

University Chatbots: A Systematic Review of the Transition from Automated Retrieval to Personalised Support

Pranshul Sharma¹, Dharmendra Solanki², Sumit Khirwar³ and Anshul Kumar Singh⁴

¹⁺²⁺³ UG Students, Department of Computer Science & Engineering

⁴ Assistant Professor, Department of Computer Science & Engineering
Raja Balwant Singh Engineering Technical Campus, Bichpuri, Agra, India

Abstract - Student recruitment, engagement, and retention remain key priorities for universities, leading to increased adoption of digital support tools. Chatbots have emerged as an effective solution for managing student queries, particularly in online and hybrid environments. This paper reviews the evolution of university chatbots from rule-based systems to advanced Natural Language Processing (NLP) and generative AI models, with a focus on administrative and support services beyond classroom use. It compares existing implementations and highlights the role of development platforms in deployment. The study identifies a major challenge in the lack of high-quality, domain-specific datasets and notes that many institutions have yet to fully leverage modern chatbot capabilities. It concludes by emphasising the potential of integrating institutional data with generative AI to create more responsive and student-centred support systems.

Key Words: University chatbots, student support, Natural Language Processing, generative AI, higher education systems, conversational agents, digital learning support.

1. INTRODUCTION

Chatbots, also known as conversational agents, originated from early artificial intelligence research, with foundational ideas such as the Turing Test proposed in the 1950s. One of the earliest working systems, ELIZA, demonstrated that basic human-computer conversations were possible and laid the groundwork for future developments in this field [1]. Over time, artificial intelligence has progressed significantly, especially since the 2000s, when increased computing power and large datasets enabled the rise of modern machine learning and deep learning techniques.

Today's chatbot systems rely heavily on Natural Language Processing (NLP). This includes Natural Language Understanding (NLU), which interprets user intent, and Natural Language Generation (NLG), which produces meaningful responses, as discussed by Maroengsit et al. [2]. These advancements have made conversational systems

common in everyday tools such as Apple Siri, Google Assistant, and Amazon Alexa. As a result, users have become comfortable interacting with machines through natural conversation, leading to widespread adoption of chatbots as efficient self-service support systems, replacing or supplementing traditional communication channels like email and phone support [3].

In the context of higher education, chatbots are increasingly used to handle student-related queries, especially for academic and administrative support such as course details, schedules, and examination information. However, research by Nwankwo [20] highlights that many existing academic advising systems still focus narrowly on administrative tasks like course registration, rather than supporting broader student development needs such as learning goals, skills, and career planning.

Several practical implementations show how chatbots are being used in education. For instance, Goel and Polepeddi [21] developed "Jill Watson" to support students in a massive open online course (MOOC), where it handled repetitive questions, shared announcements, and assisted with routine interactions at scale. Similarly, Lee et al. [22] introduced "Infobot," a chatbot designed for an introductory networking course that allowed students to ask questions about lectures, schedules, and class locations through platforms like Telegram and Facebook Messenger.

Overall, existing studies suggest that most educational chatbots are primarily used for teaching support, student engagement, and answering routine queries. A systematic review by Okonkwo and Ade-Ibikola [23] further groups chatbot applications in education into areas such as administration, assessment, advising, and research support, with the majority still focused on instructional and student assistance tasks.

2. Scope of the Research

Universities are increasingly adopting digital systems to improve student engagement, retention, and service delivery. In this context, chatbot technologies are emerging as effective tools for supporting both students and administrative staff. Hussain [5] categorises institutional chatbot applications into three primary functional areas: improving access to information and guidance for prospective and current students, streamlining routine campus services, and enhancing teaching, learning, and assessment processes.

These applications are particularly relevant in higher education environments where institutions face growing competition, rising student expectations, and pressure to deliver high-quality services under constrained budgets. As a result, the optimisation of student experience—whether in physical or online learning environments—has become a key strategic priority [9].

Based on Hussain's framework [5], chatbot deployment in universities can broadly be grouped into two main operational domains:

- Admissions, student recruitment, and administrative support
- Teaching and learning support systems

These categories provide a structured lens for analysing how conversational AI systems are currently being integrated into higher education services.

2.1 Methodology

This review focuses on chatbot applications in universities that support student queries outside direct classroom teaching. The study was conducted through a structured literature search using Google Scholar, targeting peer-reviewed conference papers and journal articles across different domains where chatbots are applied.

The search was later refined to focus specifically on higher education by using keywords such as “chatbot + university” and “chatbot + education.” Although several relevant studies were identified, only publications from 2015 onwards were included, as this period marks the emergence of modern AI frameworks such as TensorFlow [10] and PyTorch, which significantly influenced chatbot development.

Selected studies were further filtered to include only those implemented in real university environments and focused on administrative or student support functions rather than classroom-based teaching or learning systems.

To capture recent developments in generative AI, additional search queries such as “ChatGPT university,” “closed-domain chatbot GPT,” “RAG chatbot university,” and “LangChain chatbot university” were used. These searches helped identify newer implementations based on retrieval-augmented generation and large language models.

Finally, all collected studies were critically reviewed to ensure relevance, with strict inclusion criteria ensuring that only chatbot systems used for non-instructional student support within universities were considered for detailed analysis.

2.2 Admissions / Student Recruitment / Administration

University admissions and student recruitment processes are critical for maintaining a steady inflow of students, making them one of the most important areas for chatbot deployment. During peak admission periods, particularly in late summer, often referred to as the “summer melt” phase, universities experience a high volume of repetitive student queries that can overload administrative staff [11].

To address this, many institutions have introduced chatbots to improve responsiveness and enhance the applicant experience. Choque-Díaz et al. [12] emphasise that timely communication during admissions is essential, noting that some systems aim to respond to student queries within 48 hours to improve engagement and satisfaction.

Several studies demonstrate practical chatbot implementations in this area. For example, Chandra and Suyanto [13] developed a WhatsApp-based chatbot that uses frequently asked questions (FAQs) from admissions data to respond to applicants efficiently. Similarly, Santoso et al. [14] implemented an FAQ-driven chatbot to support university admission services, showing that repetitive queries are well-suited for automated handling.

These systems are generally based on a simple interaction model where a user asks a question and the chatbot immediately provides a predefined response. This is known as a single-turn conversation, which works effectively for straightforward and repetitive queries. However, more advanced systems support multi-turn conversations, where

context from previous interactions is retained. Aujogue and Aussem [15] highlight this capability, explaining that it allows chatbots to understand follow-up questions more intelligently. For example, if a student first asks about studying mechatronics online and then asks about its cost, the chatbot can interpret “it” based on earlier context and provide relevant fee information.

Beyond admissions, universities also manage a large amount of distributed administrative knowledge across websites, internal systems, and staff expertise. Khin and Soe [16] developed a chatbot that consolidates such information to handle general campus queries, including details about departments, library services, research, and alumni. In a similar approach, Rana [17] designed a chatbot system that supports enrolled students using both structured and unstructured FAQ data, ensuring assistance is available even when staff members are not accessible.

