International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2s) 2026 pp. 609-617
https://doi.org/10.70917/ijcisim-2026-2101
Article

HYBRID MULTI-DOMAIN CONVERSATIONAL FRAMEWORK USING LARGE LANGUAGE MODELS FOR ADAPTIVE DIALOGUE MANAGEMENT

Anil Kumar[1], Manushree Sahay[2], Deepa Abin[3], Rupali Gangarde[4], Pankaj Agarkar[5], Anita Sachin Mahajan[6]

[1] Department of Computer Engineering, Ajeenkya D. Y. Patil School of Engineering, Pune, India.akengineer9719@gmail.com
[2] Department of Computer Engineering, G. H. Raisoni International Skill Tech University, Pune, India.sahaymanushree@gmail.com
[3] Department of CSE – Data Science, Vishwakarma Institute of Technology, Pune, Indiadeepa.abin@vit.edu
[4] Department of Computer Engineering, MIT Academy of Engineering, Pune, India.rupa.gangarde@gmail.com
[5] Professor, Department of Computer Engineering, Ajeenkya D. Y. Patil School of Engineering, Pune, India. pmagarkar@gmail.com
[6] Assistant Professor & Head, Department of Computer Engineering, Ajeenkya D. Y. Patil School of Engineering, Pune, India.anitadapke@gmail.com
Corresponding Author: Anil Kumar (akengineer9719@gmail.com)

Abstract: the last few years have seen an explosive growth in large language models which has translated to significant improvements in the field of conversational AI. However, models that are purely generative have a tendency to hallucinate and are non-factual when answering questions of a specific domain. In order to develop a solution to this problem, the authors of this paper are proposing a Hybrid Retrieval-Augmented Generation (RAG) approach to developing a multi-domain conversational assistant based on the MultiWOZ 2.1 dataset. In order to simplify the scope of the developments and also help with the accuracy of the retrieval and the correctness of the responses, the conversational assistant is limited to the hotel and train domains. This architecture is supplemented with the following: dense semantic retrieval using ChromaDB; sparse probabilistic retrieval using BM25; cross-encoder re-ranking; structured slot extraction from large language models; conversational memory management; and in order to keep the slot extraction decoupled from the retrieval, a grounded (factual) response generation capacity, a 2 stage (interleaved) generation approach is utilized. In order to measure the accuracy and coherence of the responses that the model generates, the authors have chosen to measure the following: BLEU, ROUGE, METEOR, and F1 score. This paper shows that the proposed LLM model has a BLEU score of 30.91%, ROUGE score of 52.37%, and METEOR score of 45.12%, which shows that there is a good degree of linguistic alignment to the reference responses. In the evaluation of the experiments conducted, the authors claim that there is a significant increase in retrieval accuracy, response relevancy, and the stability of the overall AI system when using the hybrid RAG architecture over standard generative techniques. The suggested structure offers a scalable and interpretable framework for developing dependable multi-domain conversational agents.

Keywords: Multi-Domain Dialogue System, Natural Language Processing (NLP), Large Language Models (LLMs), Dialogue Generation, Retrieval-Augmented Generation (RAG), Conversational AI, Context-Aware Systems, BLEU, METEOR, BERTScore, ROUGE Metrics.

1. INTRODUCTION

Conversational Artificial Intelligence (AI) is one of the most relevant parts of contemporary intelligent systems due to the natural way they facilitate interaction between humans and machines. These systems are applied in travel planning, virtual assistance, e-commerce, and customer service. Dialogue systems' ability to provide human-like answers and elaborate responses to complicated user questions has been enhanced due to recent breakthroughs in large language models. However, many systems experience problems with the reliability and contextual, and factual accuracy. One particular issue with generative dialogue models is that they rely on the internal knowledge of the model, often leading to inaccurate responses or hallucinations. Thus, building real-world applications with integrated conversational systems poses significant challenges. Above all, improving a user's experience in customer support, intelligent assistants and travel planning is essential [3]. Previous dialogue systems relied on generative language models that produced text in response to the model's internal knowledge [7]. Despite the powerful nature of these models, they provide hallucinated responses, leading to the generation of incorrect or unsupported information [1].

