VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering

Project Report on

Generation of Company Specific ADs - ADGenAI

In partial fulfillment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai
Academic Year 2023-24

Submitted by
Yashraj Mulwani - D17A - 46
Gauri Nagral - D17A - 49
Meera Sawantdesai - D17A - 60
Kritika Yadav - D17A - 72

Project Mentor
Dr. Mrs. Sharmila Sengupta

(2023-24)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
Department of Computer Engineering

Certificate

This is to certify that **Yashraj Mulwani (D17A, 46), Gauri Nagral (D17A, 49), Meera Sawantdesai (D17A, 60), Kritika Yadav (D17A, 72)** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on "**Generation of Company Specific ADs - ADGenAI**" as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor **Dr. Mrs. Sharmila Sengupta** in the year 2023-24 .

This project report entitled **Generation of Company Specific ADs - ADGenAI** by ***Gauri Nagral, Meera Sawantdesai, Kritika Yadav, Yashraj Mulwani*** is approved for the degree of **B.E. Computer Engineering.**

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date:
Project Guide:

Project Report Approval
For
B. E (Computer Engineering)

This project report entitled **Generation of Company Specific ADs - ADGenAI** by ***Yashraj Mulwani, Gauri Nagral, Meera Sawantdesai, Kritika Yadav*** is approved for the degree of **B.E. Computer Engineering.**

Internal Examiner

External Examiner

Head of the Department

Principal

Date:
Place: Mumbai

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Gauri Nagral(49)

Meera Sawantdesai(60)

Kritika Yadav(72)

Yashraj Mulwani(46)

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Dr. Mrs. Sharmila Sengupta**(Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr. (Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J. M. Nair ,** for giving us this valuable opportunity to do this project.

We extend our sincere gratitude to **Mr. Mahesh Parthak**, **CEO** at **Oneture Technology Pvt. Ltd**., and his dedicated team for their invaluable guidance.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E PROJECT

Learners will be to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solutions for the problem.
CO 4	Able to interpret the data and datasets to be utilised.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one’s competency.
CO 11	Able to develop a professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Index

List of Figures:

List of Tables:

Abstract

In an era defined by data-driven decision-making and the relentless pursuit of customer engagement, marketing and advertising have emerged as pivotal elements for businesses to thrive. However, the contemporary advertising landscape is characterized by a paradox—consumers demand personalized, relevant, and error-free advertisements, while the scale and complexity of the advertising industry make meeting these expectations a considerable challenge. In response to this challenge, this project, bearing the title ADgenAI, endeavors to merge cutting-edge artificial intelligence (AI) with the art of advertisement creation, introducing a transformative approach that embraces advanced AI models, innovative techniques, and user-friendly interfaces.

This project explores the capabilities of Generative Artificial Intelligence (GenAI) for visual element generation through advanced techniques. It analyzes leading models like StackGAN, StyleGAN, Stable Diffusion - DreamBooth, and LORA, identifying the most efficient among them. The project then focuses on implementing these models to automate advertisement creation for product-based companies, aiming to replace manual efforts in the creative department. The proposed system, ADGenAI offers insights into the transformative potential of GenAI in streamlining creative processes, fostering efficiency, and envisioning the future of the advertising industry.

Specifically designed to alleviate the pressures faced by creative teams, such as the demand for high-quality content under tight deadlines, this system enhances creativity and productivity by automating the initial stages of ad production and enabling the exploration of new advertising concepts with minimal resource investment.

Chapter 1: Introduction

1.1.Introduction:

Marketing and advertising are crucial elements for any business to reach its target audience and drive sales. AI's integration into the marketing industry has brought about significant advancements, enabling businesses to create more engaging and personalized ads. Like in so many industries, the use of data and AI is transforming advertising at a rapid pace. Consumers are seeing these changes in the personalized ads on their web browsers, the chatbots that help them make buying decisions. 48% of marketing leaders say AI is making the most significant difference in how customers interact with them. 64% of B2B marketers consider AI to be valuable in their marketing strategy.

By harnessing the power of AI, marketing and advertising professionals can streamline processes, improve efficiency, and ultimately attract more customers with engaging and relevant ads. Overall, the successful development and implementation of the generative AI model hold the potential to revolutionize the company's marketing campaigns. By leveraging AI-generated ads that resonate with the target audience, align with the brand identity, and reflect the latest marketing trends, the company aims to increase brand visibility, drive higher engagement, and achieve superior marketing campaign performance across all product launches.

1.2.Motivation:

The primary objective of this project is to pioneer a cutting-edge approach by developing a state-of-the-art generative AI model, empowered by Stable Diffusion, to autonomously generate tailored advertisements for new company products. Leveraging the company's wealth of data and historical advertising designs, this innovative solution is poised to revolutionize our creative processes. By harnessing AI's cost-efficiency, time-saving potential, and unmatched accuracy, we aim to streamline and optimize the creative endeavors of our marketing and advertising teams. The unique capabilities of AI, combined with its ability to surpass human capabilities in this context, set the stage for a transformative leap forward in our advertising endeavors.

1.3. Problem Definition:

The goal of this project is to develop an AI system capable of generating customized advertisements and captions tailored to specific company products. This system will leverage cutting-edge technologies, including Stable Diffusion and the GPT model, to autonomously create visually compelling ads and captions that resonate with target audiences. Furthermore, the system will prompt users to provide input and customization options to ensure that the generated ads align with the company's branding and marketing objectives. By harnessing the wealth of data and historical advertising designs available, this innovative solution aims to revolutionize our creative processes, enhancing efficiency, accuracy, and creativity in our advertising endeavors. Through the seamless integration of AI-driven ad generation and user customization prompts, we seek to optimize the output of our marketing and advertising teams, driving sales and brand visibility to new heights.

1.4. Existing Systems:

AdGenAI emerges as a novel solution in the AI-driven advertisement creation landscape, synthesizing the strategic content optimization seen in platforms like Google's Responsive Search Ads and the personalized copy generation of Persado, with the innovative image generation capabilities of DALL-E and Adobe Photoshop's Neural Filters. It bridges the gap between dynamic advertisement optimization and creative visual content production, aiming to revolutionize the way businesses engage with their audiences. By integrating the automation and efficiency of digital campaign management seen in systems like Albert, with the cutting-edge visual synthesis and customization offered by platforms such as Artbreeder, AdGenAI sets itself apart as a comprehensive tool for creating visually compelling and contextually relevant advertisements. This unique positioning allows AdGenAI to offer an advanced approach to capturing consumer attention and enhancing engagement in the competitive digital marketplace, making it a standout project in the convergence of AI technology and creative marketing strategies.

1.5. Lacuna of the Existing System:

1. **Limited Integration:** Existing systems often separate text and image generation, lacking a unified ad creation approach.
2. **Rigid Automation:** Many platforms restrict creative freedom, prioritizing efficiency over customization.
3. **Technical Complexity:** Advanced features in some systems require expertise, hindering accessibility for non-experts.
4. **High Cost:** Cutting-edge advertisement technologies can be cost-prohibitive for smaller businesses.
5. **Static Content Strategies:** Few platforms dynamically adapt content strategies based on real-time data and trends.
6. **Fragmented Solutions:** Campaign management and creative content production are often addressed by separate tools.

1.6. Relevance of the Project:

The ADGenAI project stands at the crossroads of innovation and necessity in the digital marketing landscape, offering a timely solution to the growing demand for high-quality, engaging content. By automating advertisement production with advanced GenAI technologies, it empowers businesses to produce compelling visuals and copy from simple text prompts, addressing the need for efficiency, scalability, and personalization in ad creation. In an era where the ability to quickly generate customized, brand-aligned advertisements can significantly impact market presence and consumer engagement, ADGenAI's integration of Stable Diffusion, LLM-ChatGPT, and OpenCV technologies provides a competitive edge, making it a highly relevant tool for product-based companies aiming to thrive in the dynamic digital marketplace.

Chapter 2: Literature Survey

A. Overview of literature survey:

This overview explores how the latest AI developments are changing the way we create advertisements and visuals. It discusses how recent studies have found new methods to quickly come up with creative ideas and make images that are both unique and high-quality. The research highlights tools that help make better image details and compares different AI tools to see which creates the most realistic images. Overall, it shows that AI is becoming an essential part of creating engaging and personalized content, offering exciting possibilities for the future of advertising and design.

2.1. Research Papers :

1. **John Ford, Varsha Jain, Ketan Wadhwani, Damini Goyal Gupta, ”AI advertising: An overview and guidelines” - Journal of Business Research, Volume 166,2023,114124,ISSN 0148-2963**

   a. **Abstract:**
   Advertising has rapidly evolved in recent years, with a significant increase in the use of artificial intelligence (AI) and its applications. While AI advertising literature dates to the 1990s, the field has experienced a surge in research attention and development in recent years, presenting varied potential avenues for future research. Despite this progress, understanding of the evolution of AI advertising research remains limited, and a state-of-the-art overview is required to advance future research. To address this gap, this review aims to map the field's evolution by conducting a bibliometric and framework-based analysis of 75 AI advertising articles published between 1990 and 2022. The study's key findings are the publication trends in AI advertising, TCCM classification, and research contexts identified through bibliographic coupling. Four themes emerged as key focus areas of AI advertising research: programmatic advertising and automation, ad planning and engagement, advertising effectiveness, and trust in AI advertising. In addition, this review offers practical guidelines and future research directions for developing AI advertising literature. Lastly, the review suggests broader implications for industry and academia, highlighting how the identified themes can inform advertising practice and contribute to the theoretical development of the field.

   b. **Inference:**
   Advertising has undergone significant transformation over recent decades, with the potential for even more profound changes through the integration of AI. This study conducts a thorough analysis of AI's role in advertising research, employing bibliometric and framework-based methodologies. The research corpus encompasses publications spanning from 1990 to 2022, and data analysis relies on the biblioshiny tool within the R package. The descriptive data analysis reveals that while AI in advertising research has roots in the past few decades, interest in this domain has experienced noticeable growth.

2. **Liuqing Chen, Pan Wang, Hao Dong, Feng Shi, Ji Han, Yike Guo, Peter R.N. Childs, JunXiao, Chao Wu, “An artificial intelligence based data-driven approach for design ideation, Journal of Visual Communication and Image Representation” - Volume 61, 2019, Pages 10-22, ISSN 1047-3203**

    a. **Abstract:**
    Ideation is a source of innovation and creativity, and is commonly used in early stages of engineering design processes. This paper proposes an integrated approach for enhancing design ideation by applying artificial intelligence and data mining techniques. This approach consists of two models, a semantic ideation network and a visual concepts combination model, which provide inspiration semantically and visually based on computational creativity theory. The semantic ideation network aims to provoke new ideas by mining potential knowledge connections across multiple knowledge domains, and this was achieved by applying “step-forward” and “path-track” algorithms which assist in exploring forward given a concept and in tracking back the paths going from a departure concept through a destination concept. In the visual concepts combination model, a generative adversarial networks model is proposed for generating images which synthesize two distinct concepts. An implementation of these two models was developed and tested in a design case study, which indicated that the proposed approach is able to not only generate a variety of cross-domain concept associations but also advance the ideation process quickly and easily in terms of quantity and novelty.

    b. **Inference:**
    This paper introduces an integrated approach featuring two distinct models: the semantic ideation network and the visual concepts combination. These models aim to inspire creativity through both semantic and visual avenues, drawing upon principles rooted in computational creativity theory. Our case study's outcomes indicate that the semantic ideation network effectively generates diverse cross-domain associations and expedites the ideation process through the utilization of "step-forward" and "path-track" algorithms.

3. **Axel Sauer, Tero Karras, Samuli Laine, Andreas Geiger, Timo Aila, “StyleGAN-T: Unlocking the Power of GANs for Fast Large-Scale Text-to-Image Synthesis” - Machine Learning (cs.LG); Computer Vision and Pattern Recognition (cs.CV), arXiv:2301.09515, 2023**

    a. **Abstract:**
    Text-to-image synthesis has recently seen significant progress thanks to large pretrained language models, large-scale training data, and the introduction of scalable model families such as diffusion and autoregressive models. However, the best-performing models require iterative evaluation to generate a single sample. In contrast, generative adversarial networks (GANs) only need a single forward pass. They are thus much faster, but they currently remain far behind the state-of-the-art in large-scale text-to-image synthesis. This paper aims to identify the necessary steps to regain competitiveness. Our proposed model, StyleGAN-T, addresses the specific requirements of large-scale text-to-image synthesis, such as large capacity, stable training on diverse datasets, strong text alignment, and

controllable variation vs. text alignment tradeoff. StyleGAN-T significantly improves over previous GANs and outperforms distilled diffusion models— the previous state-of-the-art in fast text-to-image synthesis— in terms of sample quality and speed.

b. **Inference:**
The paper introduces StyleGAN-T, a novel GAN-based model tailored specifically to enhance text-to-image synthesis. Recognizing the persistent challenge of slow generation inherent in traditional GAN architectures, StyleGAN-T aims to overcome this limitation by introducing innovative mechanisms. Notably, the model achieves significant improvements in both speed and sample quality compared to conventional diffusion models. By focusing on lower resolutions, StyleGAN-T emerges as a competitive solution particularly suited for large-scale text-to-image synthesis tasks. Its contributions mark a substantial advancement in the field, offering a promising avenue for accelerating the generation process while maintaining high-quality output.

4. **Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, In So Kweon, "Text-to-image Diffusion Models in Generative AI: A Survey" - Computer Vision and Pattern Recognition (cs.CV); Artificial Intelligence (cs.AI); Machine Learning (cs.LG), arXiv:2303.07909, 2023**

a. **Abstract:**
This survey offers an in-depth exploration of text-to-image diffusion models, situating them within the broader landscape of diffusion models' increasing popularity across diverse generative tasks. Initially, the survey provides a foundational understanding of basic operations within diffusion models for image synthesis, elucidating their fundamental principles. It then proceeds to elucidate how conditioning or guidance mechanisms augment these models, facilitating more targeted and expressive image generation. Central to the survey is a comprehensive review of state-of-the-art methods in text-conditioned image synthesis, with a primary focus on text-to-image generation. Moreover, the survey extends its purview to encompass applications beyond text-to-image generation, including text-guided creative generation and text-guided image editing, underscoring the versatility and potential of text-conditioned generative models. Alongside showcasing the progress achieved thus far, the survey conscientiously examines prevailing challenges within the field. It further delineates promising future research directions, aiming to catalyze advancements in text-to-image synthesis and related areas, thereby enriching the landscape of generative modeling.

b. **Inference:**
This paper focuses on text-to-image synthesis, a task that generates realistic images from text descriptions. It reviews the evolution of techniques in this field, emphasizing the recent dominance of diffusion models. These models, like DDPM, have sparked a surge in interest. The paper surveys these models' background, discusses pioneering works and improvements, and considers benchmarks and ethical concerns. It also touches on related tasks, ultimately highlighting the challenges and prospects in text-to-image synthesis with diffusion models

5. **Prafulla Dhariwal, Alex Nichol, “Diffusion Models Beat GANs on Image Synthesis” - 35th Conference on Neural Information Processing Systems (NeurIPS 2021)**

    a. **Abstract:**
    It demonstrates the remarkable capabilities of diffusion models in achieving image sample quality that surpasses the current state-of-the-art in generative models. We accomplish this feat in unconditional image synthesis by iteratively refining the model architecture through a series of ablations, leading to significant improvements. Furthermore, in the realm of conditional image synthesis, we elevate sample quality even further by integrating classifier guidance. This innovative approach involves leveraging a straightforward yet computationally efficient method to balance diversity and fidelity, utilizing gradients from a classifier. Our experimentation yields impressive results, showcasing an FID (Fréchet Inception Distance) of 2.97 on ImageNet at a resolution of 128x128, 4.59 at 256x256, and 7.72 at 512x512. Notably, we achieve comparable performance to BigGAN-deep, even with as few as 25 forward passes per sample, while simultaneously maintaining a superior coverage of the distribution. Moreover, we explore the synergy between classifier guidance and upsampling diffusion models, resulting in further enhancements in FID scores to 3.94 at 256x256 resolution and 3.85 at 512x512 resolution on the ImageNet dataset. These findings underscore the significant advancements facilitated by diffusion models in generating high-quality images and highlight the efficacy of incorporating classifier guidance for improved performance across various image synthesis tasks.

    b. **Inference:**
    In their research, the authors demonstrate that diffusion models achieve superior sample quality compared to state-of-the-art GANs. They have enhanced the architecture, enabling it to excel in unconditional image generation tasks. Additionally, they introduce a classifier guidance technique for class-conditional tasks, allowing for an adjustment in classifier gradient scale to balance diversity and fidelity. While diffusion models still require multiple forward passes during sampling, the authors show that combining guidance with upsampling further improves sample quality, especially in high-resolution conditional image synthesis.

