

OmniVerse: Local On-Device Multi AI System and Live Student Monitoring via Facial Recognition

Mohd. Ashfaque
Assistant Professor

Wael Shaikh
Student

Badreen Jaffer
Student

ABSTRACT

The OmniVerse project is a pioneering venture into the world of AI-powered web applications, promising to transform digital content interaction and artificial intelligence engagement. This report comprehensively explores OmniVerse's innovative features, its diverse user base, robust technical foundations, and meticulous development process. Notably, OmniVerse's three key features - text and code generation, image manipulation, and audio creation - harness cutting-edge machine learning algorithms to redefine user experiences. It streamlines content creation, aids creative professionals, and revolutionizes music production. OmniVerse caters to a diverse user base, emphasizing efficiency and innovation, while its precision and user-friendly design make it indispensable. In essence, OmniVerse is a gateway to heightened productivity and boundless creativity, driven by AI-driven text, image, and audio generation features, offering a transformative user experience.

General Terms

LLAMA, MusicGen, Text Generation, Image Generation Student Identity System, Facial Recognition

Keywords

AI, Generative AI, WebApp, Security, Computer Vision

1. INTRODUCTION

The OmniVerse project represents a pioneering leap into the realm of AI-powered web applications, poised to revolutionize the way users engage with digital content and artificial intelligence. This report encapsulates the core aspects of OmniVerse, delving into its innovative features, the transformative potential of its various AI capabilities, the diverse target audience, its robust technical foundations, and the meticulously executed development process. Moreover, it highlights the commitment to future enhancements and the promise of OmniVerse in the dynamic landscape of AI-driven technology. OmniVerse stands out with its trifecta of features: text and code generation, image generation, segmentation, and editing, and audio generation. Each of these features harnesses cutting-edge machine learning algorithms, redefining user experiences across multiple domains. Text and code generation simplify content creation and coding tasks, automating processes, and expediting productivity. Designers, photographers, and creative individuals benefit from the image generation and editing capabilities, allowing them to realize their artistic visions efficiently. The audio generation feature represents a game-changer in music production, enabling musicians and producers to craft bespoke sounds and compositions, unlocking new frontiers in auditory creativity.

2. RESEARCH METHODOLOGY

2.1 Surveying Existing System

Generative AI took the world by storm in the months after ChatGPT, a chatbot based on OpenAI's GPT-3.5 neural network model, was released on November 30, 2022. GPT stands for generative pretrained transformer, words that mainly describe the model's underlying neural network architecture.^[1] Language models, particularly transformer-based architectures like OpenAI's GPT (Generative Pre-trained Transformer), showcased remarkable text generation capabilities. Models like GPT-3 captured attention for their ability to generate coherent and contextually relevant text, leading to numerous creative applications. Recent advancements in generative AI have focused on multi-modal capabilities, combining text, images, and other modalities to generate rich and diverse content without pre-training or human interventions beyond prompts. AI-generated artwork, music, and virtual characters have garnered attention, demonstrating the potential of generative AI in creative domains.

GPT-3 is a massive model with 175 billion parameters, making it one of the largest language models ever created. This large scale contributes to its ability to understand context and generate coherent and contextually relevant text. The "pre-training" aspect involves training the model on a vast corpus of diverse text data, allowing it to learn the nuances of language and context. This pre-training is followed by fine-tuning on specific tasks. GPT-3 excels at understanding and generating text in context. It can maintain context over longer passages and generate human-like responses based on the input it receives. GPT-3 is a versatile model that can be fine-tuned for various natural language processing tasks, including text completion, translation, question-answering, summarization, and, in the case of ChatGPT, conversation. GPT-3 can perform tasks with zero-shot and few-shot learning. This means it can generalize to new tasks without specific training examples by leveraging instructions provided in the input.

