CASG: Creative Advertisement Script Generator for Consumer Products

Submitted by

Angara Venkata Sai Singu Rishik [CB.EN.U4CSE21104]

Battula Bharat Chandra [CB.EN.U4CSE21111]

Shaik Zakeer Ahamad [CB.EN.U4CSE21155]

Vakada Rohit [CB.EN.U4CSE21165]

in partial fulfilment of the requirements for the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Department of Computer Science and Engineering
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Coimbatore – 641 112

April 2025

AMRITA VISHWA VIDYAPEETHAM

AMRITA SCHOOL OF COMPUTING, COIMBATORE – 641112

BONAFIDE CERTIFICATE

This is to certify that the project report entitled **Creative Advertisement Script Generator for Consumer Products** submitted by Angara Venkata Sai Singu Rishik (CB.EN.U4CSE21104), Battula Bharat Chandra (CB.EN.U4CSE21111), Shaik Zakeer Ahamad (CB.EN.U4CSE21155), and Vakada Rohit (CB.EN.U4CSE21165) in partial fulfillment of the requirements for the award of the Degree **Bachelor of Technology** in Computer Science and Engineering is a bonafide record of the work carried out under our guidance and supervision at the Department of Computer Science and Engineering, Amrita School of Computing, Coimbatore.

Dr. Rajathilagam B
Guide
Professor
Department of CSE

Dr. P. Bagavathi Sivakumar
Chairperson
Department of CSE

Evaluated on: ___________________

INTERNAL EXAMINER

EXTERNAL EXAMINER

DECLARATION

We, the undersigned, solemnly declare that the project titled **CASG: Creative Advertisement Script Generator for Consumer Products** is based on our own work carried out during the course of our study under the supervision of **Dr. Rajathilagam B, Professor**, Department of Computer Science and Engineering, Amrita School of Computing, and has not formed the basis for the award of any other degree or diploma, in this or any other Institution or University. In keeping with the ethical practice in reporting scientific information, due acknowledgement has been made wherever the findings of others have been cited.

Angara Venkata Sai Singu Rishik [CB.EN.U4CSE21104]-

Bhattula Bharat Chandra [CB.EN.U4CSE21111]-

Shaik Zakeer Ahamad [CB.EN.U4CSE21155]-

Vakada Rohit [CB.EN.U4CSE21165]-

ABSTRACT

The integration of AI-driven techniques in advertisement script generation presents a transformative approach to creating tailored and contextually relevant content. This study introduces CASG (Creative Advertisement Script Generator), a novel framework that leverages advanced deep learning models to automate the generation of theme-based advertisement scripts. The system utilizes the BLIP model for image captioning and a fine-tuned LLaMA3.2-3B model, trained on a curated dataset of manually crafted ad scripts, to generate creative and engaging scripts based on user-specified themes (e.g., humor, luxury, adventure).

Unlike traditional methods that rely on generic templates or manually written scripts, CASG dynamically adapts to both visual content and narrative style, ensuring greater alignment with brand identity and campaign objectives. A comparative analysis with GPT-2 demonstrates that LLaMA3.2-3B produces more creative and contextually relevant scripts, despite slightly higher inference time. To validate the effectiveness of this approach, user feedback and evaluation metrics such as BLEU score, semantic similarity, and perplexity are incorporated to assess script quality and relevance.

Furthermore, the study addresses key ethical concerns in AI-generated advertisements, including bias mitigation, content authenticity, and responsible marketing practices. Deployed as a scalable web application, CASG provides an efficient, theme-adaptive, and automated solution for advertisement script generation, catering to diverse creative needs while enhancing the overall quality of ad content.

Keywords: Forest fires, deep learning, explainable AI, meteorological data, clustering, FRP, environmental management.

ACKNOWLEDGEMENT

We would like to express our deep gratitude to our beloved Satguru Sri Mata Amritanandamayi Devi for providing the bright academic climate at this university, which has made this entire task possible. This acknowledgment is intended to thank all those people involved directly or indirectly with our project. We would like to thank the Board of Management for providing us with the necessary infrastructure required for the completion of the project. We express our thanks to Dr. Vidhya Balasubramanian, Principal, Amrita School of Computing,Dr. Bagavathi Sivakumar P, Chairperson, Department of Computer Science and Engineering and all the faculty members of the Department of Computer Science and Engineering for their valuable help and support during our study.

We wish to record our deep sense of gratitude and profound thanks to our research supervisor **Dr. Rajathilagam B**, **Professor**, Department of Computer Science and Engineering, Amrita School of Computing, Coimbatore, for her keen interest, inspiring guidance, constant encouragement with our work during all stages, to bring this project into fruition. We feel extremely grateful to **Panel Members Name** for their feedback and encouragement which helped me to complete the project.

We would also like to thank the entire fraternity of the Department of Computer Science and Engineering. We would like to extend our sincere thanks to our family and friends for helping and motivating us during the project work. Last but not the least, we thank God for his blessings which made our project a success.

Angara Venkata Sai Singu Rishik
Battula Bharat Chandra
Shaik Zakeer Ahamad
Vakada Rohit

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ML	Machine Learning
DL	Deep Learning
ANN	Artificial Neural Networks
XAI	Explainable AI
FRP	Fire Radiative Power
SVM	Support Vector Machine
LSTM	Long Short-Term Memory
ConvLSTM	Convolutional Long Short-Term Memory
SALSTM	Spatial Attention Long Short-Term Memory
SCALSTM	Spatial Channel Attention Long Short-Term Memory
VIIRS	Visible Infrared Imaging Radiometer Suite
MODIS	Moderate Resolution Imaging Spectroradiometer
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
GRU	Gated Recurrent Unit
SHAP	SHapley Additive exPlanations
LIME	Local Interpretable Model-Agnostic explanations
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
SMAPE	Symmetric Mean Absolute Percentage Error

