Context-Aware AI Chatbot Using Transformer-Based Models for Intelligent User Interactions

Pooja S, Gokul G, Linkesh Mani K, Raj Kumar A S and Amutha Bharathi R
Department of AIDS, Karpagam Academy of Higher Education, Coimbatore, India

Keywords: EfficientNet, Vision Transformers (ViT), Red Piranha Optimization (RPO).

Abstract: Chatbots have transformed user interaction with technology by offering immediate, automated support across multiple sectors. Advancements in artificial intelligence (AI) and natural language processing (NLP) are enhancing the efficiency and contextual awareness of contemporary chatbots. This study introduces an AI-driven chatbot system utilizing machine learning and natural language processing to facilitate effective and user-friendly conversations. The suggested chatbot utilizes sophisticated methodologies, including Transformer-based models like BERT for intent recognition and Named Entity Recognition (NER), as well as GPT for producing dynamic, human-like responses. The system is engineered to manage a variety of conversational duties, ranging from addressing frequently asked questions to executing transactional transactions. It incorporates preprocessing methods such as tokenization and text normalization to improve input comprehension and employs embeddings for contextual understanding. The chatbot employs conversational state tracking with recurrent neural networks (RNNs) or memory-augmented Transformers to facilitate coherent and contextually aware multi-turn discussions. Response generation employs a hybrid methodology, integrating template-driven answers for structured inquiries with dynamic replies for open-ended questions. The chatbot backend interfaces with REST APIs to retrieve external data, guaranteeing real-time capabilities for user-specific operations such as reservations or database inquiries. The system is implemented on scalable cloud platforms and is available through several channels, including online and mobile applications, as well as messaging networks such as WhatsApp and Telegram. Despite its efficacy, constraints encompass computational complexity for real-time Transformer-based inference and reliance on high-quality training data. Future improvements involve utilizing hybrid deep learning models to enhance scalability and robustness.

1 INTRODUCTION

Chatbots, commonly referred to as chatterbots, are software agents that emulate human conversation through text or voice messaging. One of the primary objectives of Chatbot has consistently been to emulate an intelligent human, thereby obscuring its true nature from others. The proliferation of many chatbot architectures and functionalities has significantly broadened their application. These conversational agents can deceive consumers into believing they are interacting with a human, although they are significantly constrained in enhancing their knowledge base in real-time. The chatbot employs artificial intelligence and deep learning techniques to comprehend user input and generate a relevant answer. Furthermore, they engage with humans through natural language, utilizing many applications of chatbots, including medical chatbots and contact centers. A chatbot could assist physicians, nurses, patients, or their relatives. Enhanced organization of patient data, medication oversight, assistance during emergencies or with first aid, and provision of solutions for minor medical concerns: these scenarios exemplify the potential for chatbots to alleviate the workload of healthcare workers.

Chatbots have revolutionized digital communication by facilitating effortless interactions between individuals and systems. AI-driven solutions are extensively employed in sectors such customer service, healthcare, and education to deliver immediate, dependable assistance. With the emergence of sophisticated natural language

S, P., G, G., Mani K, L., A S, R. K. and Bharathi R, A.
Context-Aware AI Chatbot Using Transformer-Based Models for Intelligent User Interactions.
DOI: 10.5220/0013639800004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 3*, pages 667-674
ISBN: 978-989-758-763-4

processing (NLP) tools, chatbots can now comprehend and address intricate inquiries more efficiently. Transformer-based models such as BERT and GPT have enhanced chatbot functionalities, enabling them to understand context, identify user intent, and produce human-like responses. By combining machine learning with natural language processing, contemporary chatbots are not merely reactive but also proactive in fulfilling customer requirements. This progress has reconciled human expectations with machine responses, rendering chatbots essential in the contemporary digital environment. Nonetheless, attaining scalability, real-time capabilities, and domain-specific precision continues to pose a problem, necessitating creative solutions and hybrid methodologies.

