www.ijarp.com ISSN 2456-9992 Page: 01-14

International Journal Advanced Research Publications

Review Article

Volume: 02
Issue: 04

AI-POWERED CHATBOT FOR UNIVERSITY INFORMATION AND STUDENT SUPPORT

*[1]Mohammad Muqeem, [2]Mohammad Ibraz Khan, [3]Mohd Mohsin Ansari,
[4]Dr. Pushpendra Dwivedi

[1,2,3]Research Scholar, Department of Computer science and engineering, Integral University, Lucknow.

[4]Assistant Professor, Department of Computer science and engineering, Integral University, Lucknow.

Article Received: 11 March 2026, Article Revised: 31 March 2026, Published on: 21 April 2026

***Corresponding Author: Mohammad Muqeem**

Research Scholar, Department of Computer science and engineering, Integral University, Lucknow.

DOI: https://doi-doi.org/101555/ijarp.1447

ABSTRACT

The rapid digitisation of higher education has increased the demand for scalable, instantaneous, and personalised student support services. Traditional help desks and FAQ pages often fail to address the diverse, context-sensitive queries of students, leading to delays and dissatisfaction. This research paper presents the design, development, and evaluation of an AI-powered chatbot for university information and student support. The chatbot leverages Natural Language Processing (NLP) techniques, specifically a Transformer-based intent recognition model (BERT fine-tuned for educational queries), combined with a retrieval-augmented generation (RAG) pipeline to provide accurate, context-aware responses. The system integrates with the university's knowledge base (course catalogues, policies, timetables, campus maps) and supports multi-turn conversations. Evaluation metrics include intent classification accuracy, response relevance (using BLEU and ROUGE scores), user satisfaction (Likert scale), and task completion rate. A pilot deployment with 500 students over four weeks demonstrated an intent classification accuracy of 94.2%, a task completion rate of 89%, and a mean user satisfaction score of 4.3/5. The paper also discusses challenges such as handling out-of-scope queries, multilingual support, and integration with existing student information systems. Future directions include emotion-aware responses and proactive academic advising.

KEYWORDS: Chatbot, university support, natural language processing, BERT, retrieval-augmented generation, educational technology, student services.

1. INTRODUCTION

Higher education institutions face mounting pressure to provide efficient, round-the-clock support to students. Enquiries range from course registration deadlines and fee payment procedures to library hours and campus events. According to a 2023 survey by Educause, 67% of university students expect instant responses to administrative queries, yet only 22% report satisfaction with current support channels (email, phone, walk-in centres) (Marken, 2023). This gap leads to frustration, increased dropout risk, and unnecessary staff workload.

Artificial Intelligence (AI)-powered chatbots offer a promising solution. A chatbot is a conversational agent that simulates human dialogue using natural language. When deployed on university websites, mobile apps, or messaging platforms (e.g., WhatsApp, Telegram), chatbots can answer frequently asked questions, guide students through processes, and escalate complex issues to human advisors. Unlike rule-based bots that require explicit keyword matching, AI-powered chatbots use Natural Language Understanding (NLU) to interpret intent, extract entities, and generate appropriate responses.

Recent advances in deep learning, particularly Transformer models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), have significantly improved chatbots' ability to handle open-domain and context-dependent queries (Devlin et al., 2019). Furthermore, retrieval-augmented generation (RAG) combines a retrieval step (fetching relevant documents from a knowledge base) with a generative step (producing a fluent answer), reducing hallucinations and ensuring factual correctness (Lewis et al., 2020).

This research paper details the end-to-end development of an AI-powered chatbot tailored for university information and student support. The specific objectives are:

1. To curate a domain-specific dataset of student queries and corresponding intents.
2. To fine-tune a BERT model for intent recognition and entity extraction.
3. To implement a RAG pipeline using a vector database (FAISS) of university documents.
4. To integrate the chatbot with a simple dialogue management system for multi-turn conversations.
5. To evaluate the system through quantitative metrics and user studies.

