Research Paper for Ai Shastra Categorized Content Creator Using Generative AI

[1]Priyanka Sharma, [2]Shashank Srivastava, [3]Simran Arora, [4]Tanya Sharma
ABESIT College of Engineering, Ghaziabad

Abstract— In the ever-evolving world of conversational AI, large-scale generative AI models such as ChatGPT, and GPT-4 are rapidly changing the way we communicate, illustrate, and create. This research paper addresses the limitations of current conversational AI, focusing on ChatGPT, and proposes an innovative Android application, "Ai Shastra," as a solution. ChatGPT's generic responses and lack of specialization are identified as drawbacks, prompting the development of Ai Shastra. The app aims to enhance user interactions by offering personalized assistance through specialized AI models and image search capabilities. Ai Shastra enables users to select specific domains and utilize image queries for comprehensive problem-solving. The integration of advanced AI technologies, such as specialized categorization and image search, enhances the app's contextual relevance. Through detailed methodology analysis and performance assessments, the paper demonstrates how Ai Shastra sets new standards for AI applications, providing users with specialized help and ushering in a future where personalized AI assistance becomes the norm

Keywords— Generative AI, ChatGPT, GPT, LLM, OpenAI, Content Generation, Mapping, Introduction

I. INTRODUCTION

In the digital age, content creation has transcended traditional boundaries and evolved into a sprawling, multi-faceted ecosystem. This evolution is significantly influenced by the rapid advancements in technology and notably, the advent of generative AI. Categorized content creators, a thriving subculture within this ecosystem, have harnessed the potential of AI to enhance and innovate their creative processes. This research paper delves into the dynamic world of categorized content creators who leverage generative AI, examining the implications, challenges, and opportunities presented by this symbiotic relationship.[1]

Generative AI, powered by deep learning models such as GPT-3 and its successors, has revolutionized content creation across numerous categories. It has enabled content creators to automate, augment, and revolutionize their work, leading to the birth of AI-assisted content creation in areas such as art, literature, music, video production, and more.[2] These AI-augmented creators have given rise to a paradigm shift in the way we think about human creativity, as they collaborate with machines to generate content that captivates audiences and pushes the boundaries of what is possible.[3]

II. RELATED WORK

1. LLM

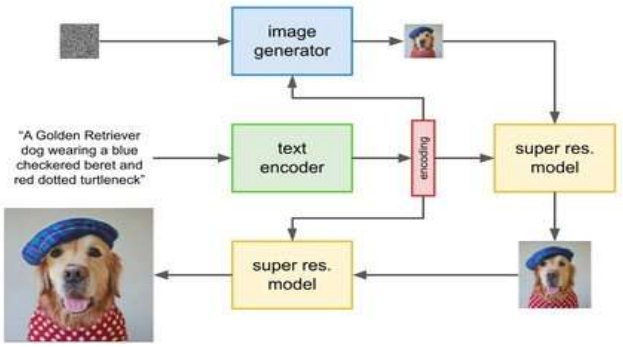
Large-scale language models (LLMs) are fundamental machine learning models that use deep learning algorithms to process and understand natural language. In simpler terms, an LLM (Language Model) is a computer program that learns to understand and interpret human language or other intricate data by being exposed to numerous examples. Often, LLMs are trained on vast amounts of text data collected from the Internet, amounting to thousands or millions of gigabytes. However, the effectiveness of LLMs in learning a natural language depends on the quality of the samples, so the programmers might opt for carefully selected datasets. These models are trained on large amounts of text data to learn patterns and entity relationships within the language. LLMs can perform many different types of language tasks, including: For example, language translation, sentiment analysis, chatbot conversations, etc.[5]

Language models vary in complexity from simple N-gram models to more sophisticated neural network models. However, the term "large language model" typically refers to models that use deep learning techniques and have a large number of parameters, ranging from millions to billions. These models can capture complex patterns in language and produce text that is indistinguishable from human-written text. These models undergo training using extensive textual data derived from diverse sources such as books, articles, websites, and various written consent forms. Through the analysis of statistical relationships among words, phrases, and sentences during this training, the model gains the ability to produce coherent and contextually relevant responses to prompts and queries.

