstrates how multimodal AI, natural language processing, computer vision, retrieval-augmented generation, and web intelligence can be integrated into a unified workflow for context-aware marketing assistance. The work contributes a practical architecture for AI-driven campaign automation and identifies future directions such as multilingual generation, brand voice learning, automatic publishing, analytics, and AI-assisted video production.

Keywords: Multimodal AI, Generative AI, Digital Marketing, Web Intelligence, Vision-Language Models, Retrieval-Augmented Generation, Campaign Automation, Customer Persona, Content Strategy.

1 Introduction

Digital marketing has become a central component of modern commerce. Online sellers, social media creators, small businesses, and startups depend on product visibility across e-commerce marketplaces, short-video platforms, social networks, blogs, and search engines. A product's success is influenced not only by quality but also by how effectively it is presented to the intended audience. Product images, captions, customer reviews, hashtags, blog narratives, short videos, and influencer-style messages collectively shape purchase decisions.

Creating high-quality marketing content is a multidisciplinary task. It requires product interpretation, knowledge of customer psychology, awareness of market trends, competitor analysis, platform-specific writing ability, and creative storytelling. Large companies often rely on dedicated marketing teams, analytics tools, and creative agencies. In contrast, small sellers and independent creators frequently lack the resources and expertise needed to conduct systematic market research and produce polished campaign material. Their promotional content may therefore become

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

generic, inconsistent, or poorly aligned with customer expectations.

Recent progress in large language models (LLMs) has made it possible to generate fluent marketing copy from natural language prompts [11, 9, 10]. At the same time, vision-language models have improved the ability of AI systems to interpret images and connect visual information with text [3, 4, 5, 6]. Retrieval-augmented generation and web intelligence techniques further allow AI systems to ground their outputs in external information such as reviews, trends, news, and competitor activity [7]. These developments create an opportunity to move beyond simple text generation and toward complete, context-aware campaign generation.

This paper proposes *Ad Genie*, a multimodal AI agent for intelligent campaign creation. The system accepts two primary inputs: a product image and a textual campaign description. It analyzes the product visually and semantically, retrieves relevant market information from online sources, extracts customer and trend insights, and generates a structured campaign package. The current prototype demonstrates this workflow through a web interface where users upload a product image, describe a campaign goal, and receive organized outputs including social media posts, a blog concept, a video script, audience persona, and market analysis.

The central research question addressed in this paper is:

> *How can multimodal product understanding and web intelligence be integrated with generative AI to automate the creation of context-aware digital marketing campaigns?*

The main contributions of this work are:

1. A unified multimodal campaign generation pipeline that combines product image understanding, textual campaign intent, web intelligence, and generative AI.
2. A modular system architecture consisting of input validation, multimodal feature extraction, query generation, market insight extraction, audience strategy generation, and content rendering.
3. A practical prototype that demonstrates campaign creation from a product image and campaign topic through a user-friendly web interface.
4. A structured output design that produces promotional copy, market trends, persona insights, blog concepts, video scripts, and machine-readable intermediate results.
5. A research-oriented evaluation framework for comparing multimodal web-grounded generation with text-only LLM prompting and manual campaign drafting.

2 Background and Related Work

2.1 Generative AI in Digital Marketing

Generative AI has rapidly entered the marketing domain because LLMs can produce fluent text, summarize information, rewrite content in different tones, and generate creative ideas [2, 11, 12]. Marketing applications include ad copywriting, product descriptions, email campaigns, blog outlines, customer support responses, and social media captions. LLMs are particularly useful for reducing the time required for brainstorming and first-draft creation.

Despite these advantages, prompt-based content generation has limitations. If an LLM receives only a short product description, it may generate content that sounds polished but lacks product-specific detail, current market context, or audience precision. The generated text may also hallucinate unsupported claims. For marketing use cases, unsupported claims can mislead customers or damage brand trust. Therefore, marketing generation systems benefit from grounding mechanisms that connect generated content to product attributes and external evidence.