2.3 Teaching And Learning

Teaching and learning in higher education involve several interconnected activities, including student engagement, feedback delivery, academic support, and learning analytics, as outlined by Singh [18]. Within this ecosystem, AI-based teaching assistants are increasingly being explored as tools to enhance learning experiences. Satow [19] proposes a six-level framework for AI teaching assistants, ranging from basic functions such as welcoming students with personalised messages to advanced capabilities like providing individualised academic coaching.

Despite these advancements, the application of chatbots in education remains complex. Yang and Evans [7] argue that educational requirements are often more diverse than those in other domains, which can limit the effectiveness of standardised chatbot solutions. Their work highlights varied implementations, including simulation-based chatbots for online programs, library recommendation assistants, and helpdesk support systems.

Similarly, Cunningham-Nelson et al. [6] present an FAQ-based chatbot designed to answer routine academic queries such as assessment guidelines, examination schedules, and class timings. However, Nwankwo [20] points out a limitation in many academic advising systems, noting that they tend to focus mainly on course selection and registration rather than broader academic development, including learning outcomes, skill-building, and career

guidance. More advanced implementations demonstrate the growing capabilities of educational chatbots. Goel and Polepeddi [21] introduced “Jill Watson,” developed for a massive open online course (MOOC), which was capable of handling large-scale student interaction by responding to introductions, sharing weekly updates, and answering frequently asked questions autonomously.

In another example, Lee et al. [22] developed “Infobot,” a chatbot designed to support students in an introductory networking course in Hong Kong. This system operated across platforms such as Telegram and Facebook Messenger, allowing students to access information related to course content, schedules, and classroom locations.

Collectively, these studies illustrate a wide range of chatbot applications in teaching and learning, from simple FAQ systems to more advanced AI-driven assistants. A systematic review by Okonkwo and Ade-Ibikola [23] further categorises these applications into areas such as administration, assessment, academic advising, and research support. Notably, around 66% of the reviewed systems focus primarily on teaching and learning functions, particularly student engagement, content delivery, and personalised assistance.

3. CURRENT STATE OF LANGUAGE

Chatbot systems in higher education typically operate within a **closed-domain setting**, meaning they are designed to respond only to queries related to a specific area such as admissions, academic services, or student support. These systems differ from open-domain chatbots, which can engage in general conversation across virtually any topic. Open-domain models are often more conversational and human-like, making them closer to passing benchmarks such as the Turing Test [24]. In contrast, closed-domain chatbots are purpose-driven and restricted to predefined tasks within a specific context, which makes them more suitable for university environments. This review focuses specifically on such closed-domain implementations.

Modern chatbot systems are built using a range of language model architectures that have evolved significantly over time. Earlier approaches include rule-based systems such as Artificial Intelligence Markup Language (AIML) [25], followed by neural approaches like Sequence-to-Sequence (Seq2Seq) models [26]. More recent advancements include transformer-based architectures such as BERT (Bidirectional

Encoder Representations from Transformers) [27], [28], which have significantly improved language understanding capabilities. These models have been widely applied beyond education, including in sectors such as e-commerce customer support, insurance claim processing, financial advisory systems, and even healthcare applications like ophthalmology consultation chatbots.

In recent years, Generative Pre-trained Transformer (GPT)-based models have emerged as state-of-the-art (SOTA) solutions for conversational AI. These models enable more natural, context-aware responses and are increasingly being explored in both open-domain and closed-domain settings. While university-specific implementations remain limited, several recent studies have demonstrated the use of GPT-based and hybrid approaches in higher education environments for student support and administrative assistance [29], [30], [31], [32].

3.1 Rule-Based Chatbots

Rule-based chatbots, also known as pattern-based systems, represent one of the earliest and most widely used approaches in conversational AI. These systems operate by defining a fixed set of **intents**, which represent categories of user queries. In a university context, these intents often include areas such as admissions, registration, fees, timetables, and examinations.

Within each intent, multiple possible user questions are mapped to predefined responses. For example, under the “registration” category, a chatbot may handle queries such as “When should I register?”, “What documents are required for registration?”, or “Is on-campus registration necessary?”. Each of these questions represents different patterns expressing the same intent.

The chatbot responds by matching user input to predefined rules in a structured hierarchy. However, its intelligence is limited strictly to the knowledge encoded in its rule set. While rule-based systems work well in closed domains with predictable queries, their performance depends heavily on how complete the rule database is [33].

A key limitation of this approach is scalability. As user queries grow, developers must manually define and maintain large sets of intents, responses, and linguistic variations. If a query does not match any predefined rule, the chatbot fails to respond, reducing its flexibility and overall usefulness in real-world scenarios.

3.1.1 AIML

AIML (Artificial Intelligence Markup Language) is an open-source, XML-based framework used for developing rule-based chatbots. It represents chatbot knowledge in the form of **patterns (user inputs)** and corresponding **responses**, structured using interpretable tags processed by an AIML engine [34]. In AIML, the fundamental unit of knowledge is a **category**, which consists of a user input pattern and a matching response template. The <pattern> tag captures possible user queries, while the <template> tag defines the chatbot’s response. However, AIML does not naturally support multiple synonyms within a single category, meaning variations of a question must be explicitly defined. For instance, different greetings such as “Hello,” “Hi,” “Hey,” and “How are you” must each be separately included as patterns to ensure correct matching.

Despite being introduced in the early 2000s, AIML remains in use today and is supported by various open-source interpreters and platforms, including Pandorabots. With a large user base and long-standing presence, Pandorabots is one of the most established AIML-based chatbot platforms, which may explain its continued use in academic chatbot implementations.

Due to its rule-based structure, AIML is particularly suitable for FAQ-driven systems. For example, Khin and Soe [16] developed a university chatbot using an AIML knowledge base containing 970 question-answer pairs across nine categories, covering queries from students, faculty, and parents. This system was deployed using the Pandorabots platform.

Similarly, Ranoliya et al. [35] implemented an AIML-based chatbot designed specifically for university FAQs, covering a broad range of student queries. In another approach, Nwankwo [20] proposed an AIML-driven academic advising chatbot named Adviserbot, aimed at supporting student guidance services within universities.

3.2 AI-Based Chatbots

AI-based chatbots, also known as neural or deep learning-based chatbots, use machine learning models—often multi-layer neural networks—to generate or predict responses from user input using training data and Natural Language Processing (NLP) techniques [36].

Unlike rule-based systems, these chatbots are not restricted to predefined responses; instead, they learn patterns from

data and generate suitable outputs based on context and probability. The choice of model depends on the application and the availability of training data.

AI chatbots are generally classified into two main types: **retrieval-based** and **generative-based** systems. Retrieval-based models select responses from an existing database of question-answer pairs.

Although they can be improved through training and expanded datasets, they cannot generate entirely new responses. Such systems are commonly implemented using platforms like Microsoft QnA Maker and Google Dialogflow.

