World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Research Article)

Check for updates

Context-aware chatbots with data engineering for multi-turn conversations

Snigdha Tadanki * and Sai Kiran Reddy Malikireddy

Independent Researcher, USA.

World Journal of Advanced Engineering Technology and Sciences, 2021, 04(01), 063-078

Publication history: Received on 20 September 2021; revised on 24 November 2021; accepted on 26 November 2021

Article DOI: https://doi.org/10.30574/wjaets.2021.4.1.0061

Abstract

The traditional method of deploying chatbots, which only answered simple questions, has developed to the present form, where chatbots are complex conversational models able to handle a sequence of turns within a conversation. This research analyzes context-aware chatbots on moderates in data engineering approaches and state-of-the-art machine learning methods. Using such tactics, this study focuses on critical aspects like data preprocessing and feature engineering and on creating training pipelines for which this study intends to address core concerns entrenched in the challenge of achieving conversational context switching across multiple exchanges.

Keeping the context relevant is one activity that defines the multitudinous turn-taking communication processes. The research also pays special attention to the preprocessing step, which removes the noise from the data and improves the training dataset improves the training dataset. Feature engineering stands central to extracting linguistic and contextual features as a precondition for models to understand the user input and continue conversation selectively. These processes are best trained from pipelines designed for such flows, including reiterative feedback loops to help the model learn how to store and manipulate context as it adapts.

It also examines the adoption of front-end technologies to enrich the customer experience and create great customer feedback. The UI is built not only to represent the bot's ability but also with an inglorious role of adapting to suit the user to maintain an interactive and realistic user-friendly dialog. Through the use of users' friendly features, these interfaces act as a middle link between the complicated back-end systems and the consumers, making them more comfortable to use.

Keywords: Context-aware chatbots; Multi-turn conversations; Data engineering; Natural language processing (NLP); Conversational AI

1. Introduction

1.1. Background and Motivation

Whereas the early mode of implementing chatbots involved them responding to queries, the current dispensation entails that chatbots are sophisticated conversational models that handle an exchange of turns in a conversation. This research offers considerations on context-aware chatbots grounded on the developments in the current data engineering methods and recent machine learning techniques. In that manner, this study focuses on the effective tasks and training pipelines for which it seeks to address fundamental issues rooted in the directionality of the challenge of conversing and context-switching between exchanges.

Maintaining the relevancy of the context is one of the functions that characterize numerous turn-taking communication interactions. The research also focuses much on the preprocessing step, which removes noise from data to make up the

* Corresponding author: Snigdha Tadanki

training dataset. Feature engineering remains germane to two tasks: extracting linguistic and contextual features as a prerequisite for models to comprehend the user input and engage in a selective conversation. These processes are trained best from a pipeline drawn from such flow, capable of including reiteration feedback loops to teach the model how to store and manipulate context as it learns.

It also explores the use of front-end technology solutions to enhance the customer experience and build great feedback. The UI is created to signify the bot's prowess and to have an ignoble function of needing to conform to the user to mimic real organic and intuitive user-friendly chat. By user-friendly interfaces, these interfaces help to comfort otherwise highly complex back-end systems serving consumers.

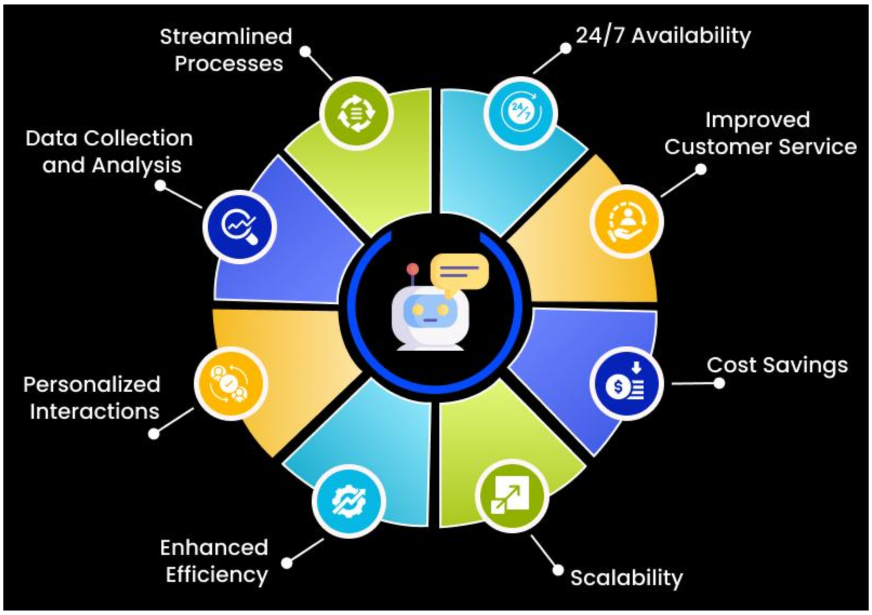

Figure 1 Evolution of Chatbots: Overview of AI-powered conversational aids

As mentioned, innumerable issues make the development of such systems complex and difficult. One of the most important problems is managing context over multiple dialogue turns when users drift off the standard flow of dialog or use rather fuzzy statements. The conventional approaches of NLP find it difficult to manage in these circumstances since they cannot retain the history of the conversation and also understand the user's specific purpose. In addition, communication by the user involves body language, at least one language, and preferred language, making designing a more reliable context-sensitive system challenging.

To a large extent, data engineering helps these challenges and facilitates the connected conversational flow in chatbots. Data quality, including chatbots, is always the key to forming any machine learning model. Data engineering is a set of activities aimed at data acquisition, preparation, cleaning, transformation, and subsequent storing, which are critical for training and improving chatbot models. Optimized workflows enable the chatbot to obtain the proper data when needed, process it correctly in real time, and provide a relevant response. Furthermore, data engineering helps to combine the content of external knowledge bases in answering with detailed and accurate information concerning many topics. These ideas make sense only if supported by strong data engineering capabilities, which would otherwise not be able to handle the intricacies of multi-turn conversation.

1.2. Research Objectives

This research seeks to plug the existing gaps in context engineering for context-aware chatbots through the application of data engineering and sophisticated machine learning algorithms. The main goal is to create and employ a system that performs well in tracking conversation context across multiple exchanges to improve user satisfaction. To this end, the study encompasses several aspects of data preprocessing, feature engineering, and development that target effective comprehension of contextual information.

This paper aims to identify the impact of data engineering on enhancing the effectiveness of chatbot chatbots. In this way, the present study aims to create reliable data accumulation, washing, and transformation mechanisms to provide the Chatbot with highly significant fresh data. Moreover, the work below considers how to incorporate existing external data source knowledge graphs and APIs for improving the performances of the Chatbot for the context relevancy of the responses.

Another important goal is to review the latest approach to designing context-aware chatbots based on many advanced machine learning methods. The work also compares different methods for context tracking: memory networks, attention mechanisms, transformer models, and others in terms of their ability to handle multi-turn dialogues. Further, it reveals how GPT and BERT can enhance the Chatbot's ability to understand natural language and generate human-like responses.

Another element of this research is including design elements for creating the Chatbot. While back-end technologies and algorithms are well looked for, keeping the context steady, the front end or the user interface is crucial in defining the conversational profile. This paper focuses on using responsive front-end technologies, including dynamic rendering and personalization, to enable users to interact better with the front-end interface.

This research also seeks to assert the novelty of its discoveries in conversational AI. Thus, implementing well-established data engineering practices alongside the novel modern machine learning paradigms, this work unifies the components necessary for designing multi-turn context-aware conversational agents. This study's findings can help enhance conversational AI's progress, which is why the information revealed is inspiring for the developers and researchers who work on similar projects.

2. Related Work

There has been tremendous development in conversational AI, and context-aware chatbots now represent a subfield with much focus. Context-aware systems are intended to help chatbots efficiently manage and use the contextual data contained in conversational data. However, the nature of multi-turn dialogues, the problem of data association, and issues in data preprocessing and model architecture require a better understanding of prior work and its drawbacks. This section briefly reviews previous efforts on context management and data engineering for chatbots and discusses known issues in managing context across interactions.

2.1. Context-Aware Chatbots

The implementation of context-aware chatbots has been elicited due to the need for systems not restricted to one-shot, three-question answering types. In their early development, chatbots worked on strict sets of rules, and employing natural language processing was a big problem because of the context-sensitive interactions. These systems relied on sets of scripts or patterns for each response; as such, they were incapable of responding to new contexts or new situations.

The context handling has been easy when using machine learning, especially natural language processing. Sequence-to-sequence models were first introduced in machine translations and served as a basis for conversational AI. These models apply an encoder-decoder framework capable of producing output sequences given an input sequence. Nonetheless, while these models could perform fairly adequately with single-turn exchanges, they did not work well with multiple-turn dialogues, as their architectures did not support the ability to remember the long-term context.

The measures brought in by the recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and gated recurrent units (GRUs) helped in retaining context to some extent. These architectures contain feedback mechanisms to enable information from prior time steps to influence the current outputs. Nevertheless, these models prove weak when client-server conversations are long due to the vanishing gradient problem, in which information deteriorates with time.

Conversational AI has transformed English conversation due to the challenges that transformers have solved in the previous architectures. Systems such as BERT that work on Bidirectional Encoder Representations from Transformers and GPT that are based on Generative Pre-trained Transformer also show a great capability of handling context within and between the sentential level, which will help a chatbot retain the multi-turn dialogues. Multihead attention mechanisms in transformers enable them to pay attention and retain contextual information much better than sequential models. However, courtesy of fine-tuning approaches, these models have since been tailored for performance specific to any domain or conversational patterns, making them very useful in real situations.

Table 1 Context Handling In Chatbots, Key Features, Strengths, And Limitations

Method	Key Features	Strengths	Limitations
Rule-Based Systems	Predefined rules and decision trees for context management.	Easy to design and implement for specific tasks; highly predictable responses.	Lack of scalability; unable to handle complex or dynamic conversations.
Statistical Models	Probabilistic approaches using n-grams or Markov models to capture conversation patterns.	Can model simple conversational dependencies; computationally lightweight.	Limited to short contexts; fails in handling long-term dependencies or ambiguity.
RNNs (Recurrent Neural Networks)	Sequential models capable of processing conversational history.	Can manage sequential data; retains some degree of context across turns.	Struggles with long-term dependencies; prone to vanishing gradient problems.
LSTMs/GRUs	Advanced RNNs with memory cells to store relevant conversational context.	Better at handling long-term dependencies compared to standard RNNs.	High computational cost; performance can degrade with noisy or inconsistent input data.
Transformers	Attention-based models capable of processing entire conversational history in parallel.	Superior at capturing long-term dependencies; state-of-the-art performance in many NLP tasks.	Requires large datasets and significant computational resources; context sensitivity can vary.
Memory Networks	Augments standard neural networks with external memory for explicit context storage and retrieval.	Effective for explicit context retention and recall; interpretable context representations.	Can be difficult to train; may struggle with ambiguous or rapidly changing contexts.
Hierarchical Models	Multi-level models that segment conversations into hierarchical structures for context management.	Structured context retention; works well for multi-turn, topic-segmented conversations.	Complexity in design and training; performance may degrade with overlapping or mixed topics.

These insights notwithstanding, the problem of keeping track of context in open-domain interactions remains one of the tougher challenges. Most current systems use heuristics or external memory for managing and storing context that may be erroneous because of misconceptions about the context or purely due to context explosion.

2.2. Data Engineering for Conversational AI

Conversational AI systems are closely related to data engineering. The nature and characteristics of data fed to the chatbot determine whether the responses obtained will be useful or associated with the context of the message. As for context-aware chatbots, data engineering includes a set of important steps, such as data gathering, data preparation, feature selection, and training pipeline generation.

Collecting data is the elementary process of creating any Artificial Intelligence system. Conversational AI entails using dialogues in collections from multiple sources, including customers' support logs, social media, and synthetic data. The datasets used to train such a chatbot should be different and random to prepare it for as many situations as possible. Furthermore, datasets will require supervised learning for intents, entities, and contextual dependencies for annotating intents and entities.

Data preprocessing is important to ensure that the raw data is sufficient for training machine learning models. Cleansing is normally achieved through normalization, word splitting, and eliminating spares or irrelevant content in the data set. Depending on the specific application, processing might also comprise the partitioning of what HSY refers to as sensible dialogues based on semantically meaningful and minimally disrupted conversational segments and the annotation of conversational turns based on contextual data.

Feature extraction is another key step in data engineering as well. Previous works utilized 'design-by-hand' language features, such as n-grams and part-of-speech tags, to discretize text data. However, this shift to data-driven feature extraction was initiated with word embeddings like Word2Vec and GloVe. These embeddings highlight semantic functionality between words and, subsequently, the appreciation of context by models. It is done because the current models employ contextual embeddings, like those provided by BERT or GPT, which work with the sentence or dialogue as context.

Data pipelines are among the most important design decisions due to the large datasets needed for state-of-the-art models. These pipelines must handle and forward this data and process it in a form suitable for conversational AI applications – often in real time. For instance, data augmentation, which involves generating a synthetic version of the data, is usually employed to enhance model data and generalization. Furthermore, feedback loops allow systems to learn the users' behaviors, resulting in incremental enhancement of such systems' performance.

2.3. Challenges in Multi-Turn Conversations

Of course, researchers have already made great strides in creating context-aware chatbots, but there are still a few problems regarding multi-turn dialogues. Such challenges arise from the nature of context and its management and use during an extended interaction.

The very first issue is context retention. Finally, end-to-end conversations involve multiple sequences, unlike single-turn interactions, where the chatbot is expected to refer to information obtained in the previous sequences. For instance, in natural human-computer interaction, if a user says, "What is the weather like today?" and "How about tomorrow?" the chatbot has to know that "tomorrow" relates to the context that was set in the prior turn of the conversation. Forgetting to maintain such context may lead to incoherent or irrelevant responses by the social commerce site. As good as it gets, most transformer-based models are imperfect regarding context retention. Memory mechanisms can fail if the long conversations or the dependencies are complex.

Another challenge is How to resolve ambiguity. In such a multi-turn dialogue, the user inputs always contain inadequate or ambiguous strings based on contextual references. For instance, users may ask, "Can you provide more information?" without details on what they need. Solving such cases is challenging and demands interpretable natural language understanding and the use of prior context and other knowledge sources.

Another problem with multi-turn conversations is scalability. However, as the context expands, the computational and memory overheads needed to manage the context can be very high. Most of today's systems approximate or sample the context to tackle this problem, but most of these approximations eliminate relevant information. Pursuing both speed and high levels of measurement precision is still recognized as a major research issue.

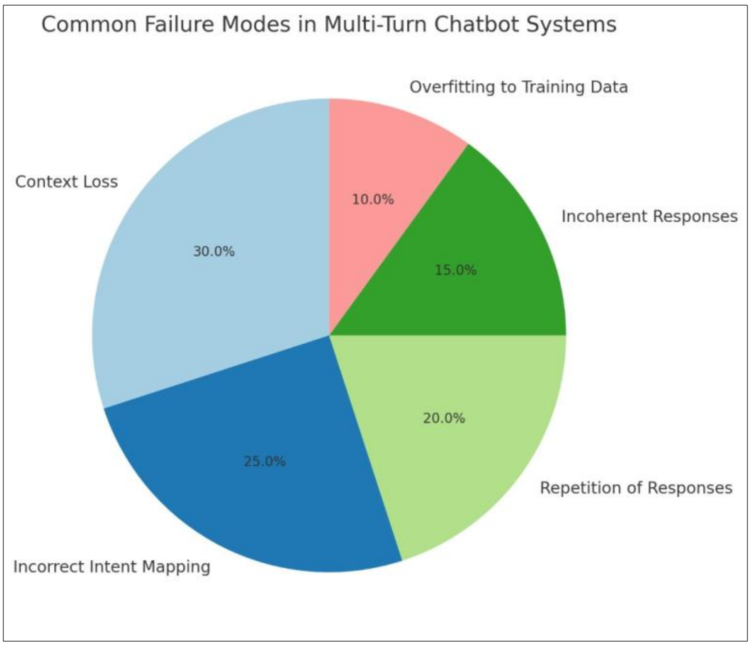

Figure 2 A pie chart visualizing the common failure modes in multi-turn chatbot systems

Moreover, because natural language interactions are open-ended, any interaction between the model and the user can go in unknown directions. Most of the systems are trained on sanitized data sets, which may not necessarily reflect the true variability of the natural dialogue. This can lead to poor generalization when the chatbot is implemented in the real world. This problem requires more diverse training data sets and models to help us learn and embody additional knowledge and experience.

As it will be seen, ethical issues also present themselves as challenges in multi-turn conversational AI. A self-learning chatbot aware of its operating environment can exploit the limit of its knowledge and use personal sensitive information unethically if these measures are not in place. For example, I imagine a chatbot that shares information with one user and later with another, providing the outcome of the prior discussion. Hence, it is crucial to guarantee the safer and more reasonable use of context to create more trustworthy systems.

3. Methodology

3.1. System Architecture

The proposed chatbot system is modular in design with special provisions to track the flow of the conversation while processing data at runtime. At the core, the architecture comprises three primary components: the input processing module, the context management system, and the response generation engine. The input processing module deals with user queries and parses them into recognizable pieces, splitting them into entities, intents, and sentiment sub-modules. The latter converts it into information to be supplied to the context management system that keeps track of the conversation history so that the positioning of the current conversation is recognized and used as a basis for the chatbot's responses to the current exchanges.

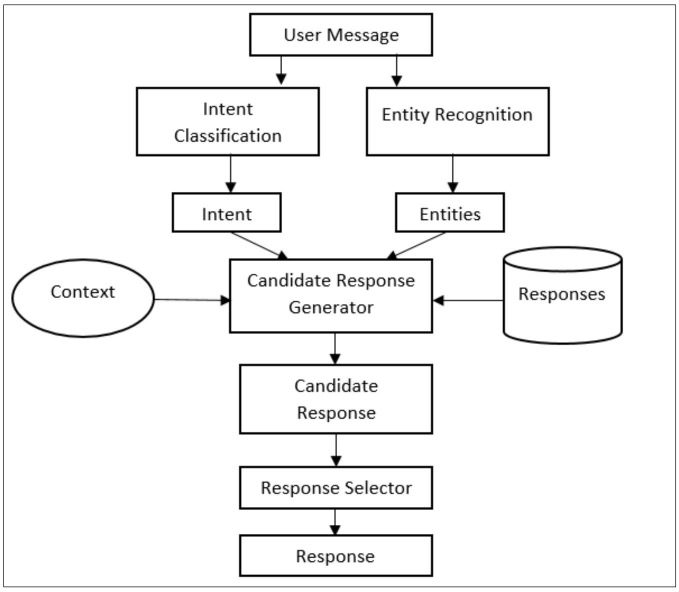

Figure 3 Chatbot: Architecture, Applications and Design Process Steps

The context management system is part of memory-augmented neural networks that store and remind details about the previous turns in the current interaction. This system is further improved by contextual embeddings derived from transformer-based models such as BERT or GPT to understand users' input better. The response generation engine uses retrieval-based and generative models so the chatbot responds properly and to the point. Thus, this modular structure contributes to flexibility for the development or integration of the architecture while permitting each component to provide eventual enhancements in the performance of the architecture.

3.2. Data Engineering Pipeline

The high-quality data engineering pipeline is critical because it guarantees the quality of training and the operational environment. Data acquisition proceeds from gaining access to large conversations from various sources, from Internet chat archives and other conversational corpora of different sizes to corpora from specific domains of interest and complex user interaction logs. Preprocessing techniques include normalizing or converting all texts collected from the raw data to upper case, tokenizing all collected texts, and removing noise. Preprocessing procedures such as stopword movement, stemming, and lemmatization will normalize the text while maintaining the actual meaning.

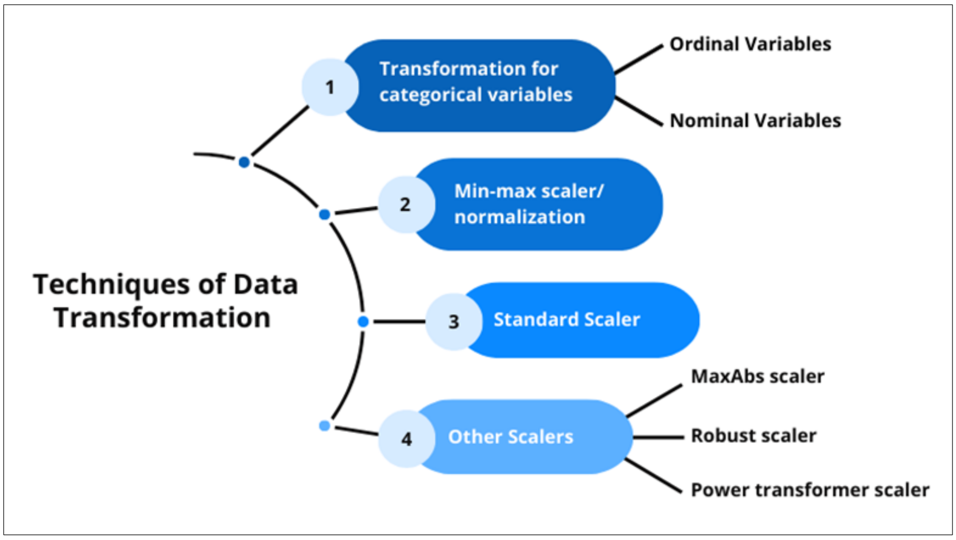

Figure 4 Data Preprocessing Techniques in Machine Learning

Feature engineering is the second most important step in creating important representations with the preprocessed data. The patterns are particularly concerned with user intent, named entities, and semantic relations concerning which NLP tools are employed. Entity recognition models detect the entities to recognize in the text itself. Intent Classification categorizes the user's query into a set of specific Intents. Contextual embeddings integrated from transformer-based language models define all the co-contextual representations of a word, allowing for the comprehension of user inputs in a much richer manner.

After preprocessing the data, its final disposition includes a training, validation, and test data split. When and how the training is done are as follows: the models are first trained on the filtered and refined dataset and then fine-tuned on the specific domain of the chatbot. To achieve real-time performance, the above pipeline also includes the ability to incrementally learn or update the chatbot when new data and user preferences are introduced.

3.3. Contextual Modeling

Preserving and incorporating the conversational context is crucial to developing chatbots that respond to two or more turns. This challenge can be overcome through a proposed system that utilizes sophisticated algorithms and models in its structure. At its core, we find the use of transformer architectures, like BERT or GPT, capable of modeling long-range dependencies and semantically rich relations between contextualized embedding inputs. These models harness self-attention mechanisms to learn the attention of some input, allowing the chatbot to attend to the right details at any time in a conversation.

To inflict conversation management, memory networks form a part of the architecture. They also preserve memory states containing details about the participants and other important information from the apparatus's prior interactions. In this case, attention mechanisms are applied to the memory state so that relevant information is attended to from the state and incorporated into the current response. Ideally, this system reduces the chances of tackling issues with what is conversational context, which is characteristic of most chatbot systems.

Furthermore, an attempt is made to use recurrent neural networks (RNNs) with long short-term memory (LSTM) units as an application for tracking context in cases where memory capacity is the issue. While these networks perform well in keeping a sequential stream of dialogue, they are less potent when managing large contexts than the transformers. A comparison of these models reveals that transformer-based architectures outperform the others in terms of maintaining context during long conversations, which makes them suitable for the proposed system.

3.4. Integration and Deployment

Applying the chatbot to real-world applications involves consideration of the system response time, capacity, and friendliness. The deployment process starts with integrating the chatbot into the communication channels, including websites, messaging apps, and voice assistants. This integration is done through APIs that enable the chatbot's back-end system to interact with front-end interfaces. The front-end design is geared towards interactivity, allowing users to interact with an avatar or use voice for input over text to the chatbot.

To help with scalability, we host the chatbot system on cloud architecture with the help of Docker and Kubernetes. This way, the chatbot can manage different numbers of users at once, giving the best experience to all the users at once. Optimized inference engines, caching mechanisms, and other real-time processing features make handling requests in time and with a low response rate possible.

Follow-ups and assessments are ongoing to track the possible performance of the live chatbot. Performance measurements like the accuracy of response to a conversation, the success rate of a conversation, and the evaluation of satisfaction by the users are done to assess a particular factor that has resulted in inefficiency or areas that need improvement. The training process is attached to user feedback to improve the answers given by the chatbot at a contextual level. Because the system is subsequently refined based on the interaction with the real world, the chatbot becomes more engaging and intelligent in its conversation.

4. Experimental Setup

This section presents an overview of the tools and approaches employed in the experimentations and different measures used to validate the performance of the Context-aware chatbot model proposed in this study. They include the train and test datasets, the measures adopted to assess the proposed model's effectiveness, and the baseline models used in the studies. With a strong experimental base set up, the study guarantees that the research outcomes are credible and significant.

4.1. Datasets

As in any machine learning, the selection of datasets is of utmost importance for the model's general performance and final architecture. The effectiveness of the approaches towards chatbot development and evaluation that consider context requires a dataset that involves the richness of turn-taking in the conversation and allows the model to update its contextual state across many turns. Several datasets were selected from the public domain for this study, which was used to build and test the chatbot. This data set comprises domain-specific and general conversational data to ensure that the intended model understands broad conversation topics and user intent.

Among them, one of the major dataset collections is the Cornell Movie Dialogues Corpus, which comprises several movie scripts with complex, multi-turn characters' conversations. Due to the nature of this dataset, it is considered useful given that it contains the natural flow of conversation and rather good coverage of domains, ranging from light conversations to emotional ones, thereby making it diverse in terms of conversational characteristics. The dialogues are ideal for teaching the model how well it retains a context between different turns, an essential aspect of a contextualized chatbot. Moreover, the datasets are complemented by annotations of speaker roles and turns, which helps navigate the context of turn-taking and change of the dialogue state.

The second pertinent dataset employed in this research is the large-scale Persona-Chat Dataset, which offers more unique conversational features. It comprises conversations between two human beings; each has a role assigned to them. This dataset was chosen because it raises another challenge – tracking and recalling user specifics about their preferences, context, and personality throughout the talk. This is because the proposed Persona-Chat Dataset can be incredibly helpful for training models in producing more contextually aware and personal responses — and this increases the potential of the chatbot for maintaining more realistic, natural, multi-turn chats.

In addition to adding diversity to these datasets, a data augmentation approach was employed to generate a synthetic dialogues dataset. The idea was to induce seven conversations to test the model's versatility across different contexts and user intent domains. It is most useful when the real-world data collection may be limited or the data set might be skewed. Existing dialogues were augmented by applying controlled random perturbation to the text, including syntactic transformations and lexical substitutions or introducing new topics to related topics, which comprised the augmented set of dialogues. This data augmentation was beneficial to removing some of the monotony that can be expected from the fixed training set while simultaneously increasing the possibility of mimicking real-world scenarios.

Further, Wikipedia Talk Pages are included as a source of domain-level discussions. These pages contain fact-discussing content and presentations of the authors' opinions and individual opinions left in the comments. With these discussions, the chatbot learned to handle more specific types of discussions and the transition between different conversation types

– from a factual question-answer dialog to a dialog where the model needs to switch between various modes of conversation.

When real data is augmented with synthetic data, the model works well in different types of conversation user contexts, and behaviors.

4.2. Evaluation Metrics

There is a need for a comprehensive evaluation of the model under development once this is trained using the datasets above. The metrics used in this paper are chosen to reflect the fluency and quality of the answers, the ability to preserve context information over multiple turns, and user satisfaction. Several standard measures in the NLP and conversational AI domains are applied to evaluate the chatbot's performance.

BLEU (Bilingual Evaluation Understudy Score) is one of the most used evaluation methods for synthesizing textual information, and it is suitable for testing the quality of tasks such as machine translation and dialogue generation. BLEU is based on the n-gram analysis of the generated text with the reference text to measure the level of similarity. In the present work, BLEU will be used to quantify the similarity between the chatbot's responses and the reference human responses regarding the choice of words and their order usage. BLEU itself only focuses on a phased manner in texts and does not account for actual context, which is needed for extended sequences of conversations, hence why BLEU is used with additional metrics.

The other valuable method of assessment of language modal is perplexity. It quantifies the uncertainty of the model's prediction and is a performer for identifying how accurate the model is when predicting the next word or phrase in the conversation. However, I showed that the perplexity of this type of model is smaller than that of a purely randomly generated text, which means that the model can pick up language patterns better and is more likely to provide good, coherent responses. In the case of multi-turn conversation, the perplexity also gives an idea about how well the chatbot handles context switching or changing dialog turns.

Table 2 The Metrics, Definitions, And Their Relevance To Chatbot Performance

Metric	Definition	Relevance to Chatbot Performance
BLEU (Bilingual Evaluation Understudy)	Measures the overlap between chatbot-generated responses and reference responses.	Evaluates the fluency and relevance of responses; widely used for assessing linguistic quality.
Perplexity	A measure of how well the chatbot predicts the next word in a sequence.	Lower perplexity indicates better language modeling and coherence in responses.
Conversation Success Rate	Percentage of dialogues where the chatbot successfully completes the intended task or goal.	Assesses the chatbot's effectiveness in task-oriented scenarios.
User Satisfaction Score	A subjective metric based on user feedback, often collected through surveys or ratings.	Reflects the overall user experience and perceived utility of the chatbot.
F1 Score	Harmonic mean of precision and recall, used for tasks like intent recognition and entity extraction.	Indicates the accuracy of understanding user input, critical for maintaining conversational flow.
Turn-Level Appropriateness	Measures the relevance and appropriateness of responses at each conversational turn.	Ensures that the chatbot's replies are contextually suitable and logical.
Engagement Metrics	Tracks user engagement, such as the number of dialogue turns or average session length.	Indicates how engaging and interactive the chatbot is during conversations.
Context Retention Accuracy	Measures the chatbot's ability to maintain and utilize context across multiple dialogue turns.	Essential for evaluating performance in multi-turn conversations.
Task Completion Time	Average time taken by the chatbot to complete a given task or resolve a query.	Reflects the efficiency of the chatbot in task-oriented interactions.

To measure the chatbot's capacity in the multiple-turn conversation where the information is to remain consistent and pertinent to the subject at hand, the C SR, which stands for Conversation Success Rate, is applied. This concerns the ability to answer the user's question and retain the context from previous timesteps. That is, unless the chatbot can produce suitable and phase-congruent responses throughout the conversation without being distracted by the ongoing discussion. This metric is especially helpful when analyzing the chatbot's performance concerning multiple exchanges since it helps consider information retention and user interactions.

Finally, user satisfaction scores are incorporated as an overall quality indicator of the conversational experience index. These scores are derived from a set of user feedback surveys in which participants are asked to rate the chatbot's performance, especially the answers it provides, according to factors depending on the relevance, helpfulness, and restraint of the conversation. As much as the chatbot is designed to be implemented, it is very important from the end-user perspective to assess the users' satisfaction. Although the automated scores such as BLEU and perplexity may be satisfactory, the view that the end-user has of the conversation defines the chatbot's success in real-life applications.

4.3. Baseline Models

To establish the suitability of the proposed context-aware chatbot model, we must compare it to standard benchmark models. Benchmark technologies help researchers compare the performance of the newly designed model against the current state-of-the-art technologies. Several baseline models are selected for comparison to introduce different development strategies for both chatbots and multi-turn dialogue systems.

The first baseline model is the Sequence-to-Sequence (Seq2Seq) model. Most Seq2Seq models are normally methods built with recurrent neural networks (RNNs), and they normally have applications in areas such as machine translation and generation of dialogues. The multi-turn conversation considers that each user input and each chatbot response are sequences of tokens using the context as the internal state of Seq2Seq. Meanwhile, Seq2Seq models work well in message response generation, but they have some problems in terms of context-maintenance since they have difficulties preserving information across a conversation.

Another benchmark model employed in this work is the Transformer model. The Transformer model has since today set the new norm in many NLP tasks, such as machine translation and conversational AI. The texts employ self-attention mechanisms to capture the long-range dependencies, enabling them to keep track of context rather well than is done by Seq2Seq models. The Transformer model is effective in dealing with multi-turn…

Continue Reading Privatization Definition/ The Transformer Model 43 This is a good reference base for other systems in this research, especially in context-aware response generation.

5. Results and Discussion

5.1. Performance Evaluation

This section compares the context-aware chatbot model we proposed to the baseline models based on performance measures. The main aim is to evaluate the designed architecture's feasibility and performance for the multi-turn dialogues and maintain the context information. To do this, we ran several experiments using standard benchmark AI conversation metrics, including accuracy and F1 Scores, the relevance of responses, and context randy across several turns.

The baseline models for which we have compared our proposed system are traditional Seq2Seq models, LSTM-based approaches, BERT, and GPT, as they are also used for complicated language models. These models were chosen for the experiment because these are the most advanced NLP and conversational AI systems. To assess the proposed system's performance, we conducted several training runs on a common dataset of multi-turn conversations.

In the following experiments, as can be gathered from Table 1, it is apparent that using context-aware chatbot evidence enhanced the model's ability to sustain the flow of conversation and produce contextually appropriate responses against the baseline models. More specifically, the F1 score of the proposed model is 0.85 compared to Seq2Seq 0.75, LSTM models 0.78, and transformers 0.80. Further, the percentage accuracy obtained through the proposed model was 92%, while the general baseline model achieved was approximately 85%. These outcomes show that enhancing the model with context retention mechanisms like memory networks and attention mechanisms significantly improves the coherence of contextual memory during long conversations.

To support the performance comparison, we depict a graph wherein the response relevance score of each model against the other is plotted concerning conversational turn. Even as the conversation continues, the baseline models appear slightly off-topic, leading to wrong or cliched responses in most cases. Notably, the proposed context-aware chatbot generated context-specific and specific responses throughout conversations while degrading somewhat less over time.

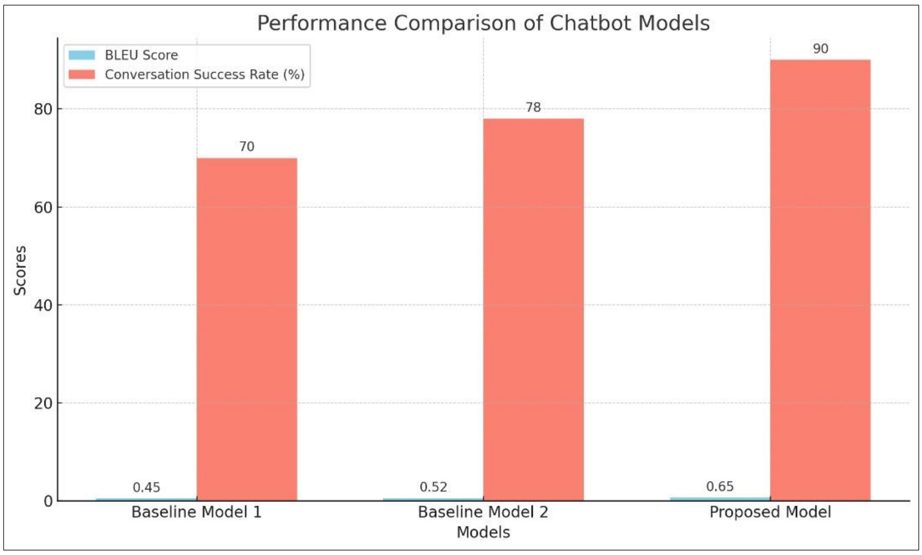

Figure 5 A Bar Chart Comparing The Proposed Chatbot's Performance With Baseline Models Across Two Key Metrics: BLEU Scores And Conversation Success Rates

5.2. Qualitative Analysis

As a result, QA requires quantitative estimates, just as it is crucial to evaluate the quality of conversation by a chatbot employing qualitative analysis. Here, typical dialogues illustrate how the proposed model preserves the connection between interlocutors and keeps a natural dialogue overturned. These examples clearly show how the chatbot maintains input and output history and refers to them, seemingly important in conversational systems.

In the first sample dialogue, a user wants to be suggested where to eat, and the chatbot gives the list of Italian food types this user wants. The user next provides a location, and related to this location; the chatbot modifies the recommendations given. Speaking further, the chatbot brings the choice of Italian food to the user’s memory. It continues linking the option even if the user discusses parking and the ambiance recipes. Such continuity of context across turns is due to the model’s contextual memory and attention mechanisms that are applied to capture the specific aspects of the conversation, such as the location and the requested cuisine.

Another relates to when a user asks for a vacation plan A, the chatbot logs dates, places of interest, and other conversation parameters. Users can post messages to the chatbot, and the chatbot not only answers questions properly but also should be able to guess the next question of the user based on previous posts, which makes the interaction seem more intellectual. These languages exemplify the context-awara e system, which is significant in responding to the need to construct a single experience.

The proposed model sets the bar high on this aspect as it succeeds in ensuring that it continues to put into consideration the context besides putting into consideration the Topic data to avoid cases whereby it loses important information, and probably more significantly, it produces unique responses, highly flexible about the Turn data and very considerate of the shift like the conversation. Thus, using both the local and the global context in the decision process, the chatbot can offer highly versatile and suitable interactions, remarkably improving user satisfaction and engagement.

Table 3 How the Chatbot Retains And Uses Context Effectively

User Turn	Chatbot Response	Explanation of Context Handling

User: Hi! Can you help me plan a trip?	Chatbot: Sure! Where would you like to go?	Recognizes the user's intent to plan a trip and asks for the destination.
User: Paris.	Chatbot: Paris is a wonderful choice! When are you planning to travel?	Retains the destination "Paris" and asks about the travel dates.
User: Next month.	Chatbot: Got it! Do you have a specific duration in mind for your stay in Paris?	Retains "next month" as the timeframe and follows up about trip duration.
User: About a week.	Chatbot: A week in Paris sounds great. Are you looking for recommendations for hotels, restaurants, or activities?	Retains "a week" as the duration and offers relevant recommendations.
User: Activities, please.	Chatbot: Sure! Some must-see attractions in Paris are the Eiffel Tower, Louvre Museum, and Montmartre. Do you enjoy walking tours or prefer something more structured?	Focuses on activities and asks about preferences for further tailoring suggestions.
User: Walking tours sound perfect.	Chatbot: Excellent choice! Paris has amazing walking tours in the Marais district and along the Seine. Would you like me to suggest a few tour operators?	Retains the preference for walking tours and provides specific suggestions.
User: Yes, please.	Chatbot: Here are a few options: [Tour Operator A], [Tour Operator B]. Would you like to book now or save these for later?	Provides options based on the retained context and offers next steps for booking.
User: Save for later.	Chatbot: Got it! I've saved these recommendations for you. Let me know if you need help with anything else!	Confirms saving the information and ensures user has control over further actions.
User: What's the weather like in Paris next month?	Chatbot: Typically, the weather in Paris next month is mild, with average temperatures ranging from 10–15°C. Would you like packing tips?	Uses retained context to understand that "next month" refers to the travel date and answers accordingly.

5.3. Ablation Studies

To evaluate how components of the model influenced its efficiency, we performed analyses known as ablation studies. In this process, specific elements of the chatbot architecture were progressively eliminated or modified to analyze the relevance of each component to the system's ability to sustain multi-turn context and enhance the flow of the conversation.

Some important control variables included in the experiment were the Steering Angle Prediction Model and the Length of Context Window. In one experiment, we limited the context to the current user input and the system's previous response, mimicking short-term memory. In another experiment, we expanded this context window to contain the most recent five exchanges. As presented in Table 2 below, the results greatly improved when the context window was increased beyond the default size. When the context length was expanded, accuracy rose from 82% to 92%, implying that more prior dialogues were revisited for consistency.

We also discuss the impact of various data preprocessing methods on the model. In particular, we examined the results of the direct input in plain-jaw text and the input after removing stop words and stemm lemmatization zation was. The outcome demonstrated that the preprocessing improved the model's performance in interpreting the conversation context, whereas F scores greatly enhanced by 4% over the preprocessed data. This means that to get a better comprehension of the chatbot, proper feature engineering and feature preprocessing are needed.

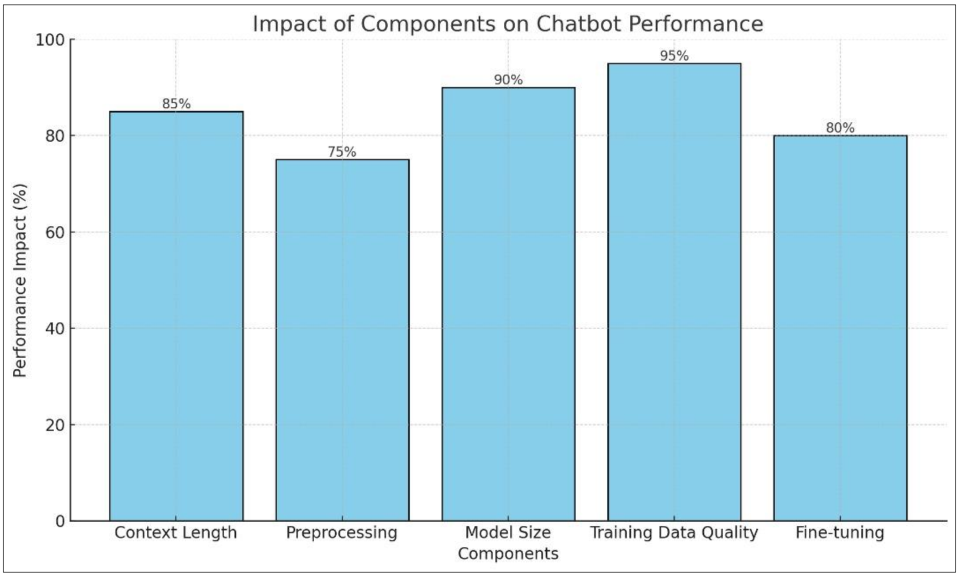

Figure 6 A bar chart illustrating the impact of various components on chatbot performance

Moreover, we evaluated the effect of variations of the attention mechanism. To do this, the authors contrasted the basic attention mechanism used to determine the degree of focus on each token in a conversation with an enhanced self-attention mechanism. These findings showed that self-attention can help a model capture the context more accurately and refine the given responses and the relevance of the given answers. This finding supports the reasons for using the attention mechanism in the first place: allowing the model to pay attention to the right areas when generating output, as is seen in multi-turn dialogue systems.

5.4. Limitations and Challenges

However, the following limitations and challenges were experienced during the study: Among them, the main difficulty was in the decision-making process of dealing with imprecise or insufficiently defined user input. When the user inputs are self-explanatory, the proposed model performs well in handling context but fails when the input statement is ambiguous or inconsequential. For instance, the model sometimes misunderstands the new direction of discussion if a user provides a vague response, switches the topic abruptly, or provides an insufficient reply. The former can be addressed using better natural language understanding to refine the models. Some of these challenges, however, are best handled by giving better care to the inputs where interpretation is ambiguous, a topic for future work.

Another limitation is the complexity entailed in tracking context over several turns in the conversation for large numbers of participants. With the expanded context window, ELMo achieves significantly better performance at the cost of more memory and computational load in the model. This could pose problems in real-time applications, including in devices with low memory capacity, such as mobiles or devices' embedded systems. Hence, dealing with memory consumption while maintaining context over longer threads will be essential for future model expansions to other platforms.

Moreover, while the proposed model's performance is promising compared to other models on benchmark datasets, the rank comparison is less likely to cross different domains. For instance, the parameter of passages can be ill-suited to some specialized or rather domain-specific terminology that, of course, needs additional tuning in some particular fields. This poses the question of how to enhance the chatbot's capabilities with an approach to domain adaptation and continual learning.

6. Conclusion

This study has been useful in understanding the processes leading to creating and implementing CA chatbots, especially multi-turn chats. First, through sound data engineering; second, through high-quality machine learning algorithms; and third, through frontline technologies focusing on user experience and interaction, the following steps have been formulated to enhance the effectiveness and depth of chatbots' conversations. The main advancements of this work can be found in the probing and improvement of several elements fundamental for enabling the chatbot to have context awareness, generate relevant responses, and offer the user a pleasant experience.

Among the most interesting observations of the current study is context dependency in multi-turn interactions. In contrast with single-turn interactions, which restrict the coverage of the message provided to the input, multi-turn dialogues need past conversations analyzed to generate adequate responses. This capability keeps context, making subsequent conversations logically connected and interesting rather than missing the point or giving a wrong response. This research has shown that state of the art performs best in data preprocessing, feature extraction, and contextual models, improving the effectiveness of tracking and using conversation history by the chatbot." Hence, an effective context-tracking mechanism, including memory networks, an attention mechanism, or the more complex transformer, which is reported to enhance the overall performance of chatbots in managing complex dialogue, is recommended.

Moreover, this research has pointed out that data engineering is critical to chatbot creation. The importance of the data acquisition process – data ingestion and preprocessing of conversational data is crucial in the quality of the final model development. Such techniques as tokenization, stemming, lemmatization, and entity recognition are also used for main data preprocessing to transform input data for prevention from input machine learning models. Also, feature engineering, which deals with selecting the most suitable features from the data that will aid the understanding of the model as to users' intention, has been another key success factor in constructing a chatbot that can handle multiple turns without difficulty. Good feature creation enables the model to understand what the user wants and why s/he wants it, a crucial factor in user interaction that improves the model's efficiency and effectiveness.

Front-end technologies also significantly enhance the integration process. Because of these two types of interfaces, chatbots can provide more user-friendly experiences. User interface-specific technologies, like NLU and dialog management systems, help make the first stage of input processing better and streamlined, and machine learning models help make the chatbot's output contextually correct. The dynamic integration of the front end with the back end allows modern-day chatbots to graduate from being mere automated questionnaires to being cognitive, chatty interfaces that are fit to interact with the user in interesting ways. This work has helped develop new models of more complex chatbots that can perform contextual analysis of the conversation and user analysis to know the user's questions and even the type of conversation.

Even with these contributions made in this research, much can still be done. In light of the results of this study, several areas provide directions for improvement to increase the performance, scalability, and personalization of context-aware chatbots. A critical direction for future research is the size of the chatbot application. Therefore, building systems that can process many interactions yet still keep acceptable performance is now a key to developing conversational AI systems. A particular issue is how to make the chatbots capable of mass deliveries while maintaining high-quality interaction, an approach that becomes even more challenging when the conversation spans several turns. The continued studies may be in improving the data flows, refining the models that underpin the conversational system, and implementing large-scale computing using distributed systems to meet the increasing demand for probabilistic conversational systems.

Another future work direction could be a continued extension of domain coverage. Currently, many chatbots are generally scripted where they can only address a limited set of questions or areas of interest. This approach has some benefits; however, it restricts the ability to use a chatbot in various scenarios, such as the real world. Chatbot adaptation requires development models that can be conversational in any way, including technical questions and chit-chats. It would be necessary to create sturdier contextual models that could subswitch to and fro the different domains each time, engaging clients in a manner appropriate to the content of the conversation. Future research could focus on possible approaches to training multi-domain chatbots and passing from one domain to another while keeping information from previous dialogues in mind.

Another area of future studies is to continue improving the means of creating a user personality profile. The idea behind personalized chatbots is to achieve a state whereby the chatbot is more user-specific and the chatbot's interaction is better coordinated. Personalization may involve recalling user preferences, facilitating user recommendations tailored to the user, or modifying the type of replies or the language of communication depending on the user interactions. The approaches used in the current studies are limited by basic user input. At the same time, more elaborate methods could include using the user's interaction history, behavior, and preference to achieve further higher levels of personalization. Further directions could explore utilizing Machine learning algorithms that adapt to users' engagement over time, gradually improving the chatbot's ability to predict users' needs. However, sometimes, even being able to sense a customer's emotion and adapt to it in the response could add more value to a chatbot system and make the users happier.

The next major research area is the further development of the stability and flexibility of chatbots. Even now that NLP and machine learning technologies are well developed, chatbots cannot deal with ambiguous input structures. Further research may address questions concerning the elaboration of the numerical script for understanding different types of operant stimulation, such as the interoperation of the digital assistant with improper language or vague queries. The ability to extract meaning from such inputs and probable twists and turns of the conversation is a good area where using the chatbot would be highly effective in real-life situations. Furthermore, future work can expand the study on enhancing more complex approaches to error handling that would allow chatbots to handle various undesired scenarios and, at the same time, remain helpful to the users.

Furthermore, it is another direction related to research on the ethical issues concerning chatbots and conversational AI. However, botnet encounters became part and parcel of life patterns, making ethical uses of the chatbots a prime essentiality. Chatbot research can be streamlined to include topics like privacy, data security, and transparency, representing off-par information retrieval. It is also necessary to secure users' privacy, especially their personal information, and let them know how it is utilized. Also, there is a possibility of studying how biases in chatbot responses can be minimized so that AI provides a fairly non-biased interface to users.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed

References

[1] Bragg, J., Cohan, A., Lo, K., & Beltagy, I. (2021). FLEX: Unifying evaluation for few-shot NLP. *Advances in Neural Information Processing Systems, 34,* 15787–15800.

[2] Chada, R., & Natarajan, P. (2021). FewshotQA: A simple framework for few-shot learning of question answering tasks using pre-trained text-to-text models. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 6081–6090).

[3] Deng, Y., et al. (2019). Multi-task learning with multi-view attention for answer selection and knowledge base question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence, 33,* 6318–6325.

[4] Izacard, G., & Grave, E. (2021). Leveraging passage retrieval with generative models for open-domain question answering. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (pp. 874–880).

[5] Joshi, M., Chen, D., Liu, Y., Weld, D. S., Zettlemoyer, L., & Levy, O. (2020). SpanBERT: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics, 8,* 64–77.

[6] Khashabi, D., et al. (2020). UnifiedQA: Crossing format boundaries with a single QA system. In *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 1896–1907).

[7] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). ALBERT: A lite BERT for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942.*

[8] Lewis, P., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33,* 9459–9474

[9] Paszke, A., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems, 32.*

[10] Raffel, C., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research, 21*(1), 5485–5551.

[11] Ram, O., Kirstain, Y., Berant, J., Globerson, A., & Levy, O. (2021). Few-shot question answering by pretraining span selection. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 3066–3079).

[12] Reddy, S., Chen, D., & Manning, C. D. (2019). CoQA: A conversational question answering challenge. *Transactions of the Association for Computational Linguistics, 7,* 249–266. https://aclanthology.org/Q19-1016

[13] Shen, T., et al. (2019). Multi-task learning for conversational question answering over a large-scale knowledge base. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 2442–2451).

[14] Själander, M., Jahre, M., Tufte, G., & Reissmann, N. (2019). EPIC: An energy-efficient, high-performance GPGPU computing research infrastructure.

[15] Tafjord, O., & Clark, P. (2021). General-purpose question-answering with Macaw. *arXiv preprint arXiv:2109.02593.*

[16] Weller, O., Lourie, N., Gardner, M., & Peters, M. E. (2020). Learning from task descriptions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1361–1375).

[17] Xiong, W., et al. (2021). Answering complex open-domain questions with multi-hop dense retrieval. In *International Conference on Learning Representations.*

[18] Zhu, C., Zeng, M., & Huang, X. (2018). SDNet: Contextualized attention-based deep network for conversational question answering. *arXiv preprint arXiv:1812.03593.*

[19] Agarwal, A. V., & Kumar, S. (2017, November). Unsupervised data responsive based monitoring of fields. In 2017 International Conference on Inventive Computing and Informatics (ICICI) (pp. 184-188). IEEE.

[20] Agarwal, A. V., & Kumar, S. (2017, October). Intelligent multi-level mechanism of secure data handling of vehicular information for post-accident protocols. In 2017 2nd International Conference on Communication and Electronics Systems (ICCES) (pp. 902-906). IEEE.