2. GPT-3 (Generative Pre-trained Transformer3)

- GPT-3, or the Generative Pre-trained Transformer3, is a state-of-the-art language model developed by Open AI. It belongs to the family of large language models and stands out for its unprecedented scale, Boasting astaggering 175 billion parameters. This immense parameter count contributes to GPT-3's exceptional ability to understand and generate human-like text across a wide range of tasks.

- Boasting a staggering 175 billion parameters. This immense parameter count contributes to GPT-3's exceptional ability to understand and generate human-like text across a wide range of tasks.
- GPT-3 engages in the processing of text input to execute various natural language tasks. Employing both natural language generation and natural language processing, it comprehends and generates text in a manner that mimics natural human language. Historically, machines have grappled with the challenge of crafting content understandable to humans, given the intricacies and nuances of language. GPT-3addresses this bygenerating articles, poetry, stories, news reports, and dialogue with minimal input text, producing substantial amounts of copy.
- Beyond human language text, GPT-3 possesses the capability to generate anything structured as text. This extends to text summarizations and even the creation of programming code.
- GPT-3 has garnered attention for its breakthroughs in natural language understanding and generation, marking a significant milestone in the field of artificial intelligence. Its vast scale and versatility make it a powerful tool for various applications, pushing the boundaries of what is possible with large language models.
- The deep learning neural network of GPT-3 comprises more than 175 billion machine learning parameters, makingit the most extensive model of its



kind. To contextualize its scale, the preceding largest language model, Microsoft's Turing Natural Language Generation NLG), had 10 billion parameters. Asof early 2021, GPT-3 stands as the most substantial neural network ever developed. Consequently, GPT-3 outperforms all previous models in generating text that convincingly resembles human-authored content.[10]

1.Text to image model

3.VertexAI

Vertex AI is an advanced machine learning (ML) platform designed to streamline the training, deployment, and customization of ML models and AI applications. With a focus on enhancing collaboration among data engineering, data science, and ML engineering teams, Vertex Ai provides a unified tool set and leverages the capabilities of Google Cloud to scale applications effectively.

Key features and options offered by Vertex AI include:

- **AutoML**: Enables The training of models on tabular, image, text, or video data without the need for coding manual data preparation. This feature is designed for ease of use and accessibility.
- **Custom Training**: Provides full control over the training process, allowing users to leverage their preferred ML frameworks, write custom training code, and choose hyperparameter tuning options for a more tailored approach.
- **Model Garden**: Offers a repository where users can discover, test, customize, and deploy Vertex AI along with select open-source (OSS) models and assets. This fosters a collaborative environment for model development.
- **Generative AI:** Grants access to Google's extensive generative AI models across various modalities such as text, code, images, and speech. Users can fine-tune these models to suit their specific requirements and deploy them for integration into AI-powered applications.

Once models are deployed, Vertex Ai provides send-to-end ML Ops tools, automating and scaling projects throughout the ML lifecycle. These ML Ops tools operate on fully managed infrastructure, allowing customization based on performance and budget considerations. In essence, Vertex Ai is a comprehensive solution that empowers organizations to efficiently manage the complexities of ML development and deployment within the Google Cloud environment.[9]

Generative AI

2. Generative AI is a revolutionary branch of artificial intelligence. It is a concept of AI that can be used to create new content, instead of looking at and using the existing data, such as expert systems. A major leap in machine learning has been made with the introduction of Generative Artificial Intelligence, enabling content from text to pictures to be autonomously generated. At the forefront of this innovation are models such as Generative Pre-trained Transformers (GPT), which have demonstrated great ability to understand and generate context-rich output.[6]

Moreover, the interpretability of generative AI results remains an important research area. As these models become more sophisticated, the ability to understand and explain the reasons for the results becomes increasingly important. Researchers are exploring ways to make generative AI more transparent and give users insight into how and why certain content is being generated.