CHAPTER 1

INTRODUCTION

1.1 Introduction:

The field of digital marketing is developing rapidly, and producing persuasive and engaging advertisement scripts is one of the most important, challenging, resource-consuming, and time-consuming parts of the process. Advertising scripts are typically produced through some aspect of human effort, and considerable human effort is involved in either using previous templates of scripts to recreate new scripts or going through entirely manual creative processes. Human-authored advertisements are often not personalized or adjustable for the individual brand, audience, or campaign themes. Many advertisers are looking for automated solutions that create quality scripts that are customized to branded campaigns, as the digital ad space has become more data-driven, and content engagement has become a greater and greater priority. However, most AI-based text generation algorithms do not adequately incorporate visual composition into their procedural output. For this reason, the output of AI-based text generation results in generic text that fails to connect to images used in the advertisements. Additionally, advertisement subject matter ranges from tone, and style, to target audience, and requires an approach that can continuously adapt based on user-defined themes, such as funny, high-end, or adventure. The demand for effective automated generation of advertisement scripts that are creative and themed has never been higher, highlighting the need for automated solutions that focus on efficiency, creative variety, and content relevance.

A range of artificial intelligence methods have been expounded for text generation in advertising, ranging from rule-based models and statistical methods, to ML and DL models. Earlier AI-based models for generating scripts subject to advertising had an AI text model based on a pre-defined text structure and therefore restricted the creativity and adaptability of the content. Advancements in NLP models such as GPT-2 and BERT further advanced language generation applications; however, these models tended to lack coherence when transforming them to advertisement categories or connecting advertisement scripts jointly with visuals. Additionally, these models faced challenges in expressing nuances in branding and generating compelling, engaging narratives that also tied in product-specific detailed information. This project illustrates an AI system entitled, CASG (Creative Advertisement Script Generator), using deep learning and combining the BLIP model for image captioning with a fine-tuned LLaMA3.2-3B model for text generation. This model adapts advertisement scripts dynamically and creatively to both image visuals and user-defined topics and themes, which, in conjunction, generates more cohesive, engaging, and contextually relevant marketing content. Lastly, a comparative study with a GPT-2 advertisement script system demonstrated that CASG generates superior ad scripts, delivering a scalable, efficient, and theme-adaptive advertising and marketing tool for modern advertisement needs.

1.2 Problem Definition:

In today's advertising landscape, generating relevant and compelling advertisement scripts is still a difficult and resource-intensive process. Scriptwriting, in its traditional form, relies on human imagination, which requires extensive manpower, time, and effort to curate content that is engaging and suitable for various products, audiences, and marketing tactics. Moreover, traditional scriptwriting procedures are not scalable, flexible, or efficient enough for wide-scale personalized advertising initiatives. In addition, advertisements often integrate visual elements, but the previous text-based generation methods do not allow for efficacious incorporation of image-related insights into the scriptwriting process. Consequently, the scripts generated for the advertisements often resulted in ads that were generic, disjointed, or thematically inconsistent, and/or did not resonate with the target audience. Besides, different brands also have their own unique tone that deserves consideration, which adds another level of demand for thematic script generation that aligns with the brand's personality and audience expectation. Without an automated way to adjust the ad scripts and optimize the content, based on the visual aspects of the ad, theme, and the brand style, modern advertising campaigns would be disorganized, and efficiency and effectiveness would suffer.

Current AI-driven text generation models like GPT-2 and BERT offer some capabilities for scriptwriting, but they often produce generic outputs that lack creative depth and contextual awareness. These models do not take product images into account, making their generated scripts disconnected from the actual advertisement visuals. Additionally, most text-generation models operate on predefined structures, limiting their ability to adapt to different themes or brand voices. Some recent advancements in deep learning and natural language processing (NLP) have improved text generation, but existing solutions still fail to effectively personalize content for specific advertising needs. To address these limitations, this project introduces CASG (Creative Advertisement Script Generator)—a deep learning-based framework that integrates image captioning and theme-based script generation. CASG utilizes the BLIP model to extract meaningful captions from product images, which are then processed by a fine-tuned LLaMA3.2-3B model to generate highly relevant, theme-adaptive advertisement scripts. This automated, scalable, and visually-aware system enhances content personalization while reducing manual effort, making it a powerful tool for modern marketing campaigns.

