IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# BUZZ CHATBOT: A SCALABLE AND RELIABLE CHATBOT BY USING MLOPS AND LEVERAGING GOOGLE CLOUD TECHNOLOGIES FOR CONVERSATIONAL AI CHATBOTS BUILT ON GCP

[1]Pratik Raghunath Sontakke, [2]Prathmesh Wadje, [3]Abhijit Sonawane, [4]Prof.Ranjana Singh

[1,2,3,4]School Of Engineering,

[1,2,3,4]Ajeenkya DY Patil University, Pune, India

***Abstract:*** Buzz Chatbot, a unique conversational agent meant to work flawlessly with the integration of image prompting to enhance the user's experience, is introduced in this research project. Buzz Chatbot uses various APIs to facilitate online communication while delivering many features like image processing, giving the desired output, and also answering queries perfectly. To demonstrate Buzz Chatbot's adaptability and versatility in a range of settings, this research attempts to investigate the integration of advanced APIs in chatbot creation and customization for better results and personal or general use.

***Index Terms*** - AI, Chatbot, Cloud Platform, MLOps, API.

## I. INTRODUCTION

The chatbot is a class of bots that have existed in the chat stages. The client can connect with them through graphical interfacing or widgets, and the drift is in this heading. They by and large give a stateful benefit i.e. the application spares information of each session. On a college's site, one regularly doesn't know where to look for a few kinds of data. It gets to be troublesome to extricate data for a individual who is not a understudy or worker there. The arrangement to these comes up with a college request chatbot, a quick, standard, and instructive gadget to improve the college website's client involvement and give successful data to the client. Manufactured Insights (AI) progressively coordinating our day by day lives with the creation and examination of brilliantly program and equipment, called cleverly specialists. Cleverly specialists can do an assortment of errands extending from labour work to advanced operations. A chatbot is a normal case of an AI framework and one of the most basic and far-reaching illustrations of brilliantly Human-Computer Interaction (HCI) [1]. Advanced chatbots have consolidated modern functionalities such as R&S advances (voice acknowledgment and union), customized interaction, integration with third-party apps, omni-channel arrangement, setting mindfulness, and multi-turn capability.

The term Chatterbot was to begin with specified in 1991[1]. It was a TINYMUD (multiplayer real-time virtual world) counterfeit player, whose essential work was to chat. Numerous genuine human players appeared to lean toward talking to Chatterbot to a genuine player [1].

Why do clients utilize chatbots? Chatbots appear to hold a huge guarantee for giving clients speedy and helpful responses, particularly to their questions. The most frequent inspiration for chatbot clients is considered to be efficiency, whereas other thought processes are amusement, social components, and contact with oddities. In any case, to adjust the inspirations specified above, a chatbot ought to be built in a way that acts as an instrument, a toy, and a companion at the same time [1]. Buzz Chatbot recognizes and tackles the

issues caused by intermittent personalization, it stands out as a trailblazing arrangement in the conversational specialist space. The capabilities of this chatbot are fundamental for providing a smooth and user-friendly encounter in the current advanced environment when the network is not continuously guaranteed. It takes advantage of different APIs from Google Cloud to donate buyers energetic, context-aware discourse. Buzz Chatbot's capacity empowers it to communicate with buyers in real-time, giving custom-fit reactions and altering to their changing requests. By coordinating with Gemini AI, the chatbot can create and comprehend dialect more like a human, which makes discussions with it feel more consistent and natural.

The Buzz Chatbot speaks to a cutting-edge approach to chatbot innovation by seamlessly integrating both content and picture functionalities. This inquiry about venture points to creating a shrewdly chatbot competent of giving energetic reactions indeed in the nearness of a picture utilizing different highlights given by Google Cloud and its APIs. Leveraging Vertex AI, Vision API, and Gemini API, the Buzz Chabot offers a flexible client encounter by interfacing with Gemini for improved conversational capacities. Also, Buzz Chatbot can be personalized according to the prerequisites of the person or the company. For instance, utilizing the bot for client administrations can offer assistance to the trade to develop and diminish the workload from the workers, which can obviously progress efficiency, boosting the development of the trade.

