 DOI:10.5281/zenodo.17795962

Title

DESIGN AND EVALUATION OF UNIBOT: AN AI-POWERED STUDENT SUPPORT CHATBOT USING APPROVED UNIVERSITY KNOWLEDGE

Author

MUHAMMAD KAZIM SOMANI

Co-Author

Mr. JOEL MULEPA



Issued November 2026 Certificate

AR2026QYM30S



ABSTRACT

Student support offices in higher education institutions handle recurring questions about admissions, fees, registration, examinations, library services, and other university processes. When institutional information is distributed across channels or staff are unavailable, students may experience delays in finding reliable answers. This paper describes the design, implementation, and preliminary evaluation of UniBot, an AI-powered student support chatbot developed for DMI St. John the Baptist University. The system combines a lexical FAQ-matching stage with a generative response stage that is constrained by active, approved university FAQ information. It records source metadata and displays supporting FAQ references for eligible responses. The interface also provides English and Chichewa language modes and browser-based voice-to-text input.

The work was evaluated through automated software tests, browser-based checks, synthetic FAQ-matcher query sets, and observations of selected deployed interactions. The preliminary matcher results reported for a 115-query synthetic development set show an increase in precision from 63.0% to 92.2%, recall from 63.0% to 72.8%, accuracy from 60.0% to 76.5%, and a reduction in false-positive rate from 47.1% to 14.7%. These figures are preliminary: a 38-query holdout set was subsequently used during tuning and is therefore not an untouched independent test. The implementation test suite was reported to pass 592 of 592 tests, with additional browser checks reported separately.

The results demonstrate implemented software capabilities and promising preliminary FAQ-matching behavior, but do not establish answer correctness across real student questions, Chichewa translation quality, student satisfaction, or institutional impact. The paper documents the architecture, current behavior, limitations, and further evaluation required before broader deployment.

KEYWORDS: student support chatbot, artificial intelligence, FAQ matching, grounded question answering, source attribution, Chichewa, higher education, voice-to-text.

INTRODUCTION

Higher education institutions receive repeated questions about admission requirements, tuition and other fees, registration, examinations, library services, accommodation, and administrative procedures. Students may need this information outside normal office hours or may find it difficult to identify the correct office or document. A conversational interface can provide another route to institutional information and may reduce the effort required to locate answers to routine questions.

However, a chatbot that produces fluent language is not necessarily reliable. If it answers from unsupported general knowledge, it may present incorrect institutional details with confidence. A useful student-support system must therefore make use of institutionally approved information, distinguish supported answers from unsupported questions, and provide a clear route for verification.

UniBot was developed as an AI-powered student support chatbot for DMI St. John the Baptist University. The project evolved from an initial proposal for an NLP-based chatbot into a web application that uses a lexical FAQ matcher and a grounded generative fallback. Subsequent work added source attribution in the chat interface and history, English–Chichewa language handling, and browser-based voice-to-text input. This paper documents the implemented system and the evidence currently available for its evaluation.

PROBLEM STATEMENT

Students may experience delays when seeking answers to routine academic and administrative questions, especially when information is spread across institutional sources or staff are unavailable. A chatbot that answers without reliable institutional evidence can introduce a separate risk by giving plausible but incorrect information. The problem addressed by this project is therefore twofold: improving access to university information through a conversational interface, and reducing unsupported answers by grounding the AI response path in an approved university FAQ knowledge base. The project also considers language choice and alternative text-input methods as accessibility features.

RESEARCH OBJECTIVES

- Develop a student-facing chatbot interface with authenticated access and conversation handling.
- Implement lexical FAQ matching with an explicit abstention mechanism for uncertain or unsupported matches.
- Constrain generative fallback answers to active, approved university FAQ information and validate the source identifiers returned by the model.
- Display supporting FAQ sources and preserve valid source metadata in conversation history.
- Provide English and Chichewa language modes and document their current behavior and limitations.
- Implement browser-based voice-to-text input that allows users to review and edit recognized text before sending it.
- Evaluate software behavior and FAQ-matcher performance using automated tests, browser checks, synthetic queries, and selected deployed interactions.