2.2 Multimodal AI and Vision-Language Models

Product marketing is naturally multimodal. A product image communicates color, shape, aesthetic style, material, usage context, and emotional tone. A product description communicates functional details, intended use cases, technical specifications, and brand messaging. A campaign generation system that uses only text misses visual cues that are important for creating relevant content.

Vision-language models such as CLIP, BLIP, LLaVA, and GPT-4o-style multimodal systems have shown that visual and textual information can be represented and reasoned about jointly [3, 4, 5, 12]. CLIP aligns images and text in a shared embedding space, BLIP supports image captioning and vision-language understanding, and LLaVA-style systems connect visual encoders with language models to enable visual question answering and multimodal reasoning. These models make it possible to extract product-level attributes such as dominant colors, product category, style, use case, and visual mood.

Prior work on image advertisements also shows that visual information can improve the interpretation of advertising symbolism and creative intent [15]. This supports the design choice of treating product images as first-class inputs rather than optional decoration.

In Ad Genie, multimodal analysis is used to convert a product image and text prompt into a richer product representation. This representation informs both market query generation and campaign content generation.

2.3 Web Intelligence and Retrieval-Augmented Generation

Marketing content must be sensitive to current trends. Customer preferences, competitor positioning, seasonal demands, and social media discussions change frequently. Static model knowledge alone is not sufficient for trend-aware marketing. Retrieval-augmented generation addresses this issue by combining generative models with external information retrieval [7].

For marketing applications, web intelligence may include search results, product reviews, frequently asked questions, competitor listings, news articles, social media trends, and video review transcripts. This information can be analyzed for sentiment, pain points, keywords, and emerging customer needs. A web-grounded campaign system can then generate content that is more relevant than a purely prompt-based system.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

2.4 Sentiment Analysis, Personas, and Strategy

Effective campaigns require an understanding of target audiences. Customer personas summarize demographic, psychographic, behavioral, and motivational characteristics of potential buyers [13, 14]. Sentiment analysis identifies whether customer discussions are positive, negative, or neutral [8]. Keyword extraction and trend mining identify the language customers use when searching for or discussing products.

In Ad Genie, the system generates an audience persona using product semantics, market trends, and campaign intent. The persona includes likely age group, user interests, needs, pain points, psychographics, and recommended channels. This makes the generated output more strategic than simple ad copy.

2.5 Connection to SOMONITOR

The base paper used for this project, SOMONITOR, presents a framework for marketing analytics using explainable AI, CTR prediction, LLM-based content pillar extraction, persona mining, communication theme mining, and data-driven story generation [1]. SOMONITOR demonstrates how LLMs can support marketing workflows by processing large amounts of advertising content, identifying audience segments, and creating actionable content briefs.

Ad Genie is conceptually related to SOMONITOR but differs in its primary objective. SOMONITOR focuses on monitoring and analyzing existing marketing content and campaign performance. Ad Genie focuses on generating a new campaign package from product-level input. In this sense, Ad Genie adapts the broader idea of AI-assisted marketing intelligence into a product-centered, multimodal campaign generation workflow.

3 State-of-the-Art Positioning

The current state of the art in AI-assisted marketing is shaped by four converging research directions: large language models for natural language generation, vision-language models for multimodal product understanding, retrieval-augmented generation for grounding responses in external knowledge, and explainable marketing analytics for converting data into actionable strategy. Ad Genie is positioned at the intersection of these directions.

Table 1 summarizes how the proposed framework differs from adjacent approaches. Text-only LLM tools are fast and useful for first drafts, but they depend heavily on prompt quality and may miss product-specific visual signals. Vision-language models can describe product appearance, but they do not independently produce a complete marketing strategy. Retrieval-augmented generation grounds outputs in external sources, but it must be connected to domain-specific insight extraction to become useful for marketing. SOMONITOR and related explainable advertising systems focus on analyzing existing campaigns and competitor content. Ad Genie combines these streams into a product-level campaign generation workflow.