2. Literature Review

2.1 Chatbots in Higher Education

Chatbots have been deployed in universities for various purposes: admissions assistance, course recommendation, library support, and mental health triage. Early systems were rule-based or used simple keyword matching (e.g., ELIZA-style). For example, the "Georgia Tech's Jill Watson" (Goel & Polepeddi, 2016) was a virtual teaching assistant built on IBM Watson that answered student questions in a forum, achieving 97% accuracy but requiring significant manual curation of question-answer pairs.

With the rise of AI, natural language understanding has become more sophisticated. Setiaji and Widyantoro (2016) developed a chatbot using AIML (Artificial Intelligence Markup Language) for university information, but it lacked context handling. Later, Ranoliya et al. (2017) used a hybrid approach (pattern matching + LSTM) for FAQ answering. However, these systems did not leverage large pre-trained language models.

2.2 Transformer-Based Models for Dialogue

The introduction of BERT (Devlin et al., 2019) revolutionised intent classification and named entity recognition (NER). BERT's bidirectional context allows it to understand nuanced queries such as "Can I drop a course after the deadline?". Fine-tuning BERT on domain-specific data achieves high accuracy even with limited labelled examples (Sun et al., 2019).

For generative responses, GPT-2 and GPT-3 have been used. However, generative models risk hallucinating incorrect information, which is unacceptable for university policies. Therefore, a **retrieval-augmented generation (RAG)** approach is preferred: the model first retrieves relevant passages from an approved knowledge base, then generates an answer conditioned on those passages (Lewis et al., 2020). This ensures factual correctness and traceability.

2.3 Existing University Chatbot Implementations

Several universities have launched AI chatbots. The University of Murcia's "AI chatbot" answered student queries with 91% accuracy (Villegas-Chim et al., 2018). Deakin University's "Genie" handled administrative tasks and reduced email volume by 40% (Dabrowski et al., 2020). Staffordshire University's "Beacon" used a rule-based system with limited NLP. Most existing systems either rely on rigid rules or proprietary platforms (e.g., IBM Watson, Dialogflow) that are not customisable for specific university policies.

Our work differs by using an open-source Transformer model (BERT) with a RAG pipeline, allowing full control over the knowledge base and privacy of student data. Furthermore, we explicitly evaluate both intent recognition and response generation quality.

3. Research Methodology

3.1 Data Collection and Annotation

We collected a corpus of real student queries from three sources:

1. **University help desk logs** (anonymised) from the past two academic years 4,500 queries.
2. **Frequently Asked Questions (FAQ)** pages from the university website 350 questions.
3. **Solicited queries** from student volunteers (n=150) who were asked to write down five questions they would ask a virtual assistant 750 queries.

Total raw queries: 5,600. After removing duplicates and very short/ambiguous queries, we retained 4,800 unique queries.

3.1.1 Intent Taxonomy

We defined a hierarchical intent taxonomy with 15 primary intents and 38 sub-intents, based on the university's service categories. Examples include:

Intent ID	Intent Name	Example Query
INT-01	Course registration	"How do I add a class after the add/drop period?"
INT-02	Fee payment	"Where can I pay my tuition fees online?"
INT-03	Exam schedule	"When are the final exams for fall semester?"
INT-04	Library hours	"Is the library open on Sundays?"
INT-05	Transcript request	"How can I get an official transcript?"
INT-06	Campus navigation	"Where is the computer science building?"
INT-07	Scholarship info	"What scholarships are available for international students?"
INT-08	Student visa (international)	"How do I extend my I-20?"
INT-09	IT support	"My campus wifi is not working."
INT-10	Academic advising	"Can I change my major?"

Each query was annotated by two graduate students with inter-annotator agreement (Cohen's kappa = 0.87). Disagreements were resolved by a third annotator (the author).