## II. METHODOLOGY

Buzz Chatbot’s implementation is grounded in a modular architecture that seamlessly combines Google Cloud, ML Ops, and APIs, enhancing the user's experience and providing customization facilities through which one can integrate Buzz Chatbot with a website, providing additional support to the website and its users.

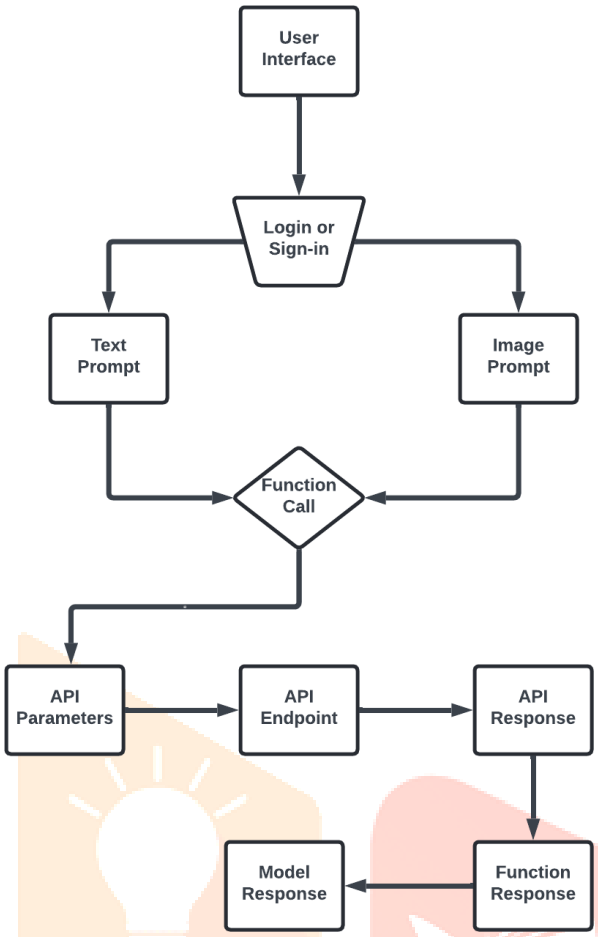
Buzz Chatbot is inclined towards the Gemini API instead of Chat-GPT because it has certain features, such as image prompting, and it also blends well with cloud operations, making it easier for Buzz Chatbot to integrate into any website and personalize it reveals the distinct attributes and capabilities of these advanced AI models. Gemini AI shows promise with unique strengths in areas such as language understanding. It emerges as a strong competitor to ChatGPT, suggesting a dynamic and evolving landscape in AI language models. Both models exhibit exceptional capabilities but differ in various aspects of language processing and response generation.

The analysis underlines the fact that each AI model, including ChatGPT, GPT-4, Bard, and Gemini AI, possesses unique strengths and weaknesses, making them suitable for different applications and use cases. It is important to note that further advancements are necessary before the use of AI chatbots in clinical settings [8].

Trust in chatbots appears to be dependent on the anthropomorphic impression and communication competence of the chatbot. Considering the significant negative interaction between anthropomorphism and the ICC component of closeness, we have grounds to think that those two measures capture parts of the same trust-predicting construct, which leads to a reduction of their individual effects [16].

Below is the backend flowchart indicating how the process takes place and why it is extremely efficient and beneficial when compared to other models:

Figure 1: Backend Flow of Buzz Chatbot




Here is the breakdown of Figure 1:

• User Interface: As seen above, the user first sees an interface where they can see two options. either log in for previous users or sign up for the new

• Login or sign up: The user is given two choices for accessing the bot where they can use G-Mail or email accordingly, making it secure.

• Text prompt and image prompt: After logging in, users have to choose between text or image to pass queries and receive

• Function Call: The system then makes a function call, which could initiate the process of answering the query requested by the

• API Parameters: This helps in verifying if the presented query could fit between the parameters of the APIs to process and return a value.

• API Endpoint: The API endpoint then accepts the request and sends back the response to the user.

• API Response: It gives the output to the user after it receives the information from the API.