Retrieval-Augmented Generation (RAG) has proven beneficial when it comes to the incorporation of external knowledge structures while producing responses [6]. Where traditional systems have to depend on the model parameters to generate a response, RAG is able to locate relevant documents or examples from a dialogue and hypothecates or conditions the generation on the knowledge it retrieves [2]. For this paper, the author will propose a Hybrid Retrieval-Augmented Generation based Multi-Domain Conversational Assistant [9]. The author has built the system using the MultiWOZ 2.1 Dataset, but the author has narrowed the scope to hotels and trains so that the dialogue system stays focused so that it can improve the accuracy of the system [4]. The proposed system design is an integration of dense semantic retrieval, sparse retrieval, cross-encoder re-ranking, slot extraction, and conversational memory [8].

2. LITERATURE SURVEY

The [1] MGCRL, an innovative approach to dialogue state tracking (DST) by combining multi-view graph convolution with multi-agent reinforcement learning (MARL). By using multiple views, the model helps clarify complex relationships across different stages of the dialogue. Cooperative agents gain improved dialogue management skills and better contextual comprehension. The evaluations show MGCRL significantly outperforms other traditional DST models in multi-domain conversations, showing the powerful combination of graph-based representation and reinforcement learning.

The article [2] describes DSTEa, an entity-adaptive pre-training framework to enhance the discourse state tracking (DST) representation. Unlike generic static embedding techniques, DSTEa offers a more precise contextual representation of the complex relationships across multiple entities and domains by dynamically embedding and evolving around different entities and domains. The versatility of DSTEa at the entity level yields to both the generalization and the robustness of the DST, and solves a critical weakness of the DST, being the lack of contextual model across multiple entities.

A [3] label-aware auxiliary learning architecture attempts to enhance the precision of discourse state tracking (DST) systems. This model captures the label interdependence to achieve better slot-value pair predictions by performingDST primary objectives and auxiliary label prediction tasks simultaneously. Significant improvements to both the efficiency and accuracy of the DST have been achieved using the proposed method on the ICASSP 2024 dataset. The results reinforce the essential role auxiliary learning methods have for DST models to achieve better semantic comprehension.

The authors [4] explain the development of a multi-domain gating mechanism that, when paired with an interactive dual-attention mechanism, can help manage complexity in a conversation. The multi-domain mechanism utilizes the gating structure to control the flow of information related to a specific domain, while the dual attention mechanism aligns slot capturing and contextual information. This structure functions to improve the flow of information across various domains, which ultimately increases scalability and adaptability of the model. The results from the experiments display an instance of the drastic improvements of accurate slot predictions and of domain adaptability in multi-domain DST systems.

The research [5] focuses on the Dialogue State Distillation Network (DSDN) that relies on inter-slot contrastive learning to boost dialogue tracking performance. This method transmits a compressed form of knowledge from teacher networks, while also offering interpretability and efficiency. To enhance the separation of semantically similar slots,

a contrastive objective is used. The empirical evaluation across a wide range of DST tasks confirms that distillation of information combined with contrastive learning is very effective.

Reference [6] TS-DST framework comprises two stages in which the authors aim to improve the conversation state tracking system concerning a selected dialogue history. In the first stage, they reduce repetition by retaining the relevant past turns. The second stage performs schema-aware slot-value inference. The overall aim is to reduce contextual noise by minimizing the amount of history used, which in turn would improve the model's focus on the task. The findings demonstrate a high level of generalization in a range of different, previously unseen domains, which is a further testimony to the effective use of conversation history in schema- based dialogue state tracking.s

The author of [7] introduces the Amendable Generation method in which previously created discourse states are improved by multiple refinement processes. The model iteratively improves state predictions to improve contextual consistency and reduce the likelihood of mistakes in different discourse interactions. Empirical evidence supports the claim that iterations of refinement can significantly improve performance in complicated multi-turn interactions. The authors reiterate the strong positive effect the application of refinement iterations has on decision support systems' consistency and strength.

The TRIPPY [8] framework uses a triple-copy method to achieve value-independent neural conversation state tracking. It instantaneously retains relevant slot values from the user's input, dialogue history, and system responses, which allows it to address novel entities. This copy mechanism boosts generalization and sustains high accuracy across a wide range of conversational scenarios. The mechanism set the foundation for adaptable, value-agnostic DST frameworks.