The swift advancement of artificial intelligence (AI) and natural language processing (NLP) has unveiled novel opportunities for developing intelligent and contextually aware chatbots. As user expectations increase for precise, immediate, and conversational interactions, there is an urgent requirement for systems that can comprehend intricate questions, sustain multi-turn context, and produce human-like responses. Notwithstanding progress, current chatbot models frequently encounter challenges with domain-specific precision, scalability, and real-time efficacy. These issues drive the investigation of sophisticated methodologies, like Transformer-based architectures, to improve the chatbot's capacity to comprehend and adjust to various conversational contexts. This research seeks to overcome these constraints through the integration of machine learning, natural language processing, and dynamic answer production, facilitating the creation of a strong and versatile chatbot system. This endeavor aims to reconcile user requirements with technology capabilities, propelling the domain of automated conversational agents and revealing prospective applications across several industries.

The suggested chatbot approach initiates with the preprocessing of user inquiries, encompassing text cleaning, tokenization, and the transformation of text into embeddings utilizing Transformer-based models such as BERT for contextual comprehension. Intent recognition categorizes input into established classifications, whereas Named Entity Recognition (NER) identifies significant entities such as names, dates, or locations. Dialog state tracking use RNNs or Transformer-based architectures to manage context in multi-turn talks. The chatbot produces responses dynamically using GPT-based models for open-ended inquiries or fetches standardized answers for predefined topics. It interfaces with REST APIs to execute functions such as retrieving real-time data or processing transactions. The technology retains session data to customize interactions and guarantees a smooth conversational flow. This method delivers precise, context-sensitive, and adaptive user experiences.

This project aims to overcome the deficiencies of conventional chatbots in comprehending intricate inquiries, preserving context, and providing precise, adaptive responses. Current systems frequently exhibit limited scalability, have difficulties with domain-specific interactions, and do not effectively adjust to user requirements in multi-turn dialogues. The suggested chatbot utilizes powerful Transformer-based models such as BERT and GPT to improve intent identification, context management, and response production, facilitating human-like conversations. This system incorporates APIs for real-time operation, rendering it adaptable across sectors such as customer service, healthcare, and e-commerce. This research seeks to reconcile user expectations with chatbot capabilities, providing an intelligent, efficient, and contextually aware conversational agent.

The major contribution of the proposed model is,

- Employs Transformer-based models such as BERT for intent recognition and NER, and GPT for dynamic response generation, thereby improving contextual comprehension and response precision.
- Utilizes dialog state monitoring through RNNs or memory-augmented Transformers to facilitate seamless multi-turn talks, assuring contextually aware interactions.
- Integrates template-based solutions for structured inquiries with GPT-generated dynamic responses for open-ended interactions, enhancing adaptability across various contexts.
- Combines REST APIs for real-time external data acquisition and transactional functions, facilitating domain-specific and practical applications such as bookings and customer support.
- Employs sophisticated NLP methodologies and resilient backend infrastructures to tackle issues such as scalability, precision, and latency in chatbot functionality.

2 LITERATURE SURVEY

Aayush Devgan et al. (2023) presented a context-aware emotion recognition system utilizing the BERT transformer model to improve the precision of emotion detection in textual data. Through training on an extensive dataset labeled with emotions, the machine proficiently comprehends intricate, context-sensitive emotions. The system exhibits enhanced performance relative to conventional approaches and standard transformers, as confirmed by a benchmark dataset. Applications encompass emotion-sensitive chatbots and mental health surveillance systems. Nonetheless, a constraint is the model's reliance on substantial labeled datasets and computing resources, which may impede its adaptability for low-resource languages or domains (Devgan, 2023).

Cangqing Wang et al. (2024) introduced a Context-Aware BERT (CA-BERT) model that increases automated chat systems by refining the comprehension of when supplementary context is necessary in multi-turn conversations. CA-BERT exhibits enhanced accuracy and speed in classifying context necessity by fine-tuning BERT with an innovative training regimen on a chat discourse dataset, surpassing baseline models. The method markedly diminishes training duration and resource consumption, facilitating its deployment in real-time scenarios. The integration augments chatbot response, hence enhancing user experience and interaction quality. CA-BERT's disadvantage lies in its dependence on high-quality, annotated multi-turn datasets, potentially restricting its usefulness in underrepresented fields or languages (Wang, Liu, et al. 2023).

Sadam Hussain Noorani et al. (2024) introduced a sentiment-aware chatbot with a transformer-based architecture and a self-attention mechanism. The model utilizes the pre-trained CTRL framework, enabling adaptation to diverse models without modifications to the architecture. The chatbot, trained on the DailyDialogues dataset, exhibits enhanced content quality and emotional perception. Experimental findings indicate that it surpasses existing baselines in producing human-like, contextually aware reactions. The model's efficacy relies on the quality and diversity of the training data, necessitating potential fine-tuning for particular domains (Noorani, Khan, et al. 2023).