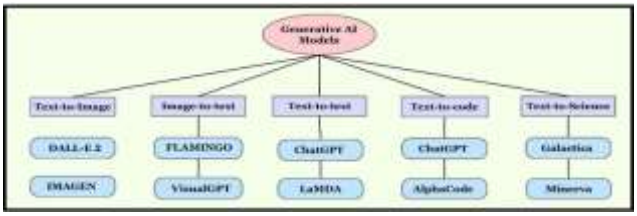
II. GENERATIVE AI

Despite these challenges, the transformative potential of generative AI is undeniable. Beyond its creative use in content production ,its usefulness in problem-solving in a variety of fields is increasingly being recognized. Continuous improvement and optimization of generative AI not only contributes to its creative performance but also makes it a valuable tool for tackling real-world challenges. Fundamentally, generative AI sits at the intersection of creativity and practicality, changing our expectations of what artificial intelligence can do in both the imaginative and problem-solving realms.

Generative Language Models

Generative models area class of artificial intelligence models designed to create new data that is similar to training data they were exposed to. The primary objective of these models is to

understand and capture the underlying patterns and structures present in the data, enabling them to generate realistic outputs. Among generative models, one notable category is Variational Autoencoders



(VAEs), which work by learning the probability distribution of the training data and then generating new samples from this distribution.

Another prominent type of generative model is Generative Adversarial Networks (GANs). GANs consist of two neural networks – a generator and a discriminator – engaged in continuous competition. The generator create s data to mimic the training set, while the discriminator's role is to distinguish between real and generated data. This dynamic inter play pushes both networks to improve continuously, resulting in the generation of increasingly realistic and diverse outputs.

As generative models continue to evolve, their role in shaping the future of AI becomes /increasingly pronounced, with potential implications for fields ranging from creative arts to scientific research which are mentioned below.

2.1.1 DALL-E2

DALL·E 2, an evolution of the original DALL·E, is a cutting-edge generative model developed by Open AI that demonstrates the amazing capabilities of modern artificial intelligence. A key feature of DALL· E2 is the ability to understand and interpret complex text prompts. Users can go beyond simple instructions and provide text explanations that include subtle details and abstract concepts. The model uses this information to create images that not only reflect the specified content but also exhibit a certain level of creativity and imagination.

2.2 Imagen

Imagen stands out as a text-to-image model that operates on the principles of a diffusion model, leveraging large transformer language models. In essence, it excels at transforming textual descriptions into visual representations. The core of Imagen's functionality lies in its complex use of transformerarchitecture. These are sophisticated models known for their ability to understand complex patterns in language and images. Imagen uses this feature to create a connection between the written description and the generation of the corresponding visual content.

What is unique about the Imagen approach is the use of a diffusion model. This means that the model gradually improves its understanding of the text, simulating a step-by-step process rather than an instantaneous transformation. The careful handling ensures that the images produced are in line with the precise details of the inputted text. The implications of this are significant as it enables the interface between natural language processing and computer vision to advance the potential of artificial intelligence by addressing the gap between textual narrative and visual display.

2.3 Image-to-text Models

2.3.1 Flamingo

Flamingo is a family of visual language models(VLM) introduced in the Deep Mind article" Flamingo: A Visual Language Model for Few-Shot Learning ". They establish a new state of the art in learning in just a few steps for a wide range of open visual and language tasks.

Flamingo outperformed fine-tuned state-of-the-art techniques in six cases (out of 16) considered by the authors. As GPT-3 demonstrates, it takes inspiration from large-scale generative language models(LMs) that are well-suited for few-shot learning. Flamingo models fuse(combine) large language models with powerful visual embeddings (each pre-trained and individually frozen) by adding new architectural components in between. Flamingo accepts images/videos and nested text and outputs free text. It can handle both open-ended tasks, such as visually answering questions or writing subtitles, and closed-ended tasks, such as categorization.

FOR TASKS SOME WHERE ON THIS SPECTRUM, A SINGLE FLAMINGO MODEL CAN REACH A NEW STATE-OF-THE-ART STATE IN A FEW SHOTS BY SIMPLY STIMULATING THE MODEL WITH TASK-SPECIfiC EXAMPLES. ACROSS NUMEROUS BENCHMARKS, FLAMINGO OUT PERFORMS MODELS TUNED TO TASK-SPECIfiC DATA BY A FACTOR OF THOUSANDS.