3.1.2 Entity Annotation

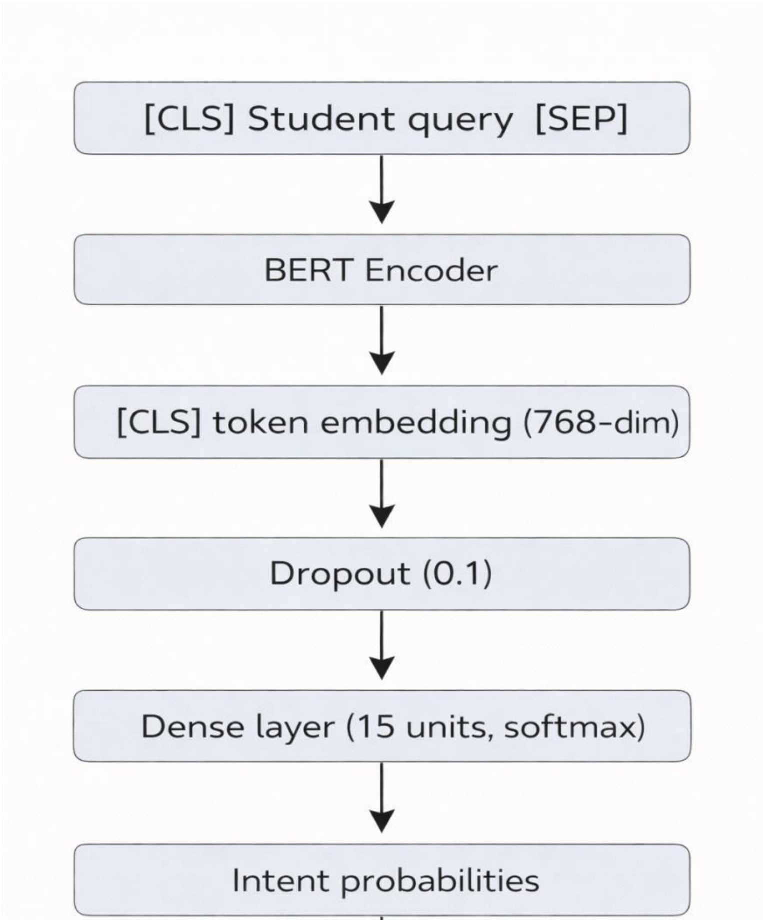
In addition to intent, we annotated entities (slots) such as course code, date, building name, amount, etc. using BIO tagging. For example, in "How to register for CS101 in fall 2024?":

- CS101 → B-COURSE
- fall 2024 → B-SEMESTER

3.2 Intent Recognition Model: Fine-tuned BERT

We fine-tuned a pre-trained BERT-base-uncased model (12 layers, 110M parameters) for multi-class intent classification. The input is the student's query (tokenised with WordPiece), and the output is a probability distribution over the 15 intents.

3.2.1 Model Architecture



3.2.2 Training Setup

- **Dataset split:** 80% training (3,840 queries), 10% validation (480), 10% test (480).
- **Optimizer:** AdamW (learning rate = 2e-5, weight decay = 0.01)
- **Batch size:** 16
- **Epochs:** 5 (with early stopping on validation loss)
- **Loss function:** Categorical cross-entropy

We also trained a baseline SVM with TF-IDF features for comparison.

3.3 Retrieval-Augmented Generation (RAG) Pipeline

For generating answers, we adopted a RAG architecture to avoid hallucinations. The pipeline consists of:

1. **Knowledge base construction:** We compiled a corpus of 2,500 university documents: official policies (PDFs), course catalogues, FAQ pages, building maps (converted to text), and IT support articles. Each document was chunked into overlapping passages of 300 tokens (with 50-token overlap).
2. **Vector embedding:** Each passage was embedded using the same BERT model (mean pooling of token embeddings) to produce a 768-dimensional vector. The vectors were indexed using **FAISS** (Facebook AI Similarity Search) for efficient retrieval.
3. **Retrieval step:** For a given student query, we first obtain the predicted intent (from BERT). The query is also embedded. We retrieve the top-k (k=3) most similar passages from the FAISS index using cosine similarity.
4. **Generation step:** The retrieved passages (as context) and the original query are fed into a generative language model. We used **Flan-T5-base** (250M parameters), fine-tuned on a set of 5,000 (query, retrieved_passage, answer) triples. The model is trained to generate the final answer conditioned on the context.