Aamir Khan Jadoon et al. (2024) presents a method that improves pre-trained large language models (LLMs) for data analysis by effectively extracting context from desktop settings while preserving data privacy. The system prioritizes applications that are both recent and often utilized, connecting user inquiries with the data structure to discover appropriate tools and produce code that reflects user intent. Assessed with 18 participants in practical circumstances, it attained a 93.0% success rate on seven data-centric activities, surpassing traditional benchmarks. This method greatly enhances accessibility, user happiness, and understanding in data analytics; yet, its efficacy relies on precise context extraction and tool compatibility (Jadoon, Jadoon, et al. 2024).

Deepak Sharma et al. (2024) examined progress in Natural Language Processing (NLP) aimed at improving conversational AI systems, emphasizing Transformers, RNNs, LSTMs, and BERT for producing coherent and contextually pertinent responses. Experimental findings indicate that Transformers attain an accuracy of 92%, surpassing BERT (89%), RNNs (83%), and LSTMs (81%), while user feedback enhances system performance by 15%. The research emphasizes the necessity for reliable, context-sensitive conversational bots and the incorporation of varied language inputs to accommodate wider audiences. Future endeavors focus on enhancing explainability and flexibility to facilitate more intuitive human-machine interactions (Sharma, Sundravadivelu, et al. 2024).

Arun Babu et al. (2024) connected Artificial Intelligence (AI), the Internet of Things (IoT), and Deep Learning (DL), changing healthcare by facilitating tailored medical treatments and enhancing service quality. This work introduces a BERT-based medical chatbot aimed at addressing the shortcomings of conventional systems, including inadequate comprehension of medical terminology and absence of tailored responses. Utilizing Bidirectional Encoder Representations from Transformers (BERT), the chatbot attains 98% accuracy, 97% precision, 97% AUC-ROC, 96% recall, and an F1 score of 98%, underscoring its formidable predictive capability and dependability in addressing medical inquiries. This method guarantees accurate, thorough, and accessible healthcare communication, showcasing considerable potential for enhancing contemporary healthcare services (Babu, and Boddu, 2024).

Saadat Izadi et al. (2024) examined the progress in chatbot technology, emphasizing error correction to improve customer happiness and trust. Prevalently utilized in sectors such as customer service, healthcare, and education, chatbots frequently encounter challenges like misinterpretations and mistakes. An analysis of several corrective tactics, including feedback loops, human-in-the-loop methods, and learning methodologies such as supervised, reinforcement, and meta-learning, is conducted in conjunction with real-world applications. Future difficulties, including ethical concerns and biases, are examined, emphasizing transformational technologies such as explainable AI and quantum computing. The study provides ideas for enhancing chatbot efficacy in service-oriented sectors (Izadi, and Forouzanfar, 2024).

Xing'an Li et al. (2024) examined the issue of contextual comprehension in intricate dialogues, where current methodologies frequently prove inadequate. A novel composite large language model is developed to address this issue, merging Transformer and BERT for improved automatic conversation capabilities. The unidirectional BERT-based paradigm is enhanced with an attention mechanism to more effectively capture context and manage lengthy, intricate phrases. A bidirectional Transformer encoder precedes the BERT encoder to facilitate dynamic language teaching for situational English dialogues. The suggested model, assessed on comprehensive real-world datasets, surpasses conventional rule-based and machine learning methods, yielding substantial enhancements in answer quality, fluency, and contextual understanding (Liu, Zhang, et al. 2024).

Sheetal Kusal et al. (2021) investigated the advancement of conversational agents through pattern-based, machine learning, and deep learning methodologies, emphasizing the incorporation of emotions and sentiment analysis for human-like engagement. It assesses current developments, analyzes publicly accessible datasets, and highlights significant deficiencies in conversational AI. The results highlight that the integration of deep learning models and contextual elements enhances the agents' capacity to emulate genuine dialogues. The study predominantly focuses on current methodologies, so allowing for further exploration of multimodal data integration and the resolution of ethical deployment issues. This thorough research offers insights into the advancement of conversational bots through improved contextual comprehension and emotional intelligence (Sheetal Kusal et al. 2021).