2.3.2 Visual GPT

Visual GPT, an innovative image captioning model developed by OpenAI, builds upon the foundational knowledge of the pre-trained language model GPT-2.To effectively bridge these mantic gap between different

modalities, Visual GPT introduces anovelencoder-decoder attention mechanism[33], featuring an unsaturated rectified gating function. A key advantage of this model lies in its exceptional data efficiency, requiring less data compared to other image-to-text models. This efficiency is particularly valuable fors wift data curation, describing rare objects, and applications in specialized domains. Notably, the model's API is openly available on GitHub, enhancing accessibility and facilitating integration into diverse projects

At the heart of Visual GPT is the Visual Foundation Model, a fundamental computer vision algorithm that transfers standard

computer vision functionality to AI applications to handle more complex tasks.Visual GPT's prompt manager consists of 22 of his

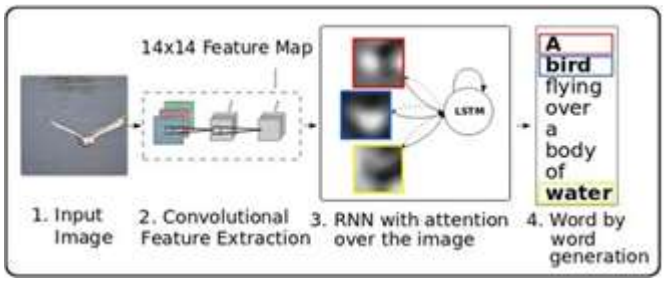


Fig 3 Famingo

VFMs, including Text-to-Image, Control Net, Edge-To-Image, and more.Visual GPT is Microsoft's latest and most innovative step in

AI development. Al though Visual GPT is still in its early stages of development, it has the potential to revolutionize the way we interact with machines.

2.3.4 Text-to-text Models Chat GPT:

ChatGPT, a widely acclaimed conversational model developed by Open AI, engages in dynamic interactions in a chat-like manner. Notably, the model exhibits a capacity to respond to subsequent questions, rectify in accuracies in assumptions, and reject inappropriate requests. Technically, Chat GPT is underpinned by a Transformer architecture, but its training methodology involves Reinforcement Learning for Human Feedback.

The training process begins with supervised fine-tuning, where human AI trainers simulate conversations by playing both user and AI assistant roles. These trainers are provided with responses generated by the model to aid in composing their replies. This dataset is then amalgamated with the Instruct GPT [3] dataset, which is transformed into a dialogue format. A detailed overview of the Chat GPT training steps can be found in Figure14, accessible on the Chat GPT demo's website.

For practical use, OpenAI provides a demo on its website, showcasing the model's conversational capabilities. Additionally, the API for Chat GPT is available on Open AI's website. It's note worthy that Chat GPT extends its capabilities beyond general conversation, demonstrating proficiency in generating code snippets and performing basic mathematical tasks.

LaMDA

LaMDA stands out as a specialized language model tailored for dialogue applications. Whatsets LaMDA apart is its unique training approach—unlike many other language models, LaMDA is specifically trained on dialogue. This family of transformer-based neural language models boasts an impressive parametercount of up to 137 billion, pre-trained on an extensive corpus of 1.56 trillion words from public dialog data and webtext.

To enhance safety and ensure factual grounding, fine-tuning is employed. Notably, only a minute fraction, 0.001%,of the training dataisused for this fine-tuning process, showcasing the model's efficiency and effectiveness. Leveraging the inherent ability of transformers to capture long-term dependencies in text, LaMDA excels india log scenarios, benefitting significantly from model scaling.

To enhance safety and ensure factual grounding, fine-tuning is employed. Notably, only a minute fraction, 0.001%,of the training dataisused for this fine-tuning process, showcasing the model's efficiency and effectiveness. Leveraging the inherent ability of transformers to capture long-term dependencies in text, LaMDA excels india log scenarios, benefitting significantly from model scaling.

2.3.5 Text-to-code Models

Codex:

Codex, an AI system developed by Open AI, is designed to translate text in to code, making it a versatile programming model applicable to a wide range of programming tasks[8]. The essence of programming involves two fundamental steps: decomposing a these sub-problems into existing code, leveraging libraries, APIs, or functions. Codex excels notably in the latter part, which is often the most time-consuming for programmers.

The model's strength lies in its proficiency at mapping textual descriptions to functional code. To achieve this capability, Codex

was trained on a substantial data set collected in May 2020 from public software repositories on GitHub. This dataset comprises 179GB of unique Python files, each under 1 MB. The training process involves fine-tuning from GPT-3, a model renowned for its robust natural language representations.

