AI-Driven FAQ Chatbot for Academic Information Retrieval in Higher Education: A Comparative Evaluation of TF-IDF and Sentence-BERT

Nurul Shazrieana binti Mad Said[1], *Mohd Shahril Izuan Mohd Zin[2]

[1]Faculty of Electronics and Computer Technology and Engineering (FTKEK), Universiti Teknikal Malaysia Melaka (UTeM)

[2]School of Technical Foundation and Diploma Studies (UTeM STEP), Universiti Teknikal Malaysia Melaka (UTeM)

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.101000541

Received: 14 October 2026; Accepted: 20 October 2026; Published: 27 October 2026

ABSTRACT

Universities increasingly rely on digital platforms to provide students with timely academic and administrative information, yet information distributed across faculty websites and handbooks can still generate repetitive enquiries and dependence on staff availability. This study develops and evaluates a closed-domain artificial intelligence FAQ chatbot for academic information retrieval in a Malaysian higher-education context. A verified knowledge base containing 70 question-answer pairs was constructed from official faculty web resources and a student handbook. Two retrieval approaches were compared using 25 paraphrased evaluation questions: a Term Frequency-Inverse Document Frequency (TF-IDF) baseline and a Sentence-BERT (SBERT) semantic retrieval model using all-MiniLM-L6-v2 embeddings and cosine similarity. SBERT achieved 56.00% retrieval accuracy compared with 40.00% for TF-IDF, representing a 16-percentage-point improvement, although standalone inference latency increased from 16.83 ms to 110.56 ms. The selected SBERT model was then integrated into a hybrid conversation-management architecture that combines finite-state dialogue routing, keyword-based topic detection, semantic retrieval, and a 0.60 similarity threshold for fallback handling. In a larger bilingual stress test of 315 queries, the deployed system achieved 73.65% overall accuracy. Performance was substantially stronger for English queries (79.04%) than for Bahasa Melayu queries (39.53%), revealing an important language-inclusivity limitation. Rule-based pathways responded in 3.74-7.48 ms, while the SBERT semantic route averaged 95.43 ms in the application-level test. The findings show that semantic retrieval can improve small-domain university FAQ matching without large training datasets, but multilingual robustness, real-user evaluation, and knowledge-base maintenance remain important priorities for responsible institutional deployment.

Keywords: academic chatbot; higher education; information retrieval; natural language processing; Sentence-BERT

INTRODUCTION

Timely access to accurate institutional information is an important component of student support in higher education. Students routinely need information about programmes, faculty structures, academic procedures, contact points, and administrative requirements. Although such information is often available on institutional websites and in handbooks, it may be dispersed across multiple pages and documents, requiring students to search manually or contact staff for clarification. Reviews of university chatbot implementations indicate that repetitive administrative enquiries remain a common motivation for deploying conversational systems, particularly where institutions seek to provide scalable and continuously available support (Peyton et al., 2025).

Educational chatbots have consequently been studied as tools for information provision, learning support, mentoring, and administrative services (Okonkwo & Ade-Ibijola, 2021; Wollny et al., 2021). In university settings, earlier systems commonly relied on predefined rules, keyword matching, or intent classification, while more recent systems increasingly use neural language models and sentence embeddings to improve semantic matching (Peyton et al., 2025; Mangotra et al., 2024). Malaysian examples also show the practical relevance of chatbot-based information services. The UNISEL Bot was designed to provide interactive university information (Ahmad et al., 2019), while AskEPTBot was developed to centralize recurrent information concerning an English Placement Test across multiple university branches (Mohamad et al., 2025). A recent Malaysian review further highlights perceived usefulness, ease of use, trust, design, and interaction as important considerations for chatbot acceptance in higher education (Idris et al., 2025).

For closed-domain FAQ retrieval, a central technical challenge is linguistic variation. Students may ask the same question with different words, grammatical forms, sentence lengths, or synonyms. Traditional Term Frequency-Inverse Document Frequency (TF-IDF) representations provide an efficient lexical baseline by weighting terms according to their frequency within a document and rarity across the collection (Salton & Buckley, 1988). However, lexical approaches can fail when a user employs semantically related terms that do not overlap with the stored question. Sentence-BERT (SBERT) addresses this limitation by producing dense sentence embeddings that can be compared efficiently using cosine similarity (Reimers & Gurevych, 2019). Previous work on student FAQ answering has reported advantages for sentence-level semantic representations over conventional chatbot platforms or lexical matching in semantically varied queries (Peyton & Unnikrishnan, 2023; Holis et al., 2025).

The present study develops a closed-domain FAQ chatbot for the Faculty of Electronics and Computer Technology and Engineering (FTKEK) at Universiti Teknikal Malaysia Melaka. Rather than using a generative large language model, the system retrieves verified answers from a bounded knowledge base. This design is intended to prioritize answer traceability and reduce unsupported free-form responses. The study also introduces a hybrid conversation-management mechanism in which simple contextual or menu-driven interactions are handled through finite-state and keyword rules, while semantically complex queries are routed to SBERT retrieval.

The study addresses four research questions: (RQ1) How does SBERT compare with a TF-IDF baseline for paraphrased FAQ retrieval in a small faculty knowledge base? (RQ2) What latency trade-off accompanies the use of semantic retrieval? (RQ3) How robust is the selected model when evaluated on a larger bilingual query set? and (RQ4) What design implications emerge for the deployment of closed-domain chatbots as student-support tools in higher education? By combining model comparison, bilingual stress testing, and application-level routing analysis, the study contributes evidence on the practical strengths and limitations of lightweight semantic retrieval for institutional student support.

LITERATURE REVIEW

Chatbots for Student Support in Higher Education

The literature on educational chatbots spans both learning-oriented and service-oriented applications. Systematic reviews identify benefits such as immediate access, scalability, personalization, and the capacity to support learners outside normal contact hours (Okonkwo & Ade-Ibijola, 2021; Wollny et al., 2021). However, evaluation quality varies substantially, and researchers have emphasized the need to align chatbot evaluation measures with the purpose of the system (Wollny et al., 2021). For an administrative or FAQ chatbot, retrieval correctness, response speed, coverage, fallback behavior, and user accessibility are therefore more relevant than measures designed for open-ended tutoring.

University-specific chatbot studies also indicate that closed-domain data are often limited and difficult to curate (Peyton et al., 2025). This makes data-efficient approaches attractive for faculties or departments that may have only dozens or hundreds of verified question-answer pairs. Mangotra et al. (2024) demonstrated an NLP-based university FAQ system using self-curated data, while Ahmad et al. (2019) showed how a rule-based university chatbot could provide interactive information in a Malaysian context. More recently, Malaysian higher-

education chatbot studies have increasingly considered usability, acceptance, and design implications (Mohamad et al., 2025; Idris et al., 2025). These developments support the use of chatbots as institutional service tools, while also highlighting that technical accuracy alone does not establish educational value or user acceptance.

Lexical and Semantic Retrieval for Closed-Domain FAQs

TF-IDF remains a useful baseline for closed-domain information retrieval because it is computationally lightweight, transparent, and easy to implement. Its performance is strongest when the user query shares informative lexical terms with the indexed question or document. The method, however, cannot inherently represent semantic equivalence between non-overlapping expressions, which can reduce robustness to paraphrases and synonyms (Salton & Buckley, 1988).

SBERT extends transformer-based language representations to efficient sentence-level similarity search by encoding sentences independently into dense vectors (Reimers & Gurevych, 2019). Once embeddings have been computed, cosine similarity can be used to rank stored questions against a new query. This architecture is particularly suitable for FAQ retrieval because the knowledge-base embeddings can be precomputed, leaving only the user query to be encoded at runtime. Prior student-support research found that sentence-BERT can outperform conventional chatbot platforms for online FAQ answering (Peyton & Unnikrishnan, 2023). Holis et al. (2025) similarly reported strong semantic FAQ retrieval using SBERT and cosine similarity for academic services.

The model used in the present study, all-MiniLM-L6-v2, belongs to a family of compact transformer encoders designed to reduce computational cost while retaining strong language-representation performance. MiniLM achieves compression through self-attention distillation (Wang et al., 2020). The use of a compact model is appropriate for local or resource-constrained deployment, but language coverage remains an important consideration. Multilingual sentence-embedding research shows that monolingual semantic spaces can be extended to additional languages through multilingual training or knowledge distillation (Reimers & Gurevych, 2020), suggesting a clear direction when bilingual institutional support is required.

Research Gap and Study Positioning

Existing studies demonstrate that university chatbots can reduce friction in information access and that semantic retrieval can improve matching performance. Nevertheless, three practical gaps remain relevant to small institutional deployments. First, many faculties have limited curated data and cannot justify expensive model training. Second, a chatbot must balance semantic accuracy with sufficiently responsive interaction on ordinary hardware. Third, multilingual higher-education environments require language-inclusive performance rather than strong results in only one language. The present study addresses these issues through a compact, retrieval-based architecture evaluated under both English and Bahasa Melayu queries, while explicitly separating fast rule-based navigation from deeper semantic matching.

METHODOLOGY

Research Design and Development Context

The study followed a two-phase development and evaluation design. Phase 1 focused on knowledge-base construction, model development, and baseline comparison. Phase 2 integrated the selected model into a web application, evaluated the hybrid routing architecture, and stress-tested the deployed system using a larger bilingual query set. Figure 1 summarizes the development and validation process.

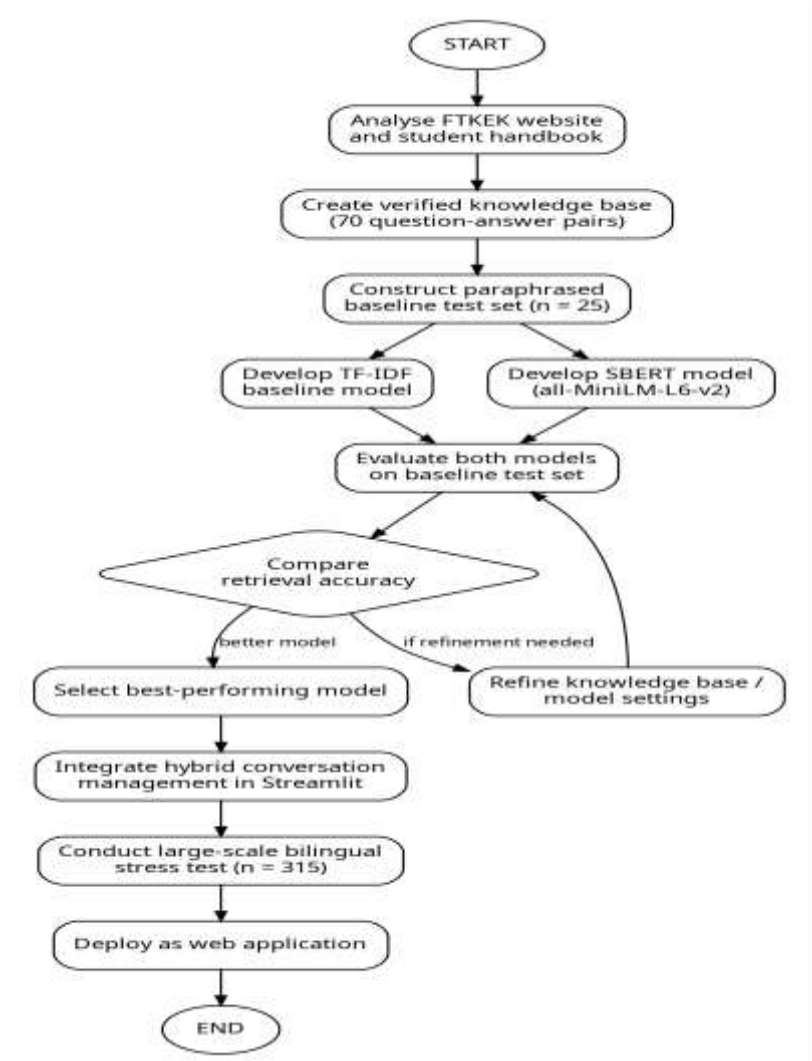

Figure 1 Development and validation process for the closed-domain FAQ chatbot

Knowledge-Base Construction

Unstructured institutional information was collected from official FTKEK web pages and the student handbook. The relevant material covered faculty organization, general descriptions, academic programmes, administrative information, and academic guidelines. The extracted content was cleaned and converted into a structured spreadsheet containing two main fields: Question and Answer. The final verified knowledge base contained 70 question-answer pairs. The use of a closed knowledge base ensures that retrieved responses originate from institutionally prepared content rather than unrestricted text generation.

Retrieval Models

Two retrieval approaches were implemented for comparison. The TF-IDF baseline was developed using Scikit-learn (Pedregosa et al., 2011). English stop-word filtering was enabled, and the resulting vocabulary contained 140 unique tokens in the baseline experiment. User queries and stored questions were represented as TF-IDF vectors and ranked using cosine similarity.

The semantic model was developed using the sentence-transformers library and the pre-trained all-MiniLM-L6-v2 encoder. The 70 knowledge-base questions were converted into dense embeddings and stored for reuse. At runtime, the user query was embedded and compared with all stored question vectors using cosine similarity. SBERT is designed for this type of sentence-level semantic similarity search (Reimers & Gurevych, 2019). Cosine similarity between query vector q and knowledge-base vector d was calculated as: $\cos(q,d) = (q \cdot d) / (||q|| \, ||d||)$.

A similarity threshold of 0.60 was used in the deployed application. When the highest similarity score was at least 0.60, the system returned the answer associated with the best-matching knowledge-base question. When the score was below 0.60, the system returned a fallback response indicating that the request could not be answered confidently and directed the user to an appropriate administrative contact. This threshold-based design limits low-confidence retrievals and provides a transparent failure pathway.

Hybrid Conversation Management

The chatbot combined semantic retrieval with a lightweight finite-state conversation mechanism. Three persistent runtime variables were used to preserve context during guided interactions: waiting for programme type, waiting for dean type, and waiting for head type. When one of these variables was active, short user inputs such as academic, degree, or engineering were interpreted relative to the current dialogue state rather than being sent directly to the embedding model.

Before semantic retrieval, the frontend also checked for root topic keywords such as programme, dean, and head. When a root keyword was detected, the relevant conversation state was initialized and the chatbot displayed guided options. Only queries that were not handled by an active conversation state or a root keyword were routed to SBERT. Figure 2 shows the resulting hybrid decision flow.

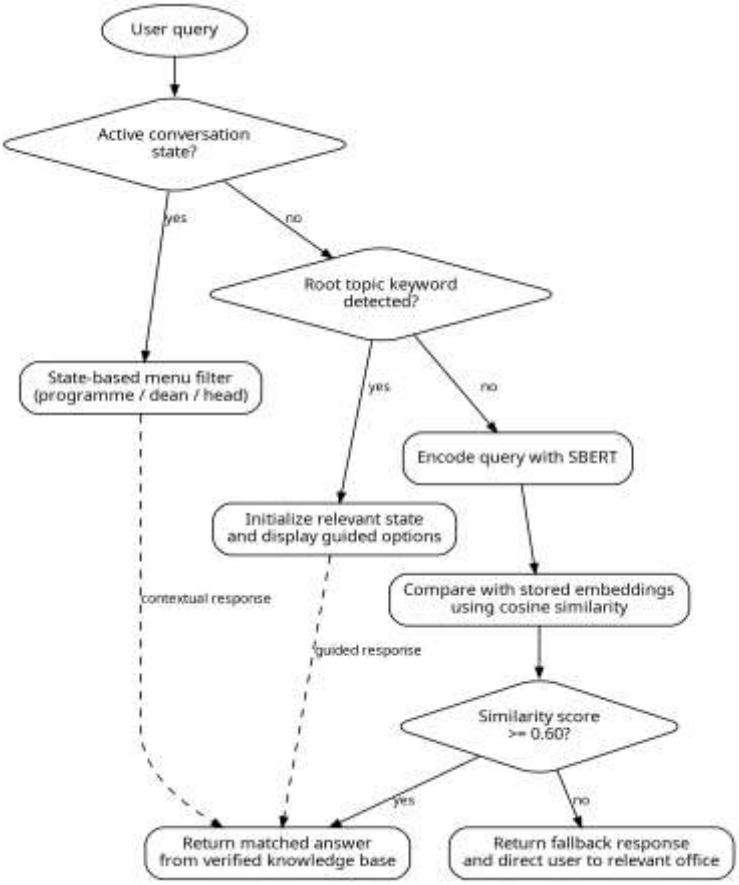

Figure 2 Hybrid conversation-management architecture combining state-based routing, keyword detection, semantic retrieval, and fallback handling

Evaluation Datasets and Metrics

For the Phase 1 comparison, a baseline evaluation set (test_set_1.csv) containing 25 paraphrased questions was constructed. Each question corresponded to a specific target answer in the knowledge base but used alternative wording, synonyms, or modified sentence structures. Retrieval accuracy was calculated as the number of correctly matched target answers divided by the total number of test questions, multiplied by 100%.

For Phase 2, a larger stress-test dataset (test_set_2.csv) containing 315 queries was used to approximate a broader range of user formulations. The dataset included short and long queries, semantic sentence variants, English questions, and Bahasa Melayu questions. The evaluation report contained 272 English queries and 43 Bahasa Melayu queries, which together account for all 315 test items. In addition to retrieval accuracy, the study recorded model inference latency and application-level routing latency.

All local deployment and latency measurements were conducted on a laptop with an Intel Core i7-10850H processor at 2.70 GHz, 16 GB RAM, a 512 GB NVMe SSD, and Windows 11 Pro (64-bit). The model-development environment used Google Colab, while the deployed application was implemented locally in Visual Studio Code with Streamlit.

RESULTS

Baseline Retrieval Accuracy

The 25-question paraphrase test showed a clear descriptive advantage for SBERT over the TF-IDF baseline. SBERT correctly retrieved 14 of 25 target answers, while TF-IDF correctly retrieved 10. The corresponding accuracies were 56.00% and 40.00%, respectively (Table 1). The 16-percentage-point difference supports the use of semantic sentence embeddings for queries that vary lexically from the stored FAQ wording, although the small evaluation set means the result should be interpreted as a development comparison rather than a definitive estimate of population-level performance.

Table 1 Baseline comparison on test_set_1.csv.

Retrieval model	Correct predictions (out of 25)	Retrieval accuracy
Sentence-BERT (SBERT)	14	56.00%
TF-IDF baseline	10	40.00%

A representative failure illustrates the difference between lexical and semantic matching. When the query used the expression faculty leader while the target FAQ used dean, TF-IDF produced a similarity score of 0.0000 because the critical terms did not overlap. SBERT, in contrast, mapped the semantic relationship between the expressions and returned the correct target. This behavior is consistent with the purpose of sentence embeddings, which represent meaning beyond exact term overlap (Reimers & Gurevych, 2019; Li et al., 2023).

Standalone Inference Latency

The lexical baseline was substantially faster than SBERT in the standalone benchmark. TF-IDF averaged 16.83 ms, while SBERT averaged 110.56 ms, an absolute difference of 93.73 ms (Table 2). Thus, the semantic model required approximately 6.6 times the inference time of the lexical baseline. Nevertheless, the measured SBERT latency remained close to one tenth of a second on the stated hardware, which is generally compatible with interactive web use. The engineering trade-off was therefore a 16-percentage-point gain in baseline retrieval accuracy at the cost of additional computation.

Table 2 Standalone inference-latency comparison.

Model	Average inference latency	Difference from TF-IDF
TF-IDF	16.83 ms	Reference
SBERT	110.56 ms	+93.73 ms

Large-Scale Bilingual Stress Test

The selected SBERT-based system was subsequently evaluated using 315 queries in test_set_2.csv. It produced 232 successful matches, corresponding to an overall retrieval accuracy of 73.65%. Performance differed sharply by language. The English subset achieved 79.04% accuracy with 215 correct matches out of 272 queries, whereas the Bahasa Melayu subset achieved 39.53% accuracy with 17 correct matches out of 43 queries. The English subset outperformed the Bahasa Melayu subset by 39.51 percentage points (79.04% vs. 39.53%), as shown in Table 3.

Table 3 Phase 2 bilingual stress-test results.

Evaluation subset	Correct matches	Total queries	Accuracy
Overall	232	315	73.65%
English	215	272	79.04%
Bahasa Melayu	17	43	39.53%

The results indicate that the chosen model and knowledge-base representation were considerably more reliable for English queries than for Bahasa Melayu queries in this dataset. Because the study did not experimentally compare multilingual sentence encoders or fine-tuned language-specific alternatives, the cause of this gap cannot be isolated from the present data. However, the finding establishes multilingual robustness as a central deployment limitation and motivates direct comparison with multilingual sentence-embedding approaches in future work (Reimers & Gurevych, 2020).

Hybrid Routing Latency

Application-level timing showed the computational benefit of routing simple interactions away from semantic inference. Keyword triggers averaged 3.74 ms, while persistent state menu filtering averaged 7.475 ms. The SBERT semantic-search route averaged 95.43 ms as shown in Table 4. These measurements were recorded for different application pathways and are therefore distinct from the standalone model benchmark reported in Table 2. The hybrid design concentrates transformer inference on queries that require semantic interpretation, while preserving near-instantaneous processing for guided navigation and state-dependent responses.

Table 4 Application-level latency of hybrid routing pathways.

Triggered mechanism	Average latency (ms)	Average latency (s)
Keyword triggers (rule-based)	3.74	0.00374
Persistent state menu filter (rule-based)	7.475	0.007475
SBERT semantic search (AI layer)	95.43	0.09543

DISCUSSION

Semantic Retrieval with a Small Knowledge Base

The Phase 1 comparison supports the central technical premise of the study: semantic sentence embeddings can improve FAQ retrieval when user wording differs from the stored question. The TF-IDF baseline remained useful for direct lexical matches and offered substantially lower latency, but its dependence on token overlap made it vulnerable to synonym substitution and paraphrasing. SBERT improved the baseline result from 40.00% to 56.00% without task-specific model training. This is important for small university units because the cost of building a large labeled training corpus can be prohibitive. The finding is directionally consistent with previous

work reporting the value of sentence-BERT for student FAQ answering and academic semantic retrieval (Peyton & Unnikrishnan, 2023; Holis et al., 2025).

At the same time, a 56.00% baseline accuracy is not sufficiently high to justify unconstrained deployment on its own. The practical contribution of the system therefore lies not only in the model choice but in the surrounding architecture. The 0.60 threshold, fallback response, guided menu states, and keyword routing collectively reduce the likelihood that every query will be forced through the semantic layer. This reflects an important design principle for institutional chatbots: a bounded retrieval model should be paired with explicit mechanisms for uncertainty and escalation rather than presenting uncertain matches as authoritative answers.

Implications for Student Support and Institutional Service Design

From a higher-education service perspective, the chatbot addresses an information-access problem rather than a teaching-content problem. Its intended value is the continuous availability of verified faculty information and the reduction of repetitive information-seeking effort. Reviews of university chatbots identify administrative and non-classroom student support as an area where conversational systems can improve service availability, particularly when institutional information already exists across websites, knowledge bases, and documents (Peyton et al., 2025). Malaysian examples likewise demonstrate the relevance of centralized conversational access to recurring university information (Ahmad et al., 2019; Mohamad et al., 2025).

The hybrid architecture also has implications for accessibility and usability. Students do not always formulate complete questions; short replies such as degree or academic may only be meaningful within the context of the preceding interaction. Preserving a limited dialogue state allows the system to interpret such responses without requiring a full semantic search. Similarly, fast menu and keyword routes reduce waiting time for common navigation tasks. These features may improve perceived effort and usefulness, factors that Malaysian literature identifies as important for chatbot acceptance (Idris et al., 2025). However, the present study did not measure actual user satisfaction, adoption, perceived usefulness, or staff workload reduction. Those outcomes remain hypotheses for future user-centered evaluation rather than demonstrated effects of the current experiment.

Multilingual Inclusion as a Deployment Requirement

The most important limitation revealed by the stress test is the substantial performance gap between English and Bahasa Melayu. In a Malaysian university, bilingual support is not a peripheral feature; it is directly related to equitable access to institutional information. A chatbot that performs well in English but inconsistently in Bahasa Melayu may reproduce information-access barriers for users who naturally formulate enquiries in the national language. The 39.53% Bahasa Melayu accuracy therefore places a clear boundary on the current system's readiness for broad institutional deployment.

This result also shows why overall accuracy can conceal subgroup weaknesses. An aggregate score of 73.65% appears substantially stronger than the Phase 1 baseline, yet the language-disaggregated analysis reveals that the improvement is concentrated in English queries. Future evaluation should therefore report language-specific metrics routinely and compare multilingual encoders, language-specific fine-tuning, bilingual knowledge-base variants, and alternative retrieval pipelines. Multilingual sentence-embedding methods provide a relevant technical foundation for this work (Reimers & Gurevych, 2020).

Practical Deployment Considerations

The latency results suggest that the system can operate on consumer-grade hardware without requiring continuous access to expensive cloud inference. Rule-based pathways responded within single-digit milliseconds, and the semantic pathway remained below approximately 0.12 seconds in the measured local tests. Nevertheless, latency values are hardware- and implementation-specific and should not be generalized to other deployment environments without benchmarking.

Operational deployment would also require a governance process for maintaining the knowledge base. Academic regulations, programme structures, staff roles, and administrative contacts can change over time. A technically

accurate retrieval model cannot compensate for outdated source content. For institutional use, the knowledge base should therefore have identified content owners, update dates, version control, and a clear escalation path for questions that fall outside the approved scope.

Limitations And Future Work

Several limitations should be considered when interpreting the findings. First, the study was conducted within a single faculty domain using only 70 question-answer pairs. The results therefore do not establish generalizability to other faculties or universities. Second, the Phase 1 comparison used only 25 paraphrased questions, which is adequate for development screening but too small for robust statistical inference. Third, the evaluation questions were constructed for testing rather than collected as a large corpus of naturally occurring anonymized student enquiries. Real-user language may contain spelling variation, code-switching, incomplete expressions, and context patterns that are not fully represented in the test sets. Fourth, the study measured retrieval accuracy and latency but did not evaluate user satisfaction, perceived usefulness, usability, trust, staff workload, or actual service outcomes. These variables are particularly important for a social-science and educational-technology assessment of institutional chatbot adoption. Fifth, the 0.60 similarity threshold was implemented as a fixed operational rule; future work should calibrate thresholds using validation data and evaluate precision-recall trade-offs. Sixth, no multilingual or Bahasa Melayu-specific sentence model was experimentally compared with all-MiniLM-L6-v2, so the observed language gap cannot be attributed to a single cause.

Future research should therefore collect anonymized real-user queries, enlarge the verified knowledge base, compare multilingual sentence encoders, and conduct a user-centered evaluation with students and administrative staff. Model evaluation should include accuracy, precision, recall, F1-score, coverage, fallback rate, and language-specific performance. The planned extension of full multi-turn context management and integration with messaging platforms such as Telegram or WhatsApp may also improve accessibility, but these extensions should be accompanied by institutional privacy, content-governance, and maintenance procedures.

CONCLUSION

This study developed and evaluated a closed-domain AI FAQ chatbot for automated academic information retrieval in a university faculty context. A comparison between TF-IDF and Sentence-BERT showed that SBERT provided better paraphrase retrieval accuracy, achieving 56.00% compared with 40.00% for the lexical baseline, while requiring greater but still interactive-level computation. The selected model was integrated into a hybrid architecture that combined finite-state context, keyword routing, semantic retrieval, threshold-based confidence handling, and a Streamlit web interface.

The larger evaluation of 315 queries produced an overall accuracy of 73.65%, but language-specific analysis revealed a major limitation: 79.04% accuracy for English compared with 39.53% for Bahasa Melayu. The hybrid routing mechanism reduced unnecessary semantic computation, with rule-based routes completing in 3.74-7.475 ms and the SBERT application route averaging 95.43 ms. Together, these results show that pre-trained sentence embeddings can provide a practical, data-efficient foundation for small-domain university information services, but technical performance must be evaluated alongside language inclusion, uncertainty handling, content governance, and user experience.

For higher-education institutions, the principal implication is that conversational access to verified information can be implemented without a large generative model or a large task-specific training corpus. However, responsible deployment requires stronger multilingual performance and real-user validation before the system can be considered a comprehensive student-support service.

Ethical Considerations

This study did not involve human participants, interventions, or personal data. The evaluation was limited to software retrieval performance using institutional information and constructed test queries; therefore, human-subject ethical approval was not required.

Data Availability

The knowledge base, evaluation datasets (test_set_1.csv and test_set_2.csv), and supporting implementation materials are available from the corresponding author upon reasonable request, subject to institutional information-governance requirements.

Conflict Of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

ACKNOWLEDGEMENT

The authors would like to thank Centre for Research and Innovation Management (CRIM), Universiti Teknikal Malaysia Melaka (UTeM) for supporting this work.

REFERENCES

1. Peyton, K., Unnikrishnan, S., & Mulligan, B. (2025). A review of university chatbots for student support: FAQs and beyond. Discover Education, 4, Article 21. https://doi.org/10.1007/s44217-025-00397-7
2. Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. Computers and Education: Artificial Intelligence, 2, 100033. https://doi.org/10.1016/j.caeai.2021.100033
3. Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet? A systematic literature review on chatbots in education. Frontiers in Artificial Intelligence, 4, 654924. https://doi.org/10.3389/frai.2021.654924
4. Mangotra, H., Dabas, V., Khetharpal, B., Verma, A., Singhal, S., & Mohapatra, A. K. (2024). University auto reply FAQ chatbot using NLP and neural networks. Artificial Intelligence and Applications, 2(2), 126-134. https://doi.org/10.47852/bonviewAIA3202631
5. Ahmad, N. A., Hamid, M. H. C., Zainal, A., & Baharum, Z. (2019). UNISEL Bot: Designing simple chatbot system for university FAQs. International Journal of Innovative Technology and Exploring Engineering, 9(2), 4689-4693. https://doi.org/10.35940/ijitee.B9067.129219
6. Mohamad, H. A., Abd Rashid, M. H., Hasan, H. R., Mohd Akhir, N., Abd Rahman, A. L., Salam, U., & Syed Khalid, S. S. A. (2025). AskEPTBot: Your virtual EPT chatbot companion! International Journal of Research and Innovation in Social Science, 9(25), 392-402. https://doi.org/10.47772/IJRISS.2025.925ILEIID000069
7. Idris, F., Idris, F., Saleh, A. A., & Ismail, M. I. (2025). A review on acceptance of chatbots for learning: Malaysian context. International Journal of Research and Innovation in Social Science, 9(9), 4260-4271. https://doi.org/10.47772/IJRISS.2025.909000348
8. Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 513-523. https://doi.org/10.1016/0306-4573(88)90021-0
9. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (pp. 3982-3992). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1410
10. Peyton, K., & Unnikrishnan, S. (2023). A comparison of chatbot platforms with the state-of-the-art sentence BERT for answering online student FAQs. Results in Engineering, 17, 100856. https://doi.org/10.1016/j.rineng.2022.100856
11. Holis, R. M., Utomo, P. E. P., & Hutabarat, B. F. (2025). Semantic FAQ chatbot using SBERT (Sentence-BERT) and cosine similarity for academic services. Brilliance: Research of Artificial Intelligence, 5(2), 915-922. https://doi.org/10.47709/brilliance.v5i2.7027
12. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

13. Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., & Zhou, M. (2020). MiniLM: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. Advances in Neural Information Processing Systems, 33, 5776-5788.
14. Reimers, N., & Gurevych, I. (2020). Making monolingual sentence embeddings multilingual using knowledge distillation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (pp. 4512-4525). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.365
15. Li, Z., Zhu, J., Cheng, X., & Lu, Q. (2023). Optimized algorithm design for text similarity detection based on artificial intelligence and natural language processing. Procedia Computer Science, 228, 195-202. https://doi.org/10.1016/j.procs.2023.11.023.