4 Research Gap

Although AI marketing tools and LLM-based copywriting systems are increasingly available, several gaps remain:

1. Many systems are text-only and do not use product images for campaign creation.
2. Many generated outputs are generic because they are not grounded in real-time market context.
3. Most tools generate isolated pieces of content rather than a complete campaign strategy.
4. Existing tools may not produce explicit audience personas, pain points, market trends, and recommended channels.
5. Small sellers require simple end-to-end workflows rather than separate tools for research, analysis, writing, and formatting.
6. Pure LLM systems may hallucinate claims when no external grounding is provided.
7. Marketing analytics frameworks often analyze existing campaign data but do not directly generate ready-to-use product campaign assets.

Ad Genie addresses these gaps by integrating multimodal product understanding, web intelligence, strategic insight extraction, and structured generative output in a single workflow.

5 Proposed System

Ad Genie is designed as a modular AI framework for automated campaign creation. The system receives multimodal input, processes it through specialized modules, and produces both human-readable and machine-readable outputs.

5.1 System Objectives

The major objectives are:

1. Automate the generation of digital marketing campaign assets.
2. Combine product image understanding with textual product or campaign descriptions.
3. Retrieve relevant market information from online sources.
4. Extract trends, keywords, sentiment, competitor insights, and audience needs.
5. Generate structured outputs such as social media posts, blog concepts, video scripts, and target personas.
6. Provide a simple interface suitable for non-technical users.

5.2 Target Users

The system is intended for small e-commerce sellers, local businesses, student entrepreneurs, freelancers, digital marketers, influencers, personal-brand builders, and marketing agencies seeking rapid campaign drafts.

5.3 High-Level Workflow

Figure 1 shows the overall workflow. The system begins with a user-provided product image and campaign description, validates the inputs, extracts visual and semantic fea-

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Table 1: State-of-the-art positioning of Ad Genie against adjacent AI marketing approaches.

Approach	Primary Capability	Limitation for Small-Seller Campaign Creation	Ad Genie Extension
Text-only LLM copywriting	Generates fluent captions, blogs, and ad copy from prompts	Often generic; lacks image grounding and live market context	Adds product image analysis, market retrieval, persona generation, and structured outputs
Vision-language product analysis	Extracts visual attributes and image captions from product photos	Does not automatically generate full campaign strategy	Uses visual features as input to query generation, audience strategy, and campaign generation
Retrieval-augmented generation	Grounds generated responses in external documents or search results	Requires domain-specific retrieval and summarization design	Converts retrieved market signals into trends, sentiment, keywords, and content angles
Explainable marketing analytics	Analyzes existing campaigns, audiences, and competitor content	Primarily analytic; not designed as a product-to-campaign generator	Adapts explainable marketing insight into an end-to-end campaign creation agent
Ad Genie full pipeline	Integrates image, text, retrieval, insight extraction, and generation	Requires broader benchmarking and production hardening	Provides a unified prototype for context-aware, product-specific campaign drafting

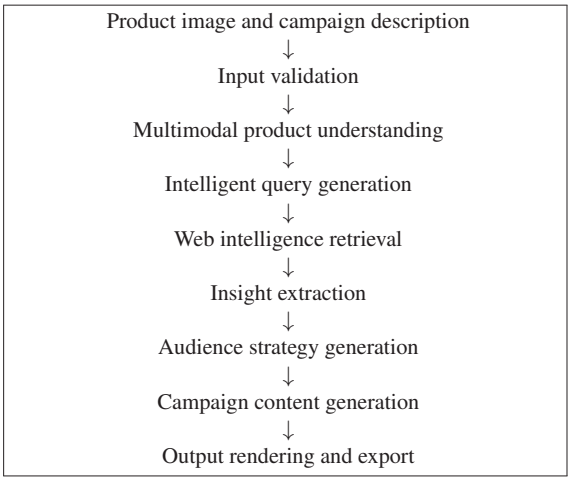

Figure 1: End-to-end workflow of the proposed Ad Genie framework.

tures, retrieves web intelligence, generates insights, and produces campaign-ready assets.