Codex's proficiency in understanding and generating code makes it a valuable tool for developers facing the challenge of translating textual requirements in to executable code. Whether for simplifying coding tasks or exploring new approaches to problem-solving, Codex offers a promising solution. For those interested in exploring its capabilities, OpenAI provides a demo and access to the API on their official website.

Alpha Code

Alpha Code distinguishes itself from other language models by specializing in code generation for complex problems that demand deeper reasoning. Its efficacy relies on three key components: a comprehensive dataset for training and evaluation, large and efficient transformer-based architectures, and a large-scale model sampling approach.

In terms of training, Alpha Code undergoes an initial pre-training phase using GitHub repositories, amassing a substantial 715.1 GB of code. This dataset surpasses the pre-training dataset of Codex, contributing to a more extensive understanding of code patterns. To refine the model, a fine-tuning data set is introduced from the Code forces platform, involving Code contests for validation, there by enhancing the model's performance

III. LITERATURE SURVEY

.	RESEARCH PAPER NAME	AUTHOR NAME	Year Of Publication	SUMMARY	PROS	CONS
1	Understanding Potential Benefits of ChatGPT in Promoting Teaching and Learning	David Baidoo-Anu	2023	ChatGPT launched a year ago, drawing mixed reactions from educators.	Personalized learning formative, Assesment support	**Mis information Risk,** **Biases Amplification**
2.	Generative AI at Work	Erik Brynjolfsson Danielle Li Lindsey R. Raymond	2023	The introduction of a generative AI conversational assistant among 5, customer support agents led to a 14% increase in productivity, particularly benefiting novice and low-skilled workers.	Productivity Boost Skill Development Positive Customer Impact Employee Retention	Potential Job Displacement Dependency Risks Implementation Challenges Ethical Considerations
3.	CHATGPT is not all you need. A state-of-the-art Review of Large Generative AI model.	YIHAN CAOSIYU LIYIXIN LIUZHILING YAN	2023	The emergence of large generative models like ChatGPT and Stable Diffusion has revolutionized various sectors, transforming tasks such as text-to-image, text-to-3D images, image-to-text, text-to-video, text-to-audio, text-to-code, and more.	Innovation Across Sectors Increased Efficiency Creative Capabilities Job Transformation	Job Displacement Concerns Complex Implementation Dependency Risks
4.	Regulating ChatGPT and other Large Generative AI models	Philipp Hacker Andreas EngelMarco Mauer	2023	The regulatory challenges posed by large generative AI models (LGAIMs) like ChatGPT propose tailored strategies for trustworthy deployment.	Technological Advancement Innovation Potential Tailored Regulatory Approach	Regulatory Gap Complex Regulatory Structure Challenges in Enforcement
5.	A History of Generative Ai from Gan to	Yihan Cao Siyu Li	2023	AIGC aims to enhance content creation efficiency	Efficient Content Creation High-Quality Output	Open Problems

	ChatGPT	Yixin Liu Philips S.Yu		by leveraging large-scale models for intent extraction, leading to improved and realistic content generation in various modalities such as text and image.	Multimodal Capabilities	Potential Bias Complexity in Multimodal Interactions Future Challenges
6.	Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence	Grant Cooper	2023	Generative AI, exemplified by ChatGPT, shows promise in science education. The study explores ChatGPT's responses to science queries, suggests ways educators can integrate it responsibly, and employs ChatGPT as a research tool.	Educational Assistance, Resource for Educators, Innovative Teaching Aid	Epistemic Authority Concerns Critical Thinking Challenges Adaptation and Evaluation
7.	What Do We Mean by GenAI? A Systematic Mapping of The Evolution, Trends, and Techniques Involved in Generative AI	Francisco José García-Peñalvo, Andrea Vázquez-Ingelmo	2023	The paper conducts a literature mapping of 631 AI-driven content generation solutions over the last five years, revealing a dichotomy in understanding.	Creative Output, Democratized Access, Innovation Catalyst	Undefined Terminology, Quality Variability, Divergent Perceptions

IV. PROBLEM STATEMENT

In the rapidly evolving field of artificial intelligence, conversational AI applications have made significant progress by enabling human-like interactions and providing advanced automated assistance. However, the current landscape of chatbot-based solutions struggles to provide an optimal experience for those who lack a specialized user focus and need expert-level guidance in a specific area. This limitation often leads to generic answers that limit the usefulness and relevance of chatbots to users with diverse needs.