In contrast, generative chatbots use deep learning models to produce new responses that are not explicitly stored in the dataset. Modern architectures such as BERT, GPT, Gemini, Claude, and LLaMA are examples of state-of-the-art (SOTA) systems that have gained popularity due to increased computational power and access to large-scale datasets [2], [37].

All AI-based systems rely on Natural Language Understanding (NLU) preprocessing to interpret user intent before generating responses. This preprocessing improves model efficiency and accuracy. Common steps include tokenization, normalization, stop-word removal, stemming, and lemmatization. These steps help convert raw text into a structured format suitable for machine learning models.

Beyond preprocessing, several NLP techniques are used for deeper understanding of user queries. **Intent classification** identifies the purpose of a user's input, while **domain classification** is used when multiple subject areas are involved. Traditional models such as Bag of Words (BoW) and TF-IDF represent text numerically by converting words into vectors based on frequency and importance [47].

For semantic understanding, techniques like Word2Vec and GloVe are widely used to generate word embeddings. Word2Vec uses neural networks to learn word relationships based on context windows, while GloVe builds word vectors using global co-occurrence statistics [48], [49]. However, both methods are context-independent, meaning a word has the same representation regardless of meaning in different sentences.

To address contextual limitations, advanced architectures such as Long Short-Term Memory (LSTM) networks were introduced. LSTMs are capable of capturing long-term

dependencies in text using gated memory structures, including forget, input, and output gates [50]. This allows models to retain important contextual information over longer sequences, improving understanding in complex sentences.

For example, in a sentence like "The student completed her exam," LSTM models can retain contextual information to correctly interpret references such as pronouns across earlier sentences.

3.2.1 Retrieval-Based Chatbots

Retrieval-based chatbots generate responses by selecting the most relevant answer from a predefined set. They are typically based on supervised learning, where models learn the mapping between user queries and corresponding responses. For new inputs, the system predicts the best match, and if confidence is low, it may ask the user to rephrase the query.

Their performance improves by updating the dataset with new question-answer pairs or refining existing ones. Data is usually obtained from FAQs, knowledge bases, or APIs and processed using techniques such as intent recognition and word embeddings (e.g., word2Vec, GloVe). Applications include systems like Leo [51] and Acabot [52], while hybrid approaches combining retrieval and generative methods have also been explored [17], [53].

Although easy to deploy, these chatbots require continuous maintenance to keep responses accurate and up to date.

3.2.2 Generative-Based Chatbots

Generative-based chatbots use self-supervised deep learning models to generate responses that are not explicitly present in the training dataset. Unlike retrieval-based systems, they are capable of producing new, context-aware outputs by learning patterns from large-scale unlabelled text data. Such data is typically collected from sources like web documents, books, code repositories, and social media, enabling large language models (LLMs) to undergo pre-training followed by task-specific fine-tuning. Common datasets used in this process include the Cornell Movie Dialogue Corpus and SQuAD, which support conversational learning and question answering tasks [37], [54].

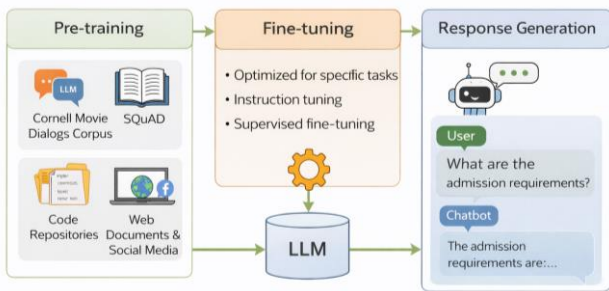


Fig.-1: Overview of Generative Chatbot Training Pipeline

Generative models are typically built using a combination of supervised learning (with labelled data) and self-supervised learning (with unlabelled data), making them highly scalable for natural language tasks. These approaches form the foundation of modern LLMs and advanced chatbot systems.

Frequent Pattern (FP) Growth is an unsupervised data mining technique used to discover frequent patterns in large datasets. It constructs a tree-based structure that compresses transactional data and identifies relationships between items. In chatbot systems, it has been used to extract user intent patterns from interaction logs. For example, Alias et al. [52] developed the “Acabot” academic chatbot using a variant called Frequent Adjacent Sequential Pattern (FASP), showing improved intent detection compared to traditional N-gram methods [55], [56].

Sequence-to-Sequence (Seq2Seq) Models is a neural network architecture based on Recurrent Neural Networks (RNNs) designed to map input sequences to output sequences. It is widely used in machine translation and conversational AI. However, standard RNNs suffer from limitations such as vanishing gradients and difficulty in handling long-term dependencies [59].

To improve performance, enhancements such as **Attention mechanisms** and **Bidirectional LSTMs (BiLSTM)** were introduced, allowing models to capture both past and future context in a sentence [60].

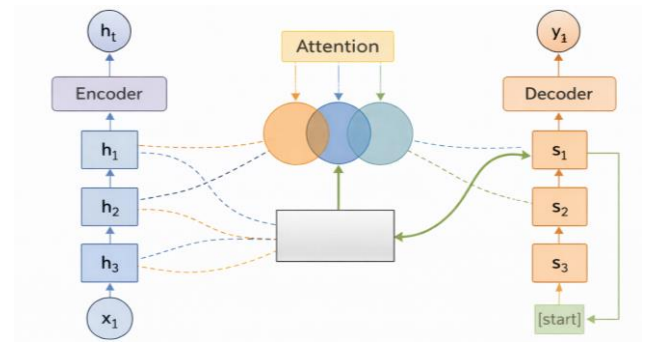


Fig. 2: Sequence-to-Sequence Model with Encoder-Decoder and Attention Mechanism

Studies show that Seq2Seq-based chatbot systems can generate meaningful responses in university settings. For instance, Khin and Soe [62] and Chandra and Suyanto [13] demonstrated improved performance using attention-based Seq2Seq models, with evaluation metrics such as BLEU scores showing better response quality compared to baseline models [63].

Transformer-Based Models (BERT, GPT, LLMs) introduced self-attention mechanisms, allowing models to analyse relationships between all words in a sentence simultaneously rather than sequentially.

This significantly improves contextual understanding [64], [65].

BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained bidirectional model that uses Masked Language Modelling (MLM) and Next Sentence Prediction (NSP) for training [66]. It is widely used for question answering and text classification tasks.

GPT (Generative Pre-trained Transformer), introduced by OpenAI, is an autoregressive model capable of generating coherent text by predicting the next token in a sequence. Newer versions, such as GPT-4, significantly improve reasoning ability and multimodal processing, including text and image understanding [67], [68]. The introduction of ChatGPT has further expanded real-world adoption in conversational AI systems.

Despite strong performance, challenges such as hallucination and bias remain in large language models, often linked to dataset quality and training limitations [70].

