**SkinGPT: An AI-Powered Diagnostic System for Skin Condition Analysis and Recommendations**

**Team Members:**

*Adithi Kallem*

*Apoorva Jayanthi*

*Pranay Reddy Ala*

*Rohith Panjala*

*Sushma Ganapuram*

**AIT-526-001 | Natural Language Processing**

**Professor: Dr. Ray Islam**

**Date: May 8, 2024**

**George Mason University**

***Abstract —* SkinGPT represents a groundbreaking advancement in dermatological diagnostics, merging state-of-the-art machine learning and natural language processing to revolutionize the way skin diseases are diagnosed and understood. At its core, SkinGPT employs YOLOv5, an advanced image classification technology, which enables precise disease identification from digital images. Complementing this is the LLaMA model, utilized for its robust natural language generation capabilities, allowing SkinGPT to provide detailed, accessible explanations and health information through a user-friendly chatbot interface. This dual functionality facilitates real-time patient interaction, significantly enhancing patient engagement and satisfaction. Moreover, SkinGPT assists dermatologists by offering reliable second opinions, thus increasing diagnostic accuracy and confidence in treatment plans. By leveraging these technologies, SkinGPT not only streamlines the diagnostic process but also enhances the accessibility of dermatological care, particularly in underserved areas where specialist access is limited. This system sets a new standard for telemedicine applications in dermatology, offering a scalable solution that could be extended to other areas of medicine in the future.**

***Keywords — SkinGPT, Dermatological Diagnostics, Machine Learning, Natural Language Processing, YOLOv5, LLaMA Model, Telemedicine, Patient Engagement.***

1. **INTRODUCTION**

Skin diseases are prevalent worldwide, affecting millions across all demographics and geographies. The traditional diagnostic method through visual examination by dermatologists often faces limitations due to the uneven availability of specialists, particularly in rural and underserved areas. SkinGPT addresses these challenges by integrating advanced machine learning and natural language processing technologies. Using YOLOv5 for precise image classification and the LLaMA model for generating comprehensible natural language responses, SkinGPT enhances the diagnosis and understanding of skin conditions.

This innovative tool not only supports dermatologists by providing accurate second opinions but also empowers users with immediate access to health information, significantly improving patient engagement and satisfaction. The inspiration for SkinGPT stems from the potential of artificial intelligence to transform healthcare, offering a scalable and cost-effective solution for early disease detection and patient education.

By bridging the gap in healthcare accessibility, SkinGPT is poised to revolutionize dermatological care, making advanced diagnostics available even in areas with limited medical resources. Its development highlights the transformative impact of technology on global health, providing a model for future advancements in telemedicine.

1. **RELATED WORK**
2. **SkinGPT-4: An Interactive Dermatology Diagnostic System with Visual Large Language Model:** SkinGPT-4 represents a significant advancement in dermatological AI systems by integrating a vision-based large language model for autonomous disease classification and detailed report generation. Unlike previous systems reliant on dermoscopic imaging and subsequent analysis by dermatologists, SkinGPT-4 processes skin images directly to identify and diagnose conditions. This not only enhances diagnostic efficiency but also improves patient interaction through immediate, understandable feedback. A notable innovation of SkinGPT-4 is its commitment to data privacy, facilitated by local deployment options that prevent data breaches and increase accessibility in remote areas. Trained on an extensive dataset of skin images, clinical concepts, and doctors' notes, SkinGPT-4 achieves accuracy comparable to professional dermatologists. Its interactive capabilities allow for dynamic user engagement, further enhancing diagnostic processes. As such, SkinGPT-4 not only extends the reach of dermatological services but also exemplifies the potential of AI to revolutionize medical diagnostics.
3. **Dermatologist-level classification of skin cancer with deep neural networks:** Recent breakthroughs in dermatological AI have been epitomized by the study "Dermatologist-level classification of skin cancer with deep neural networks," which demonstrates the use of convolutional neural networks (CNNs) to classify skin cancer at a level comparable to board-certified dermatologists. This research utilized a CNN-trained end-to-end on a dataset of 129,450 clinical images spanning 2,032 different diseases, a significantly larger set than those used in prior studies. The CNN was able to match dermatologist performance in distinguishing between keratinocyte carcinomas and benign seborrheic keratoses, as well as malignant melanomas and benign nevi. Notably, this system holds potential for deployment on mobile devices, potentially offering universal access to crucial diagnostic care. This technology underscores a significant step toward enhancing the accuracy and accessibility of skin cancer diagnostics, leveraging large datasets and advanced machine learning algorithms to possibly extend the reach of dermatological care beyond traditional clinical settings.

