iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: VII Month of publication: July 2026

DOI: https://doi.org/10.22214/ijraset.2026.84429

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

InfratGPT an Intelligent Full Stack AI Application with Image Generation Using React.JS

Gunturi Dileep Varma[1], P. Bindu Priya[2]
[1]M.Tech Student, Department of CSE, Sanketika Vidya Parishad Engineering College, P.M Palem, Madhurawada Visakhapatnam, Andhra Pradesh, India
[2]Assistant Professor, Department of CSE (M.Tech), Sanketika Vidya Parishad Engineering College, P.M Palem, Madhurawada Visakhapatnam, Andhra Pradesh, India

Abstract: InfratGPT is an intelligent AI chat application designed to deliver fast, accurate, and context-aware responses to user queries. By supporting both interactive communication and image generation, the system creates a rich, multimodal user experience that bridges the gap between text and visual communication. This paper presents a comprehensive overview of the system architecture, development process, AI service integrations, and implementation strategies, highlighting InfratGPT's potential to transform human computer interaction. To deliver a responsive user interface alongside efficient server-side processing, the system employs a React.js frontend paired with a Node.js backend. Natural language understanding and intelligent text generation are driven by the Google Gemini API, while ImageKit handles real-time image generation and media optimization enabling seamless multimodal delivery of both text and visual outputs. Data management is handled by MongoDB, which securely and efficiently stores user data and interaction histories. The system is hosted on Vercel, leveraging its high-performance cloud infrastructure to ensure low latency, high availability, and seamless scalability. Overall, the architecture emphasizes modularity, maintainability, and performance optimization establishing a robust foundation for real-world deployment and future enhancements
Keywords: Conversational AI, Google Gemini, Image Generation, React.js, Node.js, Natural Language Processing, ImageKit, Stripe, MongoDB, Real-time Response System, Full Stack Development, Multimodal AI, Vercel, Stripe Webhooks.

I. INTRODUTION

Artificial Intelligence (AI) has fundamentally transformed human–computer interaction, enabling systems to comprehend, process, and generate human-like language. The rise of Conversational AI driven by recent breakthroughs in Natural Language Processing (NLP) and Large Language Models (LLMs) has revolutionized real-time communication across diverse sectors, including education, healthcare, software engineering, and business operations. However, modern user expectations have moved beyond purely text-based interfaces. Present-day applications demand multimodal capabilities, where visual content creation seamlessly complements natural language understanding to deliver richer, more immersive interactions.

To address these evolving demands, we present InfratGPT, an intelligent, multimodal conversational platform. Powered by the Google Gemini API for deep natural language processing and integrated with ImageKit for real-time visual generation, InfratGPT enables users to converse, query, and generate custom visuals using intuitive text prompts. The system is implemented using a full-stack architecture featuring a React.js user interface, a Node.js backend, MongoDB for secure interaction storage, and Vercel for low-latency, scalable cloud hosting.

II. LITERATURE REVIEW

Recent progress in Artificial Intelligence (AI) and Natural Language Processing (NLP) has transformed conversational systems from simple rule-based programs into intelligent agents capable of nuanced, context-aware human interaction. These applications are now central to workflows across customer service, healthcare, education, and software development.

This shift was primarily catalyzed by Vaswani et al. (2017), whose landmark paper *"Attention Is All You Need"* introduced the self-attention-based Transformer architecture. By eliminating the sequence dependencies inherent to traditional Recurrent Neural Networks (RNNs), the Transformer model dramatically improved training efficiency and contextual modeling across long text sequences. Leveraging this architectural shift, Large Language Models (LLMs) like the Generative Pre-trained Transformer (GPT) family brought about a new era in natural language understanding. Trained on massive datasets, these models exhibit state-of-the-art proficiency in question answering, dialogue management, and summarization establishing the foundational technology behind modern interactive AI platforms.

Google's Gemini AI represented a key milestone in conversational intelligence by combining large-scale language understanding with native multimodal capabilities. By processing and generating information across both text and visual mediums simultaneously, Gemini enables developers to create far more interactive applications than was previously possible with text-only architectures.