RELATED WORK AND CONCEPTUAL BACKGROUND

The original project proposal identifies research suggesting that educational chatbots can provide timely assistance, support communication, and reduce repetitive staff workload. It mentions *Labadze et al. (2023)*, *Belda-Medina and Kokošková (2023)*, *Masula et al. (2025)*, and two broadly described 2024 studies. The supplied proposal contains summaries rather than complete bibliographic records. These references are therefore retained as leads for the literature review but are not presented here as verified citations.

Three concepts guide the implementation. First, FAQ matching provides a controlled route to a known answer when a student's wording sufficiently matches an approved question. Second, abstention is important: when the matcher is uncertain, the system should avoid selecting an unrelated FAQ. Third, grounded generation supplies the model with approved information and validates its cited source identifiers before accepting an answer. Source-ID validation confirms that a cited item was supplied to the model; it does not, by itself, prove that every sentence in the answer is entailed by that source.

A complete related-work section should add verified peer-reviewed studies on educational chatbots, FAQ retrieval and confidence-aware matching, grounded question answering and hallucination evaluation, source attribution, and low-resource language technologies relevant to Chichewa. Full author lists, titles, publication venues, page ranges, and DOI/URLs must be checked against the original publications before submission.

SYSTEM DESIGN AND METHODOLOGY

UniBot is implemented as a web application. The student interacts with a Next.js frontend. Requests are handled by server-side chat routes, which authenticate the user, apply request validation and rate limiting, invoke the answer-selection flow, persist the exchange, and return response and source metadata. PostgreSQL stores user, FAQ, session, message, feedback, and source-related records. The generative fallback uses the Anthropic Claude API when configured.

The current runtime is not the Rasa architecture described in the initial README. Rasa configuration files and the earlier Python/FastAPI project description are not an accurate representation of the later deployed implementation. The current answer path uses the application's lexical FAQ matcher followed, when necessary, by a Claude request supplied with active FAQ context.

Answer-generation flow

- Authentication and validation: the chat route verifies the user's session, applies rate limiting, and validates the submitted message.
- Direct FAQ matching: the matcher scores active FAQs using question wording, keywords, tags, and specificity. A sufficiently strong, unambiguous match returns the approved FAQ answer directly.
- Grounded AI fallback: if the matcher abstains, the server supplies active FAQ information to

  Claude and requests a structured result containing an answer status, answer text, and cited FAQ identifiers.

- Citation validation: the server checks that cited identifiers belong to the FAQ context supplied for that request. Invalid or unsupported model output is not accepted as a grounded answer.
- Persistence and response: the application stores message and source metadata and returns the answer, source information, and session identifiers to the frontend.

The system distinguishes direct FAQ answers, AI-generated answers grounded in supplied FAQs, and fallback or unsupported outcomes. The exact source label used in storage is implementation-specific; the publication should preserve the final code's terminology when the release version is frozen.

FAQ knowledge base

The chatbot's institutional knowledge is represented by active FAQ records containing approved questions and answers and associated metadata such as department, contact information, keywords, and tags. The available project evidence describes the production FAQ set as containing 34 records during the 30 October 2026 pre-check. This is a database count, not a count of independently verified documents. The current implementation does not provide a general document-upload, PDF-extraction, chunking, or document-retrieval pipeline. Accordingly, the paper describes approved FAQ knowledge, not unrestricted ingestion of university textbooks or documents.

FAQ matching and abstention

The matcher uses lexical signals, including question similarity, keyword overlap, tags, and specificity weighting. It can abstain when the best score is below the configured threshold, when top candidates are tied, or when query coverage is insufficient without a question match. This design aims to reduce false positive matches, where a superficially similar query is assigned an irrelevant FAQ. The matcher evaluation described in Section 7 uses synthetic queries and should be interpreted within that scope.