**III. OBJECTIVES**

The SkinGPT project is anchored by three fundamental objectives, each aimed at enhancing dermatological diagnostics using cutting-edge artificial intelligence and natural language processing technologies:

1. **Disease Identification:** The cornerstone of the SkinGPT project is to develop a sophisticated AI model that leverages deep learning algorithms, specifically utilising a robust dataset of dermatological images. This model is designed to not only identify but also classify a wide array of skin diseases with high accuracy. The focus is on ensuring that the model can handle diverse skin types and conditions, thereby reducing diagnostic disparities and improving healthcare outcomes.
2. **Information Provision:** Beyond disease identification, SkinGPT aims to implement an advanced NLP system that can interpret the AI's findings and translate them into detailed, precise, and easily comprehensible responses. This system is intended to educate patients about their diagnosed conditions by providing insights into symptoms, potential causes, and treatment options. The goal is to make medical information more accessible and understandable, thus empowering patients with knowledge about their health.
3. **User Interaction:** A key goal of the project is the design and development of a user-friendly chatbot interface. This chatbot will facilitate seamless interactions between the users and the system, allowing for easy uploading of images and efficient retrieval of health information. The interface will be intuitive, requiring minimal navigation, thereby enhancing the user experience, and encouraging engagement. This chatbot serves as the primary point of interaction, ensuring that users can effortlessly communicate with SkinGPT, making medical diagnostics more approachable and less intimidating.

**IV. THE DATASET**

The effectiveness of the SkinGPT project is heavily reliant on a robust dataset sourced from Roboflow Universe, which contains 2,384 dermatological images with 12 classes. This dataset was carefully curated to encompass a wide range of skin conditions, each accurately labeled to support effective supervised learning. It includes common ailments like acne and eczema, as well as less prevalent diseases such as psoriasis and melanoma, ensuring broad training scope for the AI model. Importantly, the dataset mimics real-world scenarios, enhancing the model's accuracy and reducing bias across different demographics and skin types. This preparation allows SkinGPT to perform reliably across varied conditions, providing dependable diagnostic support globally, thus adapting to the diverse needs of users worldwide.

