

Research Proposal

Title: Arabic Chatbots for University Admission and Registration

Student Name: Metab Hussain Jdeba

Table of Contents

Abstract.....	3
Introduction	3
Background	4
Problem Statement.....	5
Research Objectives.....	6
General Objective:	6
Specific Objectives:	6
Literary reviews.....	7
1.Tayseer: A Novel AI-Powered Arabic Chatbot Framework for Technical and Vocational Student Helpdesk Services.....	7
2. Arabic Educational Neural Network Chatbot.....	8
3. Implementing the Bashayer Chatbot in Saudi Higher Education.....	9
4. The Challenges of Using Arabic Chatbot in Saudi Universities (UFAQBot)	9
5. Nabiha: An Arabic Dialect Chatbot	10
6. Arab Chatbot's challenges and solutions: a systematic literature review	11
7. Intelligent Chatbot for Admission in Higher Education (Aloqayli & Abdelhafez, 2023)	12
8. An Inquiry Smart Chatbot System for Al-Zaytoonah University (Al-Madi et al., 2024)	13
9. Bilingual AI-Driven Chatbot for Academic Advising (Bilquise et al., 2022)	14
10. College Enquiry Chatbot using Rasa Framework (Meshram et al., 2021/22)	14
Research Methodology	15
Phase 1: Literature Review	15
Phase 2: Data Collection and Preparation	16
Phase 3: System Design and Development.....	16
Phase 4: Testing and Evaluation	17
Ethical Considerations.....	17
Limitations	17
Expected Outcomes	17
Gantt chart.....	18
References	19

Abstract

Admission and registration are often stressful for new students. Questions repeat every semester, and staff spend a lot of time giving the same answers. While some chatbot systems exist in English, Arabic chatbots are still rare and usually not tested in real contexts. This project focuses on building a simple chatbot in Arabic, with some English support, to answer admission and registration questions. The chatbot will be trained on university documents and common student queries, and it will be tested to see both accuracy and how students react to it. The goal is not a perfect system, but a working tool that saves staff effort, gives faster replies, and shows where Arabic chatbot technology still needs improvement.

Introduction

Starting university is exciting, but the first step admission and registration—often feels heavier than it should. Information sits in different places, the wording is formal, and dates move when students are already under pressure. Many of us end up asking the same questions again because it is faster than searching long pages. On the staff side, that means repeated answers during the busiest weeks. A small, focused chatbot can reduce this friction: it replies instantly, points to the right rule or form, and does so in plain language. This project takes the practical route for an Arabic-first setting, with optional English for mixed queries. It concentrates on routine tasks requirements, deadlines, add/drop steps rather than trying to solve everything at once. The plan is simple: learn the most frequent intents, extract key details like dates or term names, and keep the dialogue on track. What matters is not only accuracy but whether students actually finish the task faster and feel less confused. By keeping the scope tight and the language natural, the prototype aims to show a

workable path for Arabic admission support and to make clear where the real limits still are (dialects, informal phrasing, and small datasets).

Background

Chatbots in higher education usually follow one of three patterns: rule-based flows (fast to build but brittle), retrieval-style FAQ bots (good for exact matches), and intent-based systems that classify what a user wants and extract key details. For admission and registration, intent-based designs work best because students rarely type the question the same way twice. A typical pipeline includes intent recognition, entity extraction (e.g., term names, dates), and dialogue management to keep the conversation on track.

Arabic introduces extra friction. Students mix Modern Standard Arabic with regional dialects; they shorten words, write colloquially, or switch to English for course codes and system names. Small datasets make it hard to cover that variety, and tokenization/normalization steps matter more than in English. That is why a narrow domain (admission/registration only) and a carefully curated dataset can outperform a broad, generic bot. In this project, the training data will come from official FAQs and policy documents, plus anonymized