6. **R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 10674-10685, doi: 10.1109/CVPR52688.2022.01042**

    a. **Abstract:**
    By decomposing the image formation process into a sequential application of denoising autoencoders, diffusion models (DMs) achieve state-of-the-art synthesis results on image data and beyond. Additionally, their formulation allows for a guiding mechanism to control the image generation process without retraining. However, since these models typically operate directly in pixel space, optimization of powerful DMs often consumes hundreds of

GPU days and inference is expensive due to sequential evaluations. To enable DM training on limited computational resources while retaining their quality and flexibility, we apply them in the latent space of powerful pretrained autoencoders. In contrast to previous work, training diffusion models on such a representation allows for the first time to reach a near-optimal point between complexity reduction and detail preservation, greatly boosting visual fidelity. By introducing cross-attention layers into the model architecture, we turn diffusion models into powerful and flexible generators for general conditioning inputs such as text or bounding boxes and high-resolution synthesis becomes possible in a convolutional manner. Our latent diffusion models (LDMs) achieve new state of the art scores for image inpainting and class-conditional image synthesis and highly competitive performance on various tasks, including unconditional image generation, text-to-image synthesis, and super-resolution, while significantly reducing computational requirements compared to pixel-based DMs.

b. Inference:

The authors introduce Latent Diffusion Models (LDMs) as an efficient approach to high-resolution image synthesis. LDMs leverage a perceptually equivalent latent space obtained through a separate autoencoder training phase. This approach significantly reduces computational demands for both training and sampling, making it accessible to a wider range of researchers and users. LDMs achieve competitive performance on various tasks, including unconditional image synthesis, inpainting, and super-resolution, while lowering computational costs. They also introduce a general-purpose conditioning mechanism based on cross-attention, enabling multi-modal training for class-conditional, text-to-image, and layout-to-image models. Pretrained LDM and autoencoding models are made available for various tasks.

7. Lvmin Zhang, Anyi Rao, and Maneesh Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models", - Computer Vision and Pattern Recognition (cs.CV); Artificial Intelligence (cs.AI); Graphics (cs.GR); Human-Computer Interaction (cs.HC); Multimedia (cs.MM), arXiv:2302.05543, 2023

a. Abstract:

We propose an innovative architecture designed to integrate spatial conditioning controls into large, pretrained text-to-image diffusion models, termed ControlNet. This approach leverages existing, production-ready large diffusion models, securing their deep and robust encoding layers pretrained on vast datasets containing billions of images. These layers serve as a solid foundation, ensuring the model's capacity to effectively learn a diverse array of conditional controls. The neural architecture of ControlNet is intricately connected with "zero convolutions," initializing convolution layers with zero parameters, gradually growing them to prevent any detrimental noise from influencing the fine-tuning process. Through rigorous experimentation, we explore various conditioning controls, including edges, depth, segmentation, human pose, among others, within the framework of Stable Diffusion. We assess the efficacy of utilizing single or multiple conditions, with or without prompts, to guide the image generation process. Our findings demonstrate the robustness of ControlNet training across datasets of varying sizes, ranging from small (less than 50,000 samples) to large-scale (exceeding 1 million samples). Furthermore, extensive empirical results underscore the potential of ControlNet to enable a broader spectrum of applications for controlling image diffusion models. By providing a flexible and adaptive framework for incorporating spatial conditioning controls, ControlNet opens up new avenues for enhancing the interpretability, controllability, and diversity of generated images. This advancement holds significant promise for advancing the capabilities and practical utility of text-to-image diffusion models across diverse domains and applications.

b. Inference:

The paper introduces ControlNet, a neural network architecture designed to enhance the spatial control of large, pretrained text-to-image diffusion models. It achieves this by connecting with "zero convolutions" to progressively incorporate parameters into the architecture. ControlNet enables various conditioning controls, such as edges, depth, segmentation, human pose, and more, for Stable Diffusion models. The training of ControlNet is demonstrated to be robust with both small and large datasets. It offers wider applications for controlling image diffusion models, including multi-modal conditions and composition of multiple conditions.

8. Borji, Ali. (2022). Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. 10.48550/arXiv.2210.00586.

a. Abstract:

In recent years, the field of image synthesis has witnessed remarkable advancements, leading to the generation of images of unprecedented quality. Despite this progress, there remains a lack of fine-grained evaluation, particularly in categories such as faces. In this study, we undertake a quantitative comparison of three prominent systems—Stable Diffusion, Midjourney, and DALL-E 2—specifically focusing on their ability to generate photorealistic faces in diverse settings. Our analysis reveals that Stable Diffusion outperforms the other systems, as indicated by the FID score. Additionally, we contribute to the field by introducing a new dataset of generated faces in various real-world contexts, termed GFW, comprising 15,076 unique faces. We anticipate that our findings will not only facilitate a deeper understanding of the capabilities of generative models but also serve as a catalyst for future research aimed at enhancing their performance.

b. Inference:

This paper conducts a quantitative comparison of three popular generative models, namely Stable Diffusion, Midjourney, and DALL-E 2, in their ability to generate photorealistic faces in diverse scenes. The goal is to evaluate these models' performance in generating realistic faces within complex, cluttered scenes. The authors introduce a dataset of generated faces in the wild called GFW, consisting of 15,076 faces. They also assess the quality of these generated faces against real faces using the Fréchet Inception Distance (FID) score.

2.2. Patent Search :

1. ON-DEMAND GENERATION AND PERSONALIZATION OF VIDEO CONTENT

Inventor: Ravishankar Iyer[US], Nilesh Kumar Jain[US], Rameshkumar Illikkal[US], Carl S. Marshall[US], Selvakumar Panneer[US], Rajesh Poornachandran[US]

The patent describes a sophisticated video streaming device designed to dynamically generate and personalize advertisement content within video streams. This device utilizes interface circuitry for network communication and processing circuitry for several key functions: receiving video stream content, obtaining video analytics data (including scene content and context, and objects within the scene), and generating personalized video advertisements based on an advertisement template and the analytics data. This process involves advanced techniques such as convolutional neural network (CNN) analysis for scene recognition, generative neural network (GNN) models for advertisement content creation, and user context data for personalization. Additionally, the system incorporates digital rights management (DRM) protocols for content protection and authentication, ensuring a secure and customized advertising experience tailored to the viewer's context and the video's content. This patent outlines a comprehensive approach to integrating targeted advertisements seamlessly into video content, leveraging AI and machine learning technologies to enhance viewer engagement and advertisement effectiveness.

2. MACHINE CONTENT GENERATION

Inventor: Bao Tran[US]

The patent describes a computerized system and method for generating documents using generative artificial intelligence (AI). It involves receiving first and second text prompts, from which context-sensitive text suggestions are generated using a transformer model equipped with an encoder for the text prompts and a decoder for text expansion. This process applies AI with token biased weights to achieve zero-shot, one-shot, or some-shot generation of context-sensitive suggestions. The generated suggestions, along with user-provided text, are then combined to form a document. The document structure may include outlines, figures, and various texts such as fiction, non-fiction, code, or patent applications. Techniques like token biasing, neural network weight biasing, and the use of generative pre-trained transformers (GPT) or BERT models are employed to enhance the generation process. Additionally, the method can generate figures, part lists, and even complete patent applications by analyzing and generating context-sensitive text for various document components, showcasing a comprehensive approach to automated document creation leveraging advanced AI technologies.

2.3. Inference Drawn:

- The patents primarily focus on singular aspects of content creation, lacking an integrated approach for combining text and visuals seamlessly.
- They offer limited user interactivity and customization, relying heavily on predefined templates and prompts for content generation.
- There's a gap in making advanced AI techniques accessible to non-technical users, requiring specialized knowledge to utilize and apply the described systems effectively.
- Patents describe static content generation methods without the capability for real-time learning and adaptation based on user feedback or evolving data.

2.4. Comparison with the Existing Systems:

Other System	Our System
Typically focuses on a single AI model or technology	Utilizes a combination of GenAI, Stable Diffusion, LLM-ChatGPT, and OpenCV.
May specialize in either text or image content only.	Generates both text and image ads from simple prompts.
Varies, often less adaptive to specific user inputs.	High personalization based on user prompts and context analysis.
Efficiency varies, often with more manual steps required.	Streamlines creation and revision processes.

Chapter 3: Requirement Gathering for the Proposed System

In this chapter we are going to discuss the resources we have used and how we analyzed what the user actually needs and what we can provide. We will also discuss the functional and non-functional requirements and finally the software and hardware used.

3.1. Introduction to Requirement Gathering:

The Requirement Gathering is a process of requirements discovery or generating list of requirements or collecting as many requirements as possible by end users. It is also called as requirements elicitation or requirement capture.

The requirements gathering process consists of six steps :

- Identify the relevant stakeholders
- Establish project goals and objectives
- Elicit requirements from stakeholders
- Document the requirements
- Confirm the requirements
- Prioritize the requirements

USE CASE	DESCRIPTION
Query input	Users input a query or prompt describing the type of advertisement they want to create.
Query Conversion	The system should convert user input queries into image queries suitable for processing by the LoRAmodel.
Query Processing	The system should process image queries using a trained model.
Caption Generation	The system should present the generated captions to users in an interactive and user-friendly interface.
Image Generation	The system should present the generated images to users in an interactive and user-friendly interface.
Merging images and captions	The system should allow users to interact with the presented images and captions, selecting their preferred combination.
Final ad generation	The system should merge the selected image with the chosen caption/tagline to create the final advertisement.

Table 3.1. Requirements of the system

3.2. Functional Requirements:

- User Input and Prompts: The system should allow users to input information such as the product, target audience, key features, and any specific requirements for the ad. Users should be able to provide prompts or guidelines for the AI to create tailored advertisements.
- Prompt Suggestions: The system should provide prompt suggestions to users based on the input data, including product characteristics, industry trends, and marketing best practices. Suggestions should be generated dynamically to guide users in crafting effective prompts.
- AI-Based Ad Generation: The AI system must be capable of generating company-specific advertisements using natural language generation (NLG) techniques. Ads should be coherent, engaging, and relevant to the user's input and prompts.
- Tagline Generation: The system should be able to suggest or generate catchy and memorable taglines for the product or campaign. Taglines should align with the overall theme and goals of the ad.
- User-Friendly UI: The user interface should be intuitive and user-friendly, making it easy for users to interact with the system. Users should have the ability to customize and preview the generated ads and taglines.

3.3. Non-Functional Requirements:

- Error Reduction: The AI system should employ advanced error-checking mechanisms to minimize grammatical, factual, or stylistic errors in the generated content. It should provide feedback on potential issues and allow users to review and edit the content.
- Scalability: The system should be designed to scale to handle a growing number of users and provide a responsive service, even during peak usage.
- Performance Metrics: Implement performance monitoring to track the success of generated ads, including metrics like click-through rates, conversion rates, and user engagement.
- Continuous Learning: Incorporate machine learning algorithms to allow the AI system to learn from user interactions and improve its ad and tagline generation capabilities over time.
- Accessibility: Ensure the system is accessible to users with disabilities, complying with accessibility standards and guidelines.
- Reliability: The system should be reliable and available 24/7, with minimal downtime and system failures.

3.4. Hardware, Software, Technology and Tools Utilised:

A. Hardware Requirements:-

Resources	Minimum	Maximum
CPU	2 x 1.8 GHz 32-bit (x86)	4 x 2.4 GHz 64-bit (x64)
RAM	4 GB	8GB
Disk space	3.5 GB for new installations, 5 BG for upgrades (including temporary files required during installations)	N/A
GPU	NVIDIA RTX 3080	GTX 360 NVIDIA GTX 1660 Ti

Table 3.2. Hardware Requirements

B. Software Requirements:-

Language : Python
IDE : Jupyter notebook,IDLE,Google Colaboratory, Visual Studio Code
Front-End Tech Stack : ReactJS, Gradio, HTML, CSS, JavaScript,Typescript
Backend Tech Stack : Python, Flask
Python Libraries : 1. Transformers 2. Diffusers 3.Stable Diffusion 4. OpenAI 5.GPT2 6.NeuralNet 7. Pandas 8. Numpy
Web Scraping Tools : Web Scraper, Tab Save, Image Downloader, Vertex AI

Techniques:-

- **Text-to-Image Synthesis:** Using AI models to generate images from textual descriptions, enabling the creation of customized advertisements.
- **Contextual and Personalized Content Generation:** Leveraging GPT-4 and custom-trained models to produce content that is relevant and tailored to specific audiences or prompts.
- **Efficient Model Training and Deployment:** Employing cutting-edge AI models and cloud resources to optimize the training process and enable rapid deployment of updates.
- **Interactive and User-Friendly Design:** Integrating user interfaces and APIs that allow non-technical users to interact with the system, customizing outputs to their preferences.

Technology and Tools:-

Automatic1111 WebUI, Visual Studio Code, JupyterLab, AWS Sagemaker, ngrok, FastAPI, GPT-4 are powerful tools and libraries that can be combined to generate company-specific ads using AI. Here's how each of these tools can be utilized in the process:

- **Automatic1111 WebUI:** A user-friendly web interface for Stable Diffusion, allowing easy access to advanced image generation features without needing deep technical knowledge.

- **VSCode (Visual Studio Code):** A powerful, open-source code editor by Microsoft that supports multiple programming languages and features such as debugging, version control, and extensions for enhanced functionality.

- **JupyterLab:** An interactive web application for creating and sharing documents that contain live code, equations, visualizations, and narrative text, widely used in data analysis, scientific research, and machine learning projects.

- **AWS SageMaker:** A fully managed service by Amazon Web Services that provides every developer and data scientist with the ability to build, train, and deploy machine learning models quickly.

- **ngrok:** A tool that creates a secure tunnel to your localhost, allowing you to expose your local development server to the internet for testing, webhooks, and demos.

- **FastAPI:** A modern, fast (high-performance) web framework for building APIs with Python 3.7+, based on standard Python type hints for creating RESTful APIs in a quick and easy way.

- **GPT-4:** The fourth iteration of the Generative Pre-trained Transformer by OpenAI, a state-of-the-art language model known for its ability to generate human-like text, answer questions, summarize content, and more, based on the input it receives.

3.5. Constraints:

- **User Input Validation:** Implement user input validation to prevent inappropriate or malicious prompts. Ensure that user prompts adhere to ethical standards and are relevant to the product or service.

- **Tagline Relevance:** The generated taglines should be relevant to the product or service and align with the company's branding. Implement a mechanism to reject or modify taglines that do not meet these criteria.

- **Resource Allocation:** Manage computational resources efficiently to balance the quality of generated ads with the cost of computation. Optimize resource allocation for scalability.

- **User Interface and Experience:** Provide a user-friendly interface for users to input prompts and review generated content. Consider user experience in the design and functionality of the application.

- **Training Data Diversity:** Ensure that the training data for the AI model is diverse and representative of the target audience to prevent bias in generated content.