6 Methodology

6.1 Input Layer

The input layer accepts a product image and a textual campaign or product description. The image may be uploaded in standard formats such as JPG, JPEG, PNG, or WEBP. The text description captures the user's campaign goal or product context. For example, a user may enter: "gifting a coffee mug to a friend." The system may also provide an optional manual image description field when an image is not available or when the vision model is disabled.

The input layer validates that the image format is supported, the uploaded file is not corrupted, the text input is not empty, and the system has enough information to perform analysis.

6.2 Multimodal Product Understanding

The multimodal module analyzes both image and text. The vision component extracts product category, dominant colors, physical design, aesthetic style, mood, and usage context. The text component extracts campaign theme, subthemes, product benefits, target use case, brand tone, and constraints such as affordability, energy, elegance, sustainability, or gift suitability.

The outputs from both components are fused into a compact structured representation, as shown in Listing 1.

Listing 1: Example structured representation produced by the multimodal module.

```
{
 "topic_semantics": {
   "main_theme": "gifting",
   "subthemes": ["friendship", "coffee", "mug"],
   "brand_tone": "modern"
 },
 "visual_aesthetic": {
   "colors": ["white", "gray"],
   "style_keywords": ["minimalist", "clean"],
   "mood": "invigorating"
 }
}
```

6.3 Intelligent Query Generation

After product understanding, the system generates search queries for market intelligence. These queries are derived from the product category, campaign theme, customer use case, and visual attributes. For a coffee mug gift example, possible queries include personalized coffee mug gift trends, best gifts for coffee lovers, unique gifts for friends, coffee mug customer reviews, and trending personalized gift ideas.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Query generation improves retrieval relevance because the system searches for market-specific context rather than relying on the user's original prompt alone.

6.4 Web Intelligence Retrieval

The web intelligence module retrieves data from online sources such as search engines, review pages, news sources, social platforms, and product listings. Depending on implementation and API availability, sources may include search engine results, e-commerce reviews, product FAQs, social media discussions, news articles, and YouTube review content or transcripts.

The retrieved data is cleaned before analysis. Cleaning may include removing HTML tags, duplicate results, irrelevant links, advertisements, stopwords, and noisy text.

6.5 Market Insight Extraction

The insight extraction module transforms retrieved data into actionable marketing intelligence. It identifies market trends, customer pain points, common product expectations, customer sentiment, competitor strengths and weaknesses, SEO keywords, hashtags, and recommended content angles.

For the prototype coffee mug case, the system identified trends such as personalized gifts, experiential gifts, and subscription services. The market sentiment was marked as neutral, and the system recognized the gift as thoughtful but potentially not unique enough for some recipients.

6.6 Audience Strategy Generation

The audience strategy module generates a target persona and recommended channels. A persona may include a name, demographic profile, psychographic characteristics, needs and pain points, buying motivations, and recommended platforms. For the coffee mug example, the prototype generated a persona named "Coffee-Loving Friends," with demographics such as young adults aged 18–35, urban dwellers, coffee enthusiasts, and professionals. The recommended channels included Instagram, TikTok, Pinterest, and Facebook Groups.

6.7 Campaign Content Generation

The final generative module produces campaign-ready assets. The system generates three social media posts or tweets, a blog title and outline, a blog concept summary, a short video script with scenes and voiceover cues, a structured campaign summary, and JSON output for debugging, export, or integration.

The content generator uses the structured product representation and market insights as input. This approach reduces generic output by grounding the LLM in product-specific and market-specific context.

7 System Architecture

The system follows a modular architecture. The major components are:

Table 2: Module-wise technology mapping for the proposed system.

Module	Representative Technologies
User interface	Streamlit prototype; extensible to React-based frontend
Vision analysis	BLIP, LLaVA, CLIP-style vision-language models
Text strategy	LLaMA/GPT-style large language models
Web intelligence	DuckDuckGo, Serper API, Bing Search API, review/news sources
Insight extraction	Sentiment analysis, keyword extraction, trend mining
Output rendering	Tabbed UI, JSON data view, text/PDF export