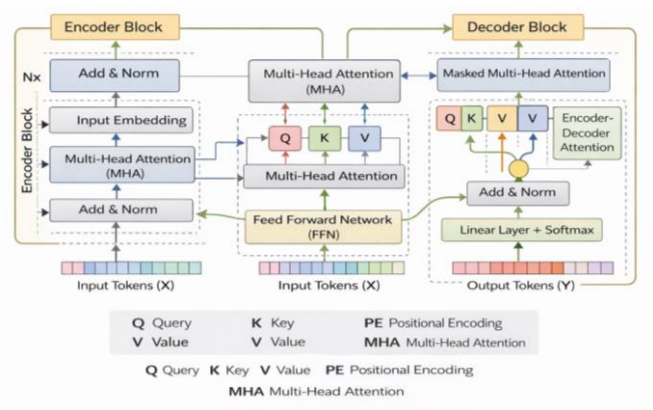


Fig. 3: Transformer-Based Language Model Architecture (Self-Attention Mechanism)

Recent studies show increasing use of GPT-based systems in universities, particularly when combined with Retrieval-Augmented Generation (RAG), which enhances responses using external knowledge sources without retraining the model [73]. Applications include academic advising, library support, and program-specific chatbots [30]–[32]. Frameworks such as LangChain are also widely used to integrate LLMs with institutional databases.

Modern LLM Ecosystem Recent advancements include models such as **Gemini (Google)**, **Claude (Anthropic)**, and **LLaMA (Meta)**. These systems improve reasoning, safety, and efficiency while supporting multimodal inputs and open-source development [74]–[76]. They collectively represent the current state-of-the-art in generative AI for conversational systems.

3.3 Model Overview

Table 1 presents a high-level comparison of closed-domain chatbot models used in university environments, selected specifically for handling student queries outside classroom teaching. These models are further evaluated in terms of performance, limitations, and applicability in later sections of this review.

As observed in Table 1, only one model follows a purely rule-based approach. Although AIML remains effective for FAQ-driven systems, its relevance has decreased with the emergence of more advanced AI-based techniques. In contrast, a majority of the reviewed models are based on encoder–decoder architectures, which form the backbone of modern generative chatbot systems.

Among these, Seq2Seq serves as an intermediate approach between retrieval-based and fully generative systems. It demonstrates that FAQ datasets can still be effectively used in generative frameworks, making it a transitional model in chatbot evolution.

With the introduction of transformer-based architectures such as BERT and GPT, chatbot development has shifted toward large-scale pre-trained models. These systems are first trained on massive datasets using high computational resources and later fine-tuned for specific closed-domain university applications using smaller task-specific datasets. This transfer learning approach has significantly improved contextual understanding and response quality.

A similar observation is also reported in existing **review studies on AI-based educational chatbots**, which highlight the transition from rule-based and retrieval-based systems to modern transformer-driven architectures, emphasising improvements in scalability, contextual understanding, and adaptability in higher education environments [77].

Table 2 summarises the comparative advantages and limitations of each model within university chatbot applications, highlighting trade-offs in complexity, scalability, and contextual performance.

3.4 Development Platforms

Chatbot development platforms play a crucial role in enabling researchers and developers to design, deploy, and evaluate conversational agents across various domains, including higher education.

These platforms support a wide range of capabilities, from simple FAQ-based retrieval systems to advanced AI-driven conversational frameworks, making them suitable for both technical and non-technical users [6].

Commonly used platforms such as Microsoft QnA Maker, Google Dialogflow, RASA, Wit.ai, and Botkit provide flexible environments for building chatbots [6]. While these tools were initially designed for retrieval-based chatbot development, they have evolved significantly to support advanced features such as intent classification, multi-turn dialogue management, and API-based integration with external systems.

Microsoft QnA Maker is often used in combination with Microsoft Language Understanding Intelligent Service (LUIS), which helps in identifying user intent and contextual

meaning. Similarly, Google Dialogflow provides a fully managed conversational AI platform that supports natural language understanding and scalable chatbot deployment using modern NLP techniques [6].

Open-source frameworks such as RASA offer full control over chatbot pipelines, allowing developers to customize natural language understanding, dialogue management, and training workflows. In contrast, Wit.ai provides an API-based development approach with support for multiple programming languages, enabling rapid prototyping of conversational agents. Botkit is another open-source toolkit designed primarily for building chatbots integrated with messaging platforms such as Slack, Facebook Messenger, and other communication channels [6].

In recent years, cloud-based and research-focused platforms such as Hugging Face and Google Colab have significantly improved accessibility to state-of-the-art machine learning models. These environments provide pre-trained models, GPU acceleration, and notebook-based development tools, allowing researchers to experiment with advanced natural language processing techniques without requiring high-performance local infrastructure.

In addition to general-purpose tools, several education-focused chatbot solutions, such as Mainstay (formerly AdmitHub) and Unibuddy, have been widely adopted in universities. These platforms are specifically designed for student engagement, admissions, enrolment, and retention processes. They also support integration with learning management systems (LMS) such as Canvas and Blackboard, as well as CRM platforms like Slate. Furthermore, they provide analytics-driven insights to measure student engagement, conversion rates, and interaction effectiveness.

Overall, chatbot development platforms have evolved from basic rule-based systems to sophisticated AI ecosystems that support scalable, intelligent, and highly customizable

conversational agents. This evolution directly aligns with the transition from traditional rule-based models to modern transformer-based architectures discussed in Tables 1 and 2, highlighting the strong relationship between chatbot models and their deployment environments [6].

Table-1: High-Level Breakdown of Closed-Domain Chatbot Models

Name	Type	Pre-training	Fine-tuning	Architecture	Model Type	Training Type
AIML	Retrieval	N/A	N/A	Rule-Based	N/A	N/A
Seq2Seq	Generative/Retrieval	X	✓	RNN/LSTM	Encoder-Decoder, Auto-regressive	Supervised
BERT	Generative	✓	✓	Transformer	Encoder-Decoder, Auto-encoding	Self-supervised
GPT	Generative	✓	✓	Transformer	Decoder, Auto-regressive	Self-supervised
FP-Growth	Retrieval	✓	✓	Frequent Pattern (FASP)	N/A	Unsupervised

Table-2: Comparative analysis of advantages and limitations of commonly used chatbot models.

Model Name	Advantages	Pitfalls
AIML	<ul style="list-style-type: none"> Well-established and widely used Simple to implement Uses XML for structured representation 	<ul style="list-style-type: none"> Requires manual data entry or editing tools Relies on pattern matching for intent detection Outdated compared to modern techniques
Seq2Seq	<ul style="list-style-type: none"> Supports both retrieval and generative tasks Attention mechanisms improve performance 	<ul style="list-style-type: none"> Struggles with long input sequences due to context loss
BERT	<ul style="list-style-type: none"> Utilises pre-training and fine-tuning Open-source availability 	<ul style="list-style-type: none"> Requires large training datasets High computational cost due to large model

		size
GPT	<ul style="list-style-type: none"> • Pre-training and fine-tuning approach • Works with relatively less task-specific data • Supports RAG-based integration • Increasing real-world implementations 	<ul style="list-style-type: none"> • Often accessible via paid APIs (though alternatives exist) • Risk of biased or incorrect outputs
FP-Growth	<ul style="list-style-type: none"> • Useful for exploring alternative, non-traditional approaches 	<ul style="list-style-type: none"> • Highly experimental with limited practical adoption

3.5 Implementation Comparison

After reviewing different chatbot models used in university environments, it is important to compare these implementations with those developed in other closed-domain sectors. This comparison primarily focuses on three aspects: dataset availability, training methodology, and evaluation techniques used to measure performance.