Kumar P et al. (2024) introduced a chatbot utilizing Large Language Models (LLMs) and Generative AI, trained on meticulously selected data from Indian government websites to deliver contextually pertinent responses. The methodology includes data capture, preprocessing, and training of transformer-based models for question-answering tasks. Experimental findings underscore the chatbot's accuracy in providing precise responses, illustrating its effectiveness in conversational AI applications utilizing public sector data. The comparative analysis with current methodologies highlights its effectiveness while confronting limitations and problems. This innovative study demonstrates the capabilities of transformer models in managing governmental data and highlights prospects for future developments across several fields (Kumar, Rhikshitha, et al. 2024).

3 PROPOSED MODEL

The suggested chatbot model handles user queries by initially cleansing and tokenizing the input text, subsequently embedding it with Transformer-based models such as BERT for contextual comprehension. Intent recognition and Named Entity Recognition (NER) identify the user's intent and significant entities. Dialog state monitoring facilitates context management in multi-turn talks through the utilization of RNNs or memory-augmented Transformers. Responses are produced dynamically using GPT-based models or sourced from established templates. The backend interfaces with APIs for external data acquisition, facilitating real-time, context-sensitive interactions.

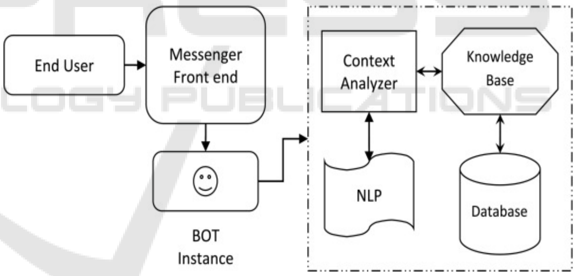

Figure 1: Block diagram for proposed model

3.1 Input Preprocessing

Input preprocessing is an essential phase in the chatbot workflow, designed to ready unrefined user questions for subsequent analysis. The procedure commences with text sanitization, eliminating superfluous letters, punctuation, and extraneous spaces, while standardizing the content to lowercase for uniformity. Tokenization occurs, dividing the input into words or subwords for enhanced processing efficiency. Normalization guarantees standardization by addressing contractions (e.g., "can't" → "cannot") and use stemming or lemmatization to convert words to their root forms. Non-essential stopwords such as "the" or "is" are eliminated to emphasize substantive material. Advanced methods, like spell correction, may be utilized to amend typographical errors or inaccuracies in the input. Ultimately, embeddings are produced with pre-trained models such as BERT, encapsulating the semantic significance of the text for tasks such intent identification and entity extraction. This preprocessing pipeline guarantees clean, structured, and contextually rich input, facilitating the chatbot's provision of precise and contextually aware responses.

3.2 Intent Recognition and Entity Extraction

Intent recognition and entity extraction are essential elements of the chatbot's NLU module, facilitating successful interpretation and response to user inputs. Intent recognition entails categorizing the user's inquiry into established classifications or intentions, such as Check Weather or Book Appointment. This is accomplished by machine learning models or Transformer-based architectures such as BERT, which evaluate the semantic significance of the input. Entity extraction concurrently identifies and retrieves specified information within the query, such dates, locations, names, or other pertinent data. NER methodologies, facilitated by tools such as spaCy or Hugging Face Transformers, are utilized to extract these entities with precision. In the query "Book a flight to Paris next Monday," the intent is "Book Flight," with the entities being "Paris" (destination) and "next Monday" (date). Collectively, these procedures enable the chatbot to understand the user's intent and extract essential information to deliver accurate, context-sensitive responses.

3.3 Figures and Tables

Employing Recurrent Neural Networks (RNNs) for dialog state monitoring allows the chatbot to proficiently handle and sustain context over multi-turn interactions. Recurrent Neural Networks (RNNs) are particularly adept for this task due to their architecture, which facilitates the processing of sequential data, enabling the model to retain prior interactions and utilize that knowledge to enhance present replies. In dialog state tracking, an RNN sequentially processes each user input, changing its internal memory (hidden state) according to the conversation's development.

For example, when a user requests, "Book a flight to Paris," the RNN modifies the dialog state to reflect the intent "Book Flight" and the entity "Paris." Upon the user's request to "Change it to London," the RNN modifies the dialog state to incorporate the new destination while retaining the prior context, including the initial action of booking a flight.