• Function Response and Model Response: The information received from the API is presented to the user on their screens when it passes through the function, and then the model displays it to the user.

## III. Results and Discussion

The evaluation of Buzz Chatbot's performance encompasses key aspects such as responsiveness, accuracy in online interactions, and effectiveness in providing relevant information in both conversational and visual modes. Additionally, user feedback and engagement metrics are analyzed to gain insights into the overall user satisfaction with the chatbot.

Usually, chatbots are primarily accessed through chat/messenger applications (e.g., Facebook, Skype, etc.). This is problematic from a security perspective, given the vulnerabilities of these systems [2]. While AI technologies can offer accurate results, as demonstrated by Hernán- dez Montilla et al. (2023) in their study on the Automatic International Hidradenitis Suppura- tiva Severity Score System (AIHS4), the essential role

of human judgment and expertise should always be considered. These include a potential overreliance on AI-generated data and a risk of undermining the irreplaceable value of human judgment and expertise [18]. Moreover, we had many technologies that we could have worked with, but the main reason we didn't choose other APIs or technologies was because we did not want to overload the bot and sacrifice its accuracy [5].

Since we are using the Gemini API for conversation and image prompting, we can review the Gemini model against ChatGPT and use the table from a research paper [9].

| Features | Gemini | ChatGPT |
|---|---|---|
| Model Type | LLM, Search integration | LLM, Focus dialogue |
| Knowledge Cutoff | Access to more up-to-date information | Knowledge cutoff around late 2021 |
| Factual Accuracy | Prioritizes accuracy, sourcing | Accuracy can vary, especially post cutoff |
| Creativity | Less emphasis on creative output | Excels in creative text formats |
| Task Completion | Strong with search-supported tasks | More conversational, open-ended tasks |
| Code Generation | Capable | Capable |
| Bias and Safety | Efforts to mitigate bias and harm | Efforts to mitigate bias and harm |
| Conciseness vs Detail | May prioritize shorter, direct responses | Can be more verbose and offer greater detail |
| Understanding complex queries | Gemini's search integration may excel in handling complex or multi-part questions | ChatGPT might be suited to parsing more conversational language even if complex |

## IV. CONCLUSION

In dynamic large language models (LLMs), the competition between Google's Gemini and OpenAI's ChatGPT has intensified. Our investigation explores the applications, performance, architecture, and capabilities of these prominent models, aiming to offer insights into their relative strengths and weaknesses. Gemini and ChatGPT showcase distinct strengths across various performance metrics. Gemini's integration with Google Search provides a notable advantage in factual accuracy, whereas ChatGPT excels in conversational fluidity and creative expression. [9]

In conclusion, Buzz Chatbot represents the revolutionary powers enabled by the deliberate integration of Google Cloud technology and MLOps concepts. Leveraging Vertex AI's comprehensive platform, Buzz Chatbot seamlessly combines AutoML features, permitting it to dynamically train and enhance its natural language understanding models based on real-time user interactions. For instance, when a user queries about local hiking trails, the chatbot employs the Gemini Conversation API's advanced NLP functionalities to understand intents, extract relevant entities like location, and give personalized recommendations tailored to the user's tastes and past interactions. Furthermore, Cloud Run provides a flexible and scalable hosting environment, ensuring Buzz Chatbot's responsiveness and reliability even during peak usage circumstances.

During events like promotional campaigns, Cloud Run's auto-scaling technology dynamically assigns resources to accommodate heightened traffic, providing a fluid conversational experience for users without compromising performance. By leveraging MLOps concepts, Buzz Chatbot continuously refines its capabilities. It auto-mates model training, evaluation, and deployment processes, ensuring the chatbot remains proficient at evolving linguistic trends and user preferences. Real-time performance monitoring helps Buzz Chatbot recognize and rectify issues swiftly, providing consistently high-quality encounters. This iterative technique enables the chatbot to automatically release updates while minimizing disturbances to the user experience, enabling a smooth transition to upgraded versions.

Moreover, Buzz Chatbot's modular architecture promotes scalability, adaptability, and customization. It effectively connects with external systems such as CRM platforms, enabling tailored interactions and efficient procedures. Additionally, integration with Gemini Vision API lets the chatbot analyze user-uploaded photographs, generating contextually relevant responses. For instance, detecting products inside photographs and easing purchasing processes.