Evaluation of natural language processing (NLP) systems has traditionally relied on standard metrics, although these continue to evolve with modern large language models (LLMs). One of the most widely used metrics is BLEU (Bilingual Evaluation Understudy), introduced by Papineni et al. [63], which is commonly applied in machine translation

tasks. In university chatbot studies, BLEU has been used in works such as Palasundram et al. [61], Chandra and Suyanto

[13], Khin and Soe [62], and Kapociute-Dzikiene [77] to evaluate response similarity between generated and reference answers. In addition to BLEU, other evaluation measures such as the F1 score are frequently adopted, particularly in intent classification and information extraction tasks. The F1 score, used in studies by Nuruzzaman and Hussain [26] and Yu et al. [28], combines precision and recall to provide a balanced measure of system performance [78]. Precision reflects correctly identified outputs, while recall measures how many relevant outputs are successfully retrieved.

More recent approaches incorporate semantic similarity metrics. For example, ROUGE-1 and BERTScore have been applied in GPT-based chatbot evaluations such as Cherumanal et al. [30]. ROUGE evaluates overlap using n-gram matching, whereas BERTScore leverages contextual embeddings to measure meaning-based similarity rather

than exact word matches [79], [80]. These methods are particularly useful for generative systems where multiple valid responses may exist.

A broader evaluation perspective is discussed by Maroengsit et al. [2], who categorise chatbot assessment into user satisfaction, system functionality, and content quality. Similarly, Liu et al. [81] argue that embedding-based and context-aware evaluation techniques are more suitable for modern conversational AI systems compared to traditional lexical metrics alone.

Table 3 presents a comparative summary of chatbot implementations across different domains, including higher education, finance, healthcare, and e-commerce. Notably, most university-based systems rely on relatively small datasets compared to other sectors.

For example, Palasundram et al. [61] used only 100 question-answer pairs for training a Seq2Seq model, while Chandra and Suyanto [13] and Khin and Soe [62] worked with a few thousand pairs collected from university admission queries and messaging platforms. In contrast, financial and commercial domains demonstrate significantly larger datasets, such as Intellibot with over 10,000 pairs [26] and AVA with more than 22,000 pairs [28].

This difference in dataset size has a direct impact on model performance and generalisation ability. For instance, Seq2Seq-based university chatbots typically achieve moderate BLEU scores (around 0.41–0.95), while larger datasets in finance-related systems result in higher F1 scores (up to 0.98) [26]. Similarly, BERT-based systems show improved robustness but still depend heavily on dataset quality and structure [28].

Across university implementations, a recurring challenge is the difficulty in accessing and curating high-quality training data. Many datasets are manually constructed from FAQs, email logs, or messaging platforms, which limits scalability.

Palasundram et al. [61] also highlight the importance of controlled dataset variation, where unseen question categories are created through modifications such as word

reordering, synonym replacement, spelling variations, and sentence shortening to test model generalisation.

In studies such as Khin and Soe [62] and Chandra and Suyanto [13], even when attention mechanisms were introduced into Seq2Seq models, improvements in BLEU scores remained limited due to linguistic complexity and dataset constraints. Alias et al. [52] further demonstrate that even small datasets can be useful for intent discovery, although performance is restricted compared to large-scale models.

More advanced systems using retrieval-augmented generation (RAG) and transformer-based models, such as Falcon and GPT, show improved adaptability.

For example, Cherumanal et al. [30] found that intent-based FAQ models performed better in structured evaluation, while RAG-based approaches were more effective for handling unseen or inferred queries. Similarly, Siragusa and Pirrone [32] observed that GPT-based systems combined with RAG provided more reliable responses than fine-tuned models alone, although hallucination issues still remain a concern in academic contexts.

In the healthcare domain, BERT-based systems such as those developed by Lee et al. [27] demonstrate potential for assisting patients, although formal evaluation remains limited.

Meanwhile, in e-commerce applications, Kapociute-Dzikiene [77] highlights that even with limited datasets, transformer-based architectures can still achieve usable performance levels for practical deployment.

Overall, a key observation from the reviewed literature is the clear disparity in dataset availability between university chatbots and those developed in commercial domains. Financial and enterprise systems benefit from large, structured datasets and sustained investment, whereas university systems often rely on limited, manually curated data sources.

This directly affects model performance, evaluation outcomes, and scalability. Despite these limitations, there is growing interest in deploying chatbot systems in universities for real-world student support, particularly in admissions, administration, and engagement services.

Table-3: Comparative Analysis of Chatbot Implementations Across Domains

Author(s)	Bot Name	Domain	Model Type	Dataset Size	Evaluation Metric	Score	Key Findings
Palasundram et al. [61]	N/A	University	Seq2Seq	100 Q-A pairs	BLEU	0.95	Word-based model outperforms character-based model
Khin and Soe [16]	N/A	University	AIML (Rule-based)	970 Q-A pairs	Human Evaluation	N/A	Strong FAQ performance; needs ML enhancement
Chandra and Suyanto [13]	N/A	University	Seq2Seq + Attention	~2900 messages	BLEU	0.4468	Attention improves response quality
Alias et al. [52]	Acabot	University	FP-Growth	537 conversation lines	N-Gram Analysis	N/A	Effective intent pattern extraction in small datasets
Khin and Soe [62]	N/A	University	Seq2Seq + Attention	5000 Q-A pairs	BLEU	0.41	Moderate performance; attention improves output
Cherumanal et al. [30]	Walert	University	Falcon-7B (RAG + Prompting)	106 Q-A + documents	BERTScore, ROUGE-1	0.771 / 0.671	RAG improves inference over structured FAQ
Siragusa & Pirrone [32]	Unipa-GPT	University	GPT (RAG + Fine-tuning)	506 docs + 269 pairs	Human Evaluation	N/A	RAG is more stable than fine-tuning alone

Rana [17]	Eaglebot	University	BERT + BiLSTM Hybrid	500+ Q-A pairs	Accuracy	0.65–0.80	BERT outperforms BiLSTM in FAQ tasks
Nuruzzaman & Hussain [26]	Intellibot	Insurance	Seq2Seq	10,115 Q-A pairs	F1 Score	0.9857	High accuracy due to a large dataset
Yu et al. [28]	AVA	Finance	BERT	22,630 Q-A pairs	F1 Score	0.754	Large dataset improves robustness
Lee et al. [27]	OCS	Medical	BERT	1,176 Q-A pairs	Not Reported	N/A	Promising, but lacks formal evaluation
Kapociute-Dzikiene [77]	N/A	E-commerce	BERT + Seq2Seq	567 Q-A pairs	BLEU	0.513	Limited dataset, but usable practical results

4. DISCUSSION

This review primarily examined the implementation of chatbots within university environments, focusing specifically on systems designed to support student queries outside of direct classroom learning activities. The objective was to understand how different chatbot architectures are applied in higher education and how they compare with solutions used in other industries.