Chapter 4: Proposed Design

4.1. Block Diagram of the proposed system:

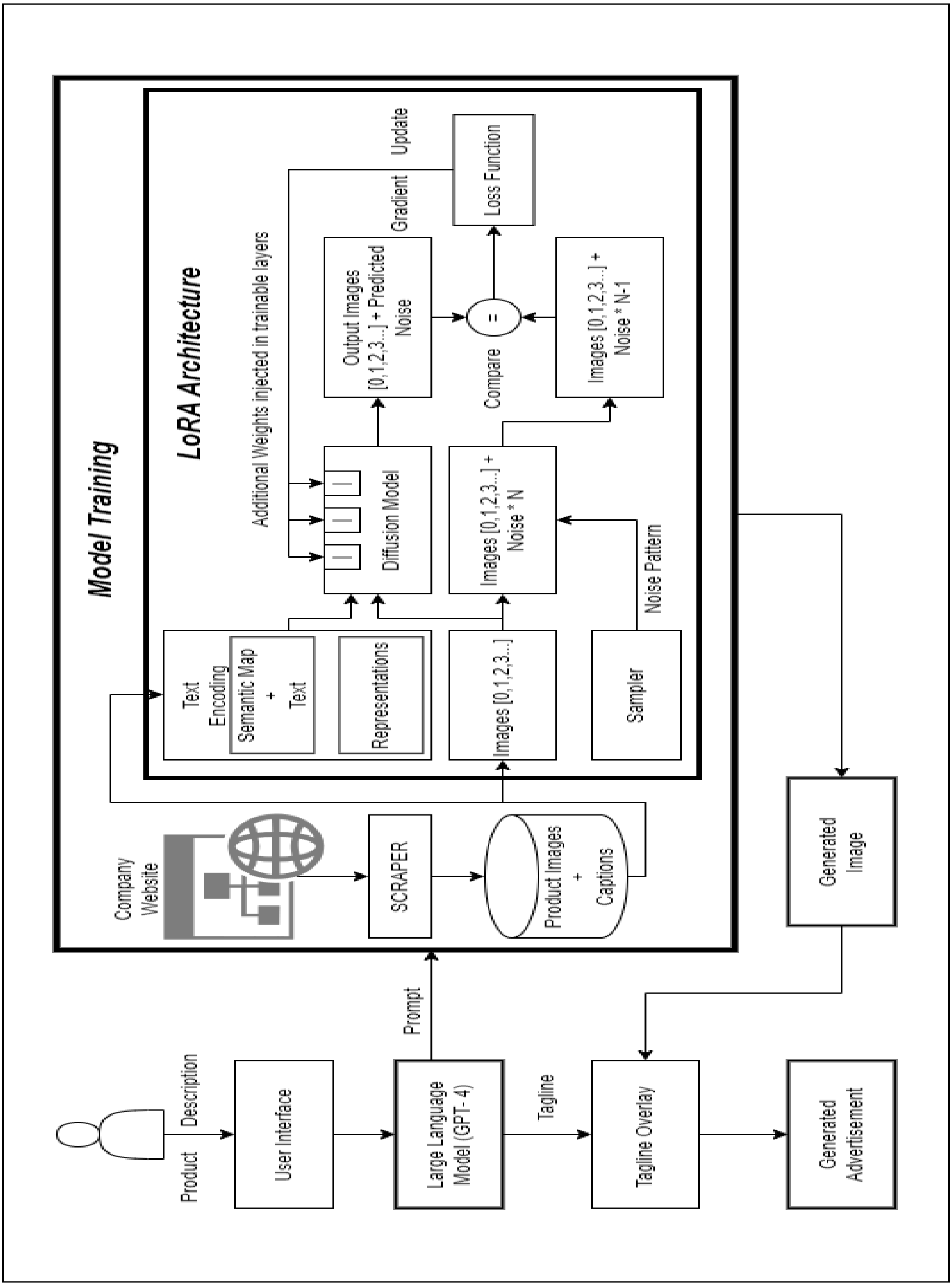


Fig 4.1. ADGenAI

4.2. Process diagram of the system:

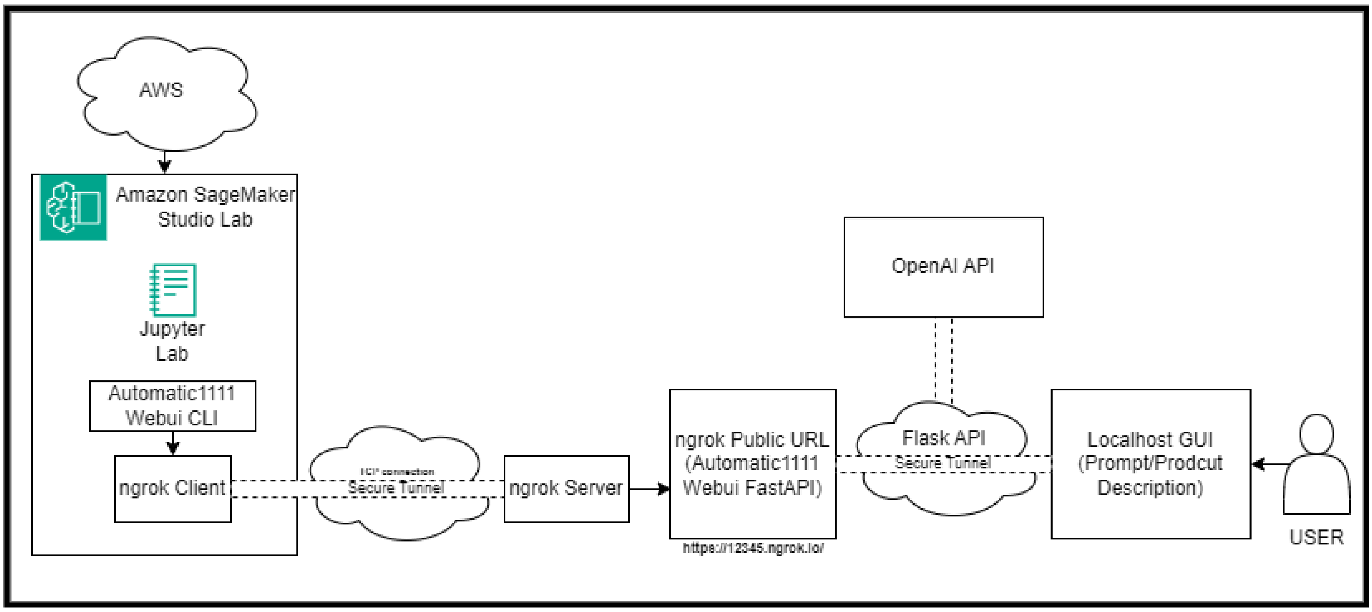


Fig 4.2. Process Diagram of ADGenAI

The product images of the target company are scraped from the "AquaPure" Company website for the purpose of research. These images, along with their corresponding captions, constitute the dataset utilized for LoRA training. The captions undergo transformation into text embeddings via established vectorization techniques such Word2VEC, BERT and GloVE. Subsequently, employing predetermined hyperparameters, the entire image set is subjected to training by introducing a specified level of noise, denoted as N. The resulting noisy images are then inputted into a pre-trained diffusion model, tasked with predicting the noise of N-1 images within the set. A comparative analysis is then conducted between the predicted noise and the actual noise of N-1 images, whereby gradient weights are directed into the injectable layers of the LoRA model. Following this training process, the LoRA model is primed for use in prompting. A user-friendly graphical user interface (GUI) is devised to facilitate user interaction with the LoRA model for generating product images.

Detailed Design

DFD Level 0:

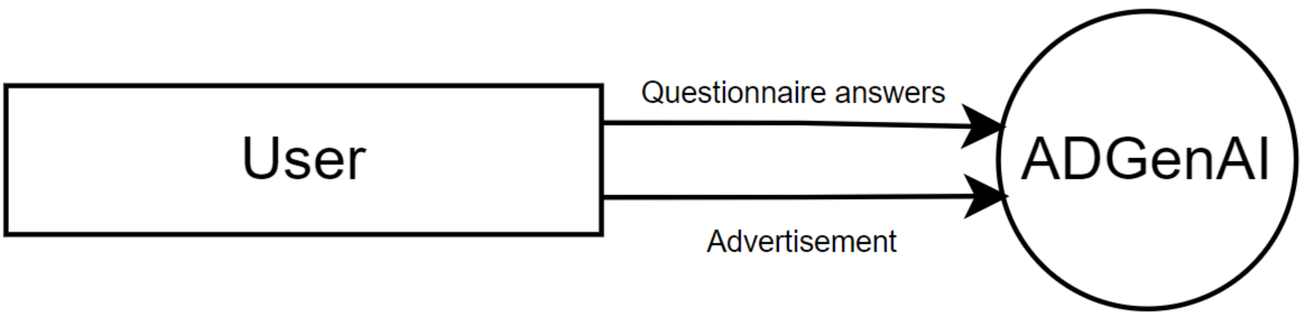


Fig 4.3. DFD Level 0

DFD Level 1:

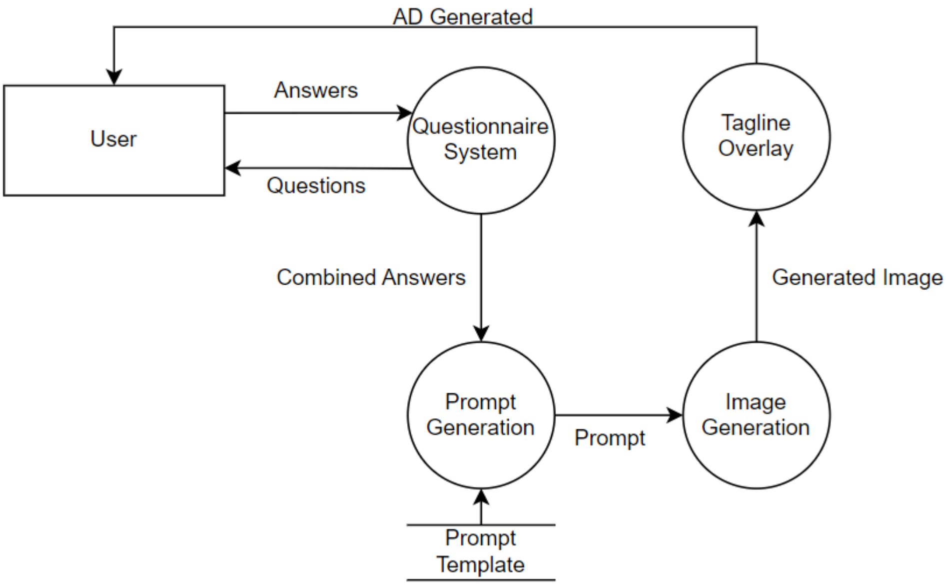


Fig 4.4. DFD Level 1

4.3. Flowchart for the proposed system :

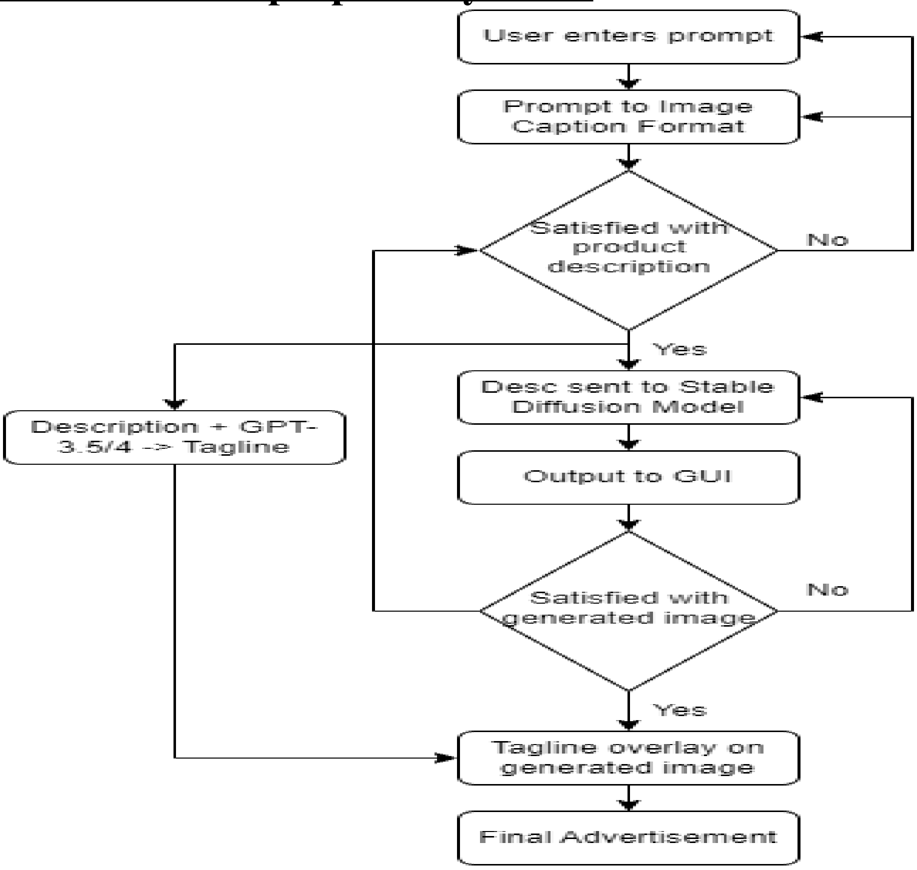


Fig 4.5. Flowchart of System

4.4. Project Scheduling & Tracking using Time line / Gantt Chart:

The Gantt chart of our project where we worked for the whole semester to create this model is shown in a timeline pattern. It is the most important part to think and design the planning of your topic and so we planned our work like the gantt chart shown.

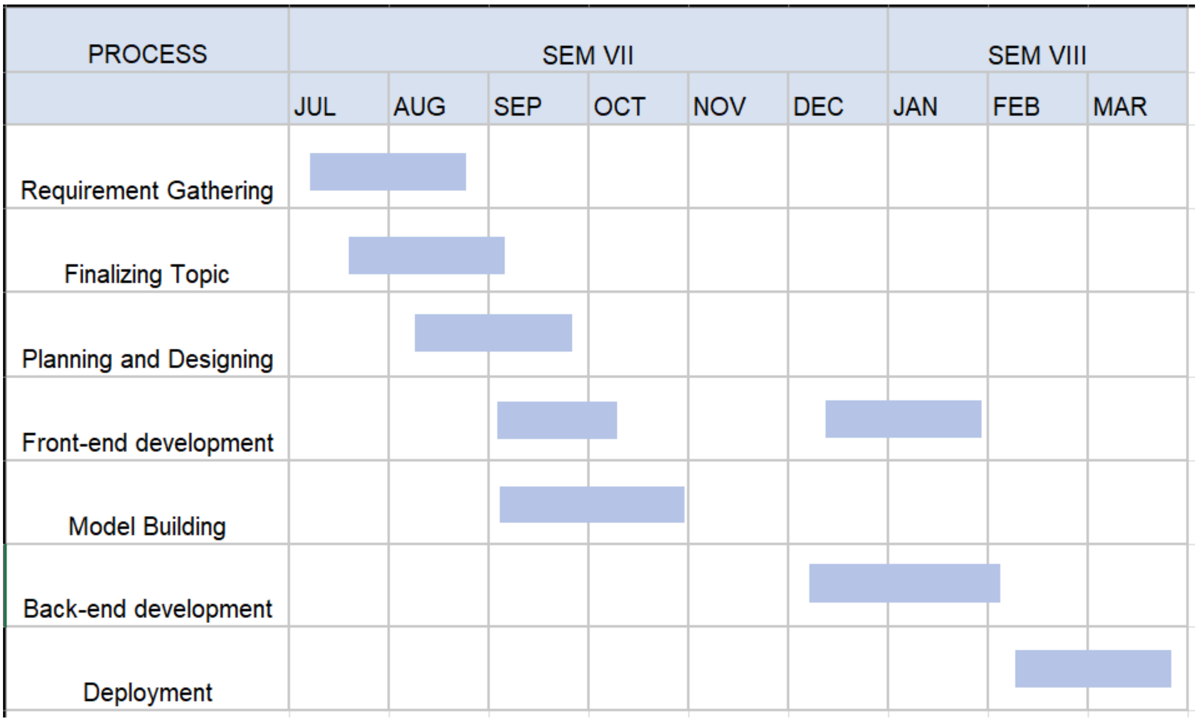


Fig 4.6. Gantt chart

Chapter 5: Implementation of the Proposed System

5.1. Methodology employed for development:

This project, AdGenAI, aims to redefine advertisement creation for businesses by leveraging the power of generative AI. With a focus on user-friendly design, it integrates a graphical user interface (GUI) to facilitate smooth operation and enhance creative workflows. The project's methodology is multifaceted, incorporating several key elements:

- **Evaluation of Text-to-Image Models:** It assesses advanced models like DALL-E and CLIP, alongside more accessible alternatives such as DeepDaze and LoRA, to identify their unique advantages and limitations. This comparison allows for a nuanced understanding of which models best serve the project's goals in the context of creative advertisement production.