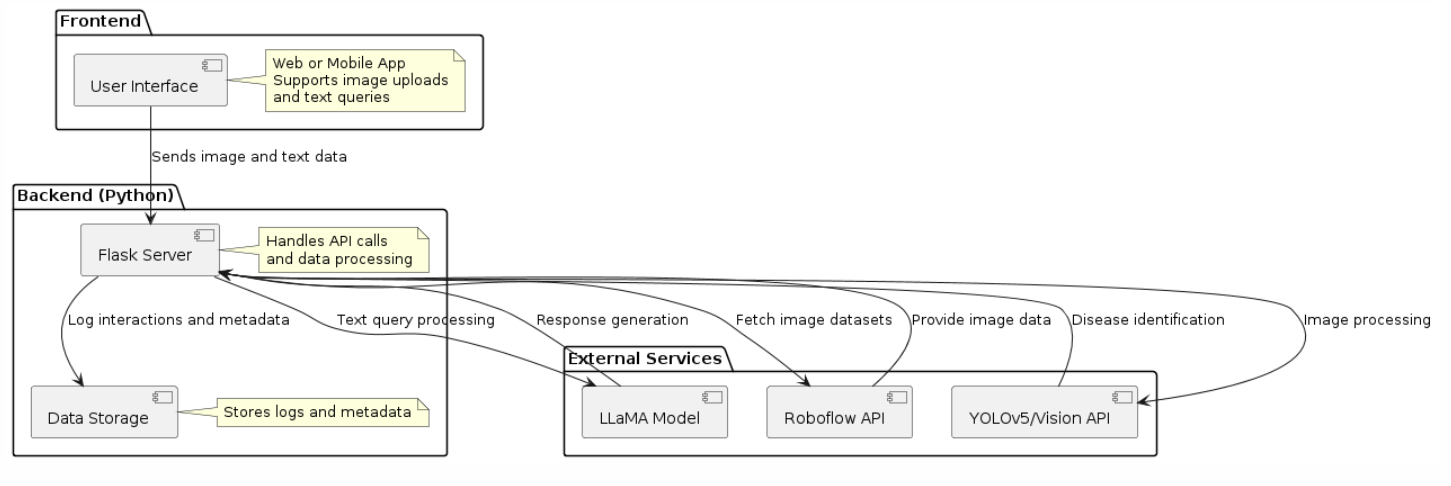
**V. SYSTEM ARCHITECTURE**

**A. Key Components:**

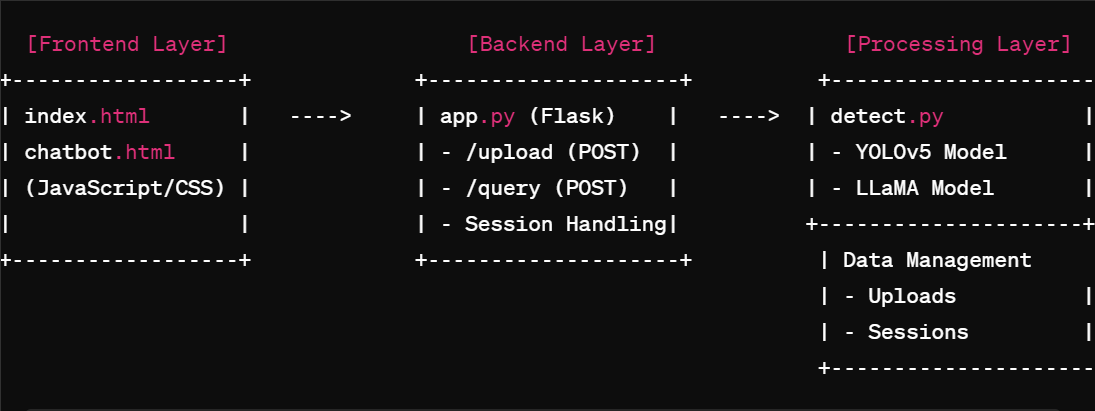
SkinGPT is engineered with a modular architecture that integrates multiple advanced technological components, meticulously structured to improve the efficiency and effectiveness of the diagnostic process. Here are the key components of the system:

1. **User Interface (UI):** The user interface (UI) layer is constructed using React, a JavaScript library that excels in building dynamic interfaces, and Bootstrap, a responsive design framework. This combination ensures that the chatbot interface is intuitive, user-friendly, and accessible across a variety of devices, enhancing user engagement through seamless navigation and interaction.
2. **Image Processing Module:** At the core of the system's diagnostic capabilities is the YOLOv5 model, a state-of-the-art deep learning framework renowned for its speed and accuracy. Optimized for high throughput, which is crucial for real-time applications, this module analyzes images to detect and classify various skin conditions, forming the primary diagnostic tool of the system.
3. **NLP Module:** The NLP capabilities are powered by the LLaMA model, utilized through the Hugging Face Transformers library, which is a leading platform for building transformer-based machine learning models. This module enhances the system’s ability to generate accurate and contextually relevant responses based on the image analysis, effectively communicating complex medical information in a user-friendly manner.
4. **Data Management:** The backend architecture, developed with Python using the Flask framework, manages the seamless flow of data between the front-end interface and the processing modules. It handles API interactions and ensures robust data security and integrity, crucial for maintaining user seamless conversation flow.