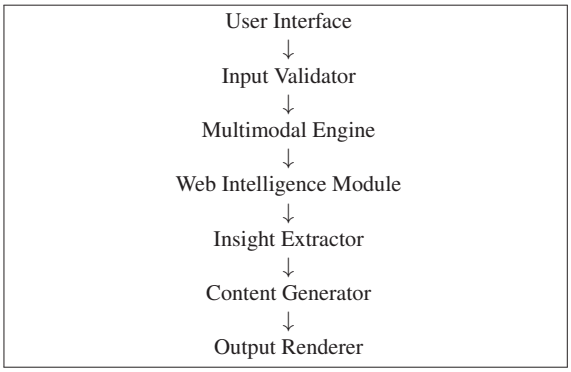

Figure 2: Modular architecture of Ad Genie.

1. **Client/UI Layer:** Provides fields for campaign topic, product image upload, optional manual image description, model settings, and result display.
2. **Input Validator:** Checks file type, input completeness, and basic constraints.
3. **Multimodal Engine:** Uses vision-language models to extract visual and semantic information.
4. **Query Generator:** Produces search queries based on the product representation.
5. **Web Intelligence Module:** Retrieves market data from online sources.
6. **NLP Insight Extractor:** Performs sentiment analysis, keyword extraction, trend detection, and competitor summarization.
7. **Content Generator:** Uses LLMs to produce campaign assets.
8. **Output Renderer:** Displays the generated insights in structured tabs and provides export options.

7.1 Deployment View

The prototype is demonstrated as a local web application. The interface shown in the project screenshots runs at `localhost:8501`, which indicates a Streamlit-based prototype. The broader architecture can also be extended to a production environment using a separate frontend, backend API layer, cloud GPU runtime, and external APIs.

A scalable deployment can consist of a client browser, web server or backend API, AI inference runtime, search and scraping APIs, external data sources, and output storage

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

or export layer.

8 Prototype Implementation

8.1 User Interface

The prototype provides a dark-themed web interface titled *Ad Genie: Where Marketing Meets Magic*. The sidebar includes model settings and indicates the active technologies: LLaMA-3-8B for text strategy, BLIP/LLaVA for image analysis, and DuckDuckGo for market trends.

The main interface includes campaign topic input, product image upload, manual image description, generate button, and result tabs for market analysis, audience strategy, creative content, and data view.

8.2 Product Example Used in Prototype

The demonstrated prototype uses the campaign topic "gifting a coffee mug to a friend" and a product image showing a white and black "Lazy Panda" coffee mug. The uploaded image contains a white mug with a black handle, panda illustration, and a small panda figure on the lid. The visual design suggests a cute, friendly, and gift-oriented product.

8.3 Prototype Output

The system generated visual palette attributes such as white and gray colors, minimalist and clean style, and an invigorating mood. It identified the core theme as gifting, with subthemes of friendship, coffee, and mug. It recognized a modern brand voice, market trends such as personalized gifts and experiential gifts, and neutral market sentiment. It also generated the persona "Coffee-Loving Friends," recommended channels such as Instagram, TikTok, Pinterest, and Facebook Groups, and produced social media posts, a blog concept, video script, and structured pipeline output.

8.4 Example Generated Content

The prototype generated post drafts such as:

1. "Fuel their friendship with a personalized coffee mug! #coffee #giftideas"
2. "Want to make a lasting impression? Try gifting an experience like a coffee-tasting tour! #experientialgifts #coffee"
3. "Ready to upgrade your gifting game? Discover unique and thoughtful presents that reflect your friends' interests! #giftinspo #coffee"

These examples show that the system does not only describe the mug but also expands the campaign toward broader gift-positioning strategies.

9 Experimental Design and Evaluation Framework

The current project bundle demonstrates a working prototype and sample outputs. For a full research evaluation, this paper proposes a structured experimental design. The evaluation should compare Ad Genie with baseline approaches across multiple product categories.

9.1 Dataset Design

A small benchmark dataset can be created using product images and descriptions from different categories: personalized gift items, wireless earbuds, eco-friendly reusable bottles, fashion accessories, skincare products, kitchen appliances, home decor items, stationery products, fitness accessories, and mobile phone accessories. Each sample should include a product image, product title, short product description, intended campaign goal, and human-written reference campaign if available.