- **Data Integration:** By combining proprietary and public datasets, the project ensures a rich source of input for image generation, maintaining data sensitivity and privacy. This approach underpins the creation of highly relevant and customized advertising content.

- **Optimization of Model Training:** Detailed captioning and the use of activation tags are explored for their efficacy in enhancing image accuracy and relevancy. The project delves into optimizing various parameters, including learning rates, batch sizes, and image resolution, to refine model performance and output quality.

- **Implementation of Additional Networks:** Advanced networks such as ControlNET and multiple LoRA models are incorporated, providing precision in image generation. These networks enable targeted modifications and refinements in the generated images, offering unparalleled control over the creative process.

- **Integration of Textual and Visual Elements:** Recognizing the impact of cohesive verbal and visual messages in advertisements, the project employs GPT-4 for generating engaging taglines. These are seamlessly integrated into the visual content using OpenCV, ensuring that advertisements not only capture attention but also communicate effectively with the target audience.

- **Comprehensive Analysis and Innovation:** The project not only showcases the capabilities of generative AI in transforming advertisement creation but also proposes innovative strategies for marketing. Through a detailed exploration of technology's potential, AdGenAI is positioned as a pioneer in advancing business growth and engaging consumers more effectively.

5.2. Algorithms and Flowcharts for the respective modules developed:

Pseudocode of Stable Diffusion

```
function text_to_image(text_description):
   # Step 1: Encode the text description into a feature vector
   text_features = encode_text(text_description)
   # Step 2: Initialize a random noise image
   noise_image = generate_random_noise()
   # Step 3: Conditionally refine the noise image using the text features
   for step in range(number_of_refinement_steps):
        # Combine the text features with the current state of the image
        combined_features = combine_features(text_features, noise_image)
        # Use a neural network to predict adjustments to make image closer to the
description
        adjustment = predict_adjustment(combined_features)
        # Apply the adjustment to the noise image
        noise_image = apply_adjustment(noise_image, adjustment)
 # Step 4: Finalize the image to ensure it's coherent and matches the text description
   final_image = finalize_image(noise_image)
        return final_image
# Helper functions (highly simplified)
# Converts text to a high-dimensional vector using a text encoder (e.g., CLIP)
function encode_text(text):
        return text_vector
# Generates an initial random image
function generate_random_noise():
  return noise_image
# Combines text features with image features, possibly using attention mechanisms
function combine_features(text_features, image):
return combined_features
# Uses a neural network (e.g., a diffusion model) to predict how to adjust the image
function predict_adjustment(features):
        return adjustment
# Applies the predicted adjustments to the image
function apply_adjustment(image, adjustment):
return updated_image
# Processes the image to ensure quality and coherence
function finalize_image(image):
        return final_image
```

Pseudocode of ControlNet

```
function text_to_image_with_control(text_description, control_signals):
    # Step 1: Encode the text description into a feature vector
    text_features = encode_text(text_description)
    # Step 2: Encode control signals into a control feature vector
    control_features = encode_control_signals(control_signals)
    # Step 3: Initialize a random noise image
    noise_image = generate_random_noise()
    # Step 4: Conditionally refine the noise image using both text and control features
    for step in range(number_of_refinement_steps):
        # Combine the text features, control features, and the current state of the image
        combined_features = combine_features(text_features, control_features, noise_image)
        # Use a neural network to predict adjustments to make the image closer to the
description and control signals
            adjustment = predict_adjustment(combined_features)
          # Apply the adjustment to the noise image
            noise_image = apply_adjustment(noise_image, adjustment)
   # Step 5: Finalize the image to ensure it's coherent and matches the text description
and adheres to control signals
    final_image = finalize_image(noise_image)
    return final_image
# Helper functions (highly simplified and conceptual)
# Converts text to a high-dimensional vector using a text encoder
function encode_text(text):
     return text_vector
# Converts control signals into a format that can be used to guide the generation
process
function encode_control_signals(control_signals):
       return control_vector
# Generates an initial random image
function generate_random_noise():
        return noise_image
# Combines text features, control features, and image features
function combine_features(text_features, control_features, image):
      return combined_features
# Uses a neural network to predict how to adjust the image based on combined
features
function predict_adjustment(features):
    return adjustment
# Applies the predicted adjustments to the image
function apply_adjustment(image, adjustment):
        return updated_image
 # Processes the image to ensure quality, coherence, and adherence to control signals
function finalize_image(image):
     return final_image
```

5.3. Datasets source and utilisation:

To commence the development of a LoRA, it necessitates access to proprietary images showcasing the enterprise's intellectual property. LoRA presents additional advantages as it leverages publicly available models trained on open datasets. This offers an avenue for incorporating Generative AI methodologies within company frameworks without divulging proprietary data to the public domain. Rather, utilizing a private LoRA model enables the integration of company-specific styles into generalized models, facilitating the utilization of image generation technologies.

Captions assume a pivotal role within text-to-image generation algorithms, serving as conduits for the conversion of textual or semantic representations into visual or pixel-based renderings. These captions not only delineate elements within the images but also encapsulate the primary subjects depicted. Each image within the dataset mandates meticulous captioning, underscoring the necessity for captions to exhibit a level of comprehension readily interpretable by AI models.

- Emphasize high-quality input to yield commensurate output.
- Prioritize quantity and diversity to enrich training efficacy.
- Quality supersedes quantity when compelled to choose.
- Exercise caution with upscaling, resorting to it only as a last recourse to preserve dataset integrity.

Images

To initiate the data preparation phase, establish a distinct thematic framework or subject matter for the model. Subsequently, curate a collection of 10 to 15 high-resolution images that align closely with the chosen theme. It is imperative that these images possess clarity and relevance, while also demonstrating diversity in content. Augmenting the dataset to encompass a range of 50 to 100 images can notably augment the efficacy of a LoRA model, particularly when aiming to emulate a specific aesthetic or thematic essence. It is essential to adhere to a total file size of less than 5GB for the compiled images.In order to mitigate the risk of overfitting, wherein the model excessively tailors its predictions to the idiosyncrasies of the training dataset, it is advisable to constrain the selection within this specified range. Furthermore, emphasis should be placed on ensuring that the images exhibit a level of dissimilarity, thereby fostering diversity within the dataset. This diversity serves to prevent the model from over-specializing on particular image attributes, thus promoting more robust and generalized learning outcomes.

Captions

Another key consideration in crafting captions is the use of rare tokens. These tokens consist of letter sequences that Stable Diffusion doesn't already associate with any particular concepts, such as "skw" or "ukj," or other random combinations of letters. Essentially, they serve as blank slates that can be linked to your specific concept during the training process. They're especially useful if you're working with something unique and you want to avoid any interference from Stable Diffusion's existing knowledge. For instance, if you're training a model on yourself or on a particular art style that Stable Diffusion isn't familiar with, using a rare token to describe those concepts is advisable. So, in such cases, you might use captions like "Photo of skw man" or "Drawing in skw style."

This framework serves as a scaffold for effective captioning:

- Globals: Incorporate rare tokens or identifiers pertinent to the training objective.
- Type/Perspective/Of a: Provide contextual descriptors encompassing medium, perspective, and subject characterization.
- Action Words: Enumerate dynamic actions or states relevant to the main subject.
- Subject Descriptions: Delve into granular details concerning the subject's attributes.
- Notable Details: Highlight unique elements warranting distinction.
- Background/Location: Furnish comprehensive depictions of the image milieu.
- Loose Associations: Integrate subjective interpretations or emotional nuances

Chapter 6: Testing of the Proposed System

6.1. Introduction to Testing :

Testing is a verification process for quality assessment and improvement. Testing is basically done to find errors, faults in the system. The basic goal of the software development process is to produce software that has very few or no errors. In an effort to detect errors soon after they are introduced each phase ends with verification activity such as reviews. However most of these verification activities in the early phase of the software development are based on human evaluation and cannot detect all the errors. Testing plays an important role in quality assurance for the software. It is a dynamic method for the verification and validation, where the system to be tested, executed and the behavior of the system is observed.

6.2. Types of tests Considered:

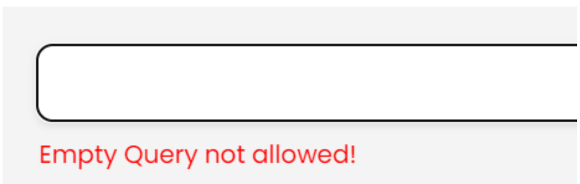
A) Pre testing phase

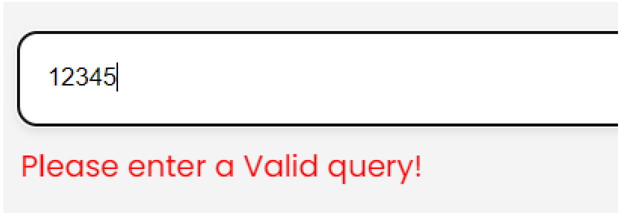
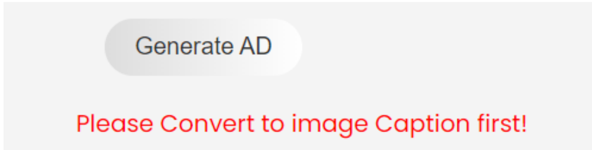
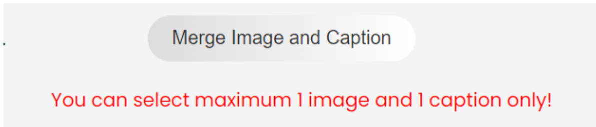
Pre testing is a type of acceptance testing; performed to identify all possible issues/bugs before releasing the product to everyday users or public. The focus of this testing is to simulate real users by using blackbox and whitebox techniques. The aim is to carry out the tasks that a typical user might perform. Alpha testing is carried out in a lab environment and usually the testers are internal employees of the organization. To put it as simply as possible, this kind of testing is called alpha only because it is done early on, near the end of the development of the software, and before beta testing.

B) Beta-Testing Phase

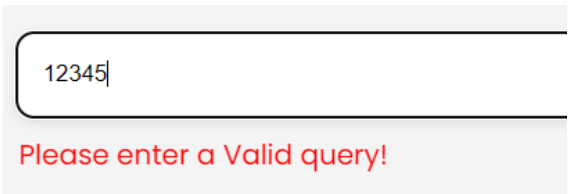
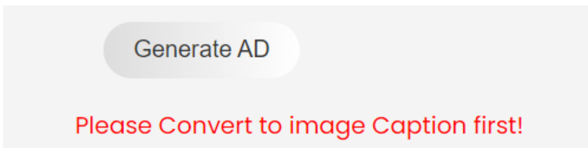
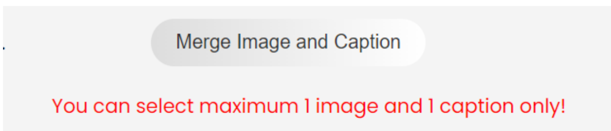
Beta Testing of a product is performed by "real users" of the software application in a "real environment" and can be considered as a form of external user acceptance testing. Beta version of the software is released to a limited number of end-users of the product to obtain feedback on the product quality. Beta testing reduces product failure risks and provides increased quality of the product through customer validation. It is the final test before shipping a product to the customers. Direct feedback from customers is a major advantage of Beta Testing. This testing helps to test the product in a real time environment. Because of time management and extensive GPU requirements of the project, testing was done with a group of reviewers only.

6.3. Various test case scenarios considered:

	Test Cases
Case 1: User cannot enter an empty query	

Case 2: User cannot enter number as a query	
Case 3: User should convert query into image caption before generating ad	
Case 4: User should select only one image and caption combination	

6.4. Inference drawn from the test cases:

	Test Cases
Case 1: User cannot enter an empty query. User should enter at least something.	
Case 2: User cannot enter number as a query. The prompt should not only just consist of numbers. It can contain both alphabets and numbers.	
Case 3: User should convert query into image caption before generating ad. Image caption is the format that is acceptable by the model.	
Case 4: User should select only one image and caption combination. More than two images or taglines should not be selected for generating an ad.	

Chapter 7: Results and Discussions

7.1. Screenshot of User Interface(UI) for the system:

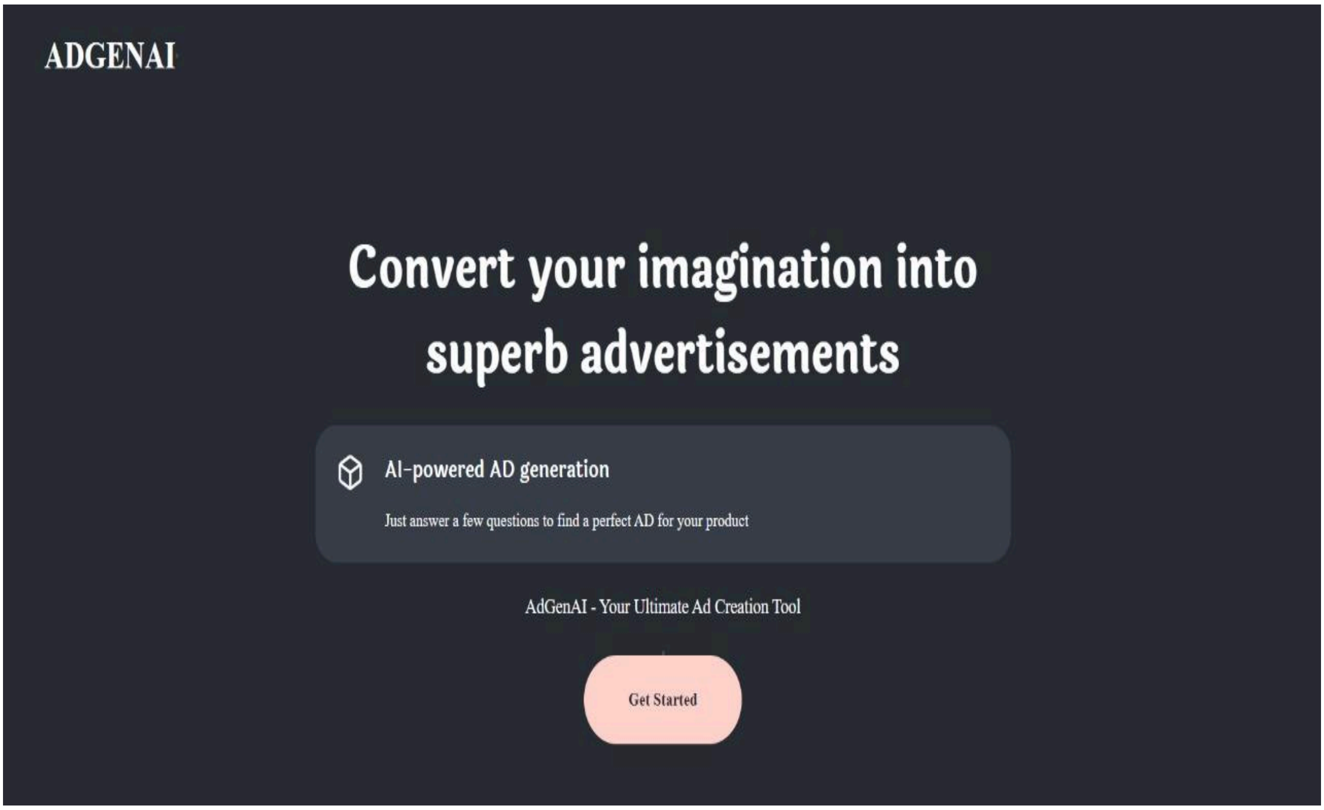


Fig. 7.1 Landing page of ADGenAI

Fig. 7.2. Questionnaire to create a prompt based on user response (Example: Question 1)