CHAPTER 2

LITERATURE SURVEY

2.1 Survey :

The recent developments in Artificial Intelligence (AI) have dramatically changed manycreative industries; among these changes, the generation of advertising scripts is perhaps themost active area of development. This paper will provide insight into a diverse range of recentacademic papers and articles into Creative Advertisement Script Generation (CASG). As digitalplatforms continue to increase and the demand for interesting, high quality, personalizedadvertising content rises, there has certainly been heightened interest in using AI in the scriptwriting process in both parseiveness and in quality. In this section we will review recentresearch studies that examine various aspects of AI in advertising, including but certainly notlimited to, natural language generation (NLG) AI, transformer models used as AI in advertising,reinforcement learning (RL) used for creative content, and AI that combines human input forcopywriting and slogan generation. Reviewing this body of literature will help us develop astrong understanding of the cutting-edge, while also identifying important trends, problems, andpotential avenues of the young field of CASG. Murakami and colleagues (2023) produced a comprehensive survey of the last ten years of Natural Language Generation (AdNLG) literature, demonstrating a remarkable shift from some attitude of simple template and extractive techniques toward more sophisticated abstractive methods powered by neural networks [Murakami et al., 2023] [1]. Their research highlights the extent to which these intensive deep learning models have now entered into the production of authentic and engaging ad copy optimized for all kinds of media channels, constituting a decisive step forward in ad creativity. This is an intellectual stride that is indicative and aligned with larger trends in AI, utilizing the proliferation of generative models to see a shift from template-style, and essentially static, rule-based systems toward more dynamic systems that employ learning to uniquely describe creative possibilities. Building on sparse possibilities, Chen and colleagues (2024) took on the complementary value of BERT and its deep semantic understanding in conjunction with the generative capabilities of GPT-4 [Chen et al., 2024] [2]. The results indicated that this approach provided better results in the development of advertising scripts that were coherent and contextually accurate, and demonstrated promise for value added in advancing effectiveness for, or in the case of CASG, these blended capabilities could enhance efficiency. This merger takes advantage of BERT's ability to capture subtle semantic meanings and GPT-4's proficiency in generating text that exhibits dialogically fluent, human-like delays, making this a performance-enhancing tool for advertisers who prefer clarity over creativity. Li et al. (2024) have presented TOLE, a new reinforcement learning algorithm that has been designed to provide reward incentives at the level of tokens for controllably generated text [Li et al., 2024] [3]. TOLE tackles longstanding issues associated with reinforcement learning-based text generation, such as overfitting and semantic collapse, positioning it as a critical resource for influencing the creative output of CASG systems in more nuanced and creative ways. Specifically, TOLE provides rewards for linguistic choices at an increased granularity and control over tone, style, and messaging, which is critical in advertising.Likewise, in a separate investigation, Ofek Redden (2024) have examined the potential of ChatGPT-4 in generating advertising slogans, learning that AI-generated advertisements are often comparable in quality if not superior to slogans produced by marketing professionals, particularly when prompts were sufficiently detailed [4]. This research presents important implications for utilizing AI to develop short, effective advertising messages intended for target customers. The performance of ChatGPT-4 indicates that prompt engineering, the stage in the human-AI interaction where humans influence the AI-generated output, are vital components of an exciting future where AI, and human creatives will work closely together. On another note, Rauf et al. (2023) investigated the use of AI for copywriting in the Nigerian advertising industry, revealing both advantages, such as time savings or content analysis, and disadvantages, such as originality and emotion [5]. Their research demonstrated the importance of the human component being able to ensure that AI-generated content has authenticity and emotional power: a consideration that remains paramount in the development of CASG frameworks. Additionally, their region-specific research further exposes how cultural and market-based perceptions impart layers of complexity in the adoption of AI globally. And, Alnajjar and Toivonen (2021), when studying the computational generation of slogans, focused on the automation of generating slogans for products or services based on predetermined attributes [6]. This method is directly related to creating creative advertising tag lines in an organized way to encourage AI outputs to integrate more with brand objectives. By establishing pre-defined parameters (e.g., tone or important attributes), this method ensures that AI-generated content is strategically aligned, which is important in advertising. Similarly, Hughes et al. (2019) explored the potential of reinforcement learning with a baseline model, which would train systems to generate search engine text advertisements tailored for user clicks [Hughes et al., 2019] [7]. Their data-driven approach illustrates how AI can improve click-through potential with advertising language, one of the key metrics for success in digital marketing campaigns. The emphasis on outcomes illustrates how AI can integrate creativity with performance in advertising. Going further, Wang et al. (2021) presented a data study with reinforcement of pretrained models to write compelling text advertisements, valuable from large corpora, as well as from transfer learning methods used to improve the functional quality and attractiveness of advertisements [Wang et al., 2021] [8]. These methods as a whole reflect the ways fundamental AI methods can be repurposed for the more complex needs of advertising creativity, merging art and data-driven optimization.

Taking the next step in this journey, Zhang et al. (2021) introduced CHASE, which adds commonsense knowledge to search engine advertising, allowing for the development of more creative and relevant ads that extend the use of external knowledge bases [Zhang et al., 2021] [9]. This illustrates how AI can be steered away from purely data-driven generation while enriching ad scripts with contextual intelligence to enhance persuasion. By tying ad content to a more extensive human understanding, CHASE demonstrates how AI can use commonsense knowledge, creating ads that are more relatable and insightful rather than simply using a mechanical approach. The PLATO-Ad framework, described by Yuan et al. (2022), is multi-task learning with prompts in a Transformer-based framework to address the variety of tasks associated with ad generation [Yuan et al., 2022] [10]. The integrated system illustrates the ability of AI to generate varied advertisement content, including scripts and taglines, with high efficiency and coherence. The multi-task approach also suggests future prospects for integrated AI systems that could produce or oversee multiple aspects of campaign creation at the same time. In addition to technological advances, there have been more expansive discourses surrounding the role of AI in advertising. Gao et al. (2023) provided a thorough review of the evolution of AI-enabled advertising, related challenges, and ethical considerations. The review discusses targeting, personalization, content generation, and ad optimization [Gao et al., 2023] [11]. Their review serves as a comprehensive view of how these systems are shaping advertising, and also raises important questions around privacy, bias, and ethical applications of automated systems. These questions arose particularly as the systems demonstrated greater autonomy and sway over the accountability and fairness of using AI systems. In the same manner, Pokhrel and Banjade (2023) analyzed the potential and consequences of AI content generator technologies based on Open AI language models, with a focus in advertising [12]. Their research emphasizes the revolutionary possibilities of these models alongside the consideration of developing standards around issues such as content authenticity and intellectual property in AI-generated materials—both are issues that could challenge brand trust and compliance with related laws. In a more tactile study, Redden (2025) took a first-hand approach to investigate the possible effectiveness of ChatGPT in generating ad taglines, and concluded that while AI can produce high-quality options, the best outcome still requires a detailed prompt and then going through multiple rounds of testing for the best tagline [13]. This result adds to the consensus that AI's generative capacity to be greater with relevant human contribution and judgement–meaning, a collaborative, iterative, and hybrid workflow where AI is employed as a generative assistant alongside human oversight instead of a substitution for creative human input. In turn, De Cremer et al. (2023) provided an insightful discussion about the possibility of generative AI to disrupt fields (including advertising) reliant on creativity and responding with questions around what it means for the future of creativity (human or AI), the impacts on employment from the technology, and whether it is even beneficial for art integrity to have AI involved [De Cremer et al., 2023] [14]. These are mounting questions and ones at the forefront of discussions relating to creativity and CASG, and so tensions between mechanisation and creativity is transportable tensions and perhaps a defining issue in this field. Li et al. (2021) ultimately presented a survey of pretrained language models for text generation, offering information on their architectures, training processes, and uses [Li et al., 2021] [15]. This visualization of techniques for CASG is an important reference point for understanding the technical nature of AI-based generation as it relates to script generation. Collectively, this body of work illustrates a field that is changing rapidly, where AI is not only contributing to the ability to produce advertising content with greater efficiency, but is also fundamentally triggering towards what is creatively possible.