Source attribution and history

For a direct FAQ response, the interface can show the matched FAQ as its source. For an AI response, the interface shows the FAQs cited by the model after server-side identifier validation. Source titles are presented using the FAQ question and department rather than exposing internal database identifiers. Source information can be restored from history only while the referenced FAQ remains active and its stored content fingerprint remains valid. This behavior reduces the risk of presenting an outdated source reference after FAQ content changes.

The application includes a safe document-URL validation helper, but the current FAQ data does not contain document URLs and no document links are available to display. Visible FAQ citations should therefore not be described as clickable links to source policy documents.

English and Chichewa interaction

English is the default language, and Chichewa is available as an additional interaction mode. The interface and answer flow include language-aware handling, including reviewed Chichewa question patterns and greeting behavior. Current behavior is not equivalent to a fully translated institutional knowledge base: some Chichewa-mode direct FAQ matches display a Chichewa explanatory note while retaining the approved FAQ answer in English. Chichewa AI-generated responses have been exercised with a fake model in tests, but the available evidence does not establish their quality with a live model across representative student questions.

Voice-to-text input

Voice input is implemented through the browser Web Speech API (SpeechRecognition or webkitSpeechRecognition). A student can dictate a question, review and edit the recognized text, and then manually press Send. The feature does not automatically submit recognized speech and does not implement a full spoken conversation or text-to-speech response. English recognition is configured for en-US. Chichewa speech recognition is not supported by the current feature and the interface communicates that limitation

rather than silently switching languages.

Microphone permission is requested only after the student activates the microphone. Recognition is stopped on manual stop, language change, or component unmount. The application does not store or upload audio as part of its own chat API flow; however, browser speech-recognition services may process audio externally depending on the browser and platform. It would therefore be inaccurate to claim that all speech processing occurs locally on the device.

IMPLEMENTATION

Student interface

- Implemented an authenticated web chat with message input, response display, and chat history functionality.
- Evidence / qualification: Browser-tested. This is a web-based interface; no mobile native application is claimed.

FAQ knowledge

- Implemented active university FAQ records for direct answers and as context for AI fallback responses.
- Evidence / qualification: 34 FAQ rows were reported in the read-only production pre-check.

AI response path

- Implemented Claude structured responses using active FAQ context, with server-side validation of cited FAQ IDs.
- Evidence / qualification: Successful deployed multi-source fee responses were observed. Production persistence was affected by a schema mismatch.

Source display

- Implemented display of FAQ source titles and departments alongside eligible answers.
- Chat history source validity is also checked.
- Evidence / qualification: No linked source documents are currently available in the FAQ records.

Language modes

- Implemented English as the default language and a Chichewa mode.
- Evidence / qualification: Some direct answers in Chichewa mode remain in English. Live Chichewa answer quality has not yet been established.

Voice input

- Implemented browser-based speech recognition that inserts editable text into the message input. Students must manually send the recognised text.
- Evidence / qualification: English only. Support varies depending on the browser and speech recognition service.

Feedback

- Implemented a feature allowing students to submit feedback on chatbot responses.
- Evidence / qualification: Historical free-text feedback may not be linked to a specific message. See deployment and data limitations.

Administration

- Implemented admin FAQ management and user/admin access controls.
- Evidence / qualification: A formal multi-stage FAQ approval workflow is not claimed. Active status is the current mechanism for controlling FAQ inclusion.

EVALUATION AND RESULTS

Evaluation approach

The available evaluation evidence consists of (a) synthetic FAQ-matcher query sets, (b) automated unit and integration tests, (c) browser-based checks at desktop and mobile viewport sizes, and (d) selected interactions with the deployed application. These evidence types answer different questions. Matcher metrics assess whether the lexical stage selects or abstains on labelled synthetic queries.

Automated tests assess defined software behavior. Browser checks assess interface behavior under selected conditions. Deployed examples demonstrate observed operation but are not a controlled user

study.