V. PROPOSED SOLUTION

The integration of these features into the "AIShastra" Android application seeks to provide users with a superior conversational AI experience. By combining specialized categorization ,image-based search capabilities, an intuitive interface, and a commitment to continuous improvement and user privacy, AIShastra aspires to address the limitations of generic chatbots and empower users with a personalized, efficient, and engaging interaction with artificial intelligence on their Android devices.

VI. CONCLUSION

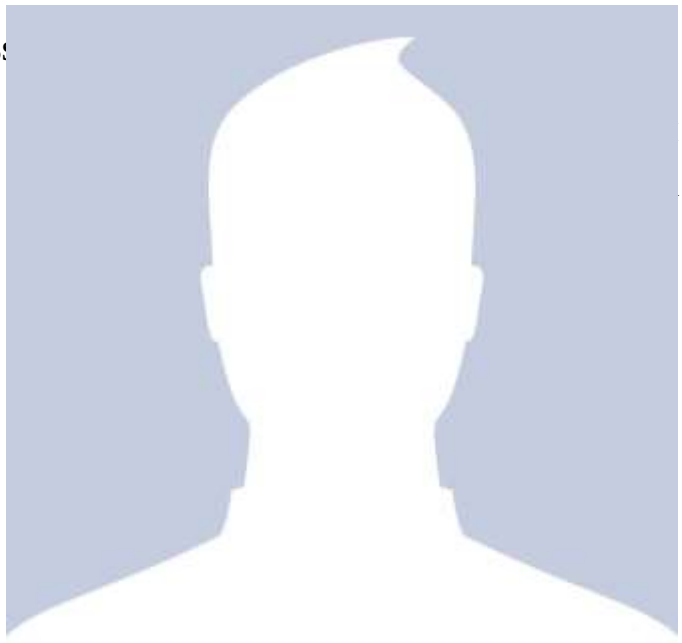
AiShastra represents an Android application that effectively addresses the short comings of traditional conversational AI. By integrating Open AI for expert classification and image processing, the app delivers nuanced user guidance and expert-level assistance across diverse domains. The addition of an innovative image search feature significantly enhances the platform's problem-solving capabilities.

As we conclude this research, it becomes evident that the implications of AI extend far beyond the immediate applications. Ai Shastra and the explored generative models collectively represent a stepping stone towards a future where AI seamlessly integrates into our daily lives, offering tailored solutions and enhancing the way we navigate intricate challenges.

VII. REFERENCES

[1] Roberto Gozalo-Brizuela, Eduardo C. Garrido-Merch´an ‘ChatGPT is not all you need. A State of the Art Review of large Generative AI models’ 11 JUNE 2023.

[2] Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan Yutong Dai, Philip S. Yu, Lichao Sun ‘A Comprehensive Survey of AI-Generated Content (AIGC): A History of GenerativeAI from GAN to ChatGPT’ 7 MARCH 2023.



[3] David Baidoo-Anu, Leticia Owusu Ansah 'Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning' MARCH 2023.

[4] Philipp Hacker, Andreas Engel, Marco Mauer 'Regulating ChatGPT and other Large Generative AI Models' 12-15 JUNE 2023.

[5] Erik Brynjolfsson, Danielle Li, Lindsey R. Raymond 'GENERATIVE AI AT WORK' APRIL 2023.

[6] Grant Cooper 'Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence' 22 MARCH 2023.

[7] Francisco José García-Peñalvo, Andrea Vázquez-Ingelmo 'What Do We Mean by GenAI? A Systematic Mapping of The Evolution, Trends, and Techniques Involved in Generative AI' 20 JULY 2023.

[8] Akademik PerspektifDerneği 'Is ChatGPT Leading Generative AI? What is Beyond Expectations?' 7 MAY 2023.

[9] Abid Haleem a, Mohd Javaid a,∗, Ravi Pratap Singh 'An era of ChatGPT as a significant futuristic support tool: A