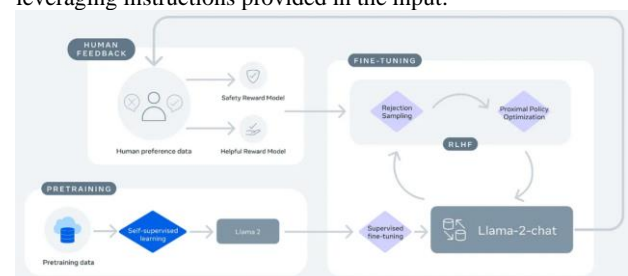


Fig 1: LLAMA-2 model architecture

2.2 Objective

The primary objectives of AI text generation are to enable natural, coherent, and contextually relevant textual interactions with users, whether in the form of chatbots, virtual assistants, or automated content generation. These systems aim to provide information, assistance, and entertainment, while continually learning and personalizing their responses to user needs. They serve diverse purposes, from facilitating communication across languages and providing educational content to offering emotional support and performing automated tasks, ultimately enhancing user experiences, and meeting a wide array of application-specific goals.

3. PROPOSED SYSTEM

OmniVerse stands out with a powerful array of features, each leveraging advanced machine learning algorithms to enhance user experiences in various domains:

Text and Code Generation: OmniVerse empowers users with AI-driven text and code generation capabilities, simplifying content creation and coding tasks. Whether it's drafting articles, crafting code snippets, or automating repetitive assignments, this feature promises to expedite and enhance productivity. One of OmniVerse's standout capabilities is its AI-driven text and code generation feature. This is a multifaceted tool that simplifies various content creation and coding tasks, catering to writers, programmers, and professionals across the board. Whether you need to draft articles, generate code snippets, or automate those tedious, repetitive assignments, OmniVerse is here to make your life easier and more productive. By taking advantage of AI technology, it not only expedites the process but also enhances the quality of the output, ensuring that your text and code are of the highest standards. **Image Generation, Segmentation, and Editing:** Designed for designers, photographers, and creative minds, OmniVerse offers the unique ability to generate images, perform intricate image segmentation, and make real-time

edits. This feature unleashes the full spectrum of creative potential while streamlining graphic design processes. This feature is a dream come true for designers, photographers, and creative minds. It enables users to effortlessly generate images, perform intricate image segmentation, and make real-time edits, all within a single platform. This functionality unleashes the full spectrum of creative potential while streamlining graphic design processes. Whether you are creating stunning visuals for a marketing campaign, need to edit photos for a project, or want to experiment with image segmentation for various applications, OmniVerse is your all-in-one creative toolkit. **Audio Generation:** Audio generation, at its core, entails the digital synthesis of sound. It empowers musicians and producers to create unique, intricate, and experimental audio compositions that traditional recording methods cannot replicate. This process is not merely on the cutting edge but at the forefront of modern music production, driving innovation in the industry. OmniVerse's audio generation feature is a game-changer in music production. Craft unique sounds and custom music compositions with ease, exploring new horizons in audio creativity. The significance of audio generation in the music industry is profound. It stimulates experimentation, the evolution of new music genres, and the ability to produce high-quality music at a lower cost. By democratizing the art of music creation, audio generation has catalyzed a new era of sonic innovation. **Live Student Monitoring via Facial Recognition:** The project focused on the development and implementation of a Live Camera Detection System designed to recognize student faces and verify their identities to monitor their attendance during scheduled classroom lectures. The primary objective was to address the challenge of ensuring punctuality and attendance compliance, with the system automatically triggering notifications to alert absent students on their registered mobile numbers within the lecture designated time slot. The project aimed to modernize attendance management, leverage technology to promote student accountability, and adapt to the ever-evolving landscape of educational delivery.

```
2023-12-17 01:35:01 INFO:Loading LLAMA-2-7B...
2023-12-17 01:35:01 INFO:llama.cpp weights detected: models/LLAMA-2-7B/llama-2-7b-chat.Q3_K_S.gguf
2023-12-17 01:35:01 INFO:Cache capacity is 0 bytes
ggml_init_cublas: found 1 CUDA devices:
  Device 0: NVIDIA GeForce RTX 3050 Ti Laptop GPU, compute capability 8.6
llama_model_loader: loaded meta data with 19 key-value pairs and 291 tensors from models/LLAMA-2-7B/llama-2-7b-chat.Q3_K_S.gguf (version GGUF V2 (latest))
llama_new_context_with_model: total VRAM used: 288.00 MB (model: 0.00 MB, context: 288.00 MB)
AVX = 1 | AVX2 = 1 | AVX512 = 0 | AVX512_VBMI = 0 | AVX512_VNNI = 0 | FMA = 1 | NEON = 0 | ARM_FMA = 0 | F16C = 1 | FP16_VA = 0 | WASM_SIMD = 0 | BLAS = 1 | SSE3 = 1 | SSSE3 = 1 | VSX = 0 |
2023-12-17 01:36:36 INFO:Loaded the model in 95.28 seconds.
```