Preliminary FAQ-matcher results

The project planning brief reports the following comparison for the original and improved matcher. The development set contained 115 synthetic queries and the holdout set contained 38 synthetic queries, constructed around a 33-FAQ fixture. The holdout set was later examined during tuning and is no longer a blind independent test. One additional production FAQ was not represented in the fixture at the time of the report.

Matcher Performance Comparison

Precision

- Original matcher: 63.0%
- Improved matcher: 92.2%

Recall

- Original matcher: 63.0%
- Improved matcher: 72.8%

False-positive rate

- Original matcher: 47.1%
- Improved matcher: 14.7%

Accuracy

- Original matcher: 60.0%
- Improved matcher: 76.5%

These figures indicate a reported improvement in precision and accuracy, alongside higher recall and a lower false-positive rate on the reported synthetic evaluation. They should be labelled preliminary until the final code, query labels, metric definitions, and evaluation command are checked and the results are reproduced. Because the holdout was used during tuning, it cannot support a claim of unbiased final generalization.

Automated and browser testing

The implementation reports 592 passing automated tests across 18 test files after a clean server restart. TypeScript checking and a production build also passed. Voice-input unit coverage included 25 tests using a fake speech recognizer. A separate Playwright run reported 27 of 27 Chromium checks passing against a production build with a fake recognizer. Desktop and mobile browser checks covered source display, fallback behavior, history persistence, and horizontal overflow at representative viewport widths.

These tests demonstrate that the tested code paths behaved as expected under the test conditions. Fake recognizers and fake model responses do not establish compatibility with every real microphone, browser speech service, or live AI response. A real deployed voice test was reported by the project team, but the exact device, browser, recognition outcome, and repeatability were not documented in the supplied evidence.

Deployed interaction observations

Selected deployed conversations on 30 October 2026 showed direct FAQ responses for library-hours questions, a multi-source response to a fees question, and an approved-information fallback for questions not covered by the knowledge base. The observed fees response included seven FAQ sources across relevant university functions. These examples demonstrate that the deployed application could return both direct FAQ answers and multi-source AI responses at the time of observation.

The same deployment logs showed database errors: the deployed code attempted to write a chat_messages.source column that was absent, and the history query referenced a feedback.message_id column that was absent. A subsequent read-only production pre-check reported that migrations 001–005 were not applied and that the starting schema matched the expected pre-migration snapshot. The supplied recovery report recorded 4 chat_messages, 15 feedback rows, and 34 FAQ rows. No production migration had been applied in the evidence available for this manuscript. Consequently, chat persistence and history retrieval were not

functioning reliably in that production state, even though answer generation was observed.

DISCUSSION

The main technical contribution of UniBot is the combination of a conservative FAQ-matching stage with a grounded generative fallback. The matcher can return a known approved answer when the query is sufficiently clear, while abstention routes uncertain questions to a second stage. Supplying active FAQs to the generative model gives it institution-specific context that would not be available from a generic prompt alone. Validating cited FAQ identifiers adds a check that the model refers only to sources supplied for that request.

The matcher results suggest that the improved scoring and abstention rules may reduce irrelevant FAQ selection on the synthetic query set. The trade-off is that more conservative matching may fail to directly answer some answerable paraphrases, which is why both precision and recall are reported. The current evaluation is not sufficient to establish how the matcher will perform on naturally occurring student questions, code-switched Chichewa-English queries, spelling variation, or new FAQ content.

Source attribution improves transparency by showing which FAQ records support an answer. Nevertheless, a valid citation is not proof that every claim in the generated text is supported. A separate human-labelled faithfulness evaluation is needed. Likewise, the presence of a Chichewa mode should not be interpreted as evidence of high-quality translation until native or suitably qualified reviewers assess meaning, grammar, naturalness, and preservation of institutional details.