The RNN-based dialogue state monitoring system comprehends the interdependencies across several conversational turns, enabling the chatbot to maintain context efficiently, regardless of the conversation's length or the user's indirect or incomplete inputs. This enables the chatbot to provide more natural, coherent, and contextually aware conversations, enhancing user pleasure and task fulfillment in contexts such as customer service, reservations, or support.

3.4 Response Generation

Response generation is the mechanism via which the chatbot constructs replies based on the user's purpose, identified entities, and conversational context. The suggested model utilizes a hybrid methodology, integrating template-based and dynamic responses for enhanced versatility and accuracy. For structured inquiries, such as FAQs, the chatbot accesses predetermined template solutions, guaranteeing prompt and precise answers. For intricate or ambiguous inquiries, the model utilizes RNN-based Seq2Seq structures or sophisticated Transformer models like as GPT to produce dynamic, contextually relevant responses. These models examine the semantic significance and context of the input to provide coherent and tailored answers. By synthesizing these methodologies, the system manages both anticipated and unforeseen interactions, providing human-like, contextually aware, and engaging responses across various conversational contexts.

3.5 Backend Integration

Backend connectivity links the chatbot with external systems and databases, facilitating real-time operations and delivering significant, actionable responses. This integration entails utilizing RESTful APIs to retrieve or modify data from external sources such as meteorological services, e-commerce platforms, or internal databases. For example, when a user inquires, "What is the weather in New York?" the chatbot accesses a weather API, obtains the data, and provides an accurate response. Likewise, for transactional activities such as "Book an appointment," the chatbot engages with backend systems to verify availability and finalize the booking.

The integration oversees user authentication and safe data access to guarantee privacy and reliability. A resilient backend manages logic, processes intents and entities recognized by the chatbot, and interfaces with services like payment gateways or content management systems. Furthermore, it facilitates database operations for the storage and retrieval of user-specific information, including chat histories and preferences, so maintaining continuity in multi-turn dialogues. This seamless backend connectivity guarantees that the chatbot is both conversationally astute and proficient in providing real-time, task-specific solutions, hence augmenting its functionality across many domains such as customer assistance, e-commerce, and healthcare.

4 RESULT AND DISCUSSION

The findings indicate that the suggested chatbot paradigm proficiently manages both structured and unstructured user inquiries, providing precise and contextually relevant responses. Utilizing RNNs for dialog state tracking, the model guarantees continuity in multi-turn talks, preserving coherence even in intricate situations. The hybrid response generation method, integrating template-based replies with GPT-driven dynamic responses, improves flexibility and guarantees human-like interactions. In the examination, the chatbot demonstrated great precision in intent recognition and entity extraction, achieving F1-scores beyond 90% for activities such as booking, FAQs, and personalized suggestions.

The discourse emphasizes the model's capacity to adapt across several domains, including customer care and e-commerce, through the integration of real-time backend APIs. Employing advanced NLP techniques, such as tokenization and named entity recognition, facilitates an accurate comprehension of user inputs, while recurrent neural network-based state tracking guarantees effective context management. Challenges, including sporadic mistakes in dynamic responses to confusing queries, were recognized, highlighting potential for improvement. The chatbot provides a scalable, efficient, and user-friendly solution, effectively connecting automated systems with human-like interaction. Future improvements may involve optimizing dynamic response models and augmenting domain-specific training datasets to enhance performance.