The change toward AI-based CASG systems illustrates a broader trend toward automation in computer-generated creative workflows, allowing agencies to generate high amounts of content at a speed previously unseen. Models like GPT-4 and ChatGPT-4 can create scripts and taglines with contextualized richness, allowing marketers to personalize campaigns to specific groups of individuals or media platforms. A marketer can create an engaging tagline for Twitter or create a full-throated script for a YouTube ad campaign with less effort than writing a well-considered process. The scale and readiness to adapt to various advertising media is especially useful with digital advertising considering that brands are competing to saturate a media presence across various behaviors (social media, search, streaming services, etc.), needing each outlet to have messaging that might be distinctly different within the same overall narrative. Nonetheless, AI usage comes with complications. Rauf et al. (2023) mention that AI-generated outputs are seen as less emotionally resonant, which is a concern in advertising as emotional connection is crucial to consumer behavior, this is attributed to an AI's basis in patterns and data as opposed to experience, so the products it generates are often formulaic. This implies a hybrid model where AI generates drafts and then a human creative rewrites the content to add empathy and cultural dimension that an algorithm would not generate. Gao et al. (2023) raise ethical concerns specifically regarding biases that perpetuate marginalizations, such as training data supposed to support certain behaviors, that could alienate particular groups, or an infringement of privacy through overpersonalization. Therefore, strict guidelines are needed as these tools become institutionalized. There is also a need to consider where multimodal generative AI tools will lead next. It is clear from literature reviews and frameworks like PLATO-Ad that AI systems linked with text, image, and audio, have the potential to innovate the CASG practice. Future AI output will include suggestions for visuals or a soundtrack alongside scripted text for cohesive campaigns. In this way, AI will create greater efficiency and innovation by associating script writing, design, and production. Real-time feedback loops that adjust content according to an audience's response may add another layer of dynamism to the process, while restrictions on thorough knowledge bases or multi-lingual models, such as CHASE by Zhang et al., might lead to transnational advertisements engaging worldwide and reducing labor-intensive human localization in multi-national ads (e.g., easily integrating the slogan into multiple languages).

In conclusion, the recent research engaged here confirms the extraordinary progress made in using machine learning to generate advertising content, with CASG as a frontier. We argue there is not only a continually evolving sophistication of machine-learning models, the emergence of reinforcement learning and transfer learning are all environmental considerations, as is the ongoing tension between automating the process and being responsible for the value of creativity, originality, and ethical appropriateness. We expect as the technologies develop and mature it will completely change the art and science of advertising, offering machines as a way to elevate, augment, and enhance human ingenuity while addressing the complexities of a rapidly changing digital environment. The future of CASG is not just in machine learning improvements, but in collaborative machine-human creativity, in order to co-create what advertising means—connecting with people.

2.2 Summary and Findings :

The survey on Creative Advertisement Script Generation (CASG) underscores the rapid evolution of artificial intelligence (AI) in transforming advertising creativity, driven by advancements in natural language generation (NLG), transformer models, and reinforcement learning (RL). The reviewed studies collectively highlight a shift from static, rule-based systems to dynamic, generative AI frameworks capable of producing personalized, high-quality advertising scripts. Early research by Murakami et al. (2023) traced a decade of AdNLG progress, showing a transition from template-driven methods to neural network-powered abstractive techniques. These developments have enabled the creation of engaging ad copy optimized across media channels, aligning with broader trends in AI-driven creativity.

Building on this, Chen et al. (2024) demonstrated the synergy of BERT's semantic depth and GPT-4's generative fluency, producing coherent and contextually relevant advertising scripts. Their findings suggest that such hybrid models enhance efficiency in CASG, offering clarity and creativity for advertisers. Similarly, Li et al. (2024) introduced TOLE, an RL algorithm rewarding token-level linguistic choices, improving control over tone and style—key for nuanced ad outputs. This addresses challenges like overfitting, positioning TOLE as a vital tool for CASG's creative precision. Ofek Redden (2024) further explored ChatGPT-4's ability to generate slogans, finding AI outputs rival human efforts when guided by detailed prompts, emphasizing the role of human-AI collaboration. Region-specific insights emerged from Rauf et al. (2023), who examined AI copywriting in Nigeria, noting time-saving benefits but highlighting limitations in emotional resonance and originality. This underscores the need for human oversight in CASG to ensure authenticity. Alnajjar and Toivonen (2021) focused on automated slogan generation, showing how predefined attributes align AI outputs with brand goals, a structured approach critical for advertising. Hughes et al. (2019) and Wang et al. (2021) leveraged RL and pretrained models to optimize click-through rates in digital ads, merging creativity with performance metrics—a core CASG objective. Advanced frameworks like CHASE by Zhang et al. (2021) integrated commonsense knowledge, enhancing ad relevance, while Yuan et al.'s (2022) PLATO-Ad showcased multi-task learning for versatile script generation. Gao et al. (2023) provided a broader review, raising ethical concerns around bias and privacy in AI-driven advertising, suggesting the need for guidelines. Pokhrel and Banjade (2023) echoed this, noting authenticity and intellectual property challenges with generative models. Redden (2025) and De Cremer et al. (2023) emphasized iterative human-AI workflows and questioned AI's impact on creative integrity, respectively.

Overall, the survey reveals significant strides in CASG through AI, yet persistent challenges remain in emotional depth, ethical deployment, and real-time adaptability. The cited works collectively call for hybrid solutions blending robust script generation, cultural sensitivity, and human judgment, which this project aims to address by advancing CASG's practical and creative potential.

2.3 Motivation :

- The advertising industry is continuously seeking innovative approaches to enhance the creation of tailored and contextually relevant content, and AI-driven techniques present a transformative solution to achieve this.

- Traditional methods of scriptwriting rely heavily on manual effort, which is time-consuming and often lacks scalability, highlighting the need for automated solutions to streamline the scriptwriting process and improve efficiency.

