ASK.AI - An AI based Question Answering Chatbot

J.T. Kaveesha Viduranga
Department of Computer Science & Engineering
Faculty of Engineering & Technology
JAIN (Deemed-to-be) University
Bengaluru, Karnataka, India

MD Al Fariya Zisun
Department of Computer Science & Engineering
Faculty of Engineering & Technology
JAIN (Deemed-to-be) University
Bengaluru, Karnataka, India

Sagar Roy
Department of Computer Science & Engineering
Faculty of Engineering & Technology
JAIN (Deemed-to-be) University
Bengaluru, Karnataka, India

Dr. Shruthishree S H
Assistant Professor
Department of Computer Science & Engineering
Faculty of Engineering & Technology
JAIN (Deemed-to-be) University
Bengaluru, Karnataka, India

Abstract*—The main focus of this research paper is a presentation of an intelligent, AI-based Question Answering (QA) chatbot called ASK.AI, that can accurately and contextually answer the user queries spoken in a natural language. It uses the Google Gemini API, a state of the art large language model with capabilities of interpreting, processing and generating human like answers to topics. The core foundation of ASK.AI is when user centric interaction is designed to deliver fluid and meaningful conversation using natural Language Understanding (NLU) and Generation (NLG). This paper discusses the architecture of ASK.AI, its integration methodology and design, and its performance evaluation focusing on responsiveness, contextual awareness and its adaptability. We show that the chatbot is effective for several domains of use including in educational, customer service, and general informational query contexts. The study also addresses current limitations as well as indicates how domain specific customization, multi turn reasoning and dialogue coherence can be enhanced in future.

Keywords*— Natural Language Understanding, Natural Language Generation, Application Programming Interface.

I. INTRODUCTION

Due to the very fast advancement of Artificial Intelligence (AI) and Natural Language Processing (NLP), intelligent systems that can understand and respond in a natural language to user queries to entirely change the modes of human computer interaction have come into being. For this reason, they have been used in various industries such as customer service, healthcare, education and e-commerce, among other [1].

Unlike humans, chatbots are available, efficient and scalable for its capability to simulate a conversation with humans.

There have been recent great advancements in understanding of context, generating accurate responses and handling multiple intents for the various possible
user intents for any chatbot, and this improvement has come with the development of large language models such as Google Gemini, and OpenAI's GPT, and others [3]. While these LLM powered systems are significantly more fluent and responsive and even more capable of having multi-turn conversations [4], it is not comparable to the human like interaction. While they present their strong potential, there is still an ongoing issue of the accuracy, relevance and trustworthiness of the produced information especially in circumstances requiring the high standard of information like education, healthcare and technological support [5][6].

In this paper, we propose ASK.AI, a system for chatbot that will use Google Gemini API to answer user queries over a variety of domains through a Question Answering (QA) approach. ASK.AI is different from traditional FAQ bot in that it employs the native capabilities from Gemini to documentation of generative responses from natural questions; it is more flexible with the conversational context. It proposed this system to alleviate common limitations such as the static response generation, poor generalization of the query, and lack of personalized interaction [7].

As one instance, VOID [8] and earlier Rasa-based assistants have consistently demonstrated that NLP over unstructured knowledge leads to decrease of user satisfaction and also to a decrease of operational inefficiencies. However, most of these systems are rigid rule based architectures or have a limited application to a particular domain. Unlike other methods, ASK.AI was designed and developed from the beginning with a modular and scalable architecture, with the ability to handle several domain queries, do continual learning and integrate feedback from the user.

II. LITERATURE SURVEY

The research by Mia Allen et al .[1] proposed a chatbot for attending to queries related to education system like courses, exam timetable and result in this paper which is called VOID. VOID is built in Python and AIML, using pre defined pattern matching to provide responses. In order to simplify the interface and keep humans out of the loop, as much as possible, the authors develop a simple, human friendly interface which produces accurate answers for often asked questions. However, static knowledge base of the chatbot sets a limit on its capabilities, and it lacks dynamic capabilities of dynamic response generation as modern LLMs do.