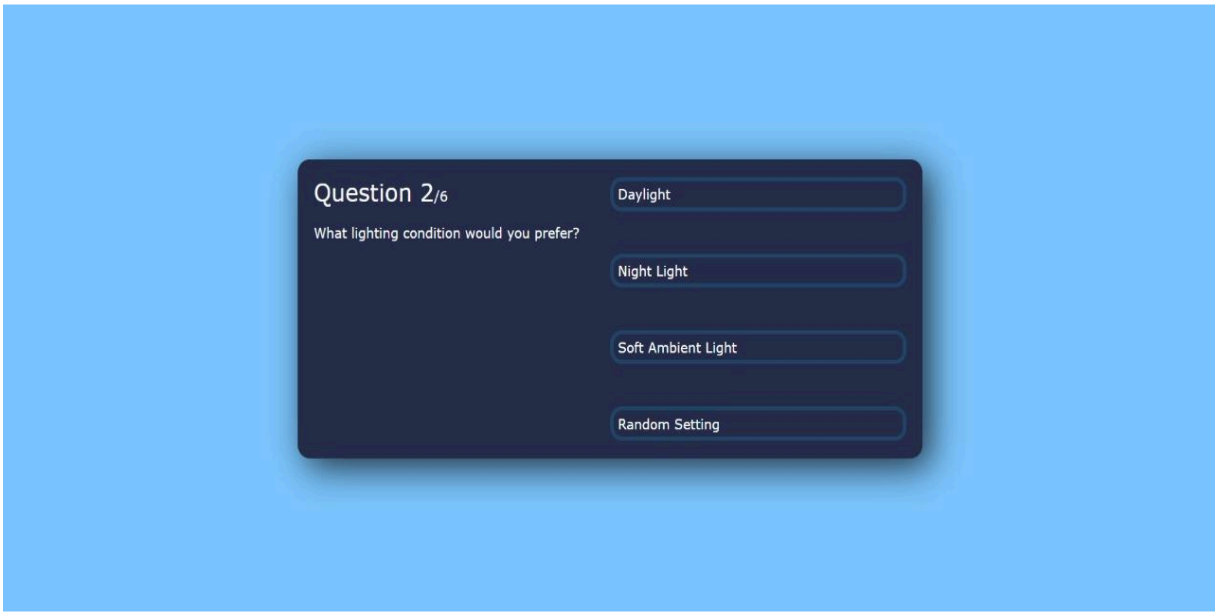


Fig.7.3. Questionnaire to create a prompt based on user response (Example: Question 2)

Fig. 7.4. Combining tagline and generated image using background detection techniques

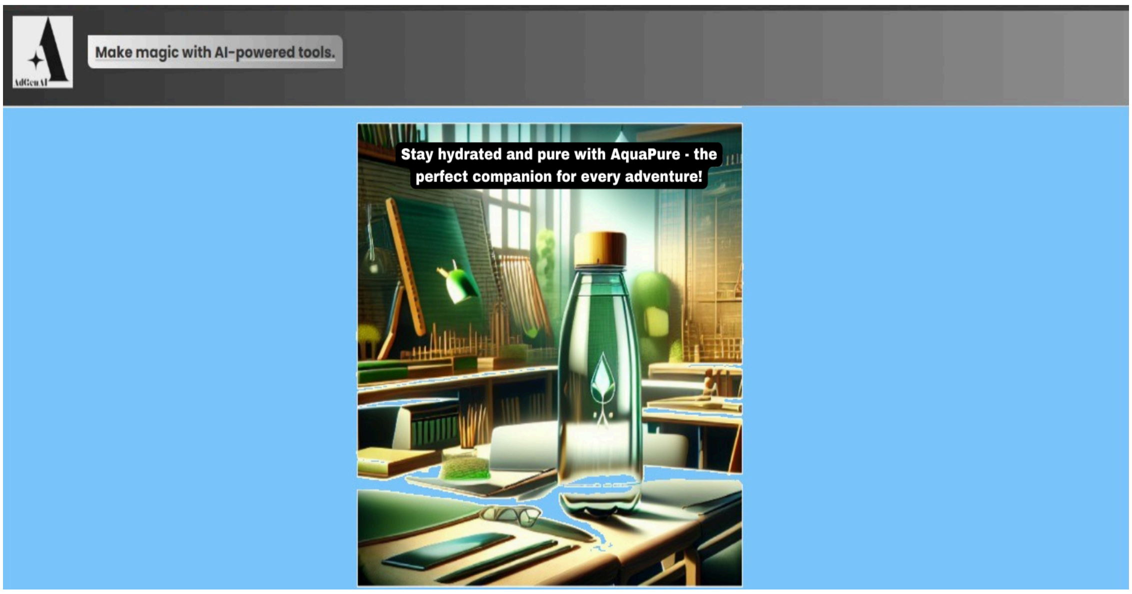


Fig. 7.5. Final Advertisement

7.2. Performance Evaluation Measures:

Qualitative Evaluation Methods

1. **Visual Inspection:** The subjective assessment by humans to judge the realism and quality of generated images is a straightforward method, albeit lacking in consistency and scalability.

2. **User Studies:** Structured surveys or experiments involving human participants provide a systematic approach to rate or compare images based on quality and realism. User studies facilitate a deeper understanding of human perceptual criteria in evaluating generated images.

Quantitative Evaluation Metrics

1. **Inception Score (IS):** The Inception Score measures both the diversity and quality of generated images, utilizing a pre-trained Inception model. High IS values indicate a model's ability to generate diverse, high-quality images recognizable by the Inception network.[21]

2. **Fréchet Inception Distance (FID):** FID assesses the similarity between the distributions of generated and real images in the feature space of an Inception-v3 network. Lower FID values indicate better quality and similarity to the target dataset, making it a widely accepted metric for evaluating image generation models.[22]

3. **Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR):** SSIM and PSNR evaluate the visual similarity between two images, often used in tasks like super-resolution or style transfer. These metrics are less relevant for generative models that create new images from scratch.

4. **Classifier Two-Sample Tests (C2ST):** C2ST measures the realism of generated images by assessing a classifier's ability to distinguish between real and generated images. Difficulty in distinguishing suggests higher realism of the generated images.[23]

7.3. Input Parameters/Features considered:

For Image Generation:

Our project implements a questionnaire system to gather user preferences for image generation, streamlining the input process for our generative AI model. By presenting structured questions, users can easily specify their needs and guide the image creation process effectively.

Users will provide general prompts specifying the type of image they want to generate, and our system will process and convert these prompts into actionable inputs for the model.

For AD Creation:

Our project aims to implement a tagline selection feature, where users can choose from a curated list of taglines generated alongside the images.

Our project aims to generate multiple images in response to user prompts, providing users with a selection to choose from. By offering diverse options, we enhance user satisfaction and creativity in the image selection process.

7.4. Comparison of Results with Existing System:

Before the advent of GenAI's success, the majority of existing systems for designing advertisements were limited to CAD software or created manually by machines or by hand.

Aspect	ADGenAI Images	Existing Systems' Images
Visual Quality	High-quality, detailed images	Quality can vary; may lack detail
Context Relevance	Highly relevant to the given prompts	May not always align closely with prompts
Creativity	Demonstrates unique, creative outputs	Creativity varies; often template-based
Personalization	Less tailored, with more generic outputs	Tailored specifically to user requests
Diversity	Wide range of styles and subjects	Limited range due to model or data constraints
Speed	Fast generation, even with complex prompts	Speed can be slower or vary with complexity
Adaptability	Static, with fixed capabilities over time	Continuously learns for improved results

7.5. Inference Drawn:

The comparison between ADGenAI and existing systems reveals that ADGenAI significantly outperforms in generating images that are not only of higher visual quality and detail but also more closely tailored to user inputs, showcasing its advanced capability to produce highly relevant and personalized content. Its outputs demonstrate a broader range of creativity and diversity, reflecting the system's flexibility and adaptability to various styles and contexts. This superiority underscores ADGenAI's enhanced understanding and application of generative AI technologies, positioning it as a more effective tool for creating engaging and context-appropriate visual content. This performance edge highlights ADGenAI's potential to revolutionize content creation by offering more dynamic, user-centered, and visually compelling solutions compared to those currently available.

Chapter 8: Conclusion

8.1. Limitations:

- Hardware and Cost Requirements: The reliance on sophisticated AI models like Stable Diffusion, LLM-ChatGPT, and others necessitates substantial computational resources. For small businesses or individuals, accessing the required GPUs and cloud services (e.g., AWS SageMaker) may present financial and logistical challenges.

- Learning Curve: Despite efforts to make the system user-friendly, the complexity of underlying technologies and the interface (Automatic1111 WebUI, Jupyter Lab) might still pose a learning curve for users without a technical background.

- Data Privacy and Security: Generating advertisements based on user inputs and data analytics raises concerns about data privacy and security. Ensuring the confidentiality and integrity of user data, especially when using cloud-based platforms like AWS, requires stringent security measures.

- Model Bias and Ethical Considerations: AI models are trained on large datasets that may contain biases, leading to skewed or insensitive content generation. Addressing these biases and ensuring the ethical use of AI in advertising content creation is a continuous challenge.

- Content Originality and Copyright: While AI can generate highly personalized and context-relevant content, there's a risk of producing content that infringes on existing copyrights or lacks originality, potentially leading to legal issues or reduced effectiveness in advertising campaigns.

- Real-time Adaptation Limits: Although ADGenAI is designed to learn and adapt from user feedback and data, the speed and efficiency of these adaptations can be limited by the underlying AI model's capabilities and the computational resources available.

- Dependency on External Platforms: The reliance on external platforms like AWS SageMaker for computational resources and ngrok for secure access can introduce dependencies on third-party services, affecting reliability and control over the system.

8.2. Conclusion:

The system introduced in this research marks a transformative approach to product-focused corporate advertising by integrating generative artificial intelligence and technologies such as Stable Diffusion (LoRA), ControlNET, Additional Networks, LLM-ChatGPT API, and OpenCV. This development offers a streamlined solution for ad production, blending automation with customization to support creative professionals in generating high-quality advertisements efficiently, while preserving brand integrity and campaign objectives.

The system's innovative architecture and functionalities demonstrate its practicality and potential to shift market dynamics by enhancing the creative capacity of its users, enabling businesses to differentiate in a competitive market and respond to changing consumer preferences.

Future prospects for this system include continued innovation in advertising strategies, enabling businesses to produce impactful, emotionally resonant ads. As technology advances, this system is set to lead the evolution of digital advertising creation and engagement.

8.3. Future Scope:

The future of ADGenAI is ripe with opportunities for growth and innovation, poised to further revolutionize AI-driven advertising through advancements in AI models, enhanced personalization, and broader market and format expansion. Key areas for development include making the platform more accessible to non-technical users, addressing ethical and bias considerations in AI, leveraging sustainable computing solutions, and ensuring content originality and copyright compliance. Additionally, exploring real-time adaptation based on user feedback, expanding into new creative industries, and adapting the system for global markets promise to enhance the effectiveness and reach of ADGenAI. These future directions not only aim to overcome current limitations but also to explore new horizons in creating engaging, personalized, and culturally relevant content across diverse platforms.

References

1. John Ford, Varsha Jain, Ketan Wadhwani, Damini Goyal Gupta, ”AI advertising: An overview and guidelines” - Journal of Business Research, Volume 166,2023,114124,ISSN 0148-2963
2. Liuqing Chen, Pan Wang, Hao Dong, Feng Shi, Ji Han, Yike Guo, Peter R.N. Childs, JunXiao, Chao Wu, “An artificial intelligence based data-driven approach for design ideation, Journal of Visual Communication and Image Representation” - Volume 61, 2019, Pages 10-22, ISSN 1047-3203
3. Axel Sauer, Tero Karras, Samuli Laine, Andreas Geiger, Timo Aila, “StyleGAN-T: Unlocking the Power of GANs for Fast Large-Scale Text-to-Image Synthesis” - Machine Learning (cs.LG); Computer Vision and Pattern Recognition (cs.CV), arXiv:2301.09515, 2023
4. Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, In So Kweon, “Text-to-image Diffusion Models in Generative AI: A Survey” - Computer Vision and Pattern Recognition (cs.CV); Artificial Intelligence (cs.AI); Machine Learning (cs.LG), arXiv:2303.07909, 2023
5. Prafulla Dhariwal, Alex Nichol, “Diffusion Models Beat GANs on Image Synthesis” - 35th Conference on Neural Information Processing Systems (NeurIPS 2021)
6. R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 10674-10685, doi: 10.1109/CVPR52688.2022.01042.
7. Lvmin Zhang, Anyi Rao, and Maneesh Agrawala, “Adding Conditional Control to Text-to-Image Diffusion Models”, - Computer Vision and Pattern Recognition (cs.CV); Artificial Intelligence (cs.AI); Graphics (cs.GR); Human-Computer Interaction (cs.HC); Multimedia (cs.MM), arXiv:2302.05543, 2023
8. Borji, Ali. (2022). Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. 10.48550/arXiv.2210.00586
9. AUTOMATIC1111, "Stable Diffusion Web UI", 2022-08-22, https://github.com/AUTOMATIC1111/stable-diffusion-webui
10. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In the International conference on machine learning (pp. 8748-8763). PMLR.
11. Bie, F., Yang, Y., Zhou, Z., Ghanem, A., Zhang, M., Yao, Z., Wu, X., Holmes, C., Golnari, P.A., Clifton, D.A., He, Y., Tao, D., & Song, S.L. (2023). RenAIssance: A Survey into AI Text-to-Image Generation in the Era of Large Model. ArXiv, abs/2309.00810.
12. [1] P. Wang, "deep-daze," 2021. [Online]. Available: https://github.com/lucidrains/deep-daze. [Accessed: 10- 3-2024].
13. Karras, T., Laine, S., & Aila, T. (2018). A Style-Based Generator Architecture for Generative Adversarial Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4396-4405.
14. Zhang, Han et al. “StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks.” 2017 IEEE International Conference on Computer Vision (ICCV) (2016): 5908-5916.
15. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation
16. Hu, J.E., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., & Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models. ArXiv, abs/2106.09685.
17. "Stable Diffusion LoRa," AITuts, 2023. [Online]. Available: https://aituts.com/stable-diffusion-lora/. [Accessed: 10- 3-2024].