9.2 Baselines

Ad Genie should be compared against four baselines: manual campaign drafting, text-only LLM prompting, image captioning followed by LLM generation, and the full Ad Genie pipeline. The full pipeline uses image, text, web intelligence, insight extraction, and content generation.

9.3 Evaluation Metrics

The evaluation should include functional, performance, and content-quality metrics.

9.3.1 Functional Metrics

Functional metrics include image upload success rate, invalid input detection, web retrieval success rate, output generation success rate, and export success rate.

9.3.2 Performance Metrics

Performance metrics include image analysis time, web retrieval time, content generation time, end-to-end response time, and failure recovery time.

9.3.3 Content Quality Metrics

Human evaluators can rate outputs on a 1–5 scale using product relevance, creativity, audience fit, market awareness, platform suitability, clarity, persuasiveness, practical usefulness, factual safety, and overall campaign quality.

10 Results and Discussion

10.1 Prototype Observation

The prototype successfully demonstrates the end-to-end concept of Ad Genie. In the coffee mug example, the system accepted a product image and campaign topic, analyzed visual and semantic attributes, identified market-oriented trends, generated a target persona, and produced structured creative content.

The generated output shows three important capabilities. First, the system performs visual grounding by recognizing color and style cues from the uploaded product image. Second, it performs semantic grounding by connecting the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Table 3: Comparison of Ad Genie with baseline campaign generation approaches.

Method	Image	Web Trends	Persona	Posts	Blog	Video Script	Expected Strength
Manual drafting	Yes	Yes	Yes	Yes	Yes	Yes	High quality but time-consuming
Text-only LLM	No	Limited	Partial	Yes	Yes	Yes	Fast but often generic
Image caption + LLM	Partial	No	Partial	Yes	Yes	Yes	Better product grounding than text-only generation
Ad Genie full pipeline	Yes	Yes	Yes	Yes	Yes	Yes	Structured, context-aware, and product-specific campaign generation

Table 4: Representative test cases for system validation.

ID	Module	Expected Output
MM-01	Input	Valid image and text are accepted
MM-02	Input	Corrupted image triggers an error
MM-03	Input	Missing text triggers validation warning
API-01	Retrieval	Valid query returns market data
API-02	Retrieval	Timeout triggers retry or fallback
IN-01	Insights	Sentiment and keywords are extracted
CG-01	Generation	Three posts are generated
CG-02	Generation	Blog concept is generated
CG-03	Generation	Video script is generated
OP-01	Output	Results are displayed in tabs

campaign topic to themes such as gifting, friendship, coffee, and mug usage. Third, it performs strategic expansion by moving from a simple mug gift into broader trends such as personalized gifts and experiential gifts.

This suggests that multimodal AI can improve campaign generation by linking product appearance, campaign intent, and market positioning.

10.2 Strengths

The main strengths of Ad Genie are its end-to-end workflow, multimodal understanding, structured strategy generation, user accessibility, extensibility, and practical relevance. The system combines tasks that are usually separate: product inspection, trend research, audience thinking, and content drafting. Even when the generated content is not final, it gives users a structured starting point.

10.3 Limitations

The current system has limitations. Web intelligence quality depends on available search results and external APIs. LLM-generated content may still require human review before publication. Sentiment analysis may be inaccurate when retrieved data is noisy or limited. The prototype screenshots demonstrate one product example; broader evaluation is required. API rate limits and model latency can affect response time. Some product categories may require domain-specific prompts or fine-tuning.

10.4 Discussion

Ad Genie should be understood as a campaign drafting assistant rather than a fully autonomous marketing manager. Its outputs can reduce ideation time and help users create a first draft of campaign material. Human review remains necessary for brand accuracy, legal safety, factual correctness, and final publishing decisions.

11 Ethical, Legal, and Practical Considerations

AI-generated marketing content must be handled carefully. The system should avoid unsupported product claims, disclose AI assistance where appropriate, protect uploaded product images, respect API and platform terms, avoid biased persona assumptions, require human approval before publishing, avoid copying protected marketing material, and prevent misleading or manipulative advertisements.