- Recent advancements in AI, particularly in computer vision and natural language processing (NLP), offer the potential to bridge the gap between visual and textual data, enabling the generation of more creative and contextually relevant advertisement scripts.

- Existing systems often lack the ability to incorporate user-defined themes and preferences, limiting their flexibility and applicability in real-world marketing scenarios.

- The integration of visual and textual data remains a challenge, particularly in ensuring that generated scripts align with both the visual content and the desired narrative style.

- By addressing these gaps, AI-driven solutions can offer a scalable and efficient approach to automated advertisement script generation, catering to diverse creative needs and enhancing the overall quality of ad content.

2.4 Objective of the work :

The advertising industry is undergoing a significant transformation with the advent of AI-driven technologies. Traditional methods of advertisement script generation are often characterized by time-consuming manual processes and limited scalability. In contrast, AI offers the potential to automate and enhance creativity in this domain. This project, ”CASG: Creative Advertisement Script Generator,” aims to address these challenges by developing a system that leverages AI to generate creative and contextually relevant advertisement scripts.

The primary objective of this project is to design and implement an AI-driven framework that automates the generation of advertisement scripts. This involves integrating computer vision techniques for visual feature extraction with natural language processing models for text generation. By doing so, the project seeks to streamline the scriptwriting process, improve efficiency, and enable the creation of more engaging and tailored ad content.

- **Dataset Curation and Processing:** A key objective is to create a custom dataset of product images and corresponding advertisement scripts. This involves collecting images from diverse sources, manually writing scripts aligned with specific themes and product attributes, and augmenting the dataset to enhance its robustness. Furthermore, the project includes preprocessing steps to ensure that the images are compatible with the AI models and that the text data is appropriately formatted for training.

- **Model Development and Fine-Tuning:** Another critical objective is to develop and fine-tune AI models for image captioning and script generation. This involves utilizing the BLIP model to extract visual features from product images and generating descriptive captions. Additionally, the Llama 3.2 model will be fine-tuned on the custom dataset to generate creative and theme-aligned advertisement scripts.

- **Evaluation and Validation:** To ensure the effectiveness of the proposed framework, a rigorous evaluation methodology will be implemented. This includes using metrics such as BLEU score, perplexity, semantic similarity, accuracy, and response time to assess the quality of the generated scripts. The performance of the fine-tuned Llama 3.2 model will be compared with other models, such as GPT-2, to demonstrate its superiority. User feedback and practical validation will also be incorporated to evaluate the real-world applicability of the system.

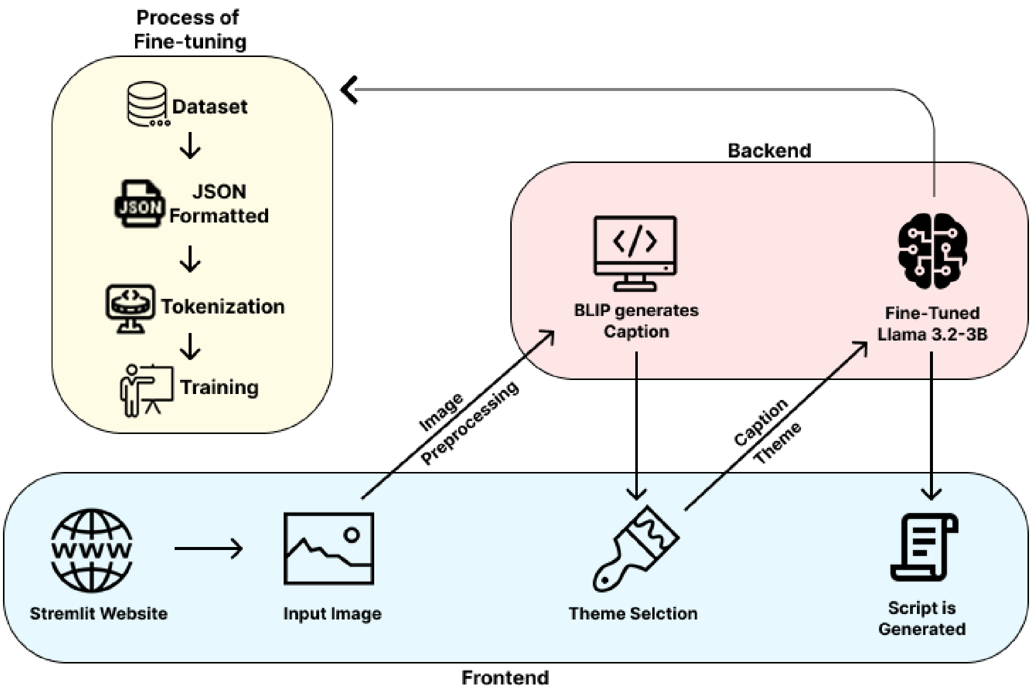
- **Deployment and Implementation:** A further objective is to deploy the developed framework as a functional application. This will involve creating a user-friendly interface that allows users to input product images and themes and generate advertisement scripts. The deployment will also address technical considerations such as scalability and efficiency to ensure that the system can be used in real-world scenarios.

CHAPTER 3

PROPOSED WORK

3.1 Architecture of the System :

The proposed architecture for the Creative Advertisement Script Generator (CASG) consists of three main modules: the Frontend, the Backend, and the Fine-Tuning Process. These modules work together to enable the generation of contextually relevant and theme-aligned advertisement scripts.



3.1.1 Frontend Module :

The Frontend Module serves as the user interface for the CASG system. The process begins with the user providing an input image, typically through the Streamlit website. In addition to the image, the user is also required to select a theme that will guide the style and content of the generated advertisement script. These two inputs, the image and the theme selection, are then passed from the Frontend module to the Backend module, where the core processing of the data takes place.