18. Hang, T., Gu, S., Li, C., Bao, J., Chen, D., Hu, H., Geng, X., & Guo, B. (2023). Efficient Diffusion Training via Min-SNR Weighting Strategy. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 7407-7417.
19.Lykon, "AnyLoRA," Hugging Face Model Repository, 2023. [Online]. Available: https://huggingface.co/Lykon/AnyLoRA. [Accessed: 10- 3-2024].
20. Junius, "Comparing LoRa Types in the Kohya SS GUI," AsheJunius.com, Mar. 10, 2022. [Online]. Available: https://ashejunius.com/comparing-lora-types-in-the-kohya-ss-gui-bfc8ccda7cf1. [Accessed: 10- 3-2024].
21. Song, Jiaming et al. “Denoising Diffusion Implicit Models.” ArXiv abs/2010.02502 (2020): n. Pag.
22. Karras, Tero et al. “Elucidating the Design Space of Diffusion-Based Generative Models.” ArXiv abs/2206.00364 (2022): n. Pag.
23. Lu, Cheng et al. “DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps.” ArXiv abs/2206.00927 (2022): n. Pag.
24. Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., & Zhu, J. (2022). DPM-Solver++: Fast Solver for Guided Sampling of Diffusion Probabilistic Models. ArXiv, abs/2211.01095.
25. Yang-Chu Lin, Yi-Chih Lee & Nu-Ting Lin (2014) The Effect of Advertising using Advertising Appeals, Pictures and Product Categories, Journal of Statistics and Management Systems, 17:1, 71-96, DOI: 10.1080/09720510.2013.867716
26. Zhang, Lvmin et al. “Adding Conditional Control to Text-to-Image Diffusion Models.” 2023 IEEE/CVF International Conference on Computer Vision (ICCV) (2023): 3813-3824.
27. kohya-ss, "sd-webui-additional-networks," 2023. [Online]. Available: https://github.com/kohya-ss/sd-webui-additional-networks. [Accessed: Day- Month- Year].

Appendix

1] Paper I details :-

Leveraging Text-to-Image Generation Models for Automated Creative Processes in Corporate Marketing

[1] Sharmila Sengupta, [2]Gauri Nagral, [3]Kritika Yadav, [4]Meera Sawantdesai, [5]Yashraj Mulwani,

[1]Associate Professor, [2]Undergrad Student, [3]Undergrad Student, [4]Undergrad Student, [5]Undergrad Student

Computer Engineering,

Vivekanand Education Society's Institute of Technology, Chembur, India

Abstract : ***In this day and age, advertisements play a huge role for businesses to flourish in competitive economies. It is said that an attractive ad is worth more than a thousand words. Creating an advertisement involves conceptualizing an idea, designing visual and textual content, and strategically placing it across media platforms to reach and influence the target audience. This entire process can take anywhere from a few hours to several months, depending on the complexity and medium of the advertisement. However, the tremendous growth in technology can turn those hours and months into just a few seconds. This paper explores the capabilities of Generative Artificial Intelligence (GenAI) for visual element generation through advanced techniques. It analyzes leading models like StackGAN, Stable Diffusion - DreamBooth, and LORA, identifying the most efficient among them. The research then focuses on implementing these models to automate advertisement creation for product-based companies, aiming to replace manual efforts in the creative department. The proposed system, ADGenAI offers insights into the transformative potential of GenAI in streamlining creative processes, fostering efficiency, and envisioning the future of the advertising industry. Specifically designed to alleviate the pressures faced by creative teams, such as the demand for high-quality content under tight deadlines, this system enhances creativity and productivity by automating the initial stages of ad production and enabling the exploration of new advertising concepts with minimal resource investment.***

IndexTerms - **GenAI, Stable Diffusion, GAN, LoRA, Automatic1111 Webui, ControlNet, Additional Networks**

I. Introduction

Marketing professionals are tasked with interpreting vast amounts of data to tailor advertising strategies effectively. The creative departments within an organization coordinate, execute, and supervise marketing initiatives using various frameworks such as functional, matrix, segment, and product centric approaches. The primary aim is to boost visibility and sales through tasks like formulating strategies, conducting market research, fostering innovation, advertising products, arranging events, and supporting sales teams. In educational resource creation, generative models generate interactive learning materials, personalized study guides, and virtual simulations to enhance the learning experience. In medical marketing, they develop visually engaging content to educate patients and healthcare professionals about new treatments. Additionally, in healthcare and vaccination campaigns, generative AI creates targeted messaging to promote public health initiatives and combat misinformation. These applications, alongside those in art, data-driven fields, natural language processing, and image synthesis, exemplify the extensive impact of generative AI across diverse sectors, fueling innovation while addressing societal and industry challenges. There is a trade-off between manual and automated processes in terms of precision, skill refinement, time and extensive labor. Achieving a balance in these processes is crucial for enhancing productivity and achieving desired results effectively. Digital technologies including programmatic and native advertising have revolutionized modern techniques, requiring marketers to adapt to evolving trends.

Generative AI is reshaping the advertising landscape by automating ad creation tasks and enabling the generation of personalized and targeted ad content at scale. Despite concerns about the quality of AI-generated content, personalized ads created through AI algorithms have shown to increase sales significantly.

The paper details the system's architecture and functionality, aiming to demonstrate its ability to streamline ad creation for creative professionals by offering ease of use, customization, efficiency, and the opportunity for innovation.

II. Literature Review

The frequently advancing landscape of GenAI has inspired several researchers to investigate their roles in several applications. The study available in paper by Ford et al. (2023) [1] highlights the growing importance of AI in advertising and its implications for both industry and academia. The descriptive data analysis reveals that while AI in advertising research has roots in the past few decades, interest in this domain has experienced noticeable growth. An integrated approach for enhancing design ideation using artificial intelligence and data mining techniques is presented in the study by Chen et al. (2019) [2]. The approach includes two models: a semantic ideation network and a visual concepts combination model. A case study demonstrates the effectiveness of these models in generating diverse cross-domain associations and expediting the ideation process. Unlike the slower iterative evaluation required by other models, StyleGAN-T offers faster generation without sacrificing sample quality. By focusing on lower resolutions, StyleGAN-T proves to be a competitive option [3] for large-scale text-to-image synthesis. The paper[4] provides an overview of text-to-image synthesis techniques, with a focus on diffusion models which have gained prominence in generative tasks. The survey delves into the advancements in text-conditioned image synthesis, exploring related tasks such as text-guided creative generation and image editing. Challenges and future directions in the field are also discussed, alongside considerations for benchmarks and ethics. The authors in paper [5] demonstrated that diffusion models achieve superior sample quality compared to state-of-the-art GANs. They introduce a classifier guidance technique for class-conditional tasks, allowing for an adjustment in classifier gradient scale to balance diversity and fidelity. While diffusion models still require multiple forward passes during sampling, the authors show that combining guidance with upsampling further improves sample quality, especially in high-resolution conditional image synthesis. The authors have introduced Latent Diffusion Models (LDMs) in paper [6] as an efficient approach to high-resolution image synthesis. This approach significantly reduces computational demands for both training and sampling, making it accessible to a wider range of researchers and users. They also introduce a general-purpose conditioning mechanism based on cross-attention, enabling multi-modal training for class-conditional, text-to-image, and layout-to-image models. Pretrained LDM and autoencoding models are made available for various tasks. A neural network architecture, ControlNet designed to enhance the spatial control of large, pretrained text-to-image diffusion models is discussed in paper [7]. ControlNet enables various conditioning controls, such as edges, depth, segmentation, human pose, and more, for Stable Diffusion models. It offers wider applications for controlling image diffusion models, including multi-modal conditions and composition of multiple conditions. In paper [8], a quantitative comparison of three popular generative models, namely Stable Diffusion, Midjourney, and DALL-E 2 is proposed to show their ability to generate photorealistic faces in diverse scenes. The goal is to evaluate these models' performance in generating realistic faces within complex, cluttered scenes.

This research paper presents a novel system that employs GenAI technologies, such as Stable Diffusion (including LoRA and ControlNET), LLM-ChatGPT API, and OpenCV, to innovate advertisement production for product-based companies. By unifying these technologies into an efficient pipeline, the system facilitates the creation of market-ready advertisements from simple text prompts.

III. Proposed System

It is built around a model trained on a diverse array of product images and captions, enabling easy interaction through prompts. Product images from the "AquaPure" company website were collected for research, forming a dataset with captions for training models with LoRA. These captions were converted into text embeddings using vectorization techniques such as Word2VEC, BERT, and GloVE. The dataset underwent training with specific hyperparameters and a defined noise level, N as shown in Fig 1. Noisy images were processed by a pre-trained diffusion model to predict the noise of N-1 images. This process allowed for a comparison between predicted and actual noise levels, adjusting the LoRA model's gradient weights accordingly. Post-training, the LoRA model supports enhanced prompting capabilities through a GUI developed to simplify user interactions with the model to generate tailored product images. A compelling tagline is an integral part of effective advertising. To achieve this, an LLM is employed to generate a tagline, which is subsequently superimposed onto the generated image.

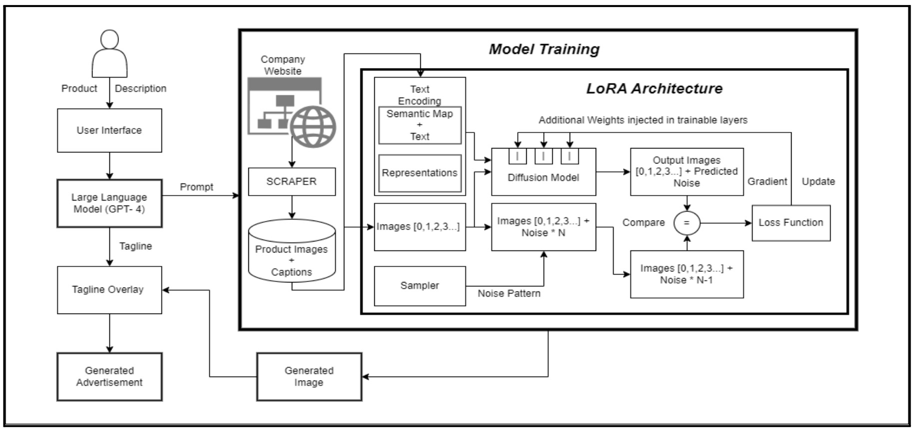


Fig.1 Architecture diagram of ADGenAI

Training and testing a LoRA model often requires more computational power than the 8GB Intel(R) Iris(R) XE Graphics Shared Memory. Therefore use of AWS SageMaker, particularly Amazon SageMaker Studio Lab, has been crucial for overcoming these limitations, especially for small businesses. It provides about 4 hours of daily runtime, 15 GB of persistent storage, and access to both CPU (T3.xlarge) and GPU (G4dn.xlarge) resources, along with enterprise-grade security and a user-friendly JupyterLab interface.

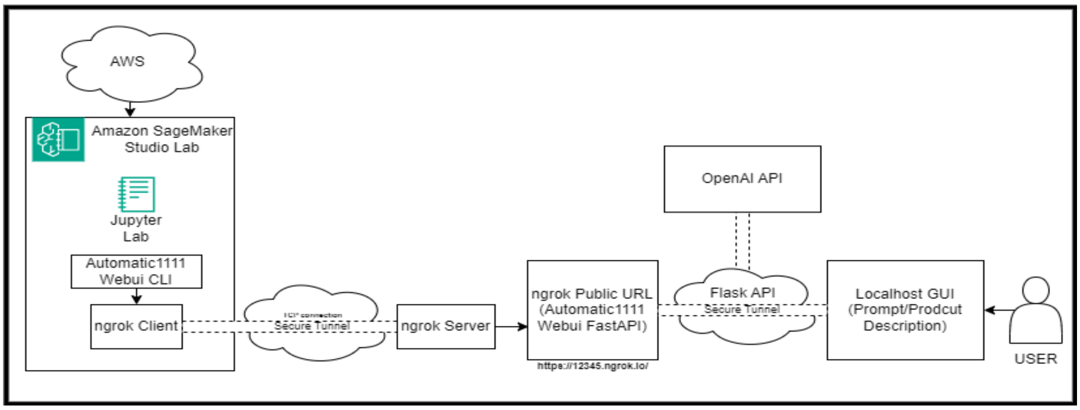


Fig.2 Process Diagram of the Proposed System

The Automatic1111 Webui as shown in Fig 2 enhances Stable Diffusion's usability with a web interface that simplifies initiating instances, fine-tuning, and navigating for improved user interaction, also enabling API integration. It supports various synthesis processes and includes extensive documentation for user guidance, promoting creativity and efficiency[9]. For secure external access, A1111 uses launch arguments for establishing a tunnel via ngrok to a public URL, with request processing via a FastAPI on the ngrok server. The OpenAI API further contributes to generating image captions and taglines, key for advertisement creation.

IV. Requirements

The recommended system requirements for the resource area are 2 x 1.8GHz 32-bit (x86) for CPU, with a recommended specification of 4 x 2.4GHz 64-bit (x64) and 4/8 GB RAM. For disk space, new installations require 3.5 GB, and upgrades need 5 GB, including the temporary files required during installation; there is no recommended specification provided for disk space. As for the GPU, the minimum requirement is an NVIDIA RTX 3080, and the recommended one is an NVIDIA GTX1660 Ti. The Software Requirements are ReactJS, Gradio, HTML, CSS, JavaScript,Typescript for Front-End Tech Stack; Transformers, Diffusers, Stable Diffusion, OpenAI, GPT2, NeuralNet, Pandas and Numpy for Backend Tech Stack and the Python Libraries used are Web Scraping Tools for example Web Scraper, Tab Save, Image Downloader, Vertex AI etc.

V. Methodology

This research paper on AdGenAI aims to develop a generative AI model focused on improving advertisement creation for business growth. This involves constructing a user-friendly graphical user interface (GUI) for seamless integration and operation, streamlining workflows for creative professionals for which the different processes for AdGenAI are explained below as in Fig. 3

A. Evaluation of Text-to-Image Generation Models

This section reviews text-to-image generation models critical for generative AI in creative fields. Highlighted models include DALL-E and DALL-E 2, noted for their large training datasets, and CLIP [10] and Imagen, emphasizing the expansive data they're built upon [11]. Due to the high costs of these advanced models, the focus shifts to accessible options like DeepDaze, DreamBooth, StyleGAN, LoRA, and SNGAN, each with unique capabilities and constraints. DeepDaze, inspired by Google DeepDream, combines CLIP and SIREN for image generation from text prompts but lacks the ability for external training, limiting customization, and achieves an Inception Score of 6.2 [12]. StyleGAN, known for its styling control, is limited in product image generation despite its success in human face datasets, marked by an FFHQ score of 4.40[13]. StackGAN [14] and DreamBooth offer solutions for photo-realism and model personalization, respectively, with DreamBooth achieving a notable Inception Score of 34.9102[15] but requiring significant storage for model updates. LoRA (Low- Rank Adaption) [16] stands out for its efficient fine-tuning and minimal computational cost, making it a preferred choice alongside Stable Diffusion for a wide range of product styles and concepts due to its lightweight and adaptable approach.

B. LoRA and creating the data

LoRA models are essential in AI, particularly for enhancing creative and design tasks with types including Character, Style, Concept, Pose, and Clothing. These models benefit from integrating proprietary and public datasets, allowing companies to leverage Generative AI without exposing sensitive data. Using a private LoRA model facilitates the blending of company-specific styles with general models, enhancing image generation capabilities

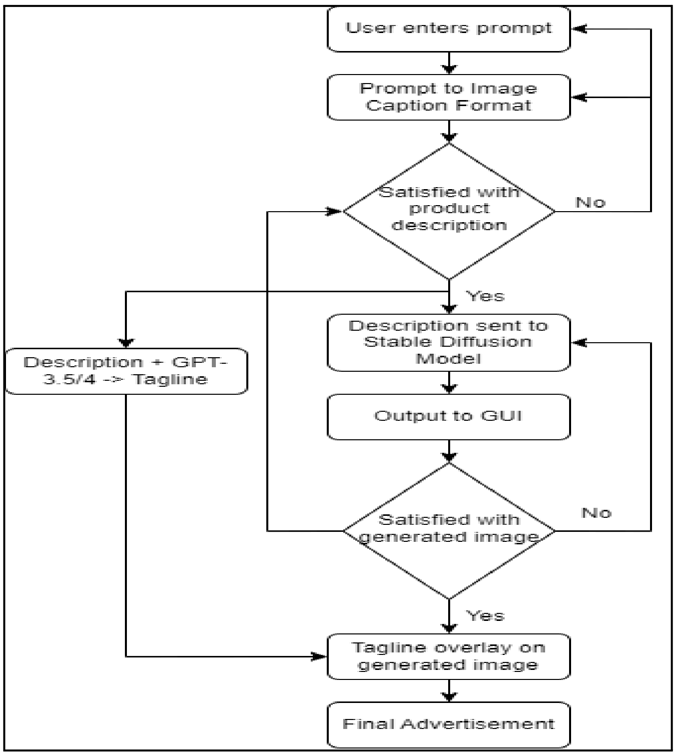


Fig.3 Flowchart of the Proposed System

C. The Dataset

Developing a LoRA model starts with assembling a proprietary image collection that represents the company's intellectual property, complemented by meticulous captioning to ensure AI interpretability. For images, the strategy is to start with 10 to 15 high-resolution images per theme, focusing on clarity, relevance, and diversity, and aim to expand the collection to 50 to 100 images to bolster the model's aesthetic range, keeping file sizes below 5GB, thus prioritizing diversity to avoid overfitting and ensure robust learning outcomes. The importance of captions are emphasized by incorporating rare tokens for unique concepts, and ensuring captions convey the intended subject matter effectively. Rare tokens like "skw" act as placeholders for specific concepts, aiding in minimizing pre-existing model biases.