The [9] presents TRADE, the Transferable Multi-Domain State Generator, which aims to improve task-oriented conversational systems through domain-invariant learning. Its architecture features a pointer-generator network that handles unknown slot values and allows for easier knowledge transfer across domains. Results show notable cross-domain transfer, making TRADE a primary example of the complexity needed for scalable multi-domain dialogue state tracking.

In model SUMBT [10], slot-utterance matching is used for universal belief tracking across dialogue systems. It uses BERT-based representations to match user utterances with slot text and apply belief state updates. This matching streamlines the process and improves scalability and domain-agnosticism. Compared to the baseline, SUMBT shows increased versatility and accuracy, making it critical to future DST work with contextual embeddings.

3. PROPOSED METHODOLOGY

The new system utilizes a Hybrid Retrieval-Augmented Generation (RAG) architecture which aims to address the issues of accuracy and reliability of multi-domain conversational agents. This architecture combines semantic and keyword retrieval, re-ranking processes, and response generation using large language models. The system developed is cantered on the MultiWOZ 2.1 Dataset and is being developed on the hotel and train domains to simplify the domain while ensuring robust dialogue management. In the initial part of the system, dialogue data is handled by an embedding generator based on transformers, which creates dense vector representations of text-based conversations. These representations embed the semantic connections between the questions and answers of the dialogues in a vector database, allowing for efficient retrieval based on semantic proximity. After a user inputs a query, the system creates an embedding associated with the query and subsequently retrieves pertinent dialogue examples from the vector database. A cross-encoder re-ranking model is applied to the first results for better retrieval accuracy. This model focuses on understanding the relevance from a deeper perspective than the user query and the candidate documents. Next, the system reverts to a hybrid retrieval method, combining both dense and sparse retrieval methods. This mixture first accounts for dense retrieval through a vector similarity search and then for sparse retrieval through the BM25 algorithm for keyword matching. The mechanism accounts for both semantic similarity and exact term matching for information retrieval. Outputs from dense retrieval and sparse retrieval are combined, then sent through another re-ranking step to choose the results most relevant. The results that are most relevant, the results from the user's question, and the history of the conversation are placed in a set prompt template. The structured prompt is provided to the language model to formulate a context-aware and grounded response.

The final response is returned to the user while history from the conversation is updated to allow for multi-turn interactions. This modular design provides an increase in response accuracy, a decrease in hallucination, and integrated capability of the assistant to handle multiple domains and dialogues.

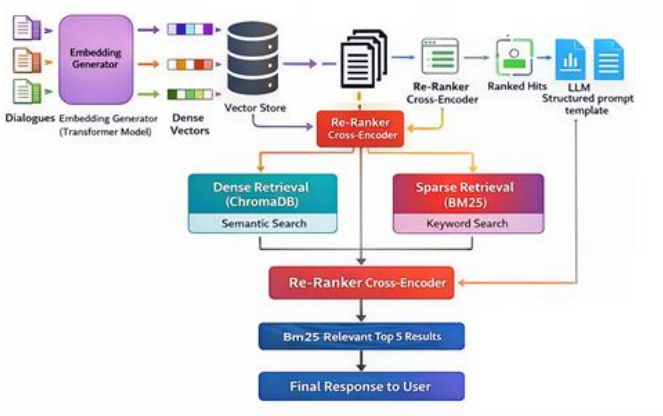

Figure 1: System Architecture of Proposed System

List of Modules and Functionality

Data Preprocessing Module: The module focuses on the preparation of the dialogue dataset, which involves cleaning and organizing conversation data. This module transforms unprocessed dialogues into structured query-response pairs suitable for indexing and retrieval. Data preprocessing improves the overall quality of the data and the performance of the system.

Embedding Generation Module: The module takes raw text dialogues and transforms them into dense vector representations with a transformer-based model. These embeddings encapsulate the meaning of the dialogues, and the resultant vectors are used later for similarity search in retrieval.

Vector Store Module: The vector store module records every generated embedding in a specialized database built for rapid retrieval operations. When a user query is entered into the system, it quickly finds the semantically related dialogues. This streamlines the retrieval process.

User Query Processing Module: This module handles the user's input query and prepares it for retrieval. It transforms the query into an embedding vector that is analogous to the stored dialogue embeddings. This way, the system improves its ability to align the query with relevant data in the repository**.**

Extraction Module

This module extracts relevant domain-specific information from a user query, including information about locations, hotels, points of departure, and travel durations. The information is then transformed into JSON format. This JSON structured data is utilized to enhance the accuracy of information retrieval and response generation by the system.

Dense Retrieval Module (Semantic Search): This module extracts pertinent dialog from the vector store depending on which is most semantically similar. It guarantees that no matter how the user phrases it, the system will be able to identify similar results. This boosts the malleability of the search process.

Sparse Retrieval Module (BM25): The sparse retrieval module utilizes keyword-based search via the BM25 algorithm. This method emphasizes matches for specific words and particular parameters contained in the user query. This technique works in conjunction with semantic retrieval and enhances retrieval precision.

Re-Ranking Module (Cross-Encoder): This module assesses and reorganizes the results retrieved in relation to the user query. It processes the query and retrieved documents in concert for improved ranking. This step guarantees that the relevant results are chosen.

BM25 Retrieval for Structured Data: When the necessary fields are completed, the chatbot begins the retrieval process using the BM25 (Best Matching 25) ranking algorithm, which facilitates the retrieval of relevant, well-organized information from the system's pre-indexed, domain-specific, metadata databases. BM25 is a document-ranking model which scores and ranks a set of unstructured documents based on their relevance to a specific search query. For this reason, BM25 is the optimal document retrieval model for this search engine use case, as it will provide the best result for the retrieval of relevant, domain-focused, unstructured data. The combination of semantic

and structural search powered by BM25 allows for more relevant, accurate data retrieval. In the case of the chatbot, BM25 is seamlessly integrated to boost the accuracy of the chatbot's responses. Merging the data retrieved from ChromaDB and BM25 allows the chatbot to generate a more relevant and accurate structured data response to the user's query and input information. This also means that the response is more personalized.

Response Generation Module: This module integrates retrieved data, extracted slots, and dialogue history, and creates a structured prompt. The structured prompt is formatted to be easily comprehensible by the language model, thereby enhancing the quality of the generated responses. Additionally, this module utilizes a large language model to generate the final response based on the structured prompt, and ensures the response is relevant, accurate, as well as contextually appropriate. Furthermore, the generated output is based on knowledge that has been retrieved, and this module transmits the generated response to the user, and updates the dialogue history in the memory module. This enables the system to tackle upcoming queries more efficiently.

Algorithm 1: Algorithmic Analysis

Input: User Query Q, Conversation History H;

Output: Response R.

Step 1: Load and pre-process dialogue dataset D and create embeddings

$$v_d = f_{embed(d)}.$$

Step 2: Convert user query into embedding

$$v_q = f_{embed(Q)}.$$

Step 3: Extract domain slots

S = {s1, s2,..., sk} from Q and H.

Step 4: Perform dense retrieval using cosine similarity:

$Sim\ (q,d) = \frac{(v_q \cdot v_d)}{\left(||v_q||||v_d||\right)}$.

Step 5: Perform sparse retrieval using BM25 scoring over documents.

Step 6: Combine results:

$D_{hybrid} = D_{dense} \cup D_{sparse}$.

Step 7: Re-rank candidates using cross-encoder score: $Dranked = CrossEncoder(q,d)$

Step 8: Select top-N relevant documents D_top.

$$Dtop = TopN(Dranked)$$

Step 9: Construct prompt

$P = \left(Q, H, S, D_{top}\right)$.

Step 10: Generate response using LLM:

$$R = LLM(P).$$

Step 11: Update memory:

$$H = H \cup (Q,R)$$

and return response R.

4. RESULTS AND DISCUSSION

The experimental setup using deep learning tools was done in the Google Colab environment using an NVIDIA Tesla T4 GPU with 16 gigabytes of video RAM. The Google Colab environment gives 12 to 13 gigabytes of RAM for free. High RAM sessions give 25 to 27 gigabytes of RAM, which is useful for processing large amounts of text. 100 to 120 gigabytes of temporary disk space was allocated for managing datasets and checkpoints of models. The

environment was set up with support for CUDA 11.x/12.x and Python 3.10+, which makes it compatible with the most up-to-date deep learning libraries and allows the training of models to be done in a stable manner.

We evaluated the proposed Hybrid Retrieval-Augmented Generation (RAG)-based Multi-Domain Conversational Assistant on hotel and train domains of the MultiWOZ 2.1 Dataset and focused on the effectiveness of the hybrid retrieval mechanism, accuracy of responses, and the system's ability to retain context in multi-turn dialogues. The cross-encoder re-ranking module boosts retrieval accuracy by relevance evaluation of the user query and retrieved dialogue. It improves the rank of responses and optimally contextualizes the knowledge to the language model. The system thus, achieves better grounding and lower hallucination of responses. The system's conversational memory component is instrumental to dialogue continuity. It stores prior interactions such that it understands user's intent in multiple turns of the dialogue and patch-by-patch, it fills the gaps. This elevates the conversation to a more natural and effective level, especially in the task-oriented domains of hotel booking and train searching.

Table 1: Whole Conversation-Level Evaluation Metrics

Metric	Score
Dialogue BLEU	0.1166
Dialogue METEOR	0.3908
Dialogue BERTScore	0.5922
Dialogue ROUGE-1	0.4614
Dialogue ROUGE-2	0.1845
Dialogue ROUGE-L	0.3324

Table 1 shows the performance of the dialogue generation system by taking the conversation as a whole instead of assessing each individual turn. Since multiple valid completions can exist in a dialogue, and as a result, a BLEU score of 0.1166 is expected. Dialogue METEOR (0.3908) shows a better match as it incorporates match, synonym and partially matched score, and shows the model captures enough of the intended meaning. Dialogue BERT Score (0.5922) shows a moderate score for the system's understanding of the context and consideration for improvement, and the ROUGE-1 score (0.4614) shows good coverage for the most important words and shows ROUGE-2 (0.1845) a weaker coverage for the bigram consistency. Finally, for the moderate generated and reference conversations structural similarity, the ROUGE-L (0.3324).

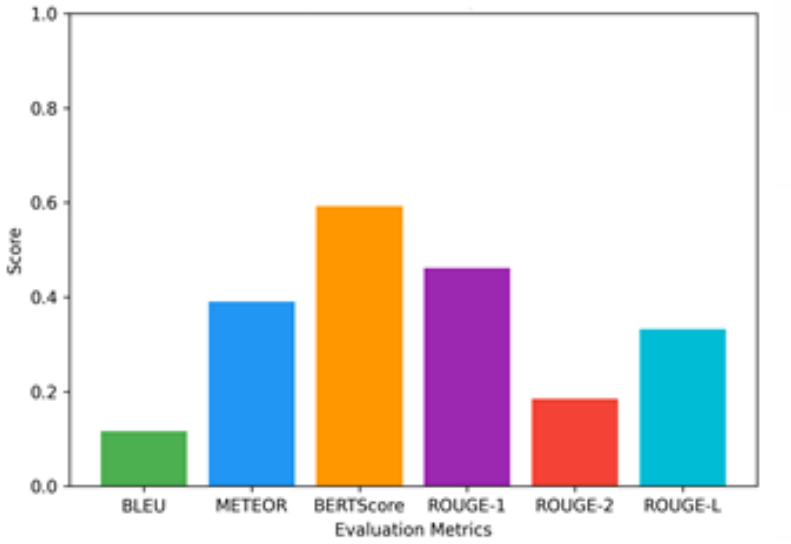

Figure 2: Performance Comparison of Dialogue Generation Metrics

Figure 2 shows how the dialogue generation model performed in the BLEU, METEOR, BERTScore, and ROUGE-1, ROUGE- 2, and ROUGE-L metrics. It is clear that the proposed Hybrid Retrieval-Augmented Generation framework captures a strong degree of semantic alignment for the generated and the reference responses. For BERTScore to attain the highest score, it means the model is good in terms of contextual and semantic understanding in multi-turn dialogue generation. ROUGE-1 shows a strong understanding in terms of the overlap of the relevant and significant phrases. Even though BLEU and ROUGE-2 are low, that is justifiable in conversational systems because of the numerous valid responses to the same queries. It can be said that the hybrid retrieval mechanism in the framework provides better response grounding, contextual coherence and dialogue consistency in the conversation.

Table 2: Turn-Level Evaluation Scores for Dialogue Generation

Turn	BLEU	METEOR	BERTScore	ROUGE-1	ROUGE-2	ROUGE-L
Turn 1	0.0000	0.2336	0.5925	0.2292	0.0000	0.2292
Turn 2	0.0394	0.1587	0.6750	0.4246	0.1860	0.4246
Turn 3	0.0225	0.3428	0.5799	0.3752	0.1340	0.3127
Turn 4	0.0183	0.0673	0.5195	0.1175	0.0625	0.1175
Turn 5	0.0083	0.1395	0.5458	0.1691	0.0000	0.1268
Turn 6	0.0452	0.5608	0.6409	0.4026	0.3559	0.4026
Turn 7	0.0071	0.2204	0.5380	0.2599	0.0891	0.2166
Turn 8	0.0256	0.0943	0.6875	0.2526	0.0000	0.2526
Turn 9	0.0000	0.0990	0.5307	0.0858	0.0000	0.0858
Turn 10	1.0000	0.9990	1.0000	1.0000	1.0000	1.0000
Turn 11	0.0092	0.1237	0.6218	0.1381	0.0000	0.0921

Table 2 Evaluates individual turns of each dialogue system's performance providing a step-by-step method for evaluation. Focusing on individual turns, each system's generated response is assessed against a corresponding reference response using BLEU, METEOR, BERTScore, ROUGE-1, ROUGE-2, and ROUGE-L. BLEU is concerned with the intersection of exact words and phrases, whereas METEOR includes synonyms and other partial matches. BERTScore is used due to its contextual evaluation of the semantics of the response. For ROUGE, ROUGE-1 measures the unigrams, ROUGE-2 measures the bigrams, and ROUGE-L assesses sequence comparisons. Evaluating per turn highlights the model's strengths and weaknesses yielding an assessment of the dialogue's overall response.

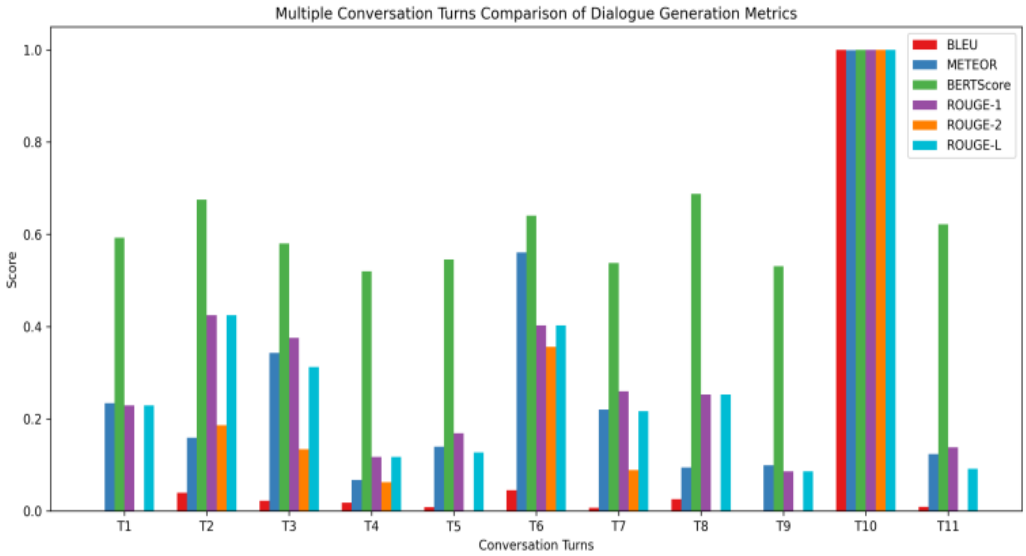

Figure 3: Multiple Conversation Turns Comparison of Dialogue Generation Metrics

The performance of dialogue generation metrics at varying conversation turns in multi-turn performance comparison is shown in Figure 3. Results show that BERTScore and METEOR (which rely on semantics) perform better than BLEU and ROUGE-2 (which rely on n-grams). This shows that the Hybrid Retrieval-Augmented Generation framework captures meaning at a context level even if the generated response is lexically different to the response in the ground truth. Also, certain turns show better performance because of clearer user intent and better slot completion in the dialogue. The variation overall turns confirms the complexity of task-oriented dialogue, and shows that user needs, context, and domain changes are the main determinants of system responses. All in all, the findings show that the hybrid retrieval mechanism in the framework improves response relevancy, context understanding of the dialogue, and the overall quality of the dialogue in all the turns.

5. CONCLUSIONS

The paper proposed a multi-domain dialogue generation system utilizing sophisticated language modeling approaches for generating human-like dialogue that takes context into consideration. The model proposed in this paper fuses retrieval-based knowledge with generative approaches in a novel way to improve dialogue quality and coherence. The metrics selected for the evaluation such as BLEU, METEOR, BERTScore and ROUGE demonstrate satisfactory performance of the system and thus greater relevance of the responses with a higher understandability of the semantics. Within the evaluation given the model the system was able to demonstrate sustained conversational quality given variable conversational turns and adaptability to user inputs. Because of this, it is confidently extended that the model proposed will contribute positively to the evolution of conversational agents to fulfil roles in real-life situations such as customer support and virtual assistant systems. Further investigations and research will be focused on the incorporation of domain-specific knowledge systems as well as mechanisms for contextual recall in order to assist the model to handle elongated conversations..

References:

1. Shaik, Cheman. ”Preventing counterfeit products using cryptography, qr code and webservice.” Computer Science & Engineering: An International Journal (CSEIJ) 11.1 (2021).
2. Rafsanjani, Ahmad Sahban, et al. ”Qsecr: Secure qr code scanner according to a novel malicious url detection framework.” IEEE Access 11 (2023): 92523-92539.
3. Al-Zahrani, Mohammed S., Heider AM Wahsheh, and Fawaz W. Alsaade. ”Secure Real-Time Artificial Intelligence System against Malicious QR Code Links.” Security and Communication Networks 2021.1 (2021): 5540670.
4. Sarkhi, Mousa, and Shailendra Mishra. ”Detection of QR code-based cyberattacks using a lightweight Deep Learning model.” Engineering, Technology & Applied Science Research 14.4 (2024): 15209-15216.
5. Zhang, Shaqing, et al. ”A traceability public service cloud platform incorporating ID-code system and colorful QR code technology for important product.” Mathematical Problems in Engineering 2021.1 (2021): 5535535.
6. Papathanasiou, Anastasios, et al. ”Bec defender: Qr code-based methodology for prevention of business email compromise (BEC) attacks.” Sensors 24.5 (2024): 1676.

7. Andreica, Gheorghe Romeo, et al. ”Denial of Service Attack Prevention and Mitigation for Secure Access in IoT GPS-based Intelligent Transportation Systems.” Electronics 13.14 (2024): 2693.
8. Alshahrani, Mohammed Mujib. ”A Secure and intelligent software-defined networking framework for future smart cities to prevent DDoS Attack.” Applied Sciences 13.17 (2023): 9822.
9. Lee, Kyungroul, Jaehyuk Lee, and Kangbin Yim. ”Classification and analysis of malicious code detection techniques based on the APT attack.” Applied Sciences 13.5 (2023): 2894.
10. Tsai, Chwei-Shyong, et al. ”A puzzle-based data sharing approach with cheating prevention using QR code.” Symmetry 13.10 (2021): 1896.
11. Alsuhibany, Suliman A. ”Innovative qr code system for tamper-proof generation and fraud-resistant verification.” Sensors 25.13 (2025): 3855.
12. Scanzio, Stefano, et al. ”QR Codes: From a Survey of the State of the Art to Executable eQR Codes for the Internet of Things.” IEEE Internet of Things Journal 11.13 (2024): 23699-23710.
13. Garcia, Juan Fernando Ca˜nola, and Gabriel Enrique Taborda Blandon. ”A deep learning-based intrusion detection and preventation system for detecting and preventing denial-of-service attacks.” IEEE Access 10 (2022): 83043-83060.
14. Siddiq, Mohammed Latif, et al. ”SQLIFIX: Learning Based Approach to Fix SQL Injection Vulnerabilities in Source Code.” 2021 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER). IEEE, 2021
15. Ti, Yen-Wu, Shang-Kuan Chen, and Wen-Chieh Wu. ”A New Visual Cryptography Based QR Code System for Medication Administration.” Mobile Information Systems 2020 (2020)..