**B. Architecture/Framework:**



***Fig 1: System Architecture Flowchart***



***Fig 2: System Architecture Layout***

The architecture of SkinGPT is meticulously designed to harmoniously integrate image processing and natural language processing (NLP) capabilities while offering a highly intuitive and user-friendly interface. The user interface (UI) layer comprises two essential pages: `index.html` and `chatbot.html`. The `index.html` page acts as the primary point of interaction, where users can upload images via a form for diagnostic analysis. Once an image is submitted, it is processed through the backend application layer, which relays diagnostic results back to the user. Meanwhile, the `chatbot.html` page provides a conversational interface for users to query the chatbot, receive diagnostic explanations, and understand treatment recommendations based on the diagnostic results.

The backend application layer is managed by `app.py`, a Python Flask application that provides essential routing and API endpoints for handling requests from the UI layer. The `/upload` endpoint processes image uploads via the `detect.py` module, which incorporates a state-of-the-art YOLOv5 model for efficient image analysis and classification of various skin conditions. The `/query` endpoint handles user queries using the LLaMA model, a sophisticated NLP model that translates diagnostic information into layman-friendly terms and provides relevant recommendations.

Within the processing layer, `detect.py` houses both the YOLOv5 and LLaMA models. YOLOv5 processes the uploaded images to detect and classify various skin conditions before returning the results to the backend for further analysis. The LLaMA model manages natural language comprehension and delivers accurate, comprehensive responses. To ensure chat continuity and personalization, the data management layer, integrated into the backend, securely stores images and retains session data across different user interactions.

Moreover, the modular architecture of SkinGPT guarantees scalability and maintainability. By isolating each layer while enabling seamless communication between them, the system provides a cohesive, smooth, and intuitive user experience. This structure ensures that future updates and improvements to the diagnostic and chat functionalities can be easily implemented without disrupting the system's overall performance or integrity.

**C. NLP and Data Analytics Approaches:**

The natural language processing (NLP) and data analytics approaches implemented in SkinGPT are integral to its ability to provide accurate and relevant diagnostic information to users. The NLP module is powered by the LLaMA model, a sophisticated transformer-based language model, leveraging the Hugging Face Transformers library. This platform is renowned for building state-of-the-art machine learning models. By utilizing LLaMA's advanced language understanding capabilities, SkinGPT can comprehend user queries and respond with contextually accurate and comprehensive information.

The LLaMA model is fine-tuned to generate responses that effectively translate complex medical information into layman-friendly terms, enabling users to understand their conditions and possible treatments clearly. Its language comprehension is enhanced with context provided by previous interactions, allowing for continuous and meaningful conversations with users.

The core of SkinGPT's image processing pipeline is built on a highly efficient deep learning framework, YOLOv5. The model summary indicates that the architecture consists of 157 layers with 7,034,398 parameters and 0 gradients, achieving a computational performance of 15.8 GFLOPs. This scale and complexity underscore the model's ability to process and analyze images accurately in real-time, enabling effective classification and diagnosis of various skin conditions. By balancing depth and computational efficiency, the model is optimized to deliver accurate results with minimal latency, crucial for a responsive and user-friendly diagnostic system.

The combination of the NLP and image analytics modules ensures that SkinGPT offers a comprehensive diagnostic experience. The NLP module complements the image processing module by offering explanations, advice, and recommendations based on the diagnostic results, thereby enhancing the system’s ability to provide accurate and contextually relevant responses. Together, these components make SkinGPT a reliable tool for translating raw diagnostic data into actionable insights for users.

**D. Software/ Hardware Development Platforms:**

SkinGPT is built on a carefully selected set of software and hardware development platforms to deliver an efficient, scalable, and user-friendly system. The software platform primarily relies on Python for backend development, leveraging the Flask framework for building web applications. Flask enables rapid development and provides a lightweight yet flexible framework that can efficiently handle API requests, manage routing, and ensure seamless communication between the UI and backend components.

For natural language processing and image analytics, SkinGPT integrates leading-edge machine learning frameworks. The LLaMA model is employed to interpret and respond to user queries effectively, wherein, for image analytics, the YOLOv5 model is used, which is implemented through the PyTorch framework. YOLOv5's high throughput and accuracy ensure that SkinGPT can process and classify skin conditions in real time.