The voice-to-text feature offers an alternative input method for students using supported browsers. Its practical value depends on browser support, microphone permissions, connectivity, and speech-recognition quality. Because the feature does not provide spoken responses and does not currently support Chichewa speech recognition, it should be described specifically as English voice-to-text input rather than a complete voice assistant.

LIMITATIONS, RISKS, AND FUTURE WORK

- The matcher evaluation uses synthetic queries. A new, independently labelled set of real or realistically collected student questions is required before making general performance claims.
- The 38-query holdout set was inspected during tuning and is not an untouched blind test.
- Citation validation checks that source IDs are valid, not that every answer statement is entailed by the cited FAQ. Human-rated factual faithfulness evaluation is required.
- The current knowledge base is FAQ-based. There is no implemented general-purpose upload and retrieval pipeline for university PDFs, textbooks, or other source documents.
- Some Chichewa-mode answers remain in English. Live-model Chichewa quality, code-switching, and fully Chichewa query coverage require evaluation.
- Voice recognition is browser-dependent and English-only in the current implementation. It does not include text-to-speech or full voice conversation.
- The chat flow is single-turn for the AI fallback; follow-up questions do not receive prior conversational context. This can limit requests such as asking for more detail about a previous answer.
- Production database migrations 001–005 were pending in the observed deployment. Until the schema is brought into line with the deployed code and persistence is verified, production history and feedback-related behavior remain operational limitations.
- The project has not conducted a controlled student usability study, satisfaction survey, workload assessment, or institutional impact evaluation.
- The literature review and bibliography remain incomplete because the references in the original proposal have not yet been verified against their original publications.

RECOMMENDED NEXT RESEARCH STEPS

- Reproduce matcher metrics from the final committed code and publish the evaluation dataset construction, labels, thresholds, and metric definitions.

- Create a fresh blind test set that is not used during development or threshold tuning.
- Evaluate AI answers with human judgements of correctness, completeness, faithfulness to cited FAQs, and appropriate abstention.
- Ask qualified Chichewa reviewers to assess translations and test naturally occurring Chichewa and code-switched student questions.
- Complete production database recovery and verify message persistence, source history, and feedback linkage before collecting further production evaluation data.
- Conduct an ethically approved usability study with students and relevant support staff, if institutional approval and access are available.
- Verify all scholarly references and add complete bibliographic details.

CONCLUSION

This paper has described UniBot, an AI-powered student support chatbot developed for DMI St. John the Baptist University. The implemented system combines lexical FAQ matching and abstention with a generative response path grounded in active university FAQ information. It also provides source attribution, English and Chichewa interaction modes, and browser-based English voice-to-text input.

The available evidence includes preliminary synthetic matcher results, automated software tests, browser checks, and selected deployed interactions. The reported matcher comparison shows promising improvements, but the evaluation dataset and tuning history limit the strength of generalization claims. The system's source validation, multilingual behavior, and voice input have been functionally tested, but factual faithfulness, Chichewa language quality, real-world speech performance, student satisfaction, and institutional impact require further study.

UniBot therefore represents an implemented technical prototype with several meaningful student-support capabilities, rather than a fully validated institutional support service. Future work should prioritize independent evaluation, reliable production persistence, improved multilingual coverage, and user-centred assessment.

The figures in this manuscript are implementation-oriented conceptual diagrams. They summarize the current application behavior and should not be interpreted as evidence of features that are planned but not implemented.

REFERENCES

1. Belda-Medina, J., & Kokošková, (2023). *[Full publication title to be verified]*. [Publication details to be confirmed].
2. Labadze, L., et al. (2023). *[Full publication title to be verified]*. [Publication details to be confirmed].
3. Masula, et al. (2025). *[Full publication title to be verified]*. [Publication details to be confirmed].
4. [Author(s) to be identified]. (2024). *[Systematic review on the relevant topic]*. [Publication details to be confirmed].
5. [Author(s) to be identified]. (2024). *[Study on student motivation]*. [Publication details to be confirmed].