Across the selected studies, eight university-based chatbot implementations were identified as suitable for detailed

analysis. These systems employed a range of techniques, including rule-based approaches, sequence-to-sequence models, and modern transformer-based architectures. However, only a small subset of these implementations utilised large language models (LLMs) or approaches that can be considered state-of-the-art (SOTA), such as GPT or similar transformer-based systems.

When compared to other domains such as customer service, finance, healthcare, and travel, the number of advanced university chatbot deployments remains relatively limited. This suggests that higher education institutions may still be in the early stages of adopting large-scale generative AI technologies for student support services. One possible explanation for this trend is the increasing reliance on commercial “turn-key” chatbot solutions within the education sector. Rather than developing systems in-house, many universities appear to adopt pre-built platforms that

require minimal technical configuration. This is supported by the presence of specialised providers such as Unibuddy and Mainstay, which offer dedicated chatbot services tailored specifically for student recruitment, admissions, and engagement processes.

These platforms reduce the need for universities to develop complex AI systems internally, while also providing integrated analytics, CRM connectivity, and learning management system (LMS) compatibility. However, this convenience may also limit innovation in developing more advanced, research-driven chatbot architectures within academic institutions.

Overall, the findings suggest a clear gap between academic experimentation and real-world deployment in university chatbot systems. While advanced NLP and LLM technologies are widely discussed in research, their practical adoption in higher education remains comparatively slow when contrasted with other commercial sectors.

4.1 Small Datasets

A key finding of this review is the consistently small size of training datasets used in university chatbot implementations. Only one study reported more than 5,000 question-answer pairs, and this was for a rule-based AIML system rather than a deep learning model.

Limited access to domain-specific data remains a major challenge in closed-domain educational chatbot development. This issue has also been highlighted by Ezen-Can [83], Sumikawa et al. [84], Kapociute-Dzikiene [85], and Adamopoulou [36], who note that suitable training datasets are often difficult to obtain or curate in this domain.

In university settings, relevant information is typically scattered across multiple sources such as emails, websites,

and internal documents, making structured dataset creation difficult. As a result, many systems rely on manually created question-answer pairs, limiting scalability and coverage.

Despite this limitation, transformer-based models still demonstrate acceptable performance even with small datasets, largely due to pre-training and transfer learning capabilities.

4.2 Absence Of Working Generative Solutions In The University Sector

Despite recent progress in generative AI, university chatbot deployments remain largely dominated by retrieval-based or rule-based systems. The continued use of conventional question-answering approaches suggests that fully generative solutions are not yet widely adopted in this sector.

This trend may be influenced by several factors. First, many advanced models such as GPT are still accessed primarily through paid APIs, which can limit large-scale institutional deployment. Second, universities often face challenges in accessing structured and high-quality knowledge bases required for effective generative model integration.

In addition, concerns regarding model reliability also play a significant role. Issues such as bias in generated responses [86] and the difficulty in distinguishing AI-generated content from human-written text [87] raise ethical and academic integrity considerations. These factors may contribute to cautious adoption within higher education environments.

Overall, while generative models are technically capable of improving chatbot performance, their practical integration in universities remains limited compared to other industries.

4.3 Use Of Sections And Paragraphs Of Text With Bert Implementations

Rana [17] observes that BERT-based chatbot systems tend to perform better when input knowledge is divided into smaller units, such as sections or paragraphs, instead of using full documents. This structure improves processing efficiency and supports faster and more scalable information retrieval.

Similarly, Yang et al. [7] report that retrieving information at the paragraph level generally produces better results than

using complete articles. Full documents often contain unrelated or less relevant content, which can reduce accuracy and introduce noise during response generation.

4.4 Non-Technical Rationale For Model Selection

Very few studies in the reviewed literature discuss model selection from a cost or implementation perspective. Alias et al. [52] indicate that their custom approach was largely driven by the high cost of commercial chatbot platforms such as Microsoft and Google services. However, their system remains more focused on intent detection rather than full-scale deployment in university environments.

Although research on ChatGPT in education is expanding, limited attention has been given to its application in university administration and support services. Nevertheless, recent studies by Bieletzke [29], Cherumanal et al. [30], Odede and Frommholz [31], and Siragusa and Pirrone [32] demonstrate increasing interest in this direction. These works explore techniques such as prompt engineering, retrieval-augmented generation (RAG), fine-tuning comparisons, and the use of frameworks like LangChain to build LLM-based systems for university use cases beyond classroom teaching.

5. LIMITATIONS

This study has certain limitations due to its focused scope on university chatbot systems for administrative and support-related tasks. While there is extensive research available on ChatGPT and other LLMs in teaching and learning contexts, comparatively fewer studies address their application in areas such as admissions, recruitment, and student services. This limited availability of directly relevant literature may have influenced the depth of comparison.

Furthermore, the increasing use of commercial, ready-to-deploy chatbot platforms in universities reduces the visibility of technical implementation details in academic research. As suggested by Yang and Evans [7], such solutions often operate as closed systems, which can limit research transparency and technical evaluation.

Another important limitation is the restricted availability of large, structured datasets in the higher education domain. Compared to sectors like finance or e-commerce, university chatbot systems typically rely on smaller and less diverse datasets, which can constrain model training, performance evaluation, and cross-study comparability.

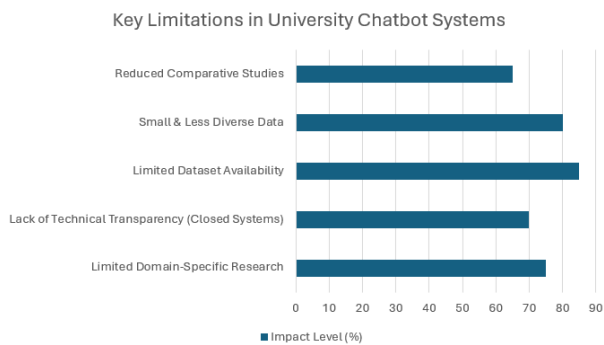


Fig-4: Key Limitations in University Chatbot Systems

6. Conclusions and Future Perspective

This review analysed state-of-the-art chatbot implementations in university environments, specifically focusing on systems designed to handle student queries outside classroom learning.

The study examined model architectures, datasets, and application domains using the classification framework proposed by Hussain [5], including Admissions, Student Recruitment, Administration, and Teaching Support. Among these, Admissions and Student Recruitment remain relatively underexplored areas in chatbot development.

The review covered both traditional rule-based systems and AIML-based approaches, alongside modern AI-driven methods such as retrieval-based and generative models.