On the front end, standard web technologies like HTML, CSS, and JavaScript are employed for the user interface. Combined with Bootstrap for responsive design, this setup provides a streamlined, intuitive, and mobile-friendly experience, ensuring that users can easily interact with the system across devices.

In terms of hardware, SkinGPT is optimized to run on standard server infrastructure and can leverage GPU acceleration for faster machine learning inference, particularly with the YOLOv5 model. This hardware configuration ensures efficient, high-throughput diagnostic processing and rapid response generation, enabling the system to handle multiple concurrent user sessions seamlessly.

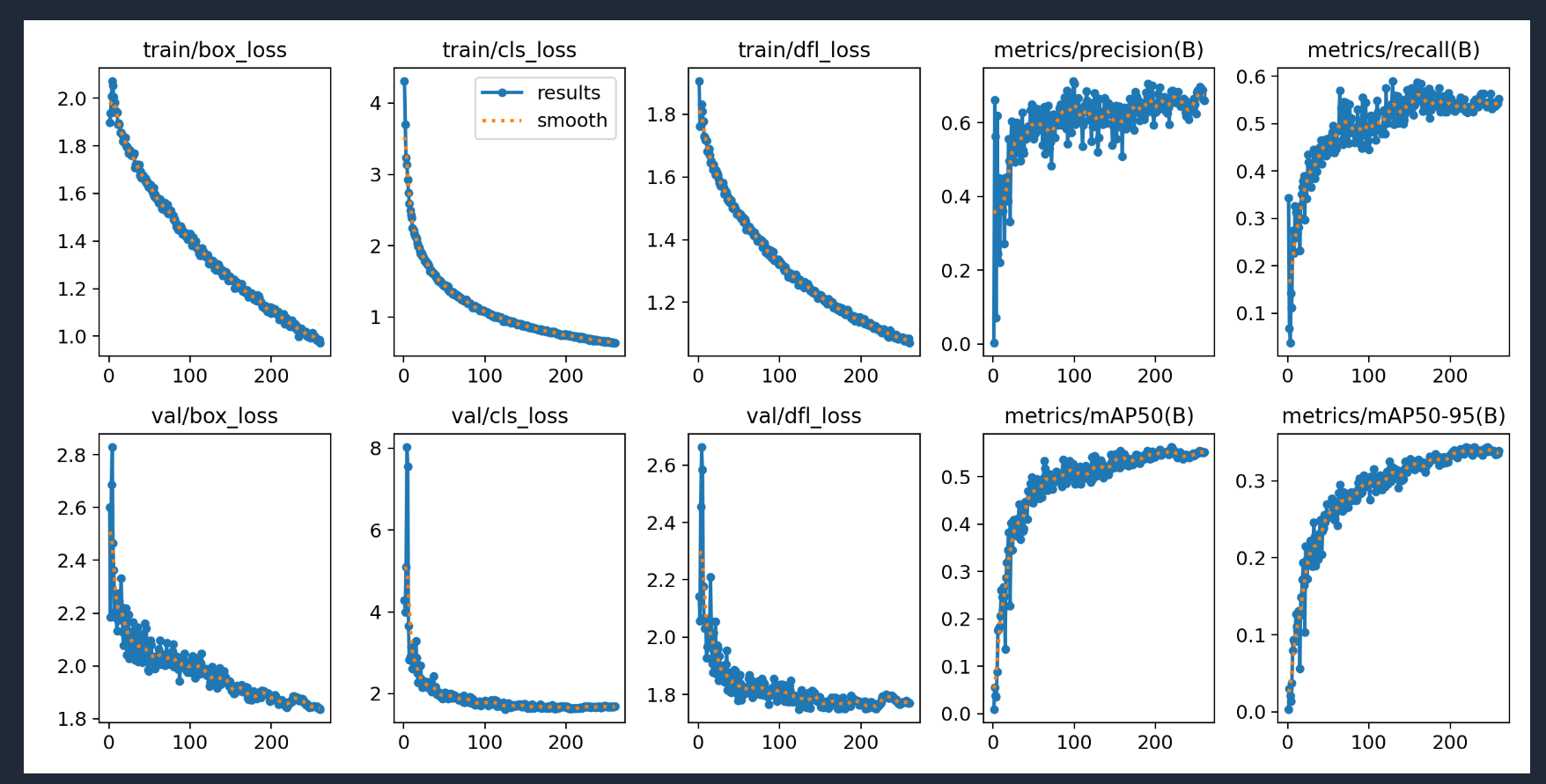
**VI. EXPERIMENTAL RESULTS AND ANALYSIS**