In parallel, advancements in deep learning have accelerated the quality and availability of automated image generation systems. Today's generative models produce high-quality visual content from simple natural language prompts, seeing widespread adoption across education, digital marketing, and design. Combining conversational dialogue with visual generation directly enhances user engagement by delivering unified text and image outputs in a single interface. Despite these technological strides, many current AI assistants continue to suffer from limitations in response accuracy, contextual memory, infrastructure scalability, and unified multimodal support. To bridge these gaps, modern application design focuses on coupling state-of-the-art multimodal APIs with scalable backend architectures and cloud-native hosting. Synthesizing these insights, **InfratGPT** was built to deliver an efficient, scalable, and real-time conversational experience. By uniting the Google Gemini API, image generation capabilities, and a full-stack web architecture (React.js, Node.js, MongoDB, and Vercel cloud hosting), InfratGPT responds directly to current research trends—demonstrating a practical, unified solution for multimodal human–computer interaction.

III. EXISTING SYSTEM

Traditional chatbots and early conversational agents rely heavily on hardcoded scripts, simple keyword matching, or shallow machine learning algorithms. Consequently, these systems fail to process complex natural language queries, retain context across multi-turn dialogues, or generate dynamic, nuanced responses. While state-of-the-art Large Language Models (LLMs) such as ChatGPT and Google Gemini—have dramatically advanced language comprehension, many consumer-facing implementations remain constrained to text-only interactions. They often lack seamless, real-time integration with visual synthesis tools within a single, unified workflow.

IV. PROPOSED METHOD

To address the limitations of conventional text-only AI assistants, we present InfratGPT a full-stack, multimodal conversational platform that unifies natural language processing and real-time image synthesis within a single environment. The platform integrates the Google Gemini API for contextual reasoning and intelligent dialogue generation, alongside ImageKit for prompt-based image creation and asset management. Built on a modular full-stack stack (React.js for the frontend and Node.js for the backend), InfratGPT ensures high responsiveness and seamless client-server communication.

V. METHODOLOGY

1) Client Interaction & Payload Framing
- User Input: The process initiates when the user enters a natural language text query or an image generation prompt via the web client.
- Frontend Processing: The React.js interface captures the input, updates local UI loading states, and packages the payload including user query, conversation history context, and request type into a structured JSON HTTP request.

2) Backend Orchestration & Decision Logic
- Client-to-Server Routing: The payload is transmitted to the Node.js/Express.js backend server.
- Validation & Dispatch: The backend validates the request, retrieves chat context, and branches execution based on query intent:
- Text Generation: If classified as a text query, the prompt and context are forwarded to the Google Gemini API, which processes the sequence and returns an intelligent text response.
- Visual Synthesis: If classified as an image prompt, the backend dispatches the request to ImageKit, which synthesizes the graphic asset and returns a direct image URL.

3) Persistence & UI Rendering
- Data Persistence: Interaction logs, including user prompts, generated outputs, session metadata, and conversation states, are recorded asynchronously in MongoDB.
- Response: The backend compiles the final output (text payload or visual URL) into a JSON packet and sends it back to the client.
- Interface Rendering: The React.js frontend dynamically renders the output presenting text in conversational bubbles or displaying visual previews while updating the active session memory.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