While retrieval-based chatbots continue to be widely used for FAQ-style interactions due to their simplicity and reliability, generative models represent a significant advancement in flexibility and conversational capability. However, their deployment still requires careful evaluation due to limitations in reliability and consistency.

The analysis of eight selected university chatbot implementations highlights that educational institutions are still gradually adopting advanced generative AI technologies. This aligns with observations by Yang and Evans [7], who noted that chatbot adoption in education is often constrained by diverse institutional requirements and the availability of easy-to-use development platforms that may still require ongoing maintenance and optimization.

Nevertheless, recent studies such as Cherumal et al. [30] and Siragusa and Pirrone [32] demonstrate growing interest in state-of-the-art LLM-based systems, including comparisons between fine-tuning approaches and Retrieval-Augmented Generation (RAG).

A recurring limitation identified across studies is the scarcity of large, high-quality datasets within universities. Compared to other sectors, educational institutions often lack structured and accessible conversational data. However, as noted by Adamopoulou [36], this challenge is common across most closed-domain chatbot applications.

Future development of university chatbots should focus on leveraging available institutional data sources such as course catalogues, FAQs, email interactions, and CRM systems to construct curated datasets. A hybrid architecture combining retrieval-based methods for structured FAQs and generative models for complex queries appears to be the most practical solution. However, effective implementation requires careful data preparation, especially for structured academic information such as fees, course duration, and curriculum details.

The rapid evolution of large language models (LLMs) since the introduction of ChatGPT in 2022 has significantly influenced research directions. Earlier transformer-based models such as BERT and its variants (e.g., RoBERTa, DistilBERT, and XLNet) remain relevant, while SBERT offers improved semantic embedding capabilities for FAQ-based systems [88]–[91].

Finally, emerging frameworks such as LangChain, along with techniques like prompt engineering, ensemble learning, and multi-agent orchestration systems (e.g., AutoGen and CrewAI), are expected to play a key role in next-generation university chatbot development. These tools enable more scalable, intelligent, and context-aware conversational systems, making them strong candidates for future research and deployment in higher education environments.

REFERENCES

- [1] J. Weizenbaum, "ELIZA—a computer program for study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, p. 36, 1966.
- [2] W. Maroengsit, P. Piyakulpinyo, K. Phonyiam, C. Pongnumkul, and P. Chaovalit, "A survey on evaluation methods for chatbots," in *Proc. 7th Int. Conf. Information and Education Technology (ICIET)*, 2019, pp. 111–119.

- [3] M. J. van der Goot and T. Pilgrim, "Exploring age differences in motivations for and acceptance of chatbot communication in a customer service context," in *Chatbot Research and Design*, Springer, 2020.
- [4] D. Baidoo-Anu and L. O. Ansah, "Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT," *Journal of AI*, vol. 7, no. 1, pp. 52–62, 2023.
- [5] A. Hussain, "5 use cases for a campus chatbot," *Chatbots Magazine*, 2018.
- [6] S. Cunningham-Nelson et al., "A review of chatbots in education: practical steps forward," in *Proc. AAEE 2019*, Engineers Australia, 2019.
- [7] S. Yang and C. Evans, "Opportunities and challenges in using AI chatbots in higher education," in *Proc. 3rd Int. Conf. Education and E-Learning*, ACM, 2019, pp. 79–83.
- [8] P. A. Smutny and P. Schreiberova, "Chatbots for learning: a review of educational chatbots for Facebook Messenger," *Computers & Education*, 2020.
- [9] S. L. Mendez et al., "Chatbots: a tool to supplement faculty mentoring," *International Journal of Doctoral Studies*, vol. 15, no. 1, p. 373, 2020.
- [10] M. Abadi et al., "TensorFlow: large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2015.
- [11] L. C. Page, J. Lee, and H. Gehlbach, "Conditions under which college students can be responsive to nudging," *EdWorkingPapers*, 2020.
- [12] M. Choque-Díaz, J. Armas-Aguirre, and P. Shiguihara-Juárez, "Cognitive technology model to enhance academic support services with chatbots," in *IEEE INTERCON*, 2018.
- [13] Y. W. Chandra and S. Suyanto, "Indonesian university admission chatbot using Seq2Seq model," *Procedia Computer Science*, vol. 157, pp. 367–374, 2019.
- [14] H. A. Santoso et al., "DINA chatbot for university admission services," in *IEEE ICAT 2018*, 2018.
- [15] J. B. Aujogue and A. Aussem, "Hierarchical recurrent attention networks for education chatbots," in *IJCNN*, IEEE, 2019.
- [16] N. N. Khin and K. M. Soe, "University chatbot using AIML," in *IEEE ICCA*, 2020.
- [17] M. Rana, "Eaglebot: chatbot-based QA system using BERT," *Electronic Theses and Dissertations*, 2019.
- [18] R. Singh, "AI and chatbots in education: what does the future hold?," *Chatbots Magazine*, 2018.
- [19] L. Satow, "Chatbots as teaching assistants," *SAP Blogs*, 2017.
- [20] W. Nwankwo, "Interactive advising with bots in education," *American Journal of Operational Management*, vol. 3, no. 1, pp. 6–21, 2018.
- [21] A. K. Goel and L. Polepeddi, "Jill Watson: a virtual teaching assistant," Georgia Institute of Technology, 2016.
- [22] L. K. Lee et al., "Multiplatform chatbot as online tutor," in *IEEE ISET*, 2020.
- [23] C. W. Okonkwo and A. Ade-Ibijola, "Chatbot applications in education: systematic review," *Computers & Education: AI*, vol. 2, 2021.
- [24] A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, no. 236, pp. 433–460, 1950.
- [25] N. T. Thomas, "E-business chatbot using AIML and LSA," in *IEEE ICACCI*, 2016.
- [26] M. Nuruzzaman and O. K. Hussain, "IntelliBot: insurance chatbot," *Knowledge-Based Systems*, vol. 196, 2020.
- [27] J. H. Lee et al., "Ophthalmic chatbot system using BERT," in *IEEE IMCOM*, 2021.
- [28] S. Yu, Y. Chen, and H. Zaidi, "Financial chatbot using transformers," *arXiv preprint*, 2020.
- [29] S. Bielezke, "AI chatbot integration in campus systems," in *EDULEARN23*, 2023.
- [30] S. P. Cherumanal et al., "Walert: LLM-powered chatbot for universities," *arXiv preprint arXiv:2401.07216*, 2024.
- [31] J. Odede and I. Frommholz, "JayBot: LLM-based admission chatbot," in *CHIIR*, 2024.
- [32] I. Siragusa and R. Pirrone, "Unipa-GPT case study," in *NL4AI Workshop*, 2023.
- [33] R. Agarwal and M. Wadhwa, "Review of chatbot design techniques," *SN Computer Science*, vol. 1, no. 5, 2020.
- [34] AIML Foundation, "AIML framework," Available: <http://www.aiml.foundation/>
- [35] B. R. Ranoliya, N. Raghuvanshi, and S. Singh, "University FAQ chatbot," in *IEEE ICACCI*, 2017.
- [36] E. Adamopoulou and L. Moussiades, "Chatbots: history, technology, and applications," *Machine Learning and Applications*, vol. 2, 2020.
- [37] C. Danescu-Niculescu-Mizil and L. Lee, "Chameleons in imagined conversations," *ACL Workshop*, 2011.
- [38] C. E. Pratt, *My Bonny Lies Over the Ocean*, 1881.
- [39] M. Honnibal, "spaCy NLP library," 2015.
- [40] S. Bird et al., *Natural Language Processing with Python*, O'Reilly, 2009.
- [41] K. Welbers, W. Van Atteveldt, and K. Benoit, "Text analysis in R," *Communication Methods and Measures*, 2017.
- [42] H. P. Luhn, "Automatic creation of literature abstracts," *IBM Journal*, 1958.
- [43] M. F. Porter, "An algorithm for suffix stripping," *Program*, 1980.
- [44] M. F. Porter, "Snowball stemming algorithm," 2001.
- [45] A. G. Jivani, "Stemming algorithms comparative study," *Int. J. Comp. Tech.*, 2011.
- [46] A. Raju et al., "Contextual language model adaptation," *arXiv*, 2018.
- [47] S. Qaiser and R. Ali, "TF-IDF text mining," *International Journal of Computer Applications*, 2018.
- [48] T. Mikolov et al., "Word2Vec," *arXiv preprint*, 2013.
- [49] J. Pennington, R. Socher, and C. Manning, "GloVe embeddings," *EMNLP*, 2014.