## REFERENCES


**[1]** Adamopoulou, E. and Moussiades, L., 2020. An overview of chatbot technology. In IFIP international conference on artificial intelligence applications and innovations (pp. 373- 383). Springer, Cham

**[2]** Cahn, J., 2017. CHATBOT: Architecture, design, & development. University of Pennsylva- nia School of Engineering and Applied Science Department of Computer and Information Science.

**[3]** Singh, G., Singh, P., Motii, A. and Hedabou, M., 2024. A secure and lightweight container migration technique in cloud computing. Journal of King Saud University-Computer and Information Sciences, 36(1), p.101887.

**[4]** Chen, A., Yao, Y., Chen, P.Y., Zhang, Y. and Liu, S., 2023. Understanding and improving visual prompting: A label-mapping perspective. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 19133-19143).

**[5]** Lin, S.Y., Jiang, C.C., Law, K.M., Yeh, P.C., Kuo, H.L., Ju, S.W. and Kao, C.H., Comparative Analysis of Generative AI in Clinical Nephrology: Assessing ChatGPT-4, Gemini Pro, and Bard in Patient Interaction and Renal Biopsy Interpretation. Gemini Pro, and Bard in Patient Interaction and Renal Biopsy Interpretation.

**[6]** Siddiqui, N., 2024. Cutting the Frame: An In-Depth Look at the Hitchcock Computer Vision Dataset. Journal of open humanities data, 10(1).

**[7]** Ahmed, I. and Islam, R., 2024. The most powerful LLM: Myth or Truth. Authorea Preprints.

**[8]** Masalkhi, M., Ong, J., Waisberg, E. and Lee, A.G., 2024. Google DeepMind's gemini AI versus ChatGPT: a comparative analysis in ophthalmology. Eye, pp.1-6.

**[9]** Rane, N., Choudhary, S. and Rane, J., 2024. Gemini Versus ChatGPT: Applications, Performance, Architecture, Capabilities, and Implementation. Performance, Architecture, Capabilities, and Implementation (February 13, 2024).

**[10]** Rao, N., Tsay, J., Kate, K., Hellendoorn, V. and Hirzel, M., 2024, March. AI for Low-Code for AI. In Proceedings of the 29th International Conference on Intelligent User Interfaces (pp. 837-852).

**[11]** Raghavendran, K.R. and Elragal, A., 2023, June. Low-Code Machine Learning Platforms: A Fastlane to Digitalization. In Informatics (Vol. 10, No. 2, p. 50). MDPI.

**[12]** Kreuzberger, D., Kühl, N. and Hirschl, S., 2023. Machine learning operations (mlops): Overview, definition, and architecture. IEEE access.

**[13]** van Riel, Z., MLOps for chatbots.

**[14]** Corberán, Á., Plana, I., Sanchis, J.M. and Segura, P., 2024. Theoretical and computational analysis of a new formulation for the Rural Postman Problem and the General Routing Problem. Computers & Operations Research, 162, p.106482.

**[15]** Lotfi, F., Beheshti, A., Farhood, H., Pooshideh, M., Jamzad, M. and Beigy, H., 2023. Storytelling with image data: A systematic review and comparative analysis of methods and tools. Algorithms, 16(3), p.135.

**[16]** Wald, R., Heijselaar, E. and Bosse, T., 2021, June. Make your own: The potential of chatbot customization for the development of user trust. In Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (pp. 382-387).

**[17]** Wang, W., Chen, Z., Chen, X., Wu, J., Zhu, X., Zeng, G., Luo, P., Lu, T., Zhou, J., Qiao, Y. and Dai, J., 2024. Visionllm: Large language model is also an open-ended decoder for vision-centric tasks. Advances in Neural Information Processing Systems, 36.

**[18]** Goktas, P., Kucukkaya, A. and Karacay, P., 2023. Leveraging the efficiency and transparency of artificial intelligence-driven visual Chatbot through smart prompt learning concept. Skin Research and Technology, 29(11).