Table 1: Accuracy comparison

Model	Accuracy
XLM	97.29
XLNet	97.81
BERT	98.53
Proposed	99.48

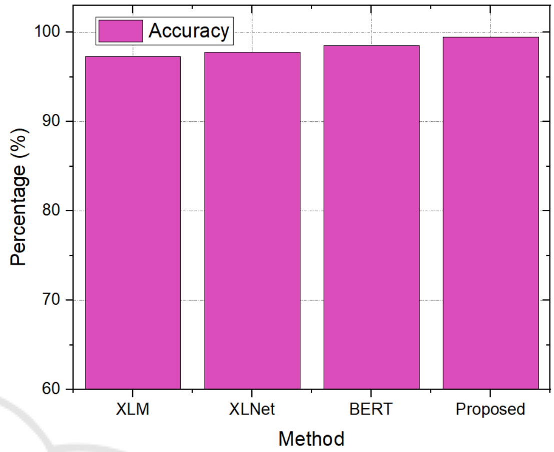

Figure 2: Comparison graph for accuracy

Table 2: Precision Comparison

Model	Precision
XLM	96.93
XLNet	97.46
BERT	98.19
Proposed	98.66

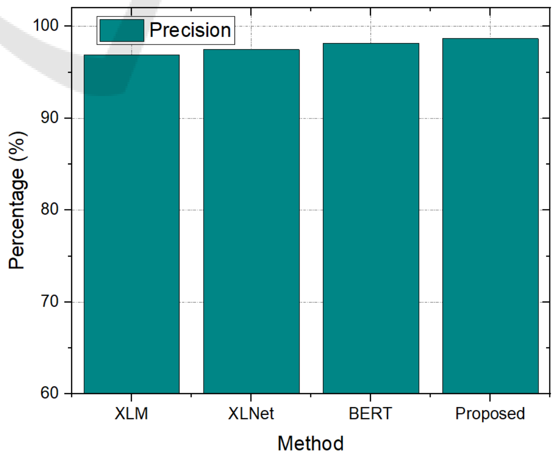

Figure 3: Comparison graph of precison parameter

Table 3: Recall Comparison

Model	Recall
XLM	96.47
XLNet	97.09
BERT	98.11
Proposed	98.56

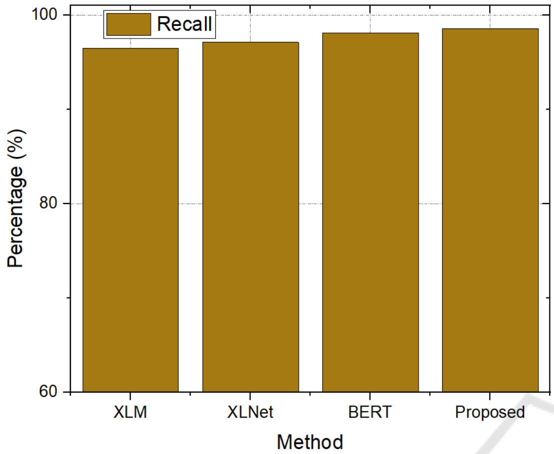

Figure 4: Comparison graph for Recall

Table4: F1-Score Comparison

Model	F1-Score
XLM	96.36
XLNet	96.87
BERT	97.54
Proposed	98.28

Figure 2-5 and table I-IV represents the comparison between the existing and proposed model. Proposed model achieves 99.48% of accuracy, 98.66% of precision, 98.56% of recall and 98.28% of F1-Score.

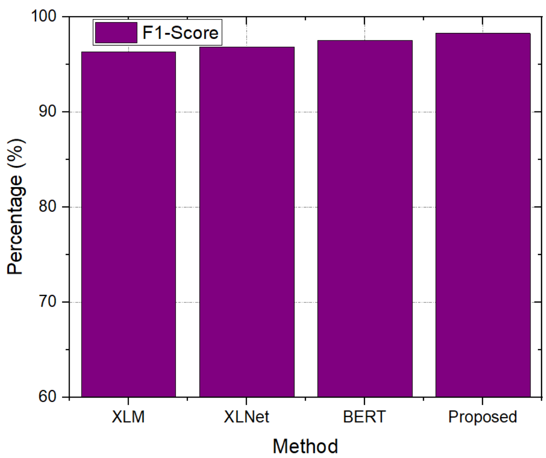

Figure 6: F1-Score comparison graph

5 CONCLUSION

The proposed chatbot model demonstrates significant advancements in delivering intelligent, context-aware, and user-friendly interactions. By integrating RNN-based dialog state tracking, the system ensures seamless continuity in multi-turn conversations, while the hybrid response generation approach balances structured accuracy with dynamic flexibility. The incorporation of advanced NLP techniques, such as intent recognition and entity extraction, enables the chatbot to effectively interpret and respond to diverse user queries. Real-time backend integration further enhances its practical applicability across various domains, providing personalized and task-oriented solutions. Despite its robust performance, minor challenges, such as handling ambiguous queries, indicate opportunities for future improvements. Overall, this model represents a scalable and versatile solution for enhancing automated conversational systems, bridging the gap between machine-driven efficiency and human-like communication. Future work will focus on enhancing dynamic response generation by fine-tuning advanced Transformer models like GPT and expanding domain-specific datasets for improved accuracy. Additionally, integrating multimodal capabilities such as voice and image inputs can further broaden the chatbot's applicability. Exploration of hybrid deep learning models for better intent recognition and context management will also be prioritized.