[50] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, 1997.

[51] A. Mislevics et al., "Chatbot implementation framework," 2018.

[52] S. Alias et al., "FP-growth chatbot intent discovery," IEEE ICETAS, 2019.

[53] S. Mathur and D. Lopez, "Neural conversational model," *Concurrency and Computation*, 2019.

[54] P. Rajpurkar et al., "SQuAD dataset," *arXiv preprint arXiv:1606.05250*, 2016.

[55] J. Han, J. Pei, and Y. Yin, "Frequent pattern mining," *ACM SIGMOD*, 2000.

[56] S. Alias et al., "Sequential pattern growth model," *Pattern Analysis and Applications*, 2018.

[57] I. Sutskever, O. Vinyals, and Q. V. Le, "Seq2Seq learning," *NeurIPS*, 2014.

[58] S. Shalev-Shwartz et al., "Failures of gradient-based learning," 2017.

[59] S. Hochreiter, "Vanishing gradient problem," *Neural Computation*, 1998.

[60] D. Chen et al., "BiLSTM attention models," arXiv, 2017.

[61] K. Palasundram et al., "Seq2Seq chatbot evaluation," *IJET*, 2019.

[62] N. N. Khin and K. M. Soe, "Seq2Seq university chatbot," IEEE O-COCOSDA, 2020.

[63] K. Papineni et al., "BLEU metric," 2002.

[64] A. Vaswani et al., "Attention is all you need," *NeurIPS*, 2017.

[65] J. Uszkoreit, "Transformer architecture overview," Google AI Blog, 2017.

[66] J. Devlin et al., "BERT model," arXiv preprint arXiv:1810.04805, 2018.

[67] A. Radford et al., "GPT pretraining," OpenAI, 2018.

[68] J. Achiam et al., "GPT-4 technical report," arXiv:2303.08774, 2023.

[69] [69] L. Floridi and M. Chiriatti, "GPT-3 analysis," *Mind & Machines*, 2020.

[70] E. M. Bender et al., "On the dangers of stochastic parrots," *FACCT*, 2021.

[71] Y. Lappalainen and N. Narayanan, "Library chatbot using GPT API," *Journal of Web Librarianship*, 2023.

[72] E. Almazrouei et al., "Falcon LLMs," arXiv:2311.16867, 2023.

[73] P. Lewis et al., "Retrieval-augmented generation," *NeurIPS*, 2020.

[74] Google Team, "Gemini multimodal models," arXiv:2312.11805, 2023.

[75] Y. Bai et al., "Constitutional AI," arXiv:2212.08073, 2022.

[76] H. Touvron et al., "LLaMA models," arXiv:2302.13971, 2023.

[77] J. Kapočiūtė-Dzikiėnė, "Low-data generative chatbot," *Applied Sciences*, 2020.

[78] L. Derczynski, "F1 score in NLP evaluation," 2016.

[79] C. Y. Lin and E. Hovy, "ROUGE metric," *NAACL*, 2003.

[80] T. Zhang et al., "BERTScore," arXiv:1904.09675, 2019.

[81] C. W. Liu et al., "Dialogue evaluation metrics study," arXiv:1603.08023, 2016.

[1] A. S. Lokman and M. A. Ameen, *Modern Chatbot Systems*, Springer, 2018.

[2] A. Ezen-Can, "BERT vs LSTM for small corpora," arXiv:2009.05451, 2020.

[3] Y. Sumikawa et al., "FAQ dataset creation," Springer, 2020.

[4] J. Kapočiūtė-Dzikiėnė, "Small data chatbot training," *Applied Sciences*, 2020.

[5] D. M. Korngiebel and S. D. Mooney, "GPT-3 healthcare risks," *npj Digital Medicine*, 2021.

[6] T. B. Brown et al., "Few-shot learning with LLMs," arXiv:2005.14165, 2020.

[7] Y. Liu et al., "RoBERTa model," arXiv:1907.11692, 2019.




[8] V. Sanh et al., "DistilBERT," arXiv:1910.01108, 2019.


[9] Z. Yang et al., "XLNet," *NeurIPS*, 2019.

[10] N. Reimers and I. Gurevych, "Sentence-BERT," arXiv:1908.10084, 2019.

[11] X. Dong et al., "Survey on ensemble learning," *Frontiers of Computer Science*, 2020.

ABOUT THE AUTHORS

	<p>Pranshul Sharma Final Year B.Tech (CSE) student at Raja Balwant Singh Engineering Technical Campus, Bichpuri, Agra Interested in Artificial Intelligence, Machine Learning, and Web Development. Currently working on a final year project in the field of A.I. Email: pranshulbharadwaj@gmail.com</p>
	<p>Dharmendra Solanki Final Year B.Tech (CSE) student at Raja Balwant Singh Engineering Technical Campus, Bichpuri, Agra. Interested in Cyber Security, Cloud Computing, and Data Science. Currently working on a final year project in the field of A.I. Email: dharmendrasolanki736@gmail.com</p>
	<p>Sumit Khirwar Final Year B.Tech (CSE) student at Raja Balwant Singh Engineering Technical Campus, Bichpuri, Agra. Interested in Internet of Things (IoT), Mobile App Development, and Software Engineering. Currently working on a final year project in the field of A.I. Email: sumitkhirwar01@gmail.com</p>

	<p>Anshul Kumar Singh Assistant Professor (CSE) at Raja Balwant Singh Engineering Technical Campus, Bichpuri, Agra. Email: akrajawat20@gmail.com</p>
---	---