3.1.2 Backend Module :

The Backend Module is responsible for the AI-driven generation of the advertisement script. This process involves two key stages. First, the input image received from the Frontend module undergoes image captioning. The image is preprocessed and then analyzed by the BLIP model to generate a descriptive caption, effectively transforming the visual information into a textual representation. Second, the generated caption from BLIP, along with the theme selected by the user, is fed into the fine-tuned Llama 3.2-3B model. This model, having been trained on a specialized dataset of advertisement scripts, utilizes these inputs to generate the final advertisement script.

3.1.3 Fine-Tuning Process Module

The Fine-Tuning Process Module is dedicated to the preparation of the Llama 3.2-3B model, ensuring its capability to generate relevant and high-quality advertisement scripts. This preparation involves several steps. Initially, a dataset of advertisement scripts is organized and formatted into JSON. The textual data within this dataset is then tokenized, a process that breaks down the text into smaller units suitable for processing by the model. Finally, the Llama 3.2-3B model is trained using this prepared dataset. This training process results in the fine-tuned Llama 3.2-3B model that is utilized within the Backend module to generate the advertisement scripts.

The architecture diagram provides a visual representation of the flow of information and the interactions between these three modules, clearly illustrating how the input image and theme are processed to produce the final advertisement script.

3.2 Algorithm Design :

The CASG framework is designed to automate the generation of advertisement scripts through a series of structured processes. This section details the algorithms and methods employed to achieve this, focusing on the integration of the BLIP model for visual feature extraction and the fine-tuned LLaMA 3.2 model for script generation.

3.2.1 System Overview:

The CASG framework comprises three primary stages, as shown in Figure 1:

- Input Image Preprocessing: The initial stage where the input image is prepared for processing.
- Visual Feature Extraction: The BLIP model extracts relevant features and generates a caption.
- Script Generation: The LLaMA 3.2 model uses the caption and theme to generate a script.

3.2.2 Input Image Preprocessing

The input image is subjected to multiple modifications to align with the BLIP model requirements. Initially, it is resized to 224×224 pixels using the formula:

$$\text{Image}_{\text{resized}} = \text{Resize}(\text{Image}_{\text{input}}, 224, 224)$$

Here, $\text{Image}_{\text{input}}$ represents the original image, while $\text{Image}_{\text{resized}}$ denotes its modified version. This resizing step is critical because models like BLIP typically require inputs of a specific size. It ensures that all images—regardless of their initial dimensions—can be effectively processed.

In this process, bicubic interpolation is employed for resizing; this method strikes an effective balance between computational efficiency and maintaining quality. Following this adjustment, the pixel values are normalized within the interval $[0, 1]$ through the equation:

$$\text{Image}_{\text{normalized}} = \frac{\text{Image}_{\text{resized}}}{255.0}$$

By dividing pixel values by 255, we scale them down to lie between 0 and 1. This normalization phase plays a vital role in enhancing neural network training stability, since networks generally perform better with smaller input values. Such normalization avoids instances where larger pixel values might overpower other attributes during learning. The pixel values are normalized and then fine-tuned with the mean and standard deviation obtained from the ImageNet dataset. This is done using the formula:

$$\text{Image}_{\text{preprocessed}} = \frac{\text{Image}_{\text{normalized}} - \text{Mean}_{\text{ImageNet}}}{\text{Std}_{\text{ImageNet}}}$$

Here, $\text{Mean}_{\text{ImageNet}}$ and $\text{Std}_{\text{ImageNet}}$ represent the average values and standard deviations for the ImageNet dataset. This refinement enhances the normalization process by ensuring that the distribution of input images aligns closely with that of the ImageNet dataset, which served as a foundation for training the BLIP model. Achieving this alignment is crucial, as it allows the BLIP model to interpret images more effectively. Adjusting based on mean and standard deviation is a type of standardization commonly employed in machine learning to boost model performance.

3.2.3 Visual Feature Extraction using BLIP

Figure displays the preprocessed image processed by the BLIP model to extract visual features and provide a description. The BLIP model's Vision Transformer (ViT) architecture extracts a feature representation defined by:

$$\text{Features}_{\text{BLIP}} = \text{BLIP}(\text{Image}_{\text{preprocessed}})$$

3.2.1.3 Spatial Channel Attention Long Short-Term Memory(SCA-LSTM):

SCALSTM is a cutting-edge extension that entails self-attention between time series and cross-attention between several correlated features.

.

- Applied SCALSTM model for learning spatial importance (time axis) and channel importance (feature axis).
- The model architecture includes:
  - LSTM layers for sequence learning.
  - Spatial attention to weigh significant time steps.
  - Direct emphasis on the significant characteristics such as temperature, rainfall, or vegetation indices.
- Data preprocessing and sequence generation followed the same pipeline as LSTM and SALSTM models.
- Each of the 7 clusters was trained individually for the model with 0.001 learning rate and Adam optimizer and MSE loss for 50 epochs.
- Following training, cluster models were saved as scalstmmodels.pth.

- This dual-process attention was imperative for performance under conditions of high climatic noise and variability

3.2.1.4 Gated Recurrent Unit (GRU):

GRU is a simpler form of LSTM but with fewer parameters, which trains faster and yet remains competitive, especially on noisy or irregular data.

- Applied GRU networks for time-series forecasting of FRP, taking into account vegetation and climate attributes.
- The GRU structure consisted of 2 stacked GRU layers and a fully connected output layer.
- Input sequences of 30 timesteps were formed for every cluster based on normalized feature values.
- The model was trained cluster-wise utilizing Adam optimizer with a learning rate of 0.001 and MSE loss for 50 epochs.
- Trained model states were stored in a file grumodels.pth for future use in prediction.
- The trained model states were saved in a file named grumodels.pth for later prediction use.
- Predictions were denormalized to the original FRP scale for analysis and visualization.

3.2.1.5 Transformers:

Transformers are advanced models created initially for NLP but currently used extensively for time-series forecasting because they have a global attention mechanism and are scalable.