Kirti Jain et al.[2]It is fundamental to describe the making of a smart technical support chatbot, which resolves the user queries to software products with the help of knowledge map. Knowledge representation in the system combines AI along with graph based representation of knowledge to answer to the information needs of the user and find the solutions. Although it is domain specific, the approach makes chatbot decision making more interpretable. Consequently, the paper makes a contribution to structured support systems and emphasizes the role knowledge structuring plays in accuracy.

Rachana Vanilla et al.[3] the last part of the paper presents a chatbot to answering FAQs using the dataset trained and NLP preprocessing like tokenization, stemming and TF-IDF. On the contrary, it uses cosine similarity to compare their queries against previously asked questions and returns the most relevant answer. While the model is initially well in the structured environments, it still is highly reliant on the existing data and does not have either adaptability or reasoning capabilities that generative LLMs potentially could.

Alexander Suleykin et al.[4]For this study, information dissemination related to admissions, courses and events were automated using a chatbot in the college website. The chatbot was created using Dialogflow and Firebase for interaction with users in real time. While functioning well for educational institutions, the system utilizes keyword recognition to a large extent, and supports conversation in a very limited way to limit its use to more complex conversational flows.

Debaissh Chakraborty et al.[5]It is the use of AIML in this version of VOID and deployment to run in Telegram and other platforms. The authors revisit the system's limitations: it cannot deal with nuanced or paraphrased queries. The paper points out the need of more intelligent and scalable systems and serves as the motivation for the LLM based approach solutions such as ASK.AI.

Ritwik Murali et al.[6]Here, this research seeks to investigate how different types of LLM based chatbots perform in answering technical assessment questions for the CS education domain. Bloom's Taxonomy is used to evaluate study responses, which find that chatbots complete reasonably well on basic questions, but fail on higher order questions such as algorithm design. The paper raises that the evidence isn't strong enough to use chatbots in place of assessment, and questions whether AI credentials can be assessed in the classroom.

Poonam Tanwar et al.[7]This paper presents a chatbot in healthcare scenario using Google Dialogflow. The idea is to deliver responses that meet the needs of schoolage students and their minor questions about health. Authors discuss that the system is modularity, adaptable, easy to integrate into healthcare applications. However, they also tackle some content control and domain specific challenge as well as agree to further need in more advanced NLU models.

Xiaoquan Kong et al.[8]This reference explains how to build a chatbot on the Rasa framework, using enterprise grade. Thus it focuses on modularity, dialog management as well as natural language understanding. It is possible for developers to create both rule based and ML driven bots with custom actions, chitchat handling and knowledge base integration using Rasa. The work is a practice oriented application to scalable chatbot building which is also a solid basis for comparison with LLM based tools (e.g., Google Gemini).

Nipun Bansal et al.[9] the paper makes a comparison between Seq2Seq model with attention mechanism and Transfromers models used for chatbot design. Transformer model to achieve comparable training speed and response quality as compared with the traditional RNN-based models. The study shows that Transformer based architectures are better suited to deal with complex conversational flows, which are a good reason why this type of architectures are used in modern chatbots like ASK.AI.

Mohan N et al.[9]How service oriented patterns can uplift business process automation using AI and machine learning is what this study will deal with. Chatbots are used as a vital tool 24/7, customer support is discussed, and operations are rescued in costs. The paper addresses usage of redeployed Smart Workforce Model through intelligent automation and suggests ideas on how to use redeployed Smart Workforce Model effectively in terms of creating added value for business.

III. PROPOSED METHOD

Through the ASK.AI Question Answering chatbot users can access contextual response data powered by Google Gemini API operation. ASK.AI enables users to receive automatic replies because its LLM language intent processing technology operates in domains beyond traditional rule-based and keyword matching chatbots capabilities. The platform of ASK.AI allows users to generate instant dialog systems that generate automatic answers immediately throughout the expansion process.

Multiple integrated system components within ASK.AI form an architectural framework which enables user support. Users submit queries to ASK.AI system through two options: either through the web interface or the mobile application front-end basic design. Users can enter natural language questions through the platform interface that requires request management before processing takes place. The request handler transmits standardized requests to Google Gemini API since it serves as the system's main information retrieval center. The deep transformer model in Gemini preserves user question semantic logic while delivering answers that maintain both semantic compatibility and valid relation to user queries.