Fig 2: Loading the model

4. METHODOLOGY

The methodology centers around the development and deployment of two distinct systems: OmniVerse, an AI-powered web application with advanced features, and a Live Student Monitoring system that utilizes facial recognition technology to enhance classroom attendance monitoring.

All AI features will run on-device and not on any cloud computer. The system will utilize CUDA via Ubuntu running on WSL (Windows Subsystem for Linux) on a Windows Machine. CUDA (Compute Unified Device Architecture) is a software framework developed by NVIDIA that allows software to use GPUs for general-purpose processing. This

significantly enhances inference time and reduces latency, leading to an overall greater user experience.

Text and Code Generation: To implement this feature, the system utilizes a text and code generation engine that leverages advanced machine learning algorithms. This engine will be capable of assisting users in drafting articles, generating code snippets, and automating repetitive tasks. Extensive algorithm development, testing, and fine-tuning needs to be conducted to ensure high-quality outputs. The size of the model, the available system RAM and GPU capacity need to be taken into consideration to ensure the system is able to handle the model.

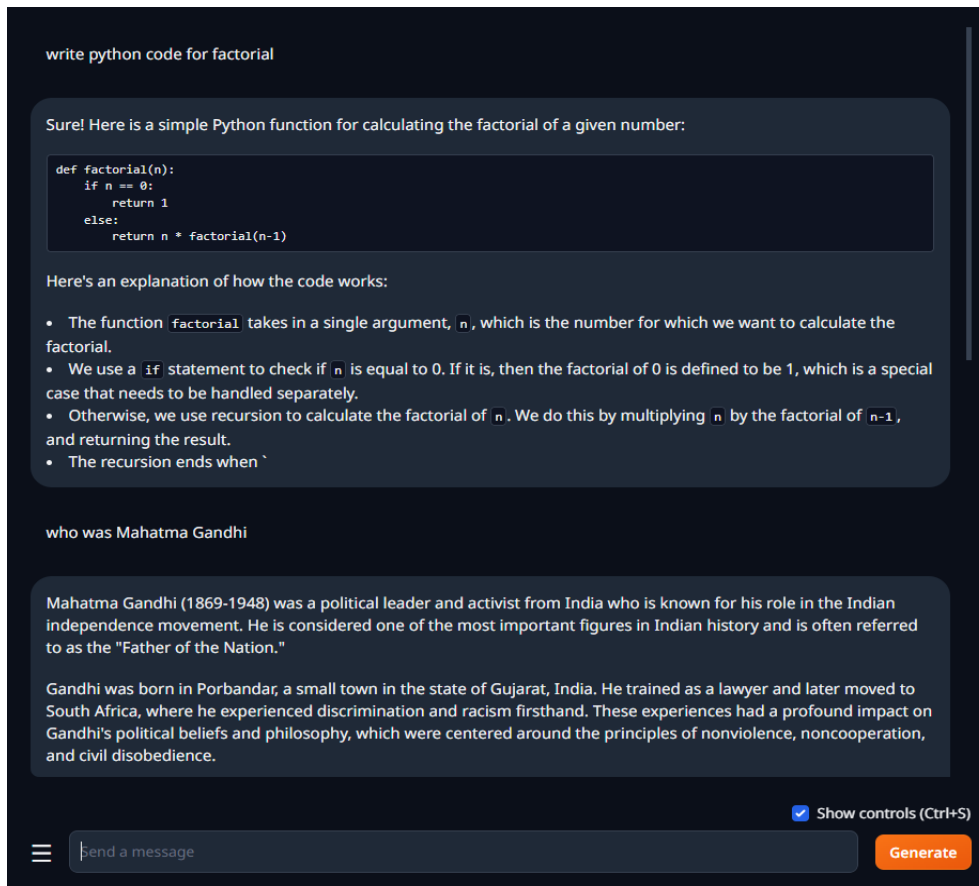


Fig 3: Local, Offline Text Generation LLM

Image Generation, Segmentation, and Editing: The image generation, segmentation, and editing feature will involve the creation of an image manipulation tool. This tool will be designed for designers, photographers, and creative professionals. Image generation algorithms will provide real-time editing capabilities to meet the creative demands of users. The tool will allow for image synthesis via a text prompt, image segmentation and modification via text prompt. The quality of the image and the time taken to generate will depend upon the size of the model and the user's system specification.

Audio Generation: The audio generation feature will focus on digital sound synthesis, allowing users to create unique audio compositions. Development will include advanced algorithms and sound generation techniques, ensuring versatility and ease of use for musicians and producers.

Facial Recognition System Development: The live student monitoring system will require the development of a robust facial recognition system. This will involve creating facial recognition algorithms and integration with camera hardware for real-time detection.