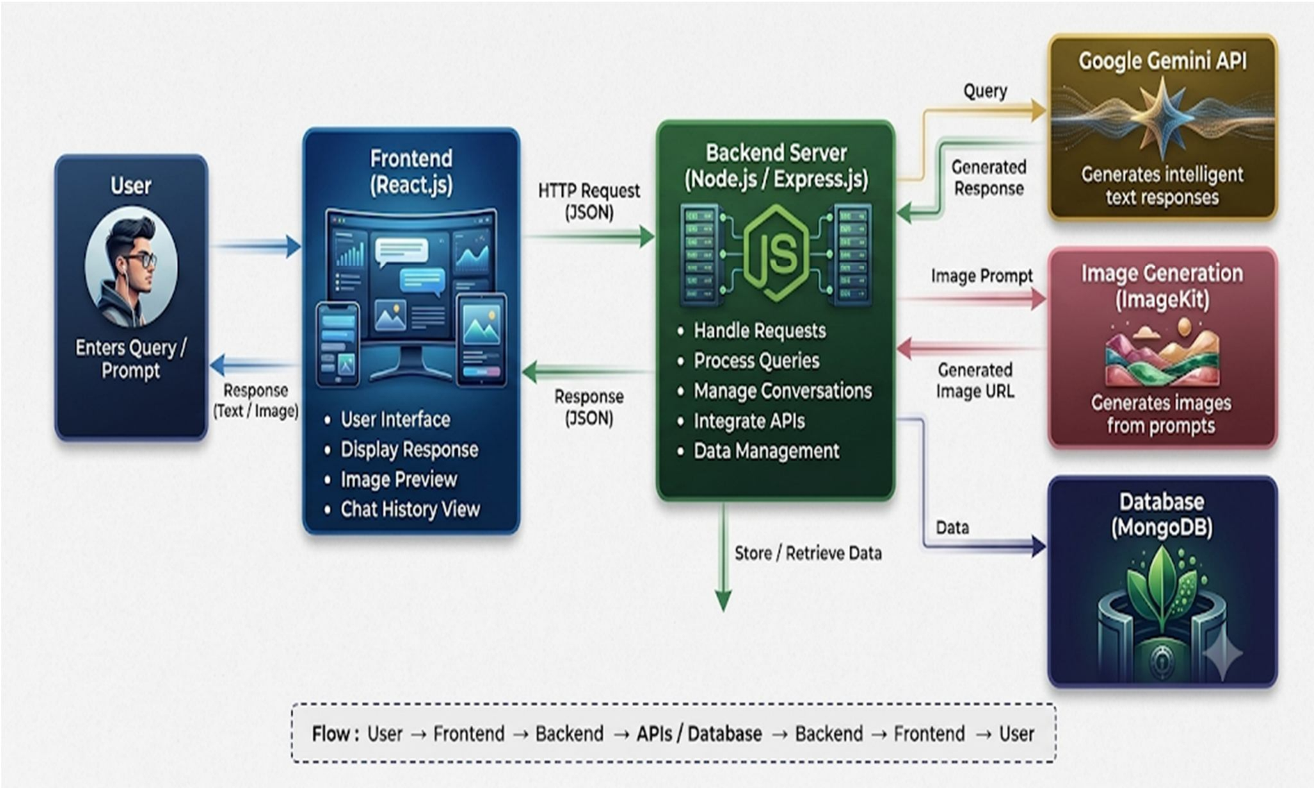


Figure 1: System Achitecture

VI. RESULTS

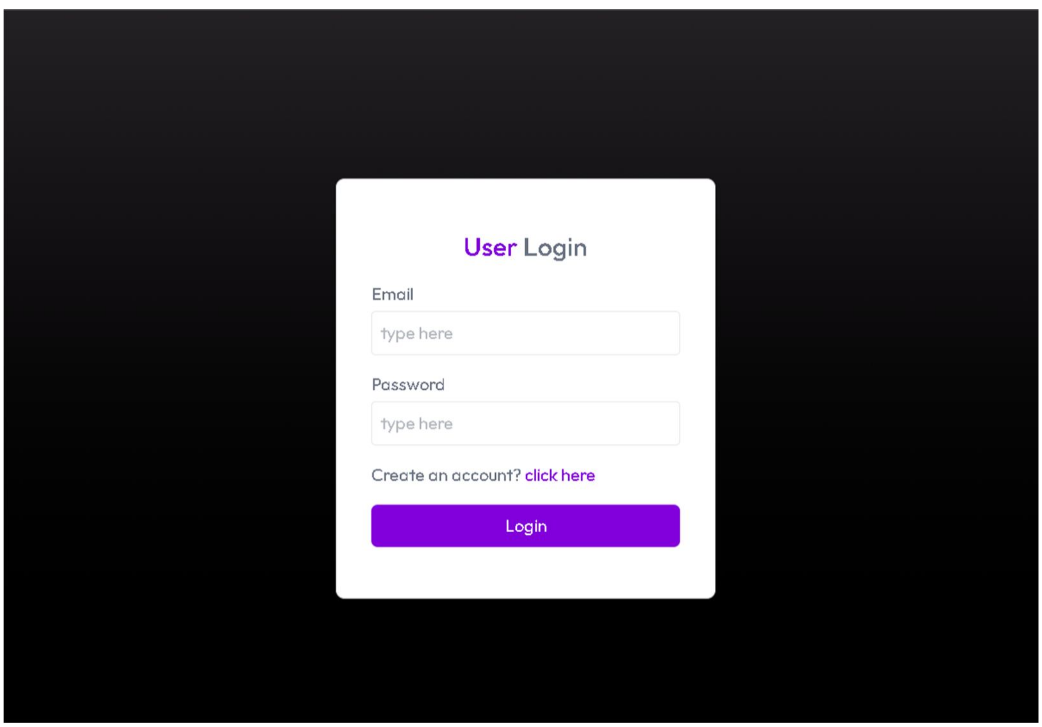


Figure2: User Login