1. **Dataset Analysis:**

The dataset used for training and evaluating SkinGPT's performance was curated to encompass a diverse range of skin conditions while ensuring sufficient samples for accurate classification. The dataset was divided into three main subsets: training, validation, and test. The training set comprised 84% of the total dataset, consisting of 2,384 images, allowing the model to learn from a wide variety of examples and improving its ability to generalize to new data. The validation set accounted for 9% of the total dataset, with 264 images used to tune hyperparameters and monitor the model’s performance during training. The final test set, constituting 7% of the dataset with 193 images, was employed to assess the model’s real-world performance and gauge its accuracy, precision, and recall in identifying various skin conditions.

The training dataset underwent preprocessing, including resizing and standardizing the images to ensure consistency. Data augmentation techniques such as rotation, flipping, and scaling were applied to artificially increase the dataset size and improve model robustness. This careful curation and augmentation of the data helped reduce overfitting and enhanced the model’s generalization ability.

The results from this analysis were used to refine the model architecture and training strategy, ensuring that SkinGPT could provide accurate diagnostics and comprehensive recommendations based on real-world data. These dataset splits enabled a thorough evaluation of the system's performance across different stages of development, ultimately contributing to the reliability of the final diagnostic tool.



***Fig 3: Training Graphs***

1. **Model Performance Evaluation:**

The experimental evaluation of SkinGPT's performance provides valuable insights into the efficacy of the system's image classification and diagnostic capabilities. Training loss curves for box loss, classification loss, and objectness loss showed consistent and significant declines throughout the training process. For instance, box loss reduced from approximately 2.0 to 1.0, classification loss dropped from 4.0 to just above 1.0, and objectness loss displayed a steady improvement. Similarly, the validation loss curves followed a downward trend, affirming the model's ability to learn and generalize effectively. The final validation losses settled at about 1.8 for box loss and below 2.0 for classification loss, indicating successful model optimization.

When it comes to performance metrics, the system's precision, a measure of the accuracy of detected results, stood at 66.7%, indicating that over two-thirds of the classifications were accurate. However, recall, which measures the model's capacity to identify positive cases, was slightly lower at 53.6%, highlighting opportunities for improvement in detection sensitivity. The mean average precision (mAP), which aggregates precision-recall curves at various IoU thresholds, achieved a score of 56.2%. This value demonstrates reliable classification performance but suggests that additional optimization could yield better results.

In conclusion, the experimental results affirm the reliability of SkinGPT in diagnosing common skin conditions, backed by steady training loss curves and promising metrics. Moving forward, refining the model architecture and incorporating more diverse data could further enhance the accuracy and response generation, leading to improved diagnostic precision.

**VII. ERRORS AND OPEN QUESTIONS**

Despite SkinGPT's promising performance, several challenges and open questions remain. False negatives occur when subtle visual differences between certain skin conditions go undetected, leading to lower recall scores. This issue is likely due to the insufficient representation of these conditions within the training dataset. Additionally, false positives are observed when certain skin conditions are misclassified because of visual similarities with others, indicating a need for more refined data augmentation and improved discriminatory features. The potential for dataset bias is also present, as the data may not represent all skin types and demographics equally, potentially affecting the model's predictions. Collecting a more diverse dataset could help enhance the system's generalization. On the natural language processing side, while the LLaMA model generates comprehensive responses, complex or ambiguous queries sometimes lead to inaccurate or irrelevant recommendations. Refining the NLP model to handle edge cases more effectively and providing clearer responses for vague queries are critical challenges that remain. Addressing these issues will ensure that SkinGPT continues to improve its diagnostic accuracy and delivers reliable recommendations across a wider range of conditions and users.