Student Identity Verification: In addition to facial recognition, a student identity verification system will be implemented to ensure that the detected face matches the registered student identity. Notification System: For notifying absent students, a notification system will be integrated. This system will send notifications to registered mobile numbers when students are not present within the designated lecture time slot.

5. CONCLUSION

OmniVerse is more than just a web app; it's an odyssey into the boundless possibilities of AI, a gateway to turbocharge productivity, and a realm to unleash the full spectrum of creativity. With AI-driven features in text, image, and audio generation, OmniVerse offers an unparalleled user experience. Its promise in a burgeoning AI market, unwavering commitment to innovation, and dedication to meeting the ever-evolving needs of its users make it a transformative tool for individuals looking to streamline tasks and explore their creative horizons.

The project further aims to develop a mobile app for Android devices that will perform all the functions on-device. Mobile hardware has come a long way and is now strong enough to hopefully run the powerful models with manageable latency with insignificant power consumption. An Android app will enable the masses to access these powerful technologies at their fingertips whenever they want in a secure, offline and free manner.

6. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards the development of the template.

7. REFERENCES

- [1] <https://ai.meta.com/resources/models-and-libraries/llama/>
- [2] <https://www.oracle.com/in/artificial-intelligence/generative-ai/what-is-generative-ai/>
- [3] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal

- Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, Guillaou. LLaMA: Open and Efficient Foundation Language Models
- [4] Baptiste Rozière, Jonas Gehring, Fabian Gloeckle, Sten Sootla, Itai Gat, Xiaoqing Ellen Tan, Yossi Adi, Jingyu Liu, Tal Remez, Jérémy Rapin, Artyom Kozhevnikov, Ivan Evtimov, Joanna Bitton, Manish Bhatt, Cristian Canton Ferrer, Aaron Grattafiori, Wenhan Xiong, Alexandre Défossez, Jade Copet, Faisal Azhar, Hugo Touvron, Louis Martin, Nicolas Usunier, Thomas Scialom, Gabriel Synnaeve. Code Llama: Open Foundation Models for Code
- [5] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models
- [6] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, Alexandre Défossez. Simple and Controllable Music Generation
- [7] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Léo Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, William El Sayed. Mistral 7B