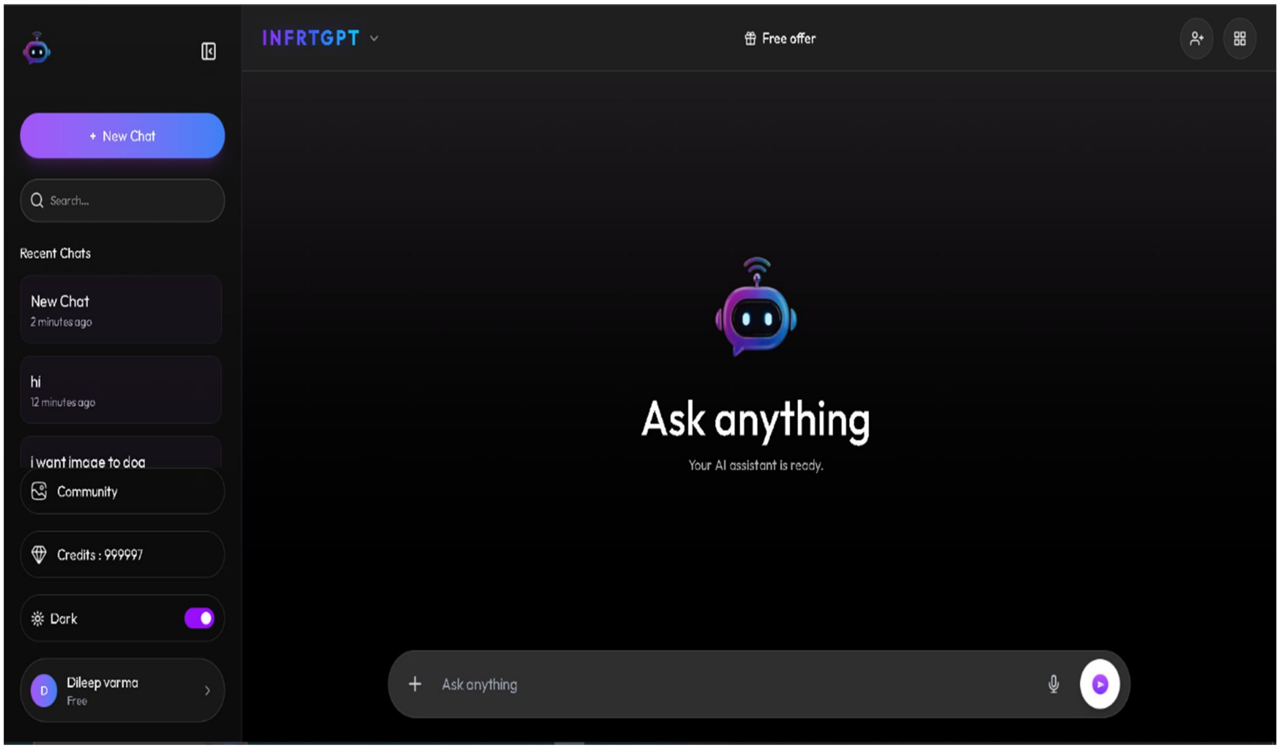


Figure3: Create or select a chat

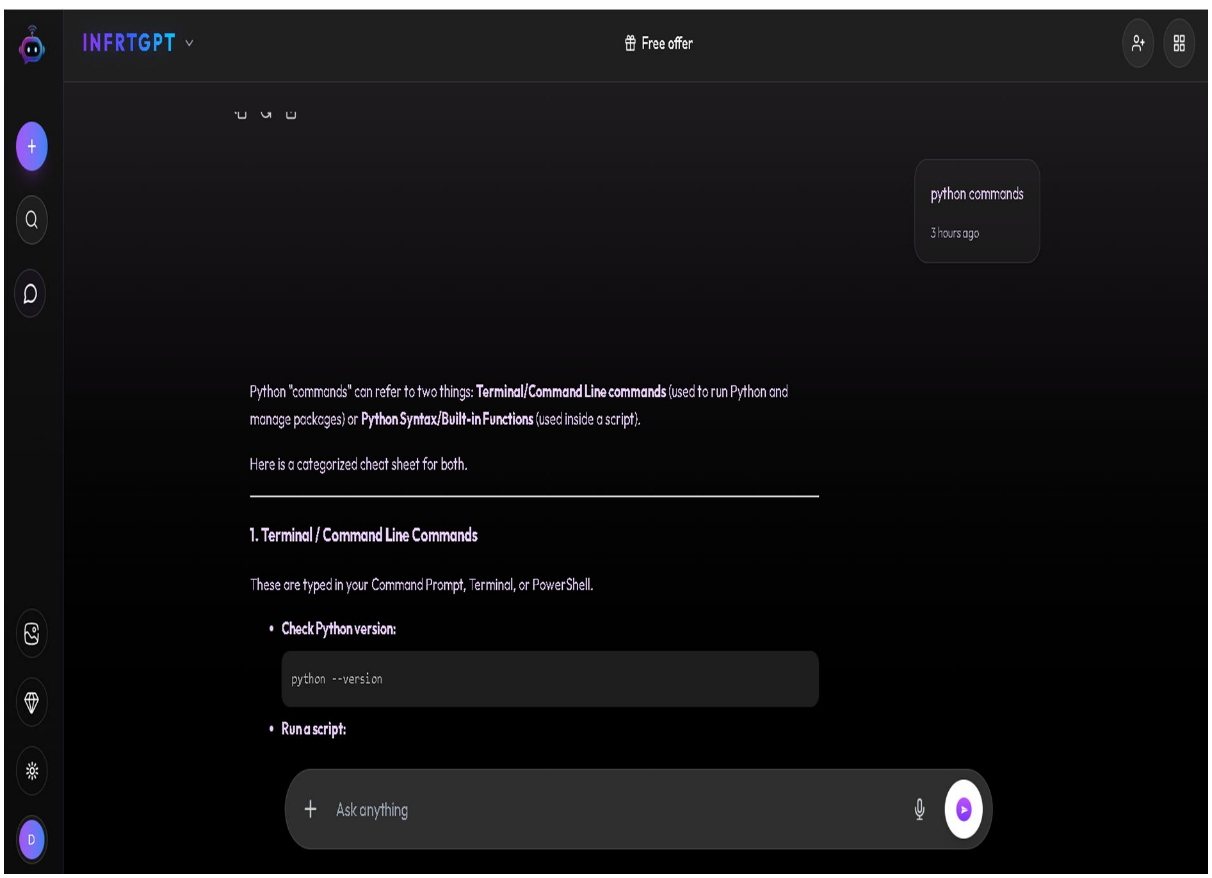


Figure4: Send a text prompt