D. Captioning Insights

The insights are high-quality, comprehensive captions are crucial for converting textual prompts into accurate visual representations, with balance in quality, quantity, and diversity in dataset curation to optimize training effectiveness and utilize rare tokens for novel or unique concepts to ensure clarity and avoid confusion with existing model knowledge. This can be visualized in Table.1.

D.1 Without Activation Tag

Prompt: <lora:bottle_no_activation_tag-20:1>glass bottle standing straight on ground, night new york times square background

Using concise captions without activation tags, the model accurately rendered the main object but struggled with complex environmental elements, such as lighting in night scenes. This underscores the model's object fidelity but highlights the need for advancements in capturing complex contexts.

	
Without Activation Tag	**With Activation Tag**

	
Short Captions	**Long Captions**

Table. 1 Impact of Captions on Image Generation

D.2. With Activation Tag

Prompt: <lora:bottle_with_activation_tag-20:1>glxbtz standing straight on ground, night new york times square background

Incorporating activation tags significantly improved the model's rendering of environmental details, like accurately depicting

D.3. With Short Length Captions

Prompt: <lora:bottle_with_activation_short_captions:1>glxbtz, glxbtz standing straight on wooden bench, autumn park background

Short captions with activation tags allowed the model to accurately interpret abstract cues, leading to precise object depiction and improved contextual detail. This showcases the effectiveness of activation tags in refining environmental understanding without affecting object accuracy.

D.4. With Long Length Captions

Prompt: <lora:bottle_with_activation_long_captions:1>glxbtz, glxbtz standing straight on wooden bench, autumn park background

Long captions with activation tags resulted in the overemphasis of detailed backgrounds at the expense of the main object's prominence. This suggests a need for balance in detailing objects and backgrounds, highlighting the importance of further research into strategies that ensure the object remains the focus in generated images.

E. Deep dive into LoRA and parameters

The hyperparameters used in training, as analyzed in Table 2, are number of epochs and repeats, the batch size, resolution of 512 pixels indicating the size of the images generated or processed and Shuffle tags to check randomizing the order of data, potentially improving generalization.

Table. 2 Basic LoRA parameters value guide

No. of Images	Repeats	Epochs	Batch Size
10	10	20	2
20	10	10	2
100	3	10	2
400	1	10	2

The UNet learning rate is 5e-4 and the text encoder learning rate is 1e-4. The Learning rate scheduler mostly used is "cosine_with_restarts" except for aq_train_6, which uses "constant". This parameter controls how the learning rate changes over time. The LR Scheduler Number is used to refer to the number of restarts for the cosine scheduler and LR Warmup Ratio of 0.05 indicates a portion of the training period during which the learning rate gradually increases to its initial value. The common parameters across all trained models are shown in Fig. 4

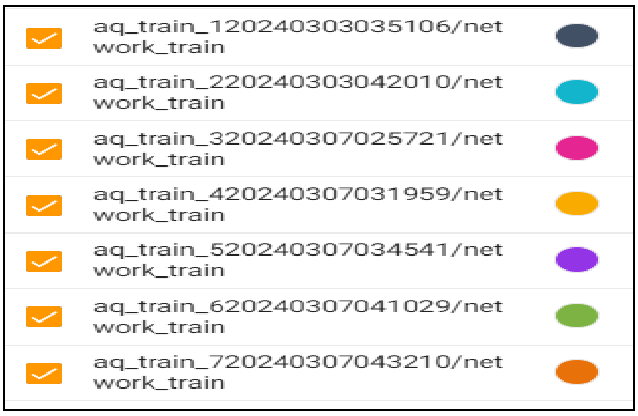


Fig.4 Color Coded Trained Models

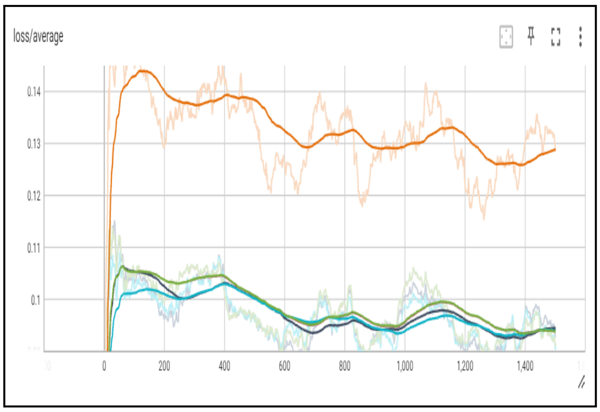


Fig.5 Loss/Average of each model

The Lora Parameters specify the type of Low-Rank Adaptation (LoRA) used, affecting how model weights are adjusted. Different configurations use "LoRA" or "LoCon", indicating different strategies for adapting pre-trained weights and Network Dim: 16, Network Alpha: 8, For LoCon: Conv Dim: 8, Conv Alpha: 4, parameters related to the dimensionality and adaptation rates in the LoRA and LoCon mechanism. Minimum SNR Gamma indicates whether a minimum signal-to-noise ratio (SNR) constraint is applied, aiming to improve model robustness. Most configurations specify "Yes," except aq_train_7, indicating an experiment with turning off this constraint. The Optimizer"none" is specified in most cases but "Prodigy" is mentioned for aq_train_3, for experimenting with a different optimization strategy.

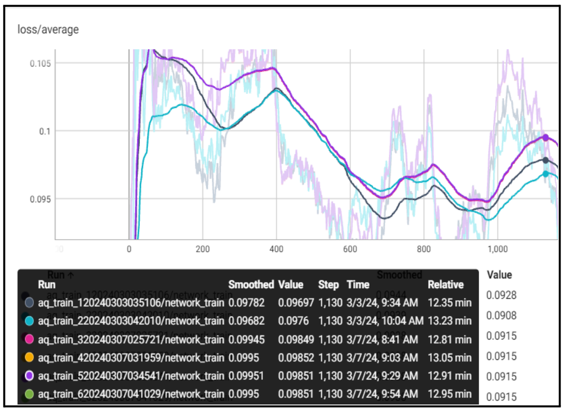


Fig. 6 Loss/Average of each model indicating convergence loss at convergence

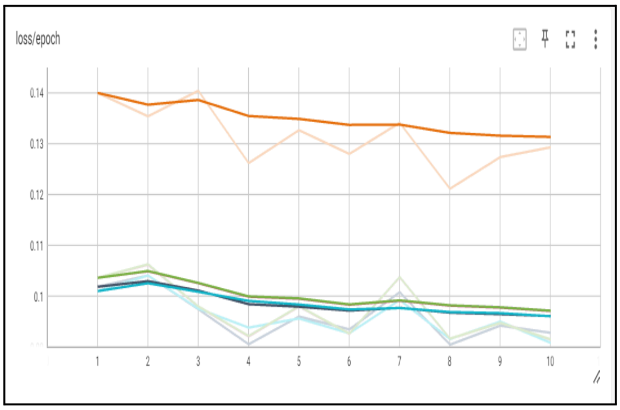


Fig.7 Loss/Epoch indicating model

As shown in Fig. 5, a downward trend indicates that the model is learning, i.e the difference between the predicted and actual values is decreasing over time. A flat line indicates that the model has stopped learning, either because it has converged (found the best solution it can) or because it is unable to improve further due to issues like being stuck in a local minimum. “aq_train7” has a very high spike of loss/average throughout training, indicating that the model is underfitting and has a very low performance. This is in line with the fact that the “min snr gamma” parameter is disabled during training of “aq_train_7”. Disabling Min-SNR-γ in denoising diffusion model training could lead to slower convergence and reduced image quality, as this strategy dynamically optimizes the training process by adjusting loss weights based on signal-to-noise ratios, enhancing learning efficiency and output fidelity. Without it, the model may not efficiently navigate the complexity of the diffusion process, affecting performance. The rest of the models other than “aq_train_7” as indicated by Fig. 6 have almost the same learning curve with “aq_train_5” whose base training model is “AnyLora (AnyLoRA_noVAE_fp16-pruned.ckpt)” which is primarily used for Anime Art Styles thus limited in performance for realistic product images, while other models have “Stable Diffusion (sd-v1-5-pruned-noema-fp16.safetensors)”. “aq_train2” however is trained with the LoRA type–LoCon. LoCON is LoRA that works on the convolutional units. All of these models converge around 1300 steps or the 9^{th} epoch.

F. Parameters in Image Generation, Testing, and X/Y/Z Plots

Testing image generation models involves assessing their ability to create images based on varied parameters, a process detailed in a specific paper. A key aspect of these models is their use of prompts for controlled image synthesis, with platforms like Automatic1111 offering customization options such as synthesis steps, sampling methods, and the Classifier-Free Guidance (CFG) scale for flexibility in image generation. Visual tools like X/Y/Z plots help analyze the effects of changing these parameters on image outcomes, highlighting the influence of choices like sampling steps or CFG scale on image quality.

In Stable Diffusion, sampling—a sequence of denoising steps—transforms a random latent space image into a coherent one. The process depends on the sampler and noise schedule, with the balance between speed and image quality hinging on the number of sampling steps. While more steps generally improve quality, there's a plateau beyond which quality doesn't increase, emphasizing the need for a balanced approach. The CFG scale determines how closely images adhere to prompts, with higher values ensuring greater fidelity to the input, though values above 7 might not enhance quality further and could introduce artifacts.Fig. 7 X/Y/Z Plot plots Sampling Steps against CFG Scale. There's obviously visible distortion in images with higher CFG values and lower steps. While steps beyond 20-30 range do not provide any additional benefits to the image quality and CFG scale range from 5-7 is the most ideal range.

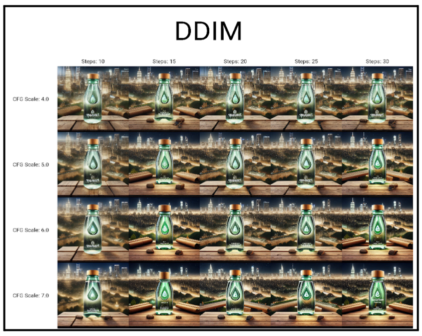
Fig. 8 CFG value vs. Number of Steps

Traditional ODE solvers such as Euler, Heun, and LMS, adapted for image generation, offer a trade-off between simplicity, speed, and accuracy. Ancestral samplers introduce randomness, affecting image consistency. The Karras noise schedule enhances image quality by optimizing noise reduction towards the sampling end. Recent advancements like DPM, DPM++, and UniPC improve efficiency and quality through adaptive steps and unified methods. Sampler choice significantly influences the generation speed, efficiency, and perceptual image quality, depending on desired outcomes for convergence, rendering time and visual fidelity.

Few sampling methods used in this paper against CFG scale are Euler, Denoising Diffusion Implicit Model (DDIM), DPM2 Karras, DPM++ SDE Karras, DPM++ 2M SDE Exponential. Euler is the most straightforward sampler, adopting a deterministic path that offers a balance between simplicity and accuracy in image generation. samplers developed for diffusion models, DDIM strikes a balance between image quality and processing speed, though newer technologies have surpassed its efficiency.

DPM2 Karras leverages a noise adjustment strategy recommended by Karras, aiming for finer image quality towards the final steps of sampling, at a trade-off of slightly reduced speed. DPM++ SDE Karras is a sophisticated approach that combines stochastic differential equations with the Karras noise schedule, aiming for superior image detail and quality. DPM++ 2M SDE Exponential represents the cutting edge in sampling technology, utilizing exponential SDEs to produce exceptionally high-quality images, optimizing both the process's efficiency and the resulting image fidelity.

As indicated by Fig. 9, DPM++ SDE Karras and DPM++ 2M SDE Exponential have a ton of variations in the image batches and many undesired distortions and artifacts. Keeping in mind the speed and accuracy, Euler and DPM2 Karras is the sampling method to be implemented for generating "AquaPure" product advertisements as they have produced consistent results.



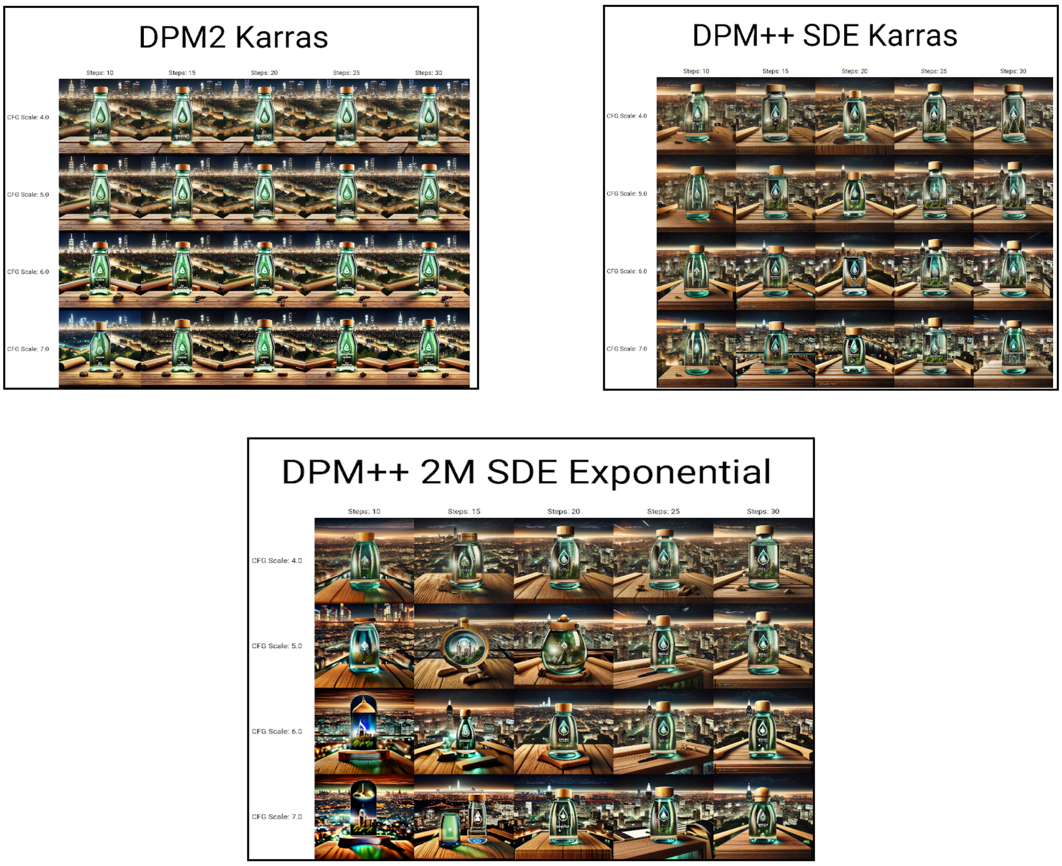


Fig. 9 Various Sampling Methods X/Y Plot

G. Prompts and Prompt S/R and Matrix

The process of creating prompts is essential for generating high-quality images, involving the strategic use of positive and negative prompts to respectively include or exclude specific terms in the generated images. Assigning different weights to elements within the prompt can influence the importance of these elements in the image generation process. The Prompt S/R (Search/Replace) feature and the prompt matrix tool on the A1111 platform exemplify advanced capabilities in prompt customization. For instance, the ability to replace elements such as "tiffin" with alternatives like "a bag" or "a towel" in a prompt enables the generation of images with diverse elements. These tools not only facilitate the creation of nuanced images but also provide insights into the impact of each prompt component on the final outcome, thereby offering enhanced creative control in creative and marketing applications.