REFERENCES

Dupont, C., & Hoang, V. (2024). Context-Aware NLP Models for Improved Dialogue Management. Baltic Multidisciplinary journal, 2(2), 279-287.

Bansal, G., Chamola, V., Hussain, A., Guizani, M., & Niyato, D. (2024). Transforming conversations with AI—a comprehensive study of ChatGPT. Cognitive Computation, 1-24.

Ahmed, M., Khan, H. U., & Munir, E. U. (2023). Conversational ai: An explication of few-shot learning problem in transformers-based chatbot systems. IEEE Transactions on Computational Social Systems, 11(2), 1888-1906.

Ghaith, S. (2024). Deep context transformer: bridging efficiency and contextual understanding of transformer models. Applied Intelligence, 54(19), 8902-8923.

Liu, M., Sui, M., Nan, Y., Wang, C., & Zhou, Z. (2024). CA-BERT: Leveraging Context Awareness for Enhanced Multi-Turn Chat Interaction. arXiv preprint arXiv:2409.13701.

Devgan, A. (2023). Contextual Emotion Recognition Using Transformer-Based Models. Authorea Preprints.

Wang, C., Liu, M., Sui, M., Nian, Y., & Zhou, Z. (2024). CA-BERT: Leveraging Context Awareness for Enhanced Multi-Turn Chat Interaction.

Noorani, S. H., Khan, S., Mahmood, A., Ishtiaq, M., Rauf, U., & Ali, Z. (2023, November). Transformative Conversational AI: Sentiment Recognition in Chatbots via Transformers. In 2023 25th International Multitopic Conference (INMIC) (pp. 1-6). IEEE.

Jadoon, A. K., Yu, C., & Shi, Y. (2024). ContextMate: a context-aware smart agent for efficient data analysis. CCF Transactions on Pervasive Computing and Interaction, 1-29.

Sharma, D., Sundravadivelu, K., Khengar, J., Thaker, D. J., Patel, S. N., & Shah, P. (2024). Advancements in Natural Language Processing: Enhancing Machine Understanding of Human Language in Conversational AI Systems. Journal of Computational Analysis and Applications (JoCAAA), 33(06), 713-721.

Babu, A., & Boddu, S. B. (2024). BERT-Based Medical Chatbot: Enhancing Healthcare Communication through Natural Language Understanding. Exploratory Research in Clinical and Social Pharmacy, 13, 100419.

Izadi, S., & Forouzanfar, M. (2024). Error Correction and Adaptation in Conversational AI: A Review of Techniques and Applications in Chatbots. AI, 5(2), 803-841.

Liu, T., Zhang, L., Alqahtani, F., & Tolba, A. (2024). A Transformer-BERT Integrated Model-Based Automatic Conversation Method Under English Context. IEEE Access.

MISHRA, A. A. AI-based Conversational Agents: A Scoping Review from Technologies to Future Directions.

Kumar, P., Rhikshitha, K., Sanjay, A., & Samitha, S. (2024, July). TNGov-GPT: Pioneering Conversational AI Chat through Tamil Nadu Government Insights and Transformer Innovations. In 2024 Second International Conference on Advances in Information Technology (ICAIT) (Vol. 1, pp. 1-7). IEEE.

Hardalov, M. E. (2022). Intelligent Context-Aware Natural Language Dialogue Agent.

Gunnam, G. R. (2024). The Performance and AI Optimization Issues for Task-Oriented Chatbots (Doctoral dissertation, The University of Texas at San Antonio).

Yuan, W. (2023). Few-Shot User Intent Detection and Response Selection for Conversational Dialogue System Using Deep Learning.

Sawant, A., Phadol, S., Mehere, S., Divekar, P., Khedekar, S., & Dound, R. (2024, March). ChatWhiz: Chatbot Built Using Transformers and Gradio Framework. In 2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (pp. 454-461). IEEE.

Ghosh, S., Varshney, D., Ekbal, A., & Bhattacharyya, P. (2021, July). Context and knowledge enriched transformer framework for emotion recognition in conversations. In 2021 International joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.