- Applied Transformer-based architecture for modeling FRP sequences with multiple climatic and vegetation parameters to learn intricate temporal and feature interactions.
- The model architecture includes an embedding layer to project input features into a higher-dimensional space, two Transformer encoder layers with 4 attention heads each, and a final fully connected output layer for next-day FRP prediction.
- Input data was normalized by MinMaxScaler, and sequences of 30 timesteps were created to model temporal dependencies between features such as temperature, rainfall, vegetation indices, and wind.
- Training was performed for every one of the seven clusters independently, with the Adam optimizer having a learning rate of 0.001 and MSE loss across 50 epochs, for regional specificity in prediction.

- Following training, state dictionaries of all cluster-specific models were merged and stored in transformermodels.pth for later use.
- During prediction, the model utilized the last 30 days of feature data to predict FRP iteratively day by day for December 2024.
- Denormalized predicted FRP values were stored in transformerpredictions.csv, and the attention mechanism allowed the model to concentrate on significant temporal points and features, enhancing prediction accuracy under highly variable environmental conditions.

3.2.2 Explainable AI Techniques :

3.2.2.1 SHapley Additive exPlanations(SHAP):

- We used SHAP to interpret and explain feature importance for all trained prediction models (GRU, LSTM, SA-LSTM, SCA-LSTM, Transformers) for multiple clusters.
- The models were initialized for each of the seven clusters, and 30-day sequential samples of input data (normalized climatic and vegetation features) were made for analysis.
- GradientExplainer from the SHAP library was used to calculate SHAP values on the input sequences of each cluster, allowing for gradient-based explanation of deep models.
- Random sampling (100 samples per cluster) was carried out for all clusters in order to get representative data for feature contribution calculations.
- The SHAP values were aggregated by calculating the mean absolute values for each time step and sample to find the most significant features on average for each cluster.
- The final output was a SHAP importance matrix, feature-wise contribution scores by clusters, saved in a structured CSV file (SHAPResults.csv).
- These findings were illustrated by dynamic bar charts (cluster-wise), giving precise and understandable descriptions to identify which environmental parameters (e.g., brightness temperature, wind speed, or vegetation indices) contributed most to forest fire prediction.

3.2.2.2 Local Interpretable Model-Agnostic Explanations(LIME):

- Local feature importance for the learned prediction models was interpreted using LIME by examining the effect of every feature on every FRP prediction at a local level.
- For each of the seven clusters, the trained GRU models were initialized and data specific to each cluster were aggregated by date and normalized for the sake of uniformity before explanation analysis.

- Random samples (100 samples per cluster) were sampled, each of which was a 30-day feature sequence, reshaped for tabular explainer compatibility with LIME.
- LimeTabularExplainer was used in regression mode with feature names being extended to correspond to each time step to allow for easier interpretation of sequential data in tabular form.
- For each sample case, LIME explanations were created using input data perturbation and noting model responses. Feature importance scores were summed over time steps and samples by summing absolute contributions.
- These importance values were normalized (divided by the number of samples) and stored in a structured dictionary for each cluster, indicating the features that most strongly influenced daily FRP predictions.
- Results were gathered and stored in a CSV file (LIMEGRU.csv), and interactive cluster-wise bar plots were generated for visualization to determine top contributing features in various clusters easily.

3.3 Data sets for the study and Platform :

3.3.1 Dataset Sources:

Forest fire forecasting can be informed by a plentiful, multi-dimensioned database documenting forest fires and environmental levels over time. Data utilized in this study were obtained from two dependable well-defined sources:

a) ERA5 Dataset:

- **Source Website:** https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview
- **Dataset Details:**
  ERA5 provides single-level hourly data from 1940 to date, combining observations and model data from the whole globe. The reanalysis approach makes the dataset temporally and spatially consistent, offering parameters most crucial to our study, including:
  – Latitude, Longitude,Date, Temperature (t2m), Wind speed (u100), Vegetation (lai-lv, lai-hv),Clouds (cvl, cvh)
- **Period Used for Study:** 2019-2024
- **Screenshot Reference:**

b) NASA FIRMS:

- **Source Website:** https://firms.modaps.eosdis.nasa.gov/
- **Dataset Details:**
  The FIRMS system provides near real-time active fire information from satellites such as MODIS (Moderate Resolution Imaging Spectroradiometer) and VIIRS (Visible Infrared Imaging Radiometer Suite). They include:
  – Latitude, Logitude, Date, Confidence Score, Brightness temperature, bright t-31, Fire Radiative Power(FRP)
- **Period Used for Study:** 2019-2024
- **Screenshot Reference:**

c) Data Integration Process:

- Upon obtaining the two datasets from the two sources, they were then merged with caution using latitude, longitude, and date-time matching. This was to guarantee that each ERA5 environmental record was matched with the closest FIRMS fire occurrence record within the same time period and geographical boundary.
- This combined dataset serves as the basis for building predictive models linking environmental conditions with fire incidence.

ERA5.jpeg

3.3.1.1 Justification for Dataset Selection:

- **Comprehensiveness:**
  ERA5 and FIRMS together provide environmental and fire event data, therefore all the factors that may affect (climatic and physical).

- **Reliability:**
  Both data sets are globally recognized, validated, and utilized by research communities and government agencies for environmental monitoring and forecasting.

- **Granularity:**

FRP.jpeg

ERA5 hourly high-resolution data and FIRMS near-real-time data allow high temporal resolution, which facilitates accurate sequential modeling.

- **Coverage of Real-time and Historical Events:**
  Utilizing data from 2019–2023 offers adequate historical variation and recent trends for training strong deep learning models.

- **Ease of Integration:**
  Both give data in clean, easily-readable formats (CSV, NetCDF), prepared to be processed and consumed by machine learning pipelines.