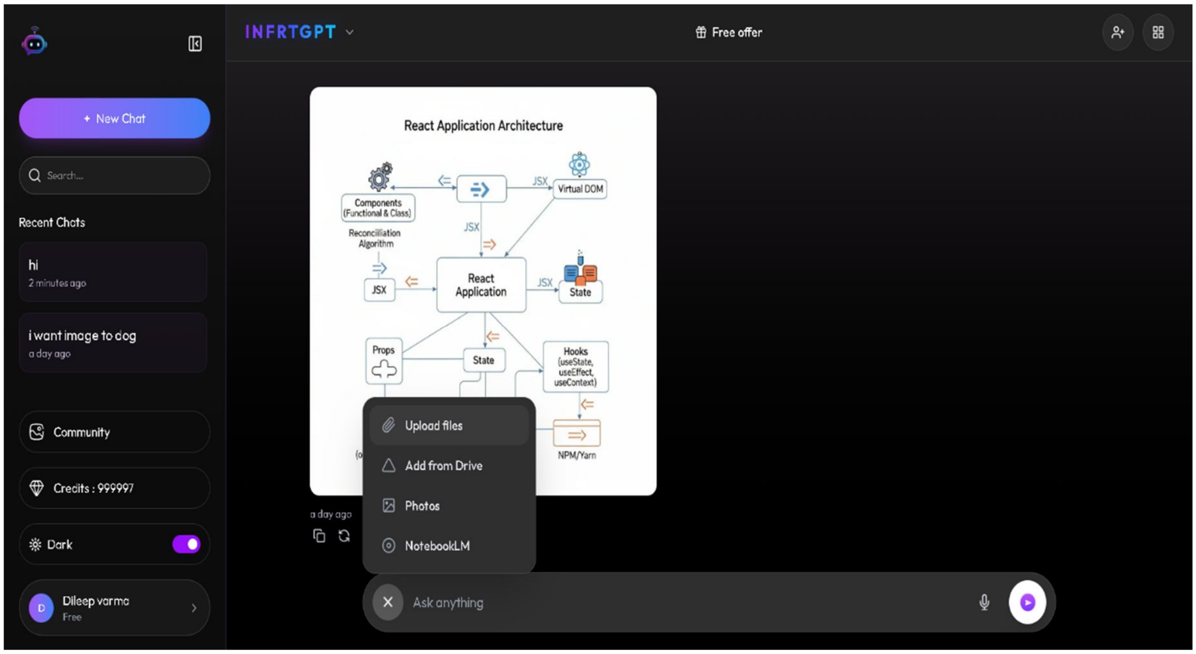


Figure5: Upload a File

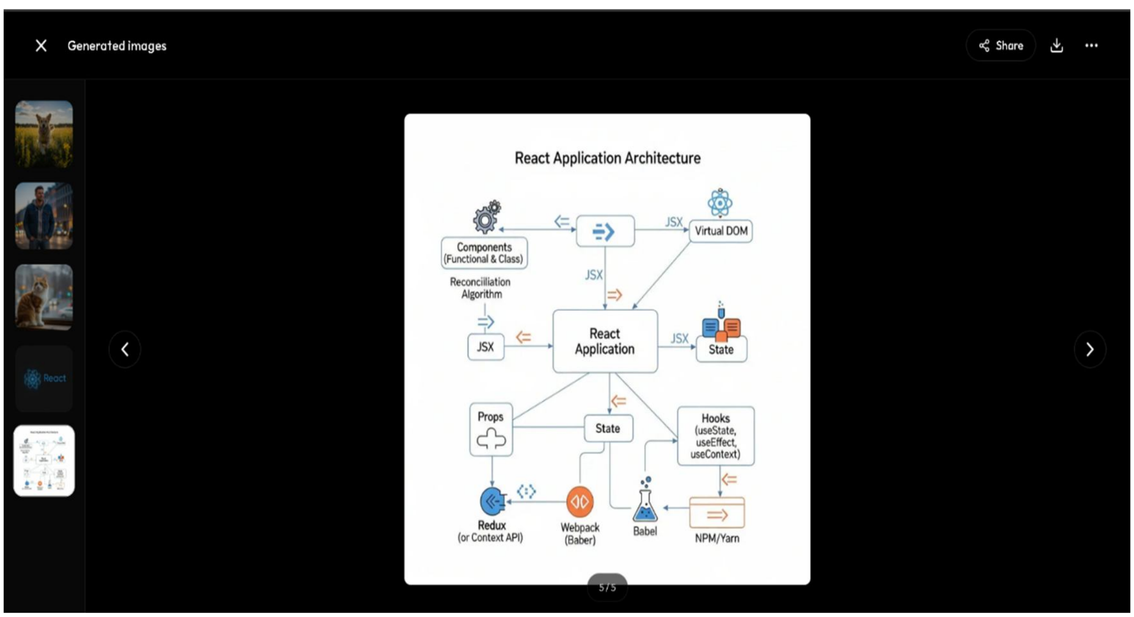


Figure4: Generate and preview images

VII. CONCLUSION

The InfratGPT AI Chat Application was successfully developed and deployed as a modern, web-based conversational platform. Key operational capabilities include:

1) User Authentication & Session Management: Secure user registration, login, and persistent session state handling.
2) Multimodal Dialogue & Media Handling: Real-time text response generation, image synthesis, file upload processing, and interactive utility actions (copying, sharing, and dynamic message regeneration).
3) Dynamic Prompt Editing & Version Control: Users can edit previously submitted prompts in-place. Modifying a prompt triggers a target updates cycle, generating a new context-aware response while retaining historical revisions accessible via intuitive previous/next version controls.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

REFERENCES

[1] Tolosana, R., Vera-Rodriguez, R., Morales, A., & Fierrez, J. (2020). DeepFakes and beyond: A survey of face manipulation and fake detection. arXiv. https://arxiv.org/abs/2001.00179

[2] Croitoru, F. A., Bogolin, S. V., Zaman, A., Leordeanu, M., & Balas, V. E. (2024). Deepfake media generation and detection in the generative AI era: A survey and outlook. arXiv. https://arxiv.org/abs/2401.10962

[3] Liu, X., Tao, D., & Zhou, J. (2024). Evolving from single-modal to multi-modal facial deepfake detection: A survey. arXiv. https://arxiv.org/abs/2402.02496

[4] Mulye, S., Shinde, S., & Thakare, D. (2024). Deepfake detection and analysis using fusion model. Global Scientific and Statistical Research Review (GSSRR). https://www.gssrr.org/index.php/JournalOfBasicAndApplied/article/view/15927

[5] Alanazi, Z., Ushaw, G., & Morgan, G. (2024). Improving detection of deepfakes through facial region analysis in images. Sensors, 24(4), 1254. https://www.mdpi.com/1424-8220/24/4/1254

[6] Soudy, M., Elsharkawy, A., & Khalil, K. (2024). Deepfake detection using convolutional vision transformers and convolutional neural networks. Multimedia Tools and Applications. https://link.springer.com/article/10.1007/s11042-023-17918-z

[7] Lu, P., & Lin, T. (2024). Robust image deepfake detection with perceptual hashing. arXiv. https://arxiv.org/abs/2403.01567

[8] Verdoliva, L. (2020). Media forensics and deepfakes: An overview. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910–932. https://ieeexplore.ieee.org/document/9043915

[9] Jiang, L., Li, C., Ju, C., He, R., & Tan, T. (2020). DeeperForensics-1.0: A large-scale dataset for real-world face forgery detection. arXiv. https://arxiv.org/abs/2001.03024

[10] Dolhansky, B., Howes, R., Pflaum, B., Baram, N., & Ferrer, C. C. (2020). The DeepFake Detection Challenge (DFDC) dataset. arXiv. https://arxiv.org/abs/2006.07397

[11] Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: A compact facial video forgery detection network. In 2018 IEEE International Workshop on Information Forensics and Security (WIFS). https://ieeexplore.ieee.org/document/8424632

[12] Li, Y., Chang, M. C., & Lyu, S. (2018). In Ictu Oculi: Exposing AI-created fake videos by detecting eye blinking. arXiv. https://arxiv.org/abs/1806.02877

[13] Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Capsule-forensics: Using capsule networks to detect forged images and videos. In ICASSP 2019: IEEE International Conference on Acoustics, Speech and Signal Processing. https://arxiv.org/abs/1810.11215

[14] Wang, S. Y., Wang, O., Zhang, R., Owens, A., & Efros, A. A. (2020). CNN-generated images are surprisingly easy to spot... for now. arXiv. https://arxiv.org/abs/1912.11035

[15] Güera, D., & Delp, E. J. (2018). Deepfake video detection using recurrent neural networks. In 2018 IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). https://ieeexplore.ieee.org/document/8639163