**VIII. CHALLENGES**

1. **Asynchronous Data Processing:** When processing image analysis results and natural language responses concurrently, asynchronous data handling may lead to inconsistencies in user-facing outputs. Analyzing high-resolution images can be time-consuming, which sometimes delay responses and cause the conversation context to fall out of sync with diagnostic results.
2. **Session State Management:** Retaining consistent session data across multiple user queries is challenging, particularly when switching between different conversation states. Misalignment between diagnostic results and follow-up queries may lead to irrelevant or inaccurate responses.
3. **Distributed System Latency:** Communication delays between the frontend and backend services, especially when multiple concurrent requests are handled, resulting in time lags and negatively affect user experience.

**IX. SOLUTIONS/ FUTURE WORK**

1. **Queue-Based Processing:** Implement message queues (e.g., RabbitMQ or Redis) to ensure that each task is handled in an ordered, first-come-first-served manner. This method helps manage multiple concurrent tasks effectively by buffering incoming requests and delivering processed data in sequence.
2. **Contextual State Management:** Use state management solutions to maintain consistent session data and align user interactions with accurate diagnostic information. Technologies like Redis or in-memory caching can help store and synchronize data efficiently between modules.
3. **Microservices and API Rate Limiting:** Decompose monolithic applications into microservices to independently scale components based on usage, preventing latency from overloading any one service. API rate limiting ensures that all data processing modules handle only the amount of data they can reasonably process, reducing inconsistencies due to overloaded services.
4. **Data Consistency Checks:** Periodically verify the integrity of data using checksum comparisons or other validation methods. Logging and monitoring tools can identify discrepancies and help resolve data sync issues quickly.

**X. CONCLUSION**

In developing SkinGPT, significant progress has been made in creating a comprehensive diagnostic tool that integrates state-of-the-art image classification and natural language processing capabilities. By leveraging the YOLOv5 model for image analysis and the LLaMA model for conversational interactions, the system efficiently detects and classifies skin conditions, while providing user-friendly explanations and treatment recommendations. The architecture, built on modular principles, ensures that each component can be scaled and maintained effectively, enabling a smooth and intuitive user experience. However, while SkinGPT demonstrates reliable diagnostic performance, there remains potential for enhancing detection sensitivity and expanding the dataset to include a broader range of skin conditions across diverse demographics. Overall, the development and evaluation of SkinGPT affirm that it is well-suited for diagnosing common skin conditions, providing actionable insights, and improving user understanding.

**XI. LESSONS LEARNT**

Throughout the development process, several key lessons emerged that will inform future improvements to the system.

1. Data diversity is critical for accurate diagnostic performance. A more representative dataset that captures a wide range of skin types and conditions will be essential for reducing bias and improving classification accuracy.
2. Careful attention must be paid to the balance between precision and recall. While high precision ensures that detected conditions are accurate, enhancing recall will minimize false negatives, leading to more comprehensive diagnostics.
3. Clear and contextual user communication is vital for maintaining trust and engagement. The LLaMA model's ability to translate medical terminology into accessible language proved valuable, yet refining responses to ambiguous queries remains a priority.
4. Finally, maintaining synchronized data between the diagnostic and conversational components is crucial for delivering consistent, relevant information.

**REFERENCES**

**[1]** J. Zhou *et al.*, “SkinGPT-4: An Interactive Dermatology Diagnostic System with Visual Large Language Model,” *arXiv.org*, Apr. 21, 2023. <https://arxiv.org/abs/2304.10691>

**[2]** A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, Jan. 25, 2017. <https://www.nature.com/articles/nature21056>

**[3]**“PSD-1-.jpg,”*Roboflow*. <https://universe.roboflow.com/skin-disease-ysval/skin-disease-detection-s7zik>

**[4]** “OpenAI.” <https://openai.com/>

**[5]** PlantUML, “Open-source tool that uses simple textual descriptions to draw beautiful UML diagrams.,” *PlantUML.com*. <https://plantuml.com/>