3.3.2 Platform and Specifications:

3.3.2.1 Platform used:

For training, testing, and development, both local infrastructure and cloud platforms were used in conjunction:

- **Cloud Platform:**
  - **Google collab Pro+**(used for model development, large data preprocessing, and experimentation)
  - GPU: NVIDIA Tesla P100 / V100
  - RAM: Up to 52 GB
  - Disk: Up to 166 GB temporary storage
  - Libraries Used: TensorFlow, Keras, PyTorch, Scikit-learn, SHAP, LIME, Matplotlib, Seaborn, Plotly
- **Local Machine Specifications:** (for preprocessing and visualization tasks)
  - Processor: Intel i7 12th Gen
  - RAM: 16 GB
  - Storage: 1 TB SSD
  - OS: Windows 11
  - Software Environment: Anaconda Distribution, Python 3.10, Jupyter Notebook, VS Code
- **Database for Storage and Handling:**
  - Google Drive (Temporary storage and backup)

3.3.2.2 Model Development Environment:

- All models (GRU, LSTM, SALSTM, SCALSTM, and Transformer) were executed and tested on Python 3.10.
- TensorFlow and PyTorch frameworks were utilized to train and optimize the model.
- SHAP and LIME explainability models were applied after model prediction to evaluate feature contributions.
- Google Colab Pro+ with GPUs were used for large-scale hyperparameter tuning and for multiple runs to enable faster training.

CHAPTER 4

RESULTS AND INFERENCES

4.1 Metrics For Evaluation :

The accuracy of forest fire prediction models was analyzed with measures that measure prediction accuracy, magnitude of error, direction of differences, and errors in percentages. These inclusive measures create a stringent check for model reliability.

- **Mean Absolute Error (MAE)**:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{1}$$

  MAE gives a straightforward interpretable measure of how much off, on average, predictions are from the actual fire radiative power (FRP). Minimizing MAE in forest fire prediction means that predictions will closely follow actual fire intensity magnitudes.

- **Root Mean Squared Error (RMSE)**:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{2}$$

  RMSE punishes large errors more than MAE, and it is therefore suitable for forest fire prediction where large prediction errors can result in extreme misallocation of resources. It emphasizes the variance and reliability of the prediction.

- **Symmetric Mean Absolute Percentage Error (SMAPE)**:

$$SMAPE = \frac{100\%}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \tag{3}$$

  SMAPE can be compared straightforwardly across different clusters or regions with varying fire intensities. As FRP measurements can differ widely between sites, applying a percentage error measure avoids scale dependence and enables equitable comparison.

4.1 Parameter Settings:

Table 2: Parameter Settings for Each Prediction Model

Model	Parameters	Values/Settings
LSTM	Hidden Layer Size	64 units
	Number of Layers	2
	Input Sequence Length	30 days
	Optimizer	Adam
	Learning Rate	0.001
	Epochs	50 (Batch Size = 64)
SALSTM	Hidden Layer Size	64 units
	Number of Layers	2
	Attention Mechanism	Spatial-attention on LSTM output
	Input Sequence Length	30 days
	Optimizer	Adam
	Learning Rate	0.001
	Epochs	50 (Batch Size = 64)
SCALSTM	Hidden Layer Size	64 units
	Number of Layers	2
	Spatial Attention	Identifies significant time steps
	Channel Attention	Highlights key features dynamically
	Input Sequence Length	30 days
	Optimizer	Adam
	Learning Rate	0.001
	Epochs	50 (Batch Size = 64)
GRU	Hidden Layer Size	64 units
	Number of Layers	2
	Input Sequence Length	30 days
	Optimizer	Adam
	Learning Rate	0.001
	Epochs	50 (Batch Size = 64)
Transformer	Embedding Dimension (d_model)	64
	Number of Attention Heads	4
	Transformer Encoder Layers	2
	Input Sequence Length	30 days
	Optimizer	Adam
	Learning Rate	0.001
	Epochs	50 (Batch Size = 64)

4.3 Results and Discussion:

4.3.1 Dataset Results:

The data used for forest fire forecasting was multi-source environmental and fire index information gathered from ERA5 (Meteorological data) and NASA FIRMS (Historical Fire data) for 2017–2024. The data contained important features such as temperature, precipitation, wind speed, vegetation indices ,brightness, confidence scores, and brightness temperature after preprocessing and latitudinal, longitudinal, and date alignment. The preprocessed dataset was massive, very heterogeneous, and normalized to be stable during training and inference.

Table 3: Sample Result of Preprocessed Dataset

Date	Lat	Long	CVL	CVH	LAI_LV	LAI_HV	T2M	U100	Brightness	Confidence	Bright_t31	frp
2017-01-06	21.25	77	0.492	0.493	0.560	0.256	0.951	0.489	0.089	0.473	0.485	9.6
2017-01-07	23	79	0.926	0.028	0.483	0.347	0.936	0.783	0.079	0.494	0.485	4.9
2017-01-08	21.25	75	0.798	0.201	0.312	0.311	0.941	0.386	0.084	0.231	0.492	4.0
2017-01-09	23	79	0.926	0.028	0.484	0.347	0.931	0.640	0.059	0.473	0.429	5.6

4.3.2 Clustering Results:

The clustering step in the current study is also important in refining model accuracy and geographic specificity. Based on the complexity and enormity of the gathered data across various years (2017–2024) within the Satpura region, where high wildfire activity has been seen in the past few years, it became necessary to spatially partition the data. Such partitioning facilitated the predictive models to understand localized patterns within regions and hence enhance accuracy and dependability.In order to accomplish this, three popular clustering algorithms were utilized and compared: K-Means, DBSCAN, and Agglomerative Clustering.

Table 4: Clustering Silhouette Scores

Clustering Method	Silhouette Score
K-Means	0.4442
Agglomerative	0.4101
DBSCAN	0.1361

- K-Means reported the highest silhouette score (0.4442), reflecting well-defined clusters.
- Agglomerative Clustering reported 0.4101, with moderately separated clusters.
- DBSCAN reported lowest (0.1361), reflecting weak clustering due to sensitivity to density.

- **K-Means** was identified as the best method via silhouette scores.
- Optimal clusters found: **9** for Agglomerative and **7** for K-Means.

Optimal Clusters.jpeg

- K-Means successfully segregated data into 7 distinct areas, reflecting peculiar climatic trends.
- Cluster optimization guaranteed significant segmentation without data dilution.
- Employing clusters enhanced model accuracy by allowing localized learning and noise reduction, improving generalization on unseen data.