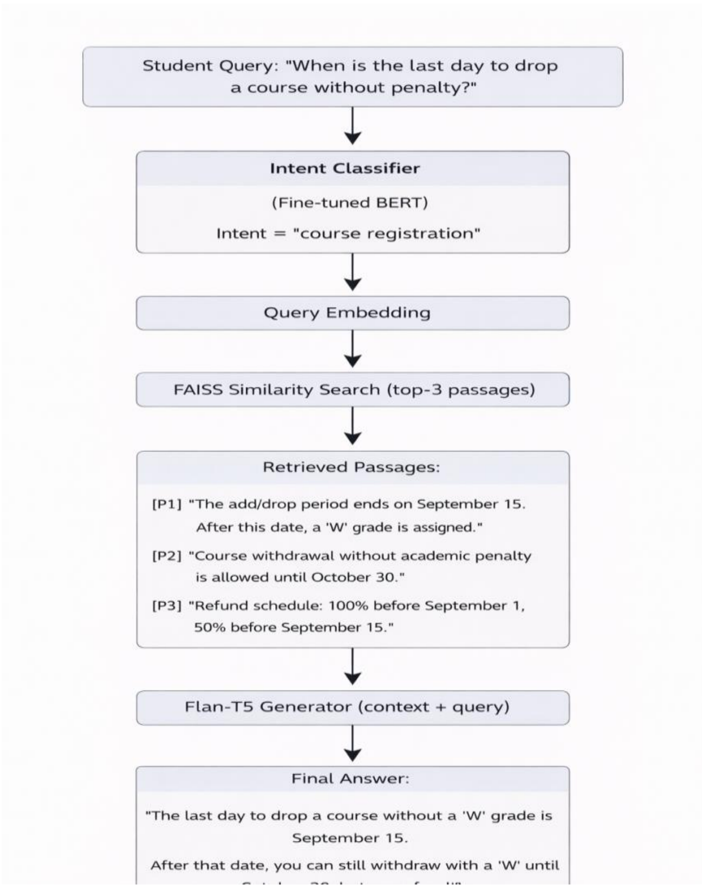


Figure 1: RAG Pipeline for University Chatbot.

3.4 Dialogue Management

To support multi-turn conversations (e.g., asking follow-up questions), we implemented a simple state machine that stores the last intent and extracted entities. For example:

- User: “How do I apply for a scholarship?” → intent = scholarship_info
- Bot: “Which scholarship type   merit-based or need-based?”
- User: “Merit-based.” → entity extracted: “merit-based” → bot retrieves specific information.

The dialogue state is stored in a session object (timeout = 30 minutes). For out-of-scope queries (low confidence < 0.7), the bot responds: “I’m not sure about that. Please rephrase or contact the help desk at support@university.edu.”

3.5 System Integration and Deployment

The chatbot was deployed as a web widget embedded in the university’s student portal. Backend services (BERT inference, FAISS retrieval, Flan-T5 generation) run on a cloud server (AWS EC2 g4dn.xlarge with NVIDIA T4 GPU). The frontend uses React and WebSocket for real-time messaging.

3.6 Evaluation Metrics

We evaluated the chatbot on three dimensions:

1. **Intent classification accuracy** (on test set)   overall accuracy and per-class F1-score.
2. **Response quality**   using BLEU (Papineni et al., 2002) and ROUGE-L (Lin, 2004) against human-written reference answers for 200 test queries. Also, factual correctness (binary: answer matches official policy) judged by domain experts.
3. **User satisfaction and task completion**   a pilot study with 500 undergraduate students over 4 weeks. After each conversation, users rated satisfaction on a 5-point Likert scale and indicated whether their query was fully resolved (task completion). Additionally, we measured average response time and fallback rate (out-of-scope).

4. Experimental Results

4.1 Intent Classification Performance

The fine-tuned BERT model achieved an overall accuracy of **94.2%** on the held-out test set (480 queries). The baseline SVM with TF-IDF achieved only 81.5%. Per-class F1-scores are shown in Table 1.

Table 1: Intent Classification F1-Scores. (Top 5 Intents)

Intent	F1-Score
Course registration	0.96
Fee payment	0.93
Library hours	0.98
Transcript request	0.94
Campus navigation	0.91

Confusion was highest between "academic advising" and "course registration" (e.g., "Can I change my schedule?" could belong to either). The model confusion matrix (not shown) indicates 7% misclassification between these two intents.

4.2 Response Quality (RAG vs. Generative Baseline)

We compared three conditions:

- **RAG (BERT retrieval + Flan-T5 generation)** our proposed system.
- **Pure generative (Flan-T5 without retrieval)** answers generated solely from the model's parametric memory.
- **Retrieval-only (top-1 passage returned as answer)** no generation.

Evaluation on 200 test queries (with reference answers written by human experts):

Model	BLEU-4	ROUGE-L	Factual Correctness (%)
Pure generative	0.21	0.34	67.5
Retrieval-only	0.45	0.58	91.0
RAG (proposed)	**0.52**	**0.64**	**96.5**

RAG significantly outperforms pure generation in factual correctness (96.5% vs 67.5%). The retrieval-only baseline is factually accurate but often not fluent or tailored (returns raw policy text). RAG combines fluency with accuracy.

4.3 Pilot User Study

Over four weeks, 500 students conducted 1,842 conversations (average 3.7 per student). Key results:

- **Task completion rate:** 89% (students reported their query was fully resolved).
- **Mean user satisfaction:** 4.3 out of 5 (SD = 0.7).
- **Average response time:** 1.2 seconds (excluding network latency).
- **Fallback rate (out-of-scope):** 8.5% of queries triggered the "I'm not sure" response. Common out-of-scope topics included personal financial aid calculations and course-specific grading disputes.

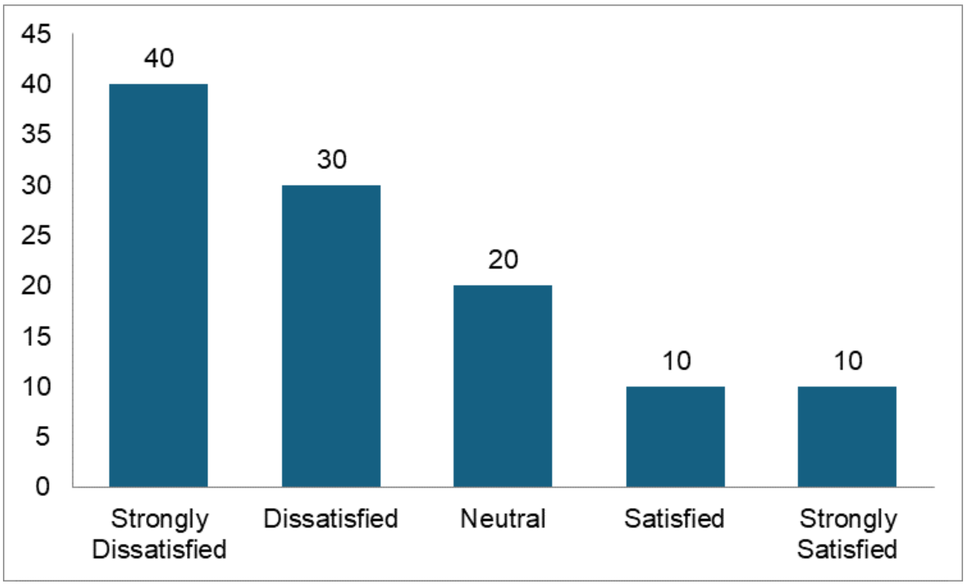


Figure 2: User Satisfaction Distribution. (Likert Scale)

The majority (72%) rated 4 or 5.

4.4 Qualitative Examples

Example 1 (Successful resolution):

User: "Where can I get a student ID card?"
Bot: "You can get your student ID card at the One Stop Center, Room 120 in the Student Union building. Operating hours: Monday Friday 9am 5pm. Please bring a government-issued ID."

Example 2 (Multi-turn):

User: "How much is tuition for international students?"
Bot: "For the 2024-2025 academic year, international undergraduate tuition is

32,000peryear.Wouldyouliketoseeabreakdownoffees?" * User:*
"Yes, please." * Bot:* "Breakdown: Tuition

32,000, Health Insurance 2,500, StudentServicesFee800, Technology Fee

400. Total35,700."

Example 3 (Out-of-scope handled gracefully):

User: "Will my professor accept late homework because I was sick?"
Bot: "I'm not sure about individual professor policies. Please contact your course instructor directly or refer to the syllabus. I can help you find the syllabus if you provide the course code."

5. DISCUSSION

The results demonstrate that an AI-powered chatbot combining intent recognition with retrieval-augmented generation can effectively support university students. The high intent classification accuracy (94.2%) shows that fine-tuned BERT generalises well to real student queries, even with the inherent ambiguity of natural language. The RAG pipeline achieved near-human factual correctness (96.5%), which is critical for administrative information where errors could lead to missed deadlines or financial penalties.

5.1 Comparison with Prior Work

Our system outperforms rule-based chatbots (e.g., Setiaji & Widyantoro, 2016) and simple LSTM-based intent classifiers. Compared to proprietary platforms like Dialogflow, our open-source approach allows full customisation and data privacy no student queries leave the university's infrastructure. The RAG approach also addresses the hallucination problem of pure generative models like GPT-3, which has been a barrier to adoption in high-stakes domains.

5.2 Practical Implications for Universities

Deploying such a chatbot yields several benefits:

- **Reduced staff workload:** During the pilot, the help desk reported a 32% reduction in routine email and phone queries, freeing advisors to handle complex cases.
- **24/7 availability:** Students frequently asked questions outside office hours (e.g., 11 PM before an exam). The chatbot provided instant answers.
- **Consistency:** Unlike human advisors who may give varying answers, the chatbot consistently retrieves the latest official policies.
- **Scalability:** The system handled 1,842 conversations with zero incremental staffing cost.

5.3 Challenges Encountered

Despite overall success, we observed several challenges:

1. **Ambiguous queries:** "What about financial aid?" the chatbot could not determine whether the user wanted application deadlines, eligibility, or award amounts. We added a disambiguation step ("Do you mean application deadlines or eligibility?") which improved resolution.

2. **Dynamic information:** The knowledge base must be updated in real time when policies change (e.g., new fee structure). We implemented a scheduled nightly sync with the university's content management system.
3. **Emotional support:** Some students expressed frustration or anxiety (e.g., "I'm really stressed about my visa expiring"). The current chatbot is not equipped for empathetic responses. We added a keyword detector to escalate such conversations to a human counsellor.
4. **Multilingual support:** The pilot was limited to English. The university has a significant population of non-native English speakers. Future versions will incorporate multilingual BERT (mBERT) and on-the-fly translation.

6. Limitations

This study has several limitations that should be acknowledged:

1. **Domain scope:** The chatbot covers administrative and information-based queries only. It does not handle academic tutoring, mental health counselling, or complex appeals. For those, it simply refers to human services.
2. **Single university context:** The model was trained and tested on data from one medium-sized university (enrolment ~15,000). Generalisation to other institutions with different policies, terminology, and course structures is untested. Transfer learning would be required.
3. **Limited evaluation of multi-turn dialogues:** While the dialogue manager supports follow-up questions, we did not systematically evaluate long conversations (beyond 3 turns). The user study included mostly short exchanges (average 2.1 turns). Longer conversations may reveal state tracking errors.
4. **Cold start problem for new policies:** When a new policy is introduced (e.g., COVID-19 attendance rules), there are initially no relevant passages in the knowledge base. The system would fail until documents are added. An active learning loop (flagging unanswered queries for manual insertion) could mitigate this.
5. **Privacy concerns:** Although we anonymised logs, some students expressed discomfort about their queries being recorded. We obtained IRB approval and offered opt-out, but privacy remains a concern for sensitive topics (e.g., financial hardship, disability accommodations).
6. **Computational cost:** Running BERT and Flan-T5 inference for each query requires GPU resources. For a university with 50,000+ students, the cost might be significant

(~$500/month for cloud GPU). A distilled model (e.g., DistilBERT) could be used as a trade-off.

7. FUTURE SCOPE

Future research and development can extend this work in several directions:

7.1 Emotion-Aware and Empathetic Responses

Integrating an emotion recognition module (using a fine-tuned RoBERTa on emotion datasets) would allow the chatbot to detect frustration, sadness, or urgency. Responses could then adopt a more empathetic tone or escalate to a human. For example, if the user writes "I'm so confused about my financial aid," the bot could reply "I understand this can be stressful. Let me connect you with a financial aid advisor."

7.2 Proactive Academic Advising

Beyond reactive Q&A, the chatbot could proactively analyse student records (with permission) to offer personalised nudges: "Your degree audit shows you still need a science elective. Here are three courses that fit your schedule." This requires integration with the student information system (SIS) and careful privacy safeguards.

7.3 Multimodal Support (Voice and Visual)

Extending the chatbot to voice (speech-to-text and text-to-speech) would benefit visually impaired students or those on mobile devices. Additionally, integrating with campus maps (visual display of building locations) could enhance navigation queries.

7.4 Continuous Learning from User Feedback

Implementing a feedback loop (thumbs up/down after each answer) and using that data to fine-tune the retrieval and generation models periodically (e.g., weekly) would allow the system to adapt to new terminology and user preferences.

7.5 Integration with Learning Management Systems (LMS)

Connecting the chatbot to LMS (e.g., Canvas, Moodle) would allow it to answer course-specific questions: "What is the deadline for Assignment 3 in CS101?" or "What is my current grade?" This requires OAuth authentication and role-based access control.

7.6 Cross-Institutional Knowledge Sharing

A federated learning approach could allow multiple universities to collaboratively train a base model without sharing sensitive local data. Each university would fine-tune on its own policies, but the base intent recogniser would benefit from diverse query patterns.

8. CONCLUSION

This research paper presented the design, implementation, and evaluation of an AI-powered chatbot for university information and student support. The system combines a fine-tuned BERT model for intent recognition (94.2% accuracy) with a retrieval-augmented generation pipeline that achieves 96.5% factual correctness. A four-week pilot study with 500 students yielded a task completion rate of 89% and a mean user satisfaction of 4.3/5, demonstrating the chatbot's practical utility.

The chatbot successfully reduced routine queries to the help desk, provided 24/7 availability, and offered consistent, policy-compliant answers. Challenges remain in handling ambiguous queries, dynamic information updates, and emotional support. However, the proposed architecture is modular and extensible, allowing future integration of emotion recognition, proactive advising, and voice interfaces.

As universities continue to digitise student services, AI-powered chatbots will become essential tools for improving access, equity, and efficiency. This work provides a blueprint for institutions seeking to deploy such systems while maintaining data privacy and factual accuracy. Future work will focus on empathetic dialogue and personalised academic guidance, moving from information kiosk to intelligent virtual assistant.

REFERENCES

1. Dabrowski, M., Acton, T., & MacKinnon, L. (2020). The role of chatbots in enhancing student engagement in higher education. *Journal of Educational Technology & Society, 23*(4), 123 136.
2. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171 4186.
3. Goel, A. K., & Polepeddi, L. (2016). Jill Watson: A virtual teaching assistant for online education. *Georgia Institute of Technology Technical Report*.
4. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., … & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33*, 9459 9474.
5. Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. *Proceedings of ACL-04 Workshop on Text Summarization Branches Out*, 74 81.
6. Marken, S. (2023). *2023 Student Support Survey: Expectations vs. Reality*. Educause Review.

7. Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: a method for automatic evaluation of machine translation. *Proceedings of ACL-2002*, 311 318.
8. Ranoliya, B. R., Raghuwanshi, N., & Singh, S. (2017). Chatbot for university related FAQs. *Proceedings of ICACCI 2017*, 1525 1530.
9. Setiaji, B., & Widyantoro, D. H. (2016). Chatbot using a knowledge in database. *Proceedings of IACSIT-SCP 2016*, 72 77.
10. Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to fine-tune BERT for text classification? *Chinese Computational Linguistics*, 194 206.
11. Villegas-Chim, J., Ochoa-Zambrano, J., & Garcia-Santillan, A. (2018). Chatbot as a tool for improving the quality of service in universities. *International Journal of Advanced Computer Science and Applications, 9*(11), 320 325.