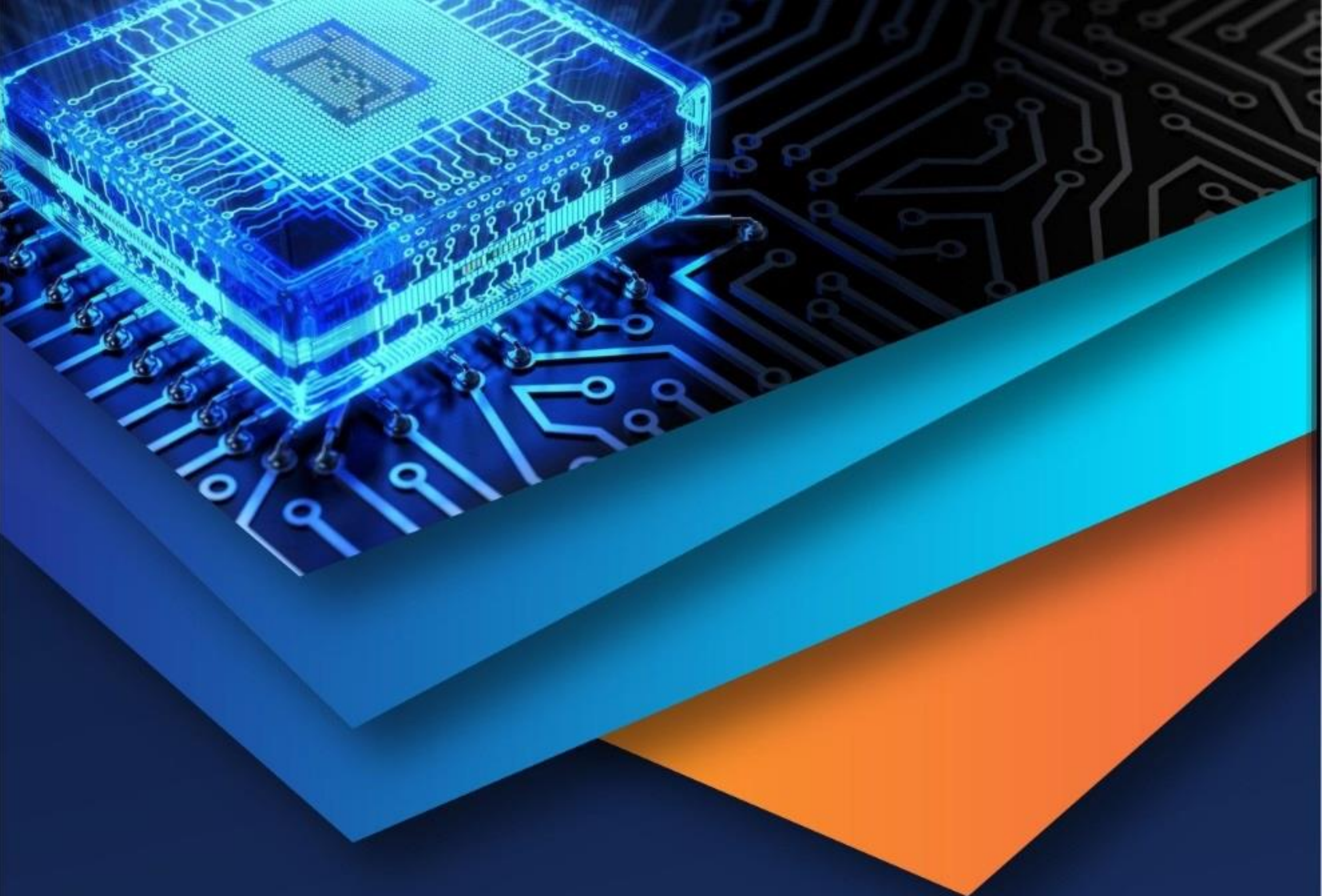
[16] Korshunov, P., & Marcel, S. (2019). Vulnerability assessment and detection of deepfake videos. arXiv. https://arxiv.org/abs/1812.08685

[17] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. arXiv. https://arxiv.org/abs/1901.08971

[18] Zhang, X., Karaman, S., & Chang, S. F. (2019). Detecting and simulating artifacts in GAN fake images. arXiv. https://arxiv.org/abs/1907.06515

[19] Haliassos, A., Vougioukas, K., Petridis, S., & Pantic, M. (2021). Lips don’t lie: A generalisable approach to face forgery detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://arxiv.org/abs/2008.04848

[20] Zhao, T., Xu, X., Xu, M., Ding, H., Xiong, Y., & Xia, W. (2021). Multi-attentional deepfake detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://arxiv.org/abs/2006.07682



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY