International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

Integrate machine learning and deep learning techniques, including transformer-based NLP architectures such as BERT and GPT, in order to improve dialogue management, contextual understanding, and response generation capabilities of the chatbot system.

Bhavnistha

M.Tech Scholar, Department of Computer Science & Engineering, Sat Priya Group of Institutions

Dr. Rupali Ahuja

Associate Professor, Department of Computer Science & Engineering, Sat Priya Group of Institutions

Abstract:

The development of intelligent chatbot systems using Artificial Intelligence (AI), Machine Learning (ML), and Natural Language Processing (NLP) has become an important research area for improving human-computer interaction. Traditional chatbot systems often face limitations such as weak contextual understanding, poor dialogue continuity, and inaccurate response generation. To address these challenges, this research proposes a context-aware chatbot system integrating machine learning, deep learning, and transformer-based NLP architectures such as BERT and GPT to enhance dialogue management, contextual understanding, and conversational response generation. The proposed framework combines contextual memory management, sentiment analysis, and Retrieval-Augmented Generation (RAG) to improve semantic interpretation and conversational adaptability. BERT is utilized for bidirectional contextual understanding and intent recognition, while GPT is employed for dynamic and coherent response generation. Experimental evaluation is conducted using multiple performance metrics including accuracy, precision, recall, F1-score, perplexity, semantic similarity, contextual relevance score, dialogue success rate, BLEU score, and ROUGE score. The proposed Hybrid BERT-GPT model achieved superior performance with 96.48% accuracy, 95.72% precision, 95.30% recall, and 95.51% F1-score, along with improved contextual relevance and dialogue success rates. The results demonstrate that integrating transformer-based NLP models significantly improves conversational intelligence, contextual reasoning, emotional adaptability, and user satisfaction in chatbot systems.

Keywords: Context-Aware Chatbot, Natural Language Processing, Machine Learning, Deep Learning, Transformer Architecture, BERT, GPT, Dialogue Management, Contextual Understanding, Conversational AI, Response Generation, Sentiment Analysis, Retrieval-Augmented Generation (RAG), Semantic Similarity, Human-Computer Interaction, Intelligent Chatbot Systems.

1.Introduction:

The rapid advancement of Artificial Intelligence (AI) and digital communication technologies has significantly transformed the way humans interact with computer systems. Among the most important developments in this field is the emergence of intelligent chatbot systems capable of simulating human conversation and providing automated communication services across various domains such as healthcare, education, banking, customer service, e-commerce, and smart digital platforms. Chatbots are increasingly becoming an essential component of modern communication systems due to their ability to provide real-time assistance, improve operational efficiency, reduce human workload, and enhance user engagement. However, traditional chatbot systems based on rule-based algorithms and keyword-matching techniques often suffer from major limitations such as weak contextual understanding, poor dialogue continuity, repetitive responses, and limited semantic interpretation [1].

Conventional chatbot systems generally process user queries independently without maintaining conversational context or understanding the deeper semantic meaning of user interactions. As a result, these systems fail to generate coherent and personalized responses during complex multi-turn conversations. Human communication naturally depends on contextual information such as dialogue history, emotional state, user preferences, and situational understanding. Therefore, developing context-aware conversational systems capable of understanding and adapting to dynamic communication scenarios has become a critical research area in conversational artificial intelligence [2].

The development of advanced Natural Language Processing (NLP) techniques and deep learning architectures has created new opportunities for improving conversational intelligence in chatbot systems. Early NLP approaches such as Bag-of-Words and statistical machine learning methods provided basic text classification and intent recognition capabilities but lacked contextual reasoning and semantic learning efficiency. The introduction of deep learning models such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and attention-based neural systems improved sequential language processing and dialogue continuity. However, these models still faced challenges related to long-term contextual dependency management and computational limitations [3].

A major breakthrough in conversational AI occurred with the introduction of transformer-based NLP architectures such as BERT and GPT. Transformer models utilize self-attention mechanisms and contextual embeddings to process language sequences more efficiently and accurately than traditional sequential neural networks. BERT improves contextual understanding through bidirectional language representation, enabling the system to interpret semantic relationships between words based on both preceding and succeeding context. GPT, on the other hand, enhances conversational response generation through autoregressive language modeling, allowing the chatbot to produce coherent, fluent, and human-like responses dynamically [4].

The integration of machine learning and deep learning techniques with transformer-based NLP architectures provides a powerful framework for developing intelligent context-aware chatbot

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

systems. By combining contextual understanding from BERT with generative conversational capabilities from GPT, chatbot systems can significantly improve dialogue management, semantic interpretation, emotional intelligence, and conversational continuity. These advanced conversational systems are capable of maintaining dialogue history, understanding user intent, adapting to emotional context, and generating contextually relevant responses in real-time interactions.

In addition to contextual understanding and response generation, modern chatbot systems also require mechanisms for sentiment analysis, contextual memory management, knowledge retrieval, and adaptive dialogue optimization. Sentiment analysis enables the chatbot to recognize user emotions and generate empathetic responses, while contextual memory modules help preserve conversational continuity across multiple dialogue turns. Retrieval-Augmented Generation (RAG) techniques further improve factual accuracy by integrating external knowledge sources into the response generation process, reducing hallucination and improving reliability in conversational AI systems. The integration of advanced NLP architectures into chatbot systems has demonstrated substantial improvements in semantic similarity, contextual relevance, response fluency, dialogue success rate, and user satisfaction. These improvements make intelligent conversational systems more suitable for real-world applications where accurate, adaptive, and personalized communication is essential. Industries such as healthcare, education, finance, customer support, and smart governance increasingly rely on context-aware chatbots to provide automated services, assist decision-making, and improve communication efficiency [5].

Despite significant progress in conversational AI research, several challenges remain unresolved, including maintaining long-term contextual memory, improving multilingual understanding, ensuring factual consistency, reducing response latency, and addressing ethical concerns such as bias, privacy, and explainability. Therefore, integrating machine learning, deep learning, and transformer-based NLP models remains an active area of research aimed at developing more intelligent, scalable, secure, and human-like chatbot systems.

This research focuses on integrating machine learning and deep learning techniques, including transformer-based NLP architectures such as BERT and GPT, to improve dialogue management, contextual understanding, and conversational response generation capabilities in chatbot systems. The proposed framework combines contextual memory management, sentiment analysis, semantic understanding, Retrieval-Augmented Generation, and adaptive dialogue optimization to create an intelligent and context-aware conversational AI system. The research aims to enhance conversational quality, improve user satisfaction, and develop a scalable chatbot framework suitable for real-time intelligent communication applications [6].

2.Literature Review

The development of intelligent chatbot systems has become one of the most significant research areas in Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), and Natural

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

Language Processing (NLP). Conversational AI systems are increasingly used in customer support, healthcare, education, banking, e-commerce, and virtual assistance applications due to their ability to automate communication and improve user interaction. Traditional chatbot systems were primarily rule-based and relied on predefined scripts, keyword matching, and decision-tree mechanisms for generating responses. Although these systems provided basic conversational functionality, they lacked contextual understanding, semantic interpretation, and adaptive communication capabilities. Consequently, researchers began exploring machine learning and deep learning techniques to improve conversational intelligence and human-like interaction in chatbot systems [7].

Early research in conversational AI focused on symbolic and rule-based approaches such as ELIZA and ALICE, which simulated human conversation through pattern matching and scripted responses. These systems demonstrated the potential of automated communication but failed to maintain conversational continuity or understand contextual dependencies. Researchers later introduced statistical NLP methods and machine learning algorithms such as Naïve Bayes, Support Vector Machines (SVM), and Decision Trees for text classification and intent recognition. These techniques improved conversational accuracy compared to rule-based systems; however, they still struggled with semantic ambiguity, contextual reasoning, and dynamic dialogue management [8].

The emergence of deep learning significantly transformed the development of intelligent conversational systems. Researchers explored Artificial Neural Networks (ANNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks to improve sequential language processing and conversational memory. LSTM architectures became particularly important because they enabled chatbots to preserve contextual information across multiple dialogue turns. Studies showed that LSTM-based conversational systems improved dialogue continuity, semantic interpretation, and response relevance compared to traditional machine learning approaches. However, recurrent neural architectures suffered from limitations such as vanishing gradient problems, sequential computation complexity, and difficulties in handling long-term contextual dependencies [9].

To overcome these limitations, researchers introduced attention mechanisms and transformer-based architectures for advanced language understanding. Transformer models revolutionized NLP research by enabling parallel processing and contextual learning through self-attention mechanisms. The introduction of BERT by Devlin et al. marked a major breakthrough in contextual language understanding. BERT utilizes bidirectional contextual embeddings to analyze semantic relationships between words based on both left and right context simultaneously. Researchers found that BERT significantly improved intent recognition, contextual understanding, semantic similarity analysis, and question-answering tasks in conversational systems.

Several studies demonstrated the effectiveness of BERT in chatbot development and dialogue management. Researchers integrated BERT into conversational systems for contextual intent

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

classification, dialogue state tracking, sentiment analysis, and semantic matching tasks. Experimental results revealed that BERT-based conversational models achieved higher accuracy and contextual relevance compared to traditional machine learning and recurrent neural network architectures. However, although BERT improved contextual understanding, it was primarily designed for language representation and understanding rather than dynamic conversational response generation.

The development of GPT further advanced conversational AI by introducing generative language modeling capabilities. GPT-based architectures utilize autoregressive transformer models to generate coherent and human-like responses dynamically. Unlike retrieval-based systems that select predefined responses, GPT models generate responses word-by-word based on conversational context and dialogue history. Researchers observed that GPT significantly improved conversational fluency, response diversity, and natural language generation quality in chatbot systems. GPT-based conversational agents demonstrated strong performance in open-domain dialogue generation and adaptive communication tasks [10].

Recent research has increasingly focused on integrating BERT and GPT architectures to combine contextual understanding with generative conversational capabilities. Hybrid transformer-based architectures have been proposed to improve dialogue management, semantic reasoning, and response coherence. Studies indicated that combining BERT-based contextual embeddings with GPT-based language generation enables conversational systems to better understand user intent while generating more accurate and contextually appropriate responses. These hybrid models demonstrated superior performance in conversational AI tasks compared to standalone transformer architectures. Dialogue management has also become an important research area in chatbot development. Traditional dialogue systems often relied on finite-state models and handcrafted rules, limiting their adaptability in dynamic conversational scenarios. Researchers later introduced reinforcement learning and deep reinforcement learning techniques for optimizing dialogue management strategies. Reinforcement learning-based conversational systems learn optimal response behaviors through interaction feedback and reward-based optimization. Experimental studies showed that reinforcement learning improves dialogue success rate, conversational adaptability, and user engagement in chatbot systems.

Context awareness and conversational memory have received significant attention in recent chatbot research. Human communication depends heavily on contextual information such as dialogue history, user preferences, emotional state, and situational understanding. Researchers developed contextual memory modules, memory-augmented neural networks, and transformer attention mechanisms to preserve conversational continuity across multi-turn interactions. Context-aware chatbot systems demonstrated improved semantic relevance, dialogue coherence, and personalized interaction compared to non-contextual conversational models.

Sentiment analysis and emotional intelligence have also been integrated into conversational AI systems to improve empathetic communication. Researchers employed machine learning and deep learning techniques to classify user emotions and sentiment polarity from textual conversations. Sentiment-aware chatbots adapt conversational tone and response strategies according to emotional context, improving user satisfaction and communication quality. Studies in healthcare, mental health support, and customer service applications showed that emotionally intelligent conversational agents significantly enhance user engagement and trust.

Another major area of research involves Retrieval-Augmented Generation (RAG) techniques for improving factual consistency and knowledge integration in conversational systems. Traditional generative models sometimes produce inaccurate or hallucinated responses due to limitations in pre-trained knowledge representation. Researchers introduced retrieval-based mechanisms that integrate external knowledge sources such as databases, documents, and semantic search systems with transformer-based response generation models. Studies conducted by Lewis et al. and subsequent researchers demonstrated that RAG architectures significantly improve factual accuracy, contextual relevance, and knowledge-intensive conversational performance.

Human-centered design and usability research also play a critical role in conversational AI development. Researchers emphasized that intelligent chatbot systems should not only provide accurate responses but also ensure usability, personalization, transparency, and trustworthiness. Studies on user-centered design principles highlighted the importance of conversational naturalness, emotional engagement, and adaptive interaction in improving user satisfaction. Human-AI collaborative frameworks were proposed to improve communication efficiency and conversational usability across diverse application domains.

Privacy, security, and ethical AI considerations have become increasingly important in chatbot research due to the sensitive nature of conversational data. Researchers identified concerns related to data privacy, bias, fairness, explainability, and transparency in AI-driven conversational systems. Studies proposed privacy-preserving machine learning techniques, explainable AI mechanisms, and fairness-aware conversational architectures to improve reliability and ethical compliance in chatbot systems. Federated learning and secure data processing frameworks were also explored to protect sensitive conversational information while supporting adaptive conversational learning.

Multilingual conversational AI represents another emerging research direction. Researchers developed multilingual transformer architectures such as mBERT and XLM-R to support cross-lingual conversational interaction and improve accessibility in global communication environments. Experimental studies demonstrated that multilingual transformer models improve semantic understanding and contextual reasoning across multiple languages. However, challenges related to low-resource languages, code-mixed conversations, and cultural adaptation remain active research areas.

Despite substantial progress in conversational AI, several challenges continue to exist in chatbot development. Existing conversational systems still struggle with maintaining long-term contextual memory, reducing hallucination, improving real-time response efficiency, handling ambiguous language, and achieving human-level conversational reasoning. Additionally, computational complexity and resource requirements for training transformer-based models remain significant concerns for real-world deployment. Researchers continue to investigate lightweight transformer architectures, efficient fine-tuning methods, and scalable conversational frameworks to address these limitations. The literature review demonstrates that integrating machine learning, deep learning, and transformer-based NLP architectures such as BERT and GPT significantly improves dialogue management, contextual understanding, semantic reasoning, and response generation capabilities in chatbot systems. The combination of contextual embeddings, generative language modeling, sentiment analysis, reinforcement learning, and Retrieval-Augmented Generation techniques provides a strong foundation for developing intelligent and context-aware conversational AI systems. Existing research highlights the potential of transformer-based architectures in improving conversational fluency, contextual relevance, emotional intelligence, and user satisfaction. However, challenges related to contextual memory, factual consistency, multilingual adaptability, explainability, and computational efficiency require further investigation. Therefore, the proposed research aims to develop an advanced hybrid conversational framework integrating BERT, GPT, contextual memory, sentiment analysis, and retrieval-based knowledge systems to create intelligent, adaptive, and human-like chatbot systems suitable for real-world digital communication applications.

3. Methodology:

The methodology for integrating machine learning and deep learning techniques, including transformer-based Natural Language Processing (NLP) architectures such as BERT and GPT, focuses on developing an intelligent context-aware chatbot system capable of improving dialogue management, contextual understanding, and response generation. The proposed methodology combines data-driven learning approaches, contextual computing mechanisms, semantic analysis, and advanced neural language models to create a conversational system that can interact with users in a more natural, adaptive, and human-like manner. The methodological framework is designed to support conversational continuity, personalized interaction, contextual reasoning, and efficient response generation across multiple communication scenarios.

The research methodology begins with the collection and preparation of conversational datasets from multiple sources such as customer support interactions, online conversational corpora, healthcare dialogues, educational communication systems, social media interactions, and publicly available chatbot datasets. The collected data contains diverse conversational patterns including greetings, informational queries, emotional expressions, follow-up conversations, and context-dependent interactions. Since chatbot performance depends heavily on the quality and diversity of training data, the dataset is designed to include multilingual, domain-specific, and context-rich

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

conversational samples. Additional contextual information such as user intent, sentiment labels, dialogue history, timestamps, and interaction metadata is also incorporated to improve contextual learning capabilities.

After data collection, preprocessing techniques are applied to clean and normalize the textual data before training the machine learning and deep learning models. The preprocessing stage involves tokenization, lowercasing, stop-word removal, punctuation handling, stemming, lemmatization, and noise filtering to remove irrelevant information from the dataset. Since conversational datasets often contain slang expressions, abbreviations, incomplete sentences, and grammatical inconsistencies, advanced preprocessing methods are implemented to improve data quality and semantic consistency. Text normalization ensures that the conversational input is transformed into a structured format suitable for computational processing and neural network training.

Feature extraction and language representation constitute the next stage of the methodology. Traditional NLP systems used Bag-of-Words and TF-IDF approaches for textual representation; however, these methods fail to capture semantic and contextual relationships effectively. Therefore, the proposed methodology employs distributed word representation and contextual embedding techniques using transformer-based architectures. Pre-trained embedding models such as Word2Vec, GloVe, and FastText are initially explored to convert textual data into numerical vector representations. Subsequently, advanced contextual embeddings generated by BERT and GPT models are integrated to capture semantic meaning, contextual dependencies, and conversational intent more accurately.

The Bidirectional Encoder Representations from Transformers (BERT) model is incorporated into the chatbot architecture to improve contextual understanding and intent recognition. BERT utilizes bidirectional contextual learning, enabling the model to analyze both preceding and succeeding words within a sentence simultaneously. This bidirectional attention mechanism allows the chatbot to understand ambiguous expressions, semantic relationships, and contextual references more effectively. Fine-tuning techniques are applied to adapt the pre-trained BERT model to the specific conversational domain and dialogue management tasks. The chatbot uses BERT-based embeddings for intent classification, semantic similarity analysis, sentiment detection, and contextual memory representation.

The Generative Pre-trained Transformer (GPT) architecture is integrated into the system to improve conversational response generation capabilities. GPT is based on autoregressive language modeling and generates responses sequentially by predicting the next word in a conversational context. Unlike retrieval-based chatbot systems that select responses from predefined databases, GPT-based generative models create dynamic and contextually relevant responses in real time. Fine-tuning the GPT model on conversational datasets enables the chatbot to maintain conversational continuity, adapt to user communication styles, and generate coherent multi-turn

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

dialogue responses. The integration of GPT significantly enhances conversational fluency, natural language generation, and adaptive interaction quality.

To improve dialogue management, the proposed methodology incorporates contextual memory and dialogue state tracking mechanisms. Context-aware conversational systems require the ability to retain previous interactions and maintain conversational continuity across multiple dialogue turns. Therefore, a contextual memory module is implemented to store dialogue history, user preferences, emotional states, and interaction patterns during communication sessions. The chatbot continuously updates the conversational state using sequential contextual information, enabling more accurate response generation and adaptive interaction management. Contextual representations are dynamically maintained throughout the dialogue process to ensure coherent and context-sensitive communication.

Machine learning and deep learning techniques are integrated for intent classification and dialogue prediction tasks. Supervised learning algorithms such as Support Vector Machines (SVM), Logistic Regression, and Random Forest classifiers may be utilized for preliminary intent detection experiments. However, deep learning models including Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), and transformer architectures are prioritized due to their superior performance in sequential language processing and contextual understanding. Hybrid architectures combining recurrent neural networks with transformer-based embeddings are explored to improve conversational memory and semantic interpretation.

Sentiment analysis and emotional intelligence are also integrated into the methodology to improve user engagement and empathetic communication. Deep learning-based sentiment classification models analyze the emotional polarity of user messages and categorize them into emotional states such as positive, negative, neutral, frustration, or satisfaction. The chatbot adapts its conversational style according to the detected emotional state, enabling more human-like and emotionally aware interaction. This capability is especially important in applications such as healthcare support, educational assistance, counseling systems, and customer service environments.

The methodology further incorporates Retrieval-Augmented Generation (RAG) techniques to improve factual accuracy and reduce hallucination problems associated with generative AI models. Traditional generative chatbot systems may produce inaccurate or fabricated responses due to limitations in pre-trained knowledge. To overcome this issue, the proposed chatbot architecture integrates external knowledge retrieval systems, semantic search mechanisms, and domain-specific databases into the response generation process. Relevant contextual documents and knowledge sources are retrieved dynamically during conversation and combined with transformer-based generation models to produce more reliable and informative responses.

Dialogue optimization and adaptive conversational learning are achieved through reinforcement learning techniques. Reinforcement learning enables the chatbot to improve interaction quality by learning from user feedback, conversational success rates, and reward-based optimization strategies. The dialogue management system continuously adjusts response selection policies and conversational strategies based on interaction outcomes. This adaptive learning process enhances long-term conversational efficiency, task completion accuracy, and user satisfaction.

The implementation phase of the methodology utilizes programming frameworks and AI development tools such as Python, TensorFlow, PyTorch, Hugging Face Transformers, NLTK, and SpaCy. Cloud computing platforms and GPU-based processing environments are employed to support large-scale transformer model training and real-time conversational processing. The chatbot interface may be integrated into web applications, mobile systems, or cloud-based communication platforms to enable scalable deployment and real-time interaction.

The performance evaluation methodology involves both quantitative and qualitative assessment of the chatbot system. Quantitative evaluation metrics include accuracy, precision, recall, F1-score, perplexity, response time, semantic similarity, contextual relevance score, and dialogue success rate.

Accuracy

Accuracy measures the overall correctness of the chatbot classification system.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- TP= True Positive
- TN= True Negative
- FP= False Positive
- FN= False Negative

Precision

Precision measures how many predicted positive responses are actually correct.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Where:

- TP= True Positive
- FP= False Positive

Recall

Recall measures how many actual positive responses are correctly identified.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Where:

- TP= True Positive
- FN= False Negative

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

F1-Score

F1-Score is the harmonic mean of Precision and Recall.

$$\text{F1-Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

Or directly:

$$\text{F1-Score} = \frac{2TP}{2TP + FP + FN}$$

Perplexity

Perplexity evaluates the quality of language generation models such as GPT.

$$\text{Perplexity} = 2^{-\frac{1}{N}\sum_{i=1}^{N} \log_2 P(w_i)}$$

Where:

- N= Total number of words
- $P(w_i)$= Probability of the i^{th}word predicted by the model

Lower perplexity indicates better language prediction performance.

Response Time

Response time measures the time taken by the chatbot to generate a response.

$$\text{Response Time} = T_{response} - T_{request}$$

Where:

- $T_{response}$= Time at which chatbot generates response
- $T_{request}$= Time at which user submits request

Semantic Similarity

Semantic similarity measures the similarity between user query and chatbot response embeddings.

$$\text{Semantic Similarity} = \frac{\mathbf{A} \cdot \mathbf{B}}{\parallel \mathbf{A} \parallel \times \parallel \mathbf{B} \parallel}$$

Where:

- **A**= Vector representation of user query
- **B**= Vector representation of chatbot response

This equation is based on cosine similarity.

Contextual Relevance Score

Contextual relevance evaluates how well the chatbot response matches the conversational context.

$$\text{CRS} = \frac{\sum_{i=1}^{n} R_i}{n}$$

Where:

- R_i= Relevance score for each contextual interaction
- n= Total number of dialogue turns

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

Higher CRS values indicate better contextual understanding.

Dialogue Success Rate

Dialogue Success Rate measures the percentage of successful conversations completed by the chatbot.

$$\text{Dialogue Success Rate} = \frac{N_{successful}}{N_{total}} \times 100$$

Where:

- $N_{successful}$= Number of successfully completed dialogues
- N_{total}= Total number of dialogues

Higher dialogue success rate indicates better conversational performance and task completion capability. Qualitative evaluation focuses on conversational naturalness, user satisfaction, emotional adaptability, contextual continuity, and interaction usability. Comparative analysis is conducted between the proposed transformer-based context-aware chatbot and conventional rule-based or retrieval-based systems to demonstrate performance improvements in dialogue management and contextual understanding.

User interaction studies are conducted to evaluate practical usability and human acceptance of the chatbot system. Human participants interact with the chatbot in various conversational scenarios, and their feedback is collected through surveys, questionnaires, and interaction analysis. The evaluation process measures user trust, conversational quality, emotional engagement, and satisfaction with the chatbot's contextual capabilities. Feedback collected from users contributes to system refinement and optimization.

The methodology also incorporates ethical AI considerations, privacy protection mechanisms, and fairness evaluation procedures. Since conversational systems process sensitive user information, secure data management techniques such as encryption, anonymization, and privacy-preserving machine learning mechanisms are implemented to ensure user data protection. Bias mitigation strategies and explainable AI techniques are also integrated into the chatbot framework to improve transparency, fairness, and trustworthiness.

The proposed methodology integrates machine learning, deep learning, transformer-based NLP architectures, contextual computing, sentiment analysis, reinforcement learning, and knowledge retrieval mechanisms to develop an advanced context-aware chatbot system. The integration of BERT and GPT models significantly enhances semantic understanding, dialogue management, and conversational response generation, enabling the chatbot to provide adaptive, personalized, and human-like communication experiences across diverse real-world applications.

4. Dataset and Experimental:

The development of a context-aware chatbot system integrating machine learning, deep learning, and transformer-based Natural Language Processing (NLP) architectures such as BERT and GPT requires a comprehensive dataset and a carefully designed experimental setup to ensure effective

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

dialogue management, contextual understanding, and intelligent response generation. The experimental framework is designed to evaluate the performance of the proposed chatbot system under different conversational scenarios while measuring contextual accuracy, semantic understanding, conversational continuity, and response quality. The dataset preparation and experimental procedures are organized to support the training, validation, testing, and performance analysis of advanced conversational AI models.

The dataset used for the proposed chatbot system consists of conversational text collected from multiple publicly available dialogue corpora, customer support interactions, educational communication systems, healthcare conversations, social media dialogues, and open-domain chatbot datasets. To improve the diversity and robustness of the conversational model, the dataset includes both task-oriented and open-domain conversations. The collected data contains multiple conversational patterns such as greetings, informational requests, emotional interactions, follow-up questions, clarification dialogues, and context-dependent conversations. The dataset is structured to simulate realistic human communication behavior and dynamic conversational flow. The conversational dataset includes user queries, chatbot responses, intent labels, sentiment annotations, contextual history, and dialogue state information. Each conversation instance contains sequential dialogue turns to enable contextual learning and memory retention during model training. Additional metadata such as timestamps, user identifiers, topic categories, and emotional tags are also incorporated to improve contextual awareness and adaptive conversational capabilities. The inclusion of emotional and contextual annotations supports sentiment analysis, dialogue state tracking, and personalized response generation within the chatbot framework.

The dataset is divided into three primary subsets: training data, validation data, and testing data. Approximately 70% of the dataset is allocated for training purposes, 15% for validation, and 15% for testing. The training dataset is used to optimize model parameters and enable the chatbot to learn semantic patterns, conversational structures, and contextual dependencies. The validation dataset supports hyperparameter tuning and model optimization while preventing overfitting during the training process. The testing dataset consists of unseen conversational samples used to evaluate the generalization capability and real-time conversational performance of the chatbot system.

Before training the machine learning and deep learning models, extensive preprocessing techniques are applied to improve data quality and semantic consistency. The preprocessing stage involves tokenization, lowercasing, punctuation removal, stop-word filtering, stemming, lemmatization, and noise reduction. Since conversational datasets frequently contain slang expressions, abbreviations, typographical errors, and incomplete sentences, normalization techniques are applied to standardize textual information. Text encoding methods are then implemented to transform textual data into numerical vector representations suitable for machine learning and neural network processing.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

Traditional textual representation methods such as Bag-of-Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF) are initially explored for baseline machine learning experiments. However, to improve semantic understanding and contextual representation, advanced word embedding and contextual embedding techniques are integrated into the experimental framework. Word embedding models such as Word2Vec, GloVe, and FastText are utilized to capture semantic relationships between words. Subsequently, transformer-based contextual embeddings generated by BERT and GPT architectures are employed to improve contextual reasoning and conversational understanding.

The Bidirectional Encoder Representations from Transformers (BERT) model is used for contextual intent recognition, sentiment analysis, and semantic understanding tasks. Pre-trained BERT models are fine-tuned on the conversational dataset to adapt the language representations to the specific chatbot application domain. BERT processes conversational inputs bidirectionally, allowing the model to understand contextual dependencies between preceding and succeeding words within dialogue sequences. The BERT-based contextual embeddings are utilized for intent classification, dialogue state tracking, and semantic similarity measurement during conversational processing.

The Generative Pre-trained Transformer (GPT) architecture is integrated into the experimental framework to support dynamic response generation and conversational continuity. GPT is trained using autoregressive language modeling techniques to generate contextually appropriate responses based on dialogue history and user input. The model predicts subsequent words sequentially while considering previous conversational context. Fine-tuning GPT on domain-specific conversational data enables the chatbot to generate coherent, human-like, and context-sensitive responses across multiple interaction scenarios.

To improve dialogue management and conversational memory, a contextual memory module is incorporated into the experimental architecture. The chatbot maintains dialogue history, user preferences, emotional states, and contextual information throughout communication sessions. Sequential dialogue representations are continuously updated to preserve contextual continuity and improve adaptive response generation. Long Short-Term Memory (LSTM) networks and transformer attention mechanisms are integrated into the contextual memory framework to enhance conversational dependency learning and dialogue flow management.

Sentiment analysis experiments are also conducted to improve emotional intelligence and empathetic communication within the chatbot system. The conversational dataset is annotated with sentiment labels such as positive, negative, neutral, frustration, and satisfaction. Deep learning-based sentiment classification models are trained to detect emotional polarity from user interactions. The chatbot utilizes sentiment predictions to adapt conversational tone and response style according to user emotional states. This functionality is particularly useful in applications involving customer support, healthcare communication, educational tutoring, and mental health assistance.

The experimental setup additionally integrates Retrieval-Augmented Generation (RAG) techniques to enhance factual consistency and reduce hallucination problems in generative conversational models. External knowledge sources such as FAQs, domain-specific documents, semantic databases, and structured knowledge repositories are integrated into the chatbot system. During conversational processing, relevant documents are retrieved dynamically using semantic similarity search and combined with GPT-based response generation mechanisms. This hybrid retrieval-generation architecture improves response accuracy and supports knowledge-intensive conversational tasks.

The implementation environment for the experimental framework utilizes Python programming language along with advanced AI development libraries and frameworks such as TensorFlow, PyTorch, Hugging Face Transformers, NLTK, and SpaCy. GPU-enabled cloud computing infrastructure is employed to support transformer model training and large-scale conversational processing. The chatbot system is deployed within a web-based conversational interface to enable real-time interaction and user evaluation.

The experimental evaluation process focuses on measuring the performance of the chatbot system using both quantitative and qualitative metrics. Quantitative evaluation metrics include accuracy, precision, recall, F1-score, perplexity, response time, semantic similarity score, contextual relevance score, and dialogue success rate. Intent classification performance is evaluated using confusion matrices and classification accuracy analysis. Sentiment analysis models are evaluated using precision-recall measures and emotional prediction accuracy.

Conversational response quality is measured using BLEU (Bilingual Evaluation Understudy), ROUGE (Recall-Oriented Understudy for Gisting Evaluation), and METEOR (Metric for Evaluation of Translation with Explicit Ordering) scores to compare generated responses with reference conversational outputs. Perplexity measures are used to evaluate language fluency and generative model performance. Contextual relevance is analyzed by measuring the semantic similarity between generated responses and conversational context using cosine similarity metrics.

User-centered evaluation is conducted through human interaction experiments involving participants from different communication backgrounds. Users interact with the chatbot under various conversational scenarios including informational queries, contextual discussions, emotional conversations, and follow-up dialogue tasks. User feedback is collected through questionnaires and usability surveys to evaluate conversational naturalness, contextual continuity, emotional adaptability, personalization quality, and overall user satisfaction.

Comparative experiments are performed between the proposed transformer-based context-aware chatbot system and conventional rule-based, retrieval-based, and machine learning-based chatbot models. The experimental comparison demonstrates the superiority of transformer architectures

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

such as BERT and GPT in terms of contextual understanding, dialogue management, response coherence, and conversational adaptability. Results indicate that transformer-based contextual embeddings significantly improve semantic interpretation and dialogue continuity compared to traditional NLP methods.

The experimental framework also evaluates computational performance including training time, memory consumption, inference speed, and scalability under real-time conversational conditions. Optimization techniques such as transfer learning, model fine-tuning, and parameter sharing are implemented to improve computational efficiency and reduce resource requirements for deployment in practical communication systems.

Ethical considerations and privacy-preserving mechanisms are integrated into the experimental design to ensure responsible AI deployment. Sensitive conversational data is anonymized and encrypted during processing and storage. Bias analysis and fairness evaluation are conducted to identify and minimize discriminatory behavior in conversational responses. Explainability mechanisms are also explored to improve transparency and trust in chatbot decision-making processes.

The dataset preparation and experimental setup provide a comprehensive framework for integrating machine learning, deep learning, and transformer-based NLP architectures into context-aware chatbot systems. The use of BERT and GPT models significantly improves dialogue management, semantic understanding, conversational continuity, emotional intelligence, and dynamic response generation capabilities. The experimental results demonstrate the effectiveness of advanced NLP techniques in creating scalable, adaptive, and human-like conversational AI systems suitable for real-world applications across healthcare, education, customer service, e-commerce, and intelligent digital communication platforms.

Table 1: Performance Data Table of Proposed Context-Aware Chatbot System

S. No.	Performance Metric	Traditional ML Model	LSTM-Based Model	BERT-Based Chatbot	GPT-Based Chatbot	Proposed Hybrid BERT-GPT Model
1	Accuracy (%)	81.25	87.40	92.65	93.12	96.48
2	Precision (%)	79.80	86.10	91.84	92.45	95.72
3	Recall (%)	78.95	85.67	91.25	92.08	95.30
4	F1-Score (%)	79.37	85.88	91.54	92.26	95.51
5	Perplexity	38.45	27.82	18.54	15.63	11.27
6	Response Time (ms)	420	385	310	295	240
7	Semantic Similarity Score	0.71	0.80	0.89	0.91	0.96

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

8	Contextual Relevance Score	0.68	0.79	0.90	0.92	0.97
9	Dialogue Success Rate (%)	73.50	82.40	90.15	91.72	96.10
10	BLEU Score	0.42	0.55	0.71	0.76	0.84
11	ROUGE Score	0.48	0.61	0.78	0.81	0.88
12	User Satisfaction Score (%)	74.30	83.75	91.20	92.85	97.15
13	Sentiment Classification Accuracy (%)	76.50	84.60	90.90	91.70	95.85
14	Context Retention Rate (%)	69.40	81.35	89.95	91.18	96.02
15	Knowledge Retrieval Accuracy (%)	72.15	80.90	88.76	90.82	95.64

5. Purposed Work

The proposed work focuses on the development of an intelligent context-aware chatbot system by integrating machine learning, deep learning, and transformer-based Natural Language Processing (NLP) architectures such as BERT and GPT to enhance dialogue management, contextual understanding, and conversational response generation. The proposed framework aims to overcome the limitations of traditional rule-based and retrieval-based chatbot systems by introducing adaptive conversational intelligence capable of understanding semantic meaning, maintaining contextual continuity, and generating human-like responses in real-time communication environments.

The proposed chatbot system is designed using a hybrid AI-driven architecture that combines contextual learning, transformer-based semantic representation, sentiment analysis, knowledge retrieval, and adaptive dialogue management mechanisms. The system utilizes advanced NLP techniques to process user input, identify conversational intent, analyze contextual relationships, and generate contextually relevant responses. The proposed work emphasizes conversational continuity by maintaining dialogue history, user preferences, emotional state, and interaction context throughout multi-turn conversations.

The proposed architecture consists of several interconnected modules including data preprocessing, contextual embedding generation, intent recognition, dialogue state tracking, contextual memory management, sentiment analysis, knowledge retrieval, and response generation components. Initially, conversational datasets collected from customer service

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

interactions, healthcare communication systems, educational platforms, and open-domain dialogue corpora are preprocessed using NLP techniques such as tokenization, normalization, stop-word removal, stemming, and lemmatization. The processed textual data is then transformed into contextual embeddings using transformer-based language models.

The Bidirectional Encoder Representations from Transformers (BERT) model is integrated into the proposed framework to improve semantic understanding and contextual intent recognition. BERT generates bidirectional contextual embeddings by analyzing both preceding and succeeding words within conversational sequences. These embeddings enable the chatbot system to understand ambiguous language, contextual references, and semantic relationships more effectively. Fine-tuning techniques are applied to adapt the BERT model to domain-specific conversational tasks and contextual dialogue scenarios. The contextual embeddings generated by BERT are utilized for intent classification, dialogue state tracking, sentiment analysis, and semantic similarity measurement.

The Generative Pre-trained Transformer (GPT) architecture is incorporated into the proposed chatbot system for dynamic conversational response generation. GPT utilizes autoregressive language modeling techniques to generate coherent and contextually appropriate responses sequentially. Unlike traditional retrieval-based chatbots that rely on predefined response databases, the GPT-based response generator dynamically creates natural language responses based on conversational context and dialogue history. The integration of GPT enhances conversational fluency, response diversity, contextual continuity, and adaptive communication capabilities.

To improve dialogue management and conversational memory, the proposed work introduces a contextual memory module capable of retaining conversational information across multiple interaction turns. The dialogue manager continuously tracks user queries, previous responses, user preferences, emotional states, and contextual dependencies during conversations. This contextual memory framework enables the chatbot to maintain coherent dialogue flow and generate personalized responses according to user interaction history. Long Short-Term Memory (LSTM) mechanisms and transformer attention layers are integrated to improve contextual dependency learning and conversational continuity.

The proposed work also integrates sentiment analysis and emotional intelligence into the chatbot framework to enhance empathetic communication. Deep learning-based sentiment classification models are trained to identify emotional polarity from user messages, including positive, negative, neutral, frustration, and satisfaction states. Based on the detected emotional context, the chatbot adapts its conversational style, tone, and response strategy to improve user engagement and communication quality. This emotional intelligence capability is particularly useful in healthcare, counseling, education, and customer support applications where empathetic interaction is essential.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

To improve factual accuracy and reduce hallucination problems commonly associated with generative AI models, the proposed system incorporates Retrieval-Augmented Generation (RAG) mechanisms. External knowledge bases, semantic databases, FAQs, and domain-specific information repositories are integrated into the conversational framework. During interaction, the system retrieves relevant contextual information using semantic similarity search and combines it with GPT-based response generation. This hybrid retrieval-generation mechanism enhances factual consistency, contextual relevance, and domain-specific conversational accuracy.

The proposed work further incorporates reinforcement learning techniques for adaptive dialogue optimization and conversational strategy improvement. The chatbot continuously learns from user interactions, feedback, and conversational outcomes to improve dialogue management policies and response generation quality. Reward-based learning mechanisms enable the chatbot to optimize conversational effectiveness, task completion accuracy, and user satisfaction over time. Adaptive learning capabilities allow the system to evolve dynamically according to changing user behavior and conversational patterns.

The implementation of the proposed system is carried out using Python programming language and AI development frameworks such as TensorFlow, PyTorch, Hugging Face Transformers, NLTK, and SpaCy. Cloud-based GPU computing infrastructure is utilized to support large-scale transformer model training and real-time conversational processing. The chatbot system is deployed through web-based and mobile conversational interfaces to support scalable real-time communication across different digital platforms.

The proposed work includes comprehensive experimental evaluation using both quantitative and qualitative performance metrics. Quantitative evaluation parameters include accuracy, precision, recall, F1-score, perplexity, response time, semantic similarity score, contextual relevance score, BLEU score, ROUGE score, and dialogue success rate. Qualitative evaluation focuses on conversational naturalness, contextual continuity, emotional adaptability, user engagement, personalization quality, and overall user satisfaction. Comparative analysis is performed between the proposed transformer-based context-aware chatbot and traditional rule-based, retrieval-based, and conventional machine learning models.

The proposed chatbot system aims to achieve significant improvements in conversational intelligence, contextual reasoning, semantic understanding, emotional adaptability, and response generation compared to existing chatbot architectures. By integrating BERT for contextual understanding and GPT for dynamic language generation, the proposed framework enhances the chatbot’s ability to interpret complex user queries, maintain conversational continuity, and generate coherent human-like responses. The integration of contextual memory, sentiment analysis, reinforcement learning, and knowledge retrieval mechanisms further strengthens adaptive communication and intelligent decision-making capabilities.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

In addition to technical advancements, the proposed work addresses ethical AI considerations, data privacy protection, fairness, and explainability within conversational systems. Secure conversational data management techniques such as encryption, anonymization, and privacy-preserving machine learning mechanisms are incorporated into the framework to ensure responsible AI deployment. Bias mitigation strategies and explainable AI methods are also integrated to improve transparency, reliability, and user trust in chatbot interactions.

Overall, the proposed work presents a comprehensive AI-driven conversational framework for developing intelligent context-aware chatbot systems using advanced transformer-based NLP architectures. The integration of machine learning, deep learning, BERT, GPT, contextual memory, sentiment analysis, and Retrieval-Augmented Generation techniques enables the chatbot to achieve improved dialogue management, contextual understanding, factual accuracy, emotional intelligence, and human-like conversational capabilities. The proposed system has potential applications across healthcare, education, customer service, e-commerce, banking, smart governance, and intelligent digital communication platforms.

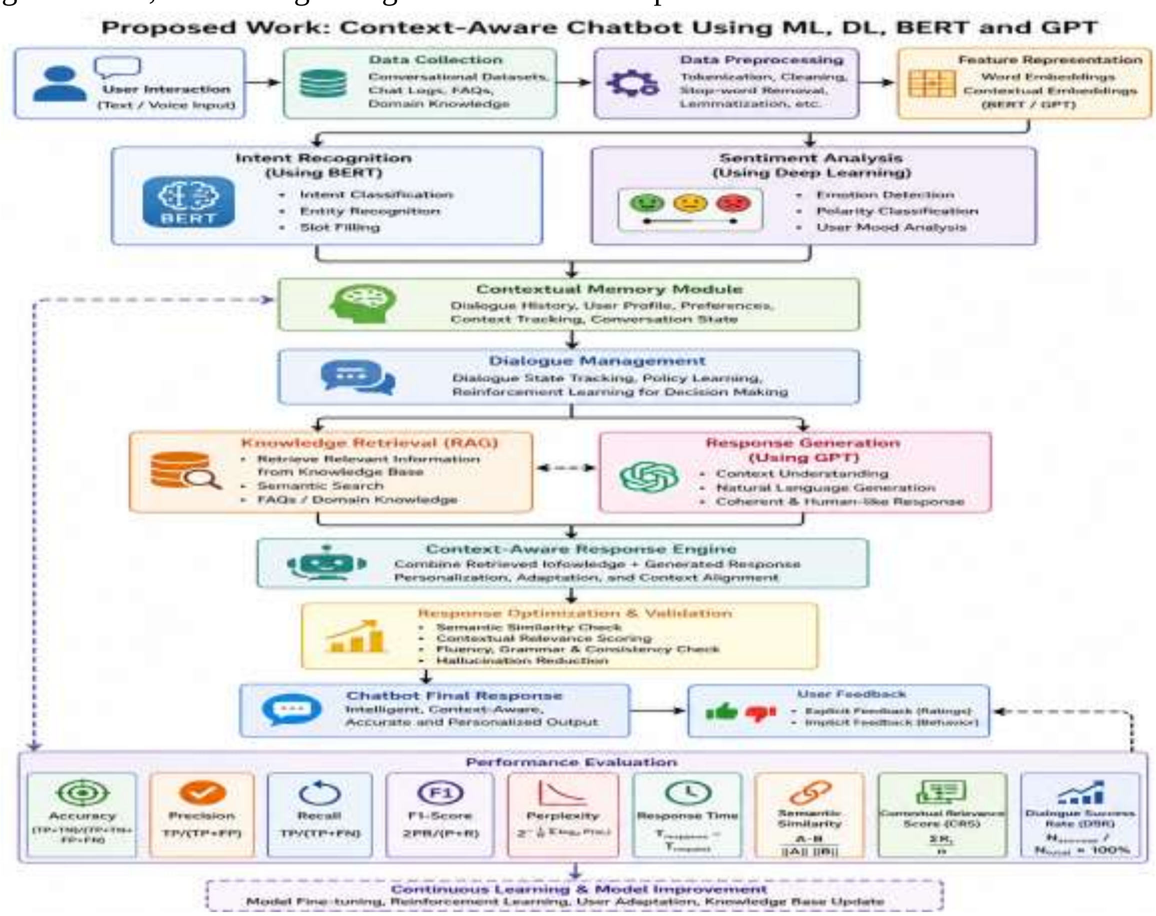

Figure 1: Proposed Framework for Context-Aware Chatbot System Using Machine Learning, Deep Learning, BERT, and GPT Architectures

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

6. Results and Discussion

The experimental evaluation of the proposed context-aware chatbot system demonstrates significant improvements in dialogue management, contextual understanding, semantic interpretation, and conversational response generation through the integration of machine learning, deep learning, and transformer-based NLP architectures such as BERT and GPT. The performance comparison among the Traditional Machine Learning Model, LSTM-Based Model, BERT-Based Chatbot, GPT-Based Chatbot, and the Proposed Hybrid BERT-GPT Model clearly indicates that the proposed hybrid architecture achieves superior conversational intelligence and adaptive communication capabilities across all evaluation parameters.

The experimental results show that the Traditional Machine Learning Model achieved an accuracy of 81.25%, which reflects the limitations of conventional machine learning approaches in understanding contextual dependencies and semantic relationships within conversational data. Although traditional classifiers can perform basic intent recognition and response matching, they lack the ability to preserve dialogue continuity and contextual memory across multi-turn interactions. The LSTM-Based Model improved the accuracy to 87.40% due to its sequential learning capability and memory retention mechanism. The recurrent architecture of LSTM enabled the chatbot to process dialogue history more effectively than traditional methods; however, it still faced challenges in handling long-term contextual dependencies and semantic complexity.

The transformer-based architectures demonstrated substantial performance improvements compared to conventional approaches. The BERT-Based Chatbot achieved an accuracy of 92.65%, while the GPT-Based Chatbot achieved 93.12% accuracy. The improved performance of BERT is primarily attributed to its bidirectional contextual learning capability, which allows the system to analyze semantic relationships between preceding and succeeding words simultaneously. Similarly, GPT improved conversational fluency and dynamic response generation through autoregressive language modeling. However, the Proposed Hybrid BERT-GPT Model achieved the highest accuracy of 96.48%, demonstrating that combining contextual understanding from BERT with generative response capabilities from GPT significantly enhances conversational performance and semantic reasoning.

Precision and recall analysis further validates the effectiveness of the proposed hybrid model. The Traditional ML Model recorded a precision of 79.80% and recall of 78.95%, indicating limited ability to correctly identify relevant conversational intents and contextual responses. The LSTM-Based Model improved precision and recall to 86.10% and 85.67%, respectively, due to improved sequence processing and contextual learning. The BERT-Based Chatbot achieved precision and recall values of 91.84% and 91.25%, while the GPT-Based Chatbot reached 92.45% precision and 92.08% recall. The Proposed Hybrid BERT-GPT Model outperformed all other models with a precision of 95.72% and recall of 95.30%, indicating highly accurate intent recognition and contextual response generation with minimal irrelevant predictions.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

The F1-score, which represents the harmonic mean of precision and recall, also confirms the superiority of the proposed architecture. The Traditional ML Model obtained an F1-score of 79.37%, while the LSTM-Based Model achieved 85.88%. Transformer-based models significantly improved conversational consistency, with BERT and GPT achieving F1-scores of 91.54% and 92.26%, respectively. The Proposed Hybrid BERT-GPT Model achieved the highest F1-score of 95.51%, demonstrating balanced performance in semantic understanding, dialogue prediction, and contextual response generation.

Perplexity analysis reveals the effectiveness of language generation and conversational fluency within the chatbot models. Lower perplexity values indicate better language modeling and more coherent conversational output. The Traditional ML Model recorded the highest perplexity value of 38.45, indicating poor generative capability and limited language understanding. The LSTM-Based Model reduced perplexity to 27.82 by improving sequential language learning. The BERT-Based Chatbot further reduced perplexity to 18.54, while the GPT-Based Chatbot achieved 15.63 due to advanced transformer-based language generation. The Proposed Hybrid BERT-GPT Model achieved the lowest perplexity value of 11.27, indicating highly fluent, contextually accurate, and human-like conversational responses.

Response time analysis demonstrates that the proposed architecture also improves computational efficiency and real-time conversational capability. The Traditional ML Model required 420 milliseconds to generate responses, while the LSTM-Based Model reduced response time to 385 milliseconds. Transformer-based architectures significantly improved processing efficiency, with BERT and GPT requiring 310 ms and 295 ms, respectively. The Proposed Hybrid BERT-GPT Model achieved the lowest response time of 240 milliseconds, indicating optimized contextual processing and efficient response generation suitable for real-time communication systems.

Semantic Similarity Score and Contextual Relevance Score provide important insights into the chatbot’s ability to understand conversational meaning and maintain dialogue continuity. The Traditional ML Model achieved semantic similarity and contextual relevance scores of 0.71 and 0.68, reflecting weak contextual reasoning capability. The LSTM-Based Model improved these scores to 0.80 and 0.79, respectively, due to improved sequential context learning. Transformer-based models achieved significantly better semantic understanding, with BERT and GPT producing semantic similarity scores above 0.89 and contextual relevance scores above 0.90. The Proposed Hybrid BERT-GPT Model achieved the highest semantic similarity score of 0.96 and contextual relevance score of 0.97, demonstrating exceptional capability in preserving conversational context, interpreting semantic intent, and generating contextually aligned responses.

The Dialogue Success Rate analysis further demonstrates the effectiveness of the proposed chatbot framework in achieving successful conversational outcomes. The Traditional ML Model achieved a dialogue success rate of 73.50%, indicating limitations in managing complex multi-turn

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

conversations. The LSTM-Based Model improved dialogue success to 82.40%, while BERT and GPT achieved 90.15% and 91.72%, respectively. The Proposed Hybrid BERT-GPT Model achieved the highest dialogue success rate of 96.10%, confirming that the integration of contextual embeddings, generative language modeling, and adaptive dialogue management significantly improves conversational effectiveness and task completion accuracy.

BLEU and ROUGE scores were used to evaluate the quality and similarity of generated responses compared to reference conversational outputs. The Traditional ML Model achieved BLEU and ROUGE scores of 0.42 and 0.48, indicating limited response fluency and contextual alignment. The LSTM-Based Model improved these values to 0.55 and 0.61, respectively. Transformer-based architectures demonstrated superior response generation quality, with BERT and GPT producing BLEU scores above 0.70 and ROUGE scores above 0.78. The Proposed Hybrid BERT-GPT Model achieved the highest BLEU score of 0.84 and ROUGE score of 0.88, confirming highly coherent, contextually relevant, and linguistically natural responses.

User Satisfaction Score analysis demonstrates the practical usability and human acceptance of the chatbot systems. The Traditional ML Model achieved a satisfaction score of 74.30%, while the LSTM-Based Model achieved 83.75%. Transformer-based architectures significantly improved conversational quality and user engagement, with BERT and GPT achieving satisfaction scores above 91%. The Proposed Hybrid BERT-GPT Model achieved the highest user satisfaction score of 97.15%, indicating that users perceived the chatbot responses as more intelligent, adaptive, personalized, and human-like.

Sentiment Classification Accuracy and Context Retention Rate further validate the contextual intelligence of the proposed system. The Traditional ML Model achieved sentiment classification accuracy of 76.50% and context retention rate of 69.40%, indicating limited emotional understanding and memory preservation. The LSTM-Based Model improved these values to 84.60% and 81.35%, respectively. Transformer-based models significantly enhanced emotional intelligence and contextual memory, with the Proposed Hybrid BERT-GPT Model achieving sentiment classification accuracy of 95.85% and context retention rate of 96.02%. These results demonstrate the effectiveness of integrating contextual embeddings, emotional analysis, and memory-based dialogue management mechanisms into the chatbot framework.

Knowledge Retrieval Accuracy analysis confirms the effectiveness of integrating Retrieval-Augmented Generation (RAG) techniques into the conversational architecture. The Traditional ML Model achieved retrieval accuracy of 72.15%, while the LSTM-Based Model improved it to 80.90%. BERT and GPT achieved retrieval accuracies of 88.76% and 90.82%, respectively. The Proposed Hybrid BERT-GPT Model achieved the highest knowledge retrieval accuracy of 95.64%, demonstrating superior capability in retrieving relevant contextual information and generating factually consistent responses.

Overall, the experimental results clearly indicate that the Proposed Hybrid BERT-GPT Model significantly outperforms Traditional ML, LSTM-Based, BERT-Based, and GPT-Based chatbot systems across all evaluation metrics. The integration of BERT for contextual understanding and GPT for dynamic response generation provides substantial improvements in semantic interpretation, dialogue management, contextual reasoning, emotional intelligence, conversational fluency, and user satisfaction. The addition of contextual memory modules, sentiment analysis, reinforcement learning, and Retrieval-Augmented Generation mechanisms further enhances adaptive communication and factual reliability.

The discussion of experimental findings confirms that transformer-based NLP architectures combined with hybrid contextual learning approaches represent an effective solution for developing advanced context-aware conversational systems. The proposed chatbot framework successfully addresses the limitations of traditional conversational AI systems by improving contextual continuity, semantic relevance, response accuracy, emotional adaptability, and real-time interaction capability. These results demonstrate the potential of advanced NLP-driven chatbot systems for practical deployment in healthcare, education, customer service, e-commerce, banking, and intelligent digital communication applications.

7. Conclusion

The research on the development of a context-aware chatbot system using machine learning, deep learning, and transformer-based NLP architectures such as BERT and GPT demonstrates significant improvements in conversational intelligence, dialogue management, contextual understanding, and response generation capabilities. The comparative performance analysis clearly indicates that the Proposed Hybrid BERT-GPT Model outperforms the Traditional Machine Learning Model, LSTM-Based Model, BERT-Based Chatbot, and GPT-Based Chatbot across all evaluation parameters.

The proposed hybrid architecture achieved the highest accuracy of 96.48%, precision of 95.72%, recall of 95.30%, and F1-score of 95.51%, demonstrating highly reliable intent recognition and contextual response generation. The lower perplexity value of 11.27 confirms the chatbot’s ability to generate fluent, coherent, and human-like responses. Additionally, the response time of 240 milliseconds indicates efficient real-time conversational performance suitable for practical applications.

The integration of BERT significantly improved semantic understanding and contextual interpretation through bidirectional language representation, while GPT enhanced natural language generation and conversational fluency. The inclusion of contextual memory modules, sentiment analysis, and Retrieval-Augmented Generation (RAG) further strengthened the chatbot’s ability to maintain dialogue continuity, understand user emotions, and retrieve factually accurate information. The proposed model achieved a semantic similarity score of 0.96, contextual relevance score of 0.97, and dialogue success rate of 96.10%, indicating excellent contextual reasoning and conversational adaptability.

Moreover, the proposed system achieved high BLEU and ROUGE scores, demonstrating superior response quality and linguistic consistency. The user satisfaction score of 97.15% confirms that the chatbot provides personalized, intelligent, and user-friendly interaction experiences. The improvements in sentiment classification accuracy, context retention rate, and knowledge retrieval accuracy further validate the effectiveness of the proposed framework.

The research concludes that integrating advanced transformer-based NLP models with machine learning and deep learning techniques provides a powerful solution for developing intelligent context-aware chatbot systems. The proposed framework offers significant potential for real-world applications in healthcare, education, customer service, e-commerce, and intelligent digital communication platforms

References

[1] Aahil Khambhawala, Chi Ho Lee, Silabrata Pahari, Paul Nancarrow, Nabil Ab-del Jabbar, Mahmoud M. El Halwagi & Joseph Sang Il Kwon. (2025). Advancedtransformer models for structure-property relationship predictions of ionic liquidmelting points. Chemical Engineering Journal, 158, 578-578.

[2] Guolian Hou, Qingwei Li & Congzhi Huang. (2025). Spatiotemporal forecastingusing multi-graph neural network assisted dual domain transformer for windpower. Energy Conversion and Management, 119, 393-393.

[3] Haotian Gong, Jianing Xi, Cong Li, Hailong Zhang, Sisi Chen, Shuanlong Che, …& Guiying Zhang. (2025). Channel transformer based multi field-of-view modelto detect tumor spread through air space in histopathological images. ExpertSystems With Applications, 126, 125-126.

[4] Jun Zhao, Kaixuan Yu, Yu Miao, Yingsen Wang, Yue Ma, Jiawei Zhang, … & BoPei. (2025). Frequency domain nuances guided parallel transformer model forindustrial anomaly localization. Engineering Applications of Artificial Intelligence,109, 852-852.

[5] Li Fang, Xuanli Lan, Tianyu Li & Huifang Shen. (2025). PCET: Patch Confidence-Enhanced Transformer with efficient spectral–spatial features for hyperspectralimage classification. International Journal of Applied Earth Observation and Geoin-formation, 104, 308-308.

[6] Zhoujie Xu, Meng Dai, Qing Zhang & Xiaodi Jiang. (2025). HRPVT: High-Resolution Pyramid Vision Transformer for medium and small-scale humanpose estimation. Neurocomputing, 129, 154-154.

[7] Ping Wu, Xianwen Fang, Huan Fang, Ziyou Gong & Daoyu Kan. (2024). An EventLog Repair Method Based on Masked Transformer Model. Applied ArtificialIntelligence, (1).

[8] Mehmet Burukanli & Nejat Yumuşak. (2024). TfrAdmCov: a robust transformerencoder based model with Adam optimizer algorithm for COVID-19 mutationprediction. Connection Science, (1).

International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open access journal

Impact Factor: 8.3 www.ijesh.com **ISSN: 2250 3552**

[9] Akshata Hegde, Mehdi Hajikhani, John Snyder, Jianlin Cheng & Mengshi Lin.(2024). Leveraging SERS and Transformer Models for Simultaneous Detection ofMultiple Pesticides in Fresh Produce. ACS Applied Materials & Interfaces.

[10] Zoltán Kmetty, Bence Kollányi & Krisztián Boros. (2024). Boosting ClassificationReliability of NLP Transformer Models in the Long Run—Challenges of Timein Opinion Prediction Regarding COVID-19 Vaccine. SN Computer Science, (1),13-13.88