These safeguards are especially important if Ad Genie is extended to automatic publishing or paid advertising workflows.

12 Future Enhancements

Future versions of Ad Genie can include multilingual campaign generation, seller dashboard integration, automatic content publishing after user approval, AI video generation, brand voice learning, analytics dashboards, A/B testing, CTR prediction, price intelligence, and a mobile application. These extensions would allow Ad Genie to evolve from a campaign drafting assistant into a broader AI-powered digital marketing platform.

13 Conclusion

This paper presented Ad Genie, a multimodal generative AI framework for automated marketing campaign creation. The system integrates product image analysis, textual campaign understanding, web intelligence, market insight extraction, audience persona generation, and LLM-based content creation. Unlike text-only copywriting tools, Ad Genie uses both visual and semantic product information and grounds campaign generation in market-oriented insights.

The prototype demonstrates a practical workflow in which a user uploads a product image, enters a campaign

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

topic, and receives structured marketing outputs including social media posts, a blog concept, a video script, persona information, and machine-readable intermediate data. The coffee mug case study illustrates how the system can transform a simple product and campaign goal into a more complete strategy involving gifting themes, audience segments, recommended platforms, and creative content.

Ad Genie is not intended to replace human marketers. Instead, it serves as an AI-assisted campaign drafting and research tool that can reduce manual effort, support small businesses, and provide structured creative direction. With further evaluation, stronger retrieval grounding, multilingual support, analytics integration, and human-in-the-loop safeguards, Ad Genie can evolve into a more complete AI-powered digital marketing assistant.

Acknowledgment

The authors thank the Department of Computer Science and Engineering, Keshav Memorial Institute of Technology, and the project guide Ms. Gouthami for their guidance and support.

References

[1] A. Farseev, Q. Yang, M. Ongpin, I. Gossoudarev, Y.-Y. Chu-Farseeva, and S. Nikolenko, "SOMONITOR: Combining Explainable AI and Large Language Models for Marketing Analytics," arXiv:2407.13117, 2024.

[2] A. Vaswani *et al.*, "Attention Is All You Need," in *Advances in Neural Information Processing Systems*, 2017.

[3] A. Radford *et al.*, "Learning Transferable Visual Models From Natural Language Supervision," in *International Conference on Machine Learning*, 2021, arXiv:2103.00020.

[4] J. Li *et al.*, "BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation," in *International Conference on Machine Learning*, 2022, arXiv:2201.12086.

[5] H. Liu *et al.*, "Visual Instruction Tuning," arXiv:2304.08485, 2023.

[6] L. Baraldi *et al.*, "The Revolution of Multimodal Large Language Models: A Survey," in *Findings of the Association for Computational Linguistics*, 2024, arXiv:2402.12451.

[7] P. Lewis *et al.*, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Advances in Neural Information Processing Systems*, 2020, arXiv:2005.11401.

[8] C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," in *International AAAI Conference on Web and Social Media*, 2014.

[9] S. Makridakis, F. Petropoulos, and Y. Kang, "Large Language Models: Their Success and Impact," *Forecasting*, vol. 5, no. 3, pp. 536–549, 2023.

[10] S. Minaee *et al.*, "Large Language Models: A Survey," arXiv preprint, 2024.

[11] OpenAI, "GPT-4 Technical Report," arXiv preprint, 2023.

[12] OpenAI, "Hello GPT-4o," OpenAI technical announcement, 2024. [Online]. Available: `https://openai.com/index/hello-gpt-4o/`

[13] A. Malik, "Persona Based Marketing Strategies: Creation of Personas Through Data Analytics," Master's thesis, 2019.

[14] D. Pelleg and A. W. Moore, "X-Means: Extending K-Means with Efficient Estimation of the Number of Clusters," in *International Conference on Machine Learning*, 2000.

[15] A. Savchenko *et al.*, "Ad Lingua: Text Classification Improves Symbolism Prediction in Image Advertisements," in *International Conference on Computational Linguistics*, 2020.