Fig. 10 S/R matrix with different associative items

H. ControlNET and Additional Networks

This study outlines the development of a LoRA model notable for achieving benchmarks with a CLIP score range of 34.9102 to 36.2137 and an FID score on par with DreamBooth, utilizing proprietary datasets. Although successful in generating product variations across diverse backgrounds, its utility is mainly confined to basic advertising requirements, lacking in delivering narrative-driven advertising that connects with consumer trends and demographics. The research highlights the critical role of engaging design elements in advertising for capturing consumer interest and fostering emotional connections. It suggests employing multiple specialized LoRA models to align with a company's design ethos or reflect current market trends, thereby improving advertisement attractiveness and relevance. The paper proposes a strategy for associative generation, planning to train additional LoRA models on related items (e.g., a glass water bottle with lunchboxes and cooler bags) to craft richer narratives. This integrated model approach seeks to produce advertisements that are visually captivating, thematically coherent, and resonate deeply with the intended audience, in line with foundational advertising principles..

ControlNet, an innovative neural network, augments text-to-image generation in combination with Stable Diffusion by leveraging auxiliary models such as OpenPose, Sketch, and Canny for creating detailed outlines that inform image synthesis. It preserves the foundational parameters of the original model while incorporating trainable encoders for responsiveness to varied inputs. Featuring a U-Net architecture with 25 blocks, including ResNet layers and Vision Transformers, ControlNet methodically enhances image detail. It preprocesses conditioning images into feature vectors through a compact network, ensuring seamless integration with Stable Diffusion for the production of high-quality images tailored to precise aesthetic and structural preferences..

The ControlNET module within the system offers various control mechanisms, including Canny edges, depth perception, normal mapping, and OpenPose for pose estimation. Users can fine-tune the degree of influence through settings such as control intensity and the number of control steps. Preprocessors for each control type allow users to preview masks used in image synthesis. By leveraging multiple ControlNET units, users can enhance image precision, particularly within the AdditionalNetworks unit, facilitating critical guidance for image customization and refinement. This system uses control_v11p_sd15_scribble [d4ba51ff] ControlNET Model for generating the outline/sketch of the original image. As shown in Table. 3, using a preprocessor t2ia_sketch_pidi to generate a preview gives the sketch output.

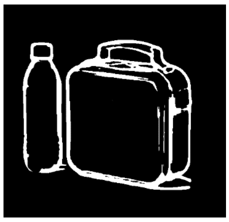
ControlNET Module	Seed Image Through ControlNET process	
1	Original Seed Image	On using Edge Detection
		

Table. 3 Preprocessor Output

In the context of Additional Networks, the AUTOMATIC1111 Stable Diffusion Web UI extension expands the original model's functionality by incorporating additional LoRA models, facilitating dynamic improvements in image generation. These models, trained with scripts from the sd-scripts repository, are configured via the "Additional Networks" panel in the Web UI, allowing the use of up to five LoRA models for detailed control over image creation. This unique integration permits precise application of models to specific image areas using mask images, offering targeted modification capabilities. Differing from alternatives like 'Latent Couple extension' and 'Composable LoRA', this extension enables fine-tuned manipulation within the U-Net architecture's layers, utilizing RGB masks to designate regions for applying multiple LoRAs' effects, enhancing the versatility and precision of image generation. RGB Mask of the same control as in the ControlNET module as shown in Fig. 11.

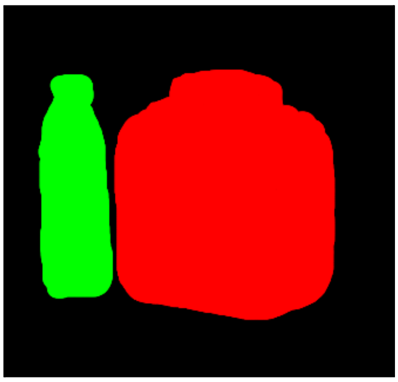


Fig.11 RGB Mask of Image

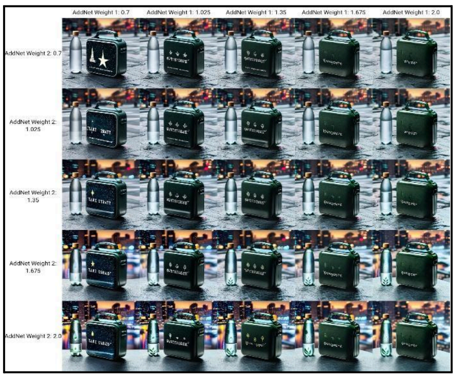


Fig. 12 Additional Network weights of two models

By varying the weights of LoRA in both the red and green channels as observed in Fig. 12, such as the LoRA for the "AquaPure" water bottle and the LoRA for the "AquaPure" lunch box, using the X/Y/Z plot, optimal weights for each can be determined. This process helps in finding the best balance between the additional network LoRAs and the prompt.
The masking process involves importing an image, transforming its color space from BGR to RGB for standard color representation, and manipulating specific colors while excluding white to maintain the image's background integrity. By calculating the Euclidean distance to identify and replace dominant non-white colors with a new color set, it creates a mask. This mask is then used with advanced networks like LoRA in a Stable Diffusion framework, facilitating targeted image manipulation for desired aesthetic or thematic outcomes. OpenCV thus acts as a conduit, enhancing neural network-based image synthesis with greater control and nuance.

I. Tagline Overlay on the generated image

A study published in the International Journal of Management, Technology, and Engineering [17] emphasizes the importance of incorporating textual information alongside visual content in advertising, particularly for new products. It involves crafting taglines inspired by the company's previous advertising campaigns, with the aid of GPT-4. These taglines are then integrated into the generated images using OpenCV, by segmenting the image to create an RGB mask and overlaying the tagline onto the image area with the largest contour. The text box's coordinates are precisely calculated for accurate placement on the image as visualized in Table. 4. This automated method significantly enriches the advertisement's communicative value by synergizing textual and visual elements, demonstrating an advanced approach to enhancing advertising effectiveness.

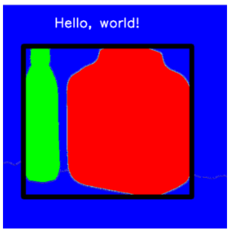
Tagline Overlay	Generated Image through tagline overlay process	
1	Background Detection	Text on Background
		

Table. 4 Combining tagline and generated image using background detection techniques

VI. Results And Discussions

The Results and Discussion section showcases the system's Graphical User Interface (GUI), designed to guide users through a detailed questionnaire aimed at capturing essential information about the product, including orientation, lighting, and background settings. Upon gathering this data, the system executes various pipelined modules to generate the final advertisement, presenting a novel solution that efficiently utilizes proprietary datasets without compromising company secrets. LoRA emerges as a cutting-edge approach, enabling the seamless incorporation of generative AI into the business landscape, thus optimizing the creative process for advertisement production. This strategic implementation demonstrates a significant advancement in automating content creation, ensuring that the generated advertisements are both high in quality and aligned with the user's specific requirements. This system has achieved satisfactory FID and IS scores. The figures below are the functional and navigational pages of the systems.

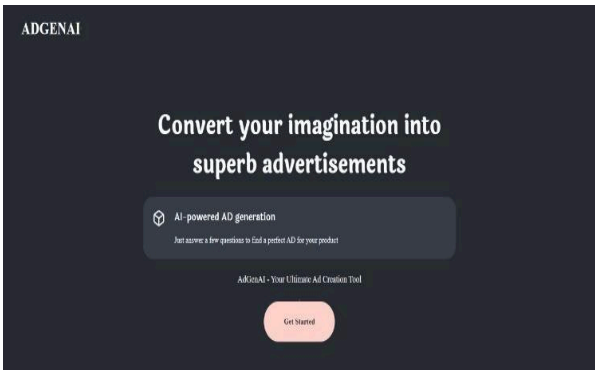


Fig. 13 Landing page of ADGenAI

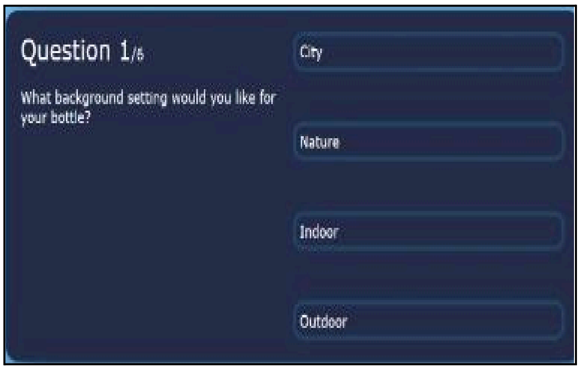


Fig. 14 Questionnaire to create a prompt based on response image

Fig. 15 Combining tagline and generated

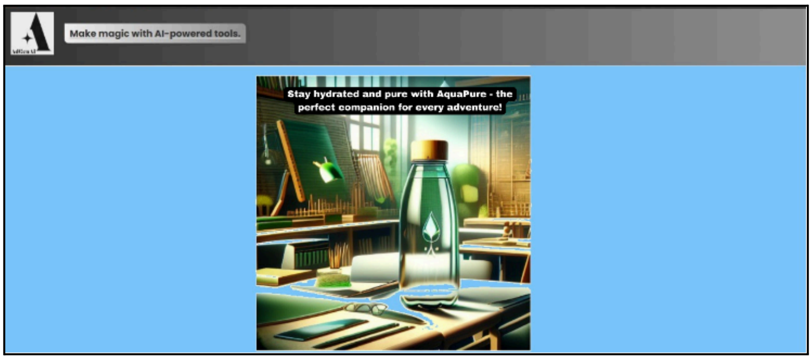


Fig. 17 Final Advertisement

VII. Conclusion

The system introduced in this research marks a transformative approach to product-focused corporate advertising by integrating generative artificial intelligence and technologies such as Stable Diffusion (LoRA), ControlNet, Additional Networks, LLM-ChatGPT API, and OpenCV. This development offers a streamlined solution for ad production, blending automation with customization to support creative professionals in generating high-quality advertisements efficiently, while preserving brand integrity and campaign objectives.

The system's innovative architecture and functionalities demonstrate its practicality and potential to shift market dynamics by enhancing the creative capacity of its users, enabling businesses to differentiate in a competitive market and respond to changing consumer preferences.

Prospects for this system include continued innovation in advertising strategies, enabling businesses to produce impactful, emotionally resonant ads. As technology advances, this system is set to lead the evolution of digital advertising creation and engagement.

Disclosure statement

No potential conflict of interest was reported by the author(s)

References

[1] John Ford, Varsha Jain, Ketan Wadhwani, Damini Goyal Gupta, "AI advertising: An overview and guidelines" - Journal of Business Research, Volume 166,2023,114124,ISSN 0148-2963

[2] Liuqing Chen, Pan Wang, Hao Dong, Feng Shi, Ji Han, Yike Guo, Peter R.N. Childs, JunXiao, Chao Wu, "An artificial intelligence based data-driven approach for design ideation, Journal of Visual Communication and Image Representation" - Volume 61, 2019, Pages 10-22, ISSN 1047-3203

[3] Axel Sauer, Tero Karras, Samuli Laine, Andreas Geiger, Timo Aila, "StyleGAN-T: Unlocking the Power of GANs for Fast Large-Scale Text-to-Image Synthesis" - Machine Learning (cs.LG); Computer Vision and Pattern Recognition (cs.CV), arXiv:2301.09515, 2023

[4] Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, In So Kweon, "Text-to-image Diffusion Models in Generative AI: A Survey" - Computer Vision and Pattern Recognition (cs.CV); Artificial Intelligence (cs.AI); Machine Learning (cs.LG), arXiv:2303.07909, 2023

[5] Prafulla Dhariwal, Alex Nichol, "Diffusion Models Beat GANs on Image Synthesis" - 35th Conference on Neural Information Processing Systems (NeurIPS 2021)

[6] R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 10674-10685, doi: 10.1109/CVPR52688.2022.01042.

[7] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala, “Adding Conditional Control to Text-to-Image Diffusion Models”, - Computer Vision and Pattern Recognition (cs.CV); Artificial Intelligence (cs.AI); Graphics (cs.GR); Human-Computer Interaction (cs.HC); Multimedia (cs.MM), arXiv:2302.05543, 2023
[8] Borji, Ali. (2022). Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. 10.48550/arXiv.2210.00586
[9] AUTOMATIC1111, "Stable Diffusion Web UI", 2022-08-22, https://github.com/AUTOMATIC1111/stable-diffusion-webui
[10] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In the International conference on machine learning (pp. 8748-8763). PMLR.
[11] Bie, F., Yang, Y., Zhou, Z., Ghanem, A., Zhang, M., Yao, Z., Wu, X., Holmes, C., Golnari, P.A., Clifton, D.A., He, Y., Tao, D., & Song, S.L. (2023). RenAIssance: A Survey into AI Text-to-Image Generation in the Era of Large Model. ArXiv, abs/2309.00810.
[12] P. Wang, "deep-daze," 2021. [Online]. Available: https://github.com/lucidrains/deep-daze. [Accessed: 10-3-2024].
[13] Karras, T., Laine, S., & Aila, T. (2018). A Style-Based Generator Architecture for Generative Adversarial Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4396-4405.
[14] Zhang, Han et al. “StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks.” 2017 IEEE International Conference on Computer Vision (ICCV) (2016): 5908-5916.
[15] DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation
[16] Hu, J.E., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., & Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models. ArXiv, abs/2106.09685.
[17] Mitchell, Andrew A. “The Effect of Verbal and Visual Components of Advertisements on Brand Attitudes and Attitude Toward the Advertisement.” Journal of Consumer Research, vol. 13, no. 1, 1986, pp. 12–24. JSTOR.

b.Plagiarism Report

ORIGINALITY REPORT

3%	1%	1%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

c.Project review sheet ;

Project review sheet 1:

Inhouse/ Industry_Innovation/Research:

Sustainable Goal: Industry, Innovation + Infrastructure (9)

Project Evaluation Sheet 2023 - 24

Class: D17 A/B/C

Group No.: 12

Title of Project: Generation of company specific Ads (ADGenAI)

Group Members: Gauri Nagral (49), Meera Sawantdesai (60), Kritika Yadav (72), Yashraj Mulwani (46)

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
4	4	4	3	4	2	2	2	2	2	3	3	3	2	4	45

Comments: fine tuning of model with UI improvement is needed.

Name & Signature Reviewer1

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
4	4	4	3	4	2	2	2	2	2	3	3	3	2	4	45

Comments: Complete your UI with logging

Date: 10th february, 2024

Name & Signature Reviewer 2

Project review sheet 2:

Inhouse/ Industry_Innovation/Research:

Sustainable Goal: 9 (Industry, inovation & infrastructure)

Project Evaluation Sheet 2023 - 24

Class: D17 A/B/C

Group No.: 12

Title of Project: Generation of company specific ADS (ADGENAI)

Group Members: Gauri Nagral (49), Meera Sawantdesai (60), Kritika Yadav (72), Yashraj Mulwani (46)

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
4	4	4	4	4	1	2	2	2	2	3	2	2	2	4	42

Comments: Prepare for your final presentation with new dataset

Name & Signature Reviewer1

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
4	4	3	3	4	1	2	2	2	2	3	2	2	2	4	40

Comments: UI integration Pending.

Date: 9th March, 2024

Name & Signature Reviewer 2