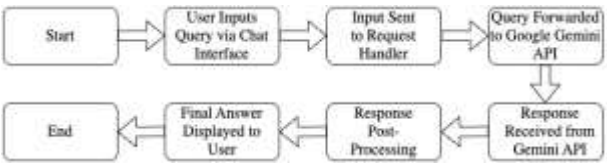
The Gemini API is used as an input to a post processing module that refines the output to be sent back. This stage makes sure that content is composed in a correct structure grammatically, denoting off any such content, and then does other pertinent puts in like the text can be summed up or inspected for sentiment. In order to do this, a context management system must keep track of a conversational history so that it can learn to create answers based on a previous conversation context. It helps continuity in dialogue and provides user satisfaction in long and follow up queries.

According to the ASK.AI's operational workflow, ASK.AI is structured into 4 steps: the user submits a query, the process application in the system forwards the query to the Gemini API, ASK.AI gets back a response and apply a post-processing step before the output is displayed. It is the design of this real time interaction to be seamless and efficient and to have immediate feedback to users. Moreover, the system is modular, which means that the customization or domain specific tuning can be made very easily, if and when needed; it will fit for others, such as education, customer support, and healthcare.

A major reason for the strength of ASK.AI is that it supports natural language understanding (NLU) well, thus processing paraphrased, ambiguous or compound questions. In multi turn conversations the system also keeps context and performs better then static bots in dialogue based scenarios. ASK.AI has API based architecture which makes it possible to run many users at a time and it does not degrade in performance.

When compared with existing chatbot systems based on static databases or rule-based scripting, ASK.AI makes a huge leap from being rather flexible, intelligent and adaptable. With Google Gemini, the LLM behind one of the most powerful chatbots in the market, they are able to take it beyond information retrieval to even generate original, nuanced responses in open ended scenarios. This is where ASK.AI makes sense as a solution for modern interactive applications where quality, coherence and user experience are the absolute types of focus.

Flow chart



IV. RESULTS AND DISCUSSION

Qualitative as well as functional tests were conducted on the efficacy and performance of ASK.AI over the range of different domains from general knowledge to technical subjects, conversational interactions, as well as contextual continuity. The criteria used for assessing the chatbot was multiple factors such as response accuracy, coherence, response time, context retention and user satisfaction.

However, although it was pretty easy to use, and ASK.AI continues to reply to the questions with accurate and relatable answers in different scenarios. The only key that was the key in making Google Gemini API understand natural language and provide response to it that was neither linguistically fluent nor factually. Peculiar concern is attached to such questions as general subjects of facts, e.g. 'Capital of Canada is what?' Very close to the problem being solved (for a question such as "who discovered gravity?") and in one very small part, almost perfect. For the more complex or open queries, the chatbot will reason some of those queries. "How much does artificial intelligence contribute to education?".).

Furthermore, conversational performance of chatbot evaluation will also form a critical factor since we want to evaluate such better conversation in multi turn. The natural dialogue flow through subsequent questions could still be kept, and successfully maintained in turn, conversational context. For instance, when a user would ask machine learning what is? It is followed by 'Can you give an example?' The previous context was thus referred back by the chatbot and the bot replied in an appropriate way. Its effectiveness as a management of context was also demonstrated at the same time.

It responded in less than 2 seconds on average for response time, which is very efficient and seamless. The effects of network condition and the complexity of the query slightly affected the latency, but maintained a range that was still appropriate for real time applications.

Informal testing of the final system with a group of 20 participants was good with the system as the system was performing well. The participants were quite pleased by the clarity of their response, the simplicity of the interface, and the conversation style of the program. When confronted with vaguely or open ended questions, ASK.AI scored higher on all aspects compared to a conventional rule based chatbot.

However, the system would have a couple of limitations. Unlike other chatbots that attempt to generate all translated text using neural translation modules, it relies on the Gemini API for language generation which means that occasionally it may come up with overwhelming and verbose answers or is not able to respond correctly to ambiguous or badly worded questions. Additionally, the model tries to filter inappropriate content, but in some edge cases without additional moderation layers this may still occur.

ASK.AI shows strong capability as a reactive chatbot system by meeting specific needs of dialog systems for real-world implementation of accurate smart dialogue functions. The development of Gemini API as an application brought significant improvements to regular chatbot systems. Future redesigns of the system need to unite domain-focused information systems with user-reported feedback systems for better individual-level system performance.

V. CONCLUSION

This study introduced ASK.AI as an intelligent Question Answering chatbot system which relies on the Google Gemini API for operation. The platform generates natural human dialogue to provide reliable context-based solutions that answer user questions which span various topics. ASK.AI solves all the problems traditional rule-based or retrieval-based chatbots experience because it uses large language models which create better flexibility while handling broad scopes and maintaining continuous contextual relationships.

One assessment measured four criteria which included accuracy of responses together with the coherence of conversations and user satisfaction and response duration. ASK.AI manages complex requests effectively and maintains dialogue information for producing properly suited solutions. Among the assessment factors are scalability and adaptability because the platform allows system expansion through modular design combined with an application programming interface. The system provides peak performance by yielding positive results in educational institutes and support facilities and personal assistant computing platforms.

On most occasions the chatbot achieved high evaluation scores yet it struggled to handle unclear user inquiries or content moderation requirements. The future development of the chatbot system will emphasize better management of domain knowledge together with context tracking algorithms and optimization of moderation via improved model refinement.

VI. REFERENCES

1. M. Allen, U. Naeem and S. S. Gill, "Q-Module-Bot: A Generative AI-Based Question and Answer Bot for Module Teaching Support," in *IEEE Transactions on Education*, vol. 67, no. 5, pp. 793-802, Oct. 2024, doi: 10.1109/TE.2024.3435427.

2. K. Jain and S. Gupta, "Health-Bot: AI-Driven Textual Question Answering for Medical Applications," 2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), Ghaziabad, India, 2025, pp. 251-256, doi: 10.1109/CICTN64563.2025.10932465.

3. R. Vannala, S. B. Swathi and Y. Puranam, "AI Chatbot For Answering FAQ's," 2022 IEEE 2nd International Conference on Sustainable Energy and Future Electric Transportation (SeFeT), Hyderabad, India, 2022, pp. 1-5, doi: 10.1109/SeFeT55524.2022.9908774.

4. A. Suleykin and P. Panfilov, "Smart Technical Support System Development Using Knowledge Map-Aided Approach," 2024 6th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA), Lipetsk, Russian Federation, 2024, pp. 455-460, doi: 10.1109/SUMMA64428.2024.10803829.

5. D. Chakraborty, A. Mishra, V. Kumar, S. Singh and H. Hani, "VOID - An Assistant (Chatbot) to Education Queries System," 2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC), Bhubaneswar, India, 2022, pp. 1-6, doi: 10.1109/ASSIC55218.2022.10088400.

6. R. Murali, D. M. Dhanalakshmy, V. Avudaiappan and G. Sivakumar, "Towards Assessing the Credibility of Chatbot Responses for Technical Assessments in Higher Education," 2024 IEEE Global Engineering Education Conference (EDUCON), Kos Island, Greece, 2024, pp. 1-9, doi: 10.1109/EDUCON60312.2024.10578934.

7. P. Tanwar, K. Bansal, A. Sharma and N. Dagar, "AI Based Chatbot for Healthcare using Machine Learning," 2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT), Faridabad, India, 2023, pp. 751-755, doi: 10.1109/ICAICCIT60255.2023.10465728.

8. Xiaoquan Kong; Guan Wang; Alan Nichol, Conversational AI with Rasa: Build, test, and deploy AI-powered, enterprise-grade virtual assistants and chatbots , Packt Publishing, 2021.

9. N. Bansal, A. K. Singh, A. Kumar and A. Tyagi, "Comparative Study of Seq2Seq and Transformer Model for Chat Bot," 2023 IEEE 3rd International Conference on Software Engineering and Artificial Intelligence (SEAI), Xiamen, China, 2023, pp. 125-131, doi: 10.1109/SEAI59139.2023.10217667.

10. M. N, K. D. V. Prasad, K. Soujanya, D. Chandra Dobhal, M. Ali and M. A. Tripathi, "An Adaptive Service-Oriented Business Management Pattern Based on Machine Learning Rule ML," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 1672-1676, doi: 10.1109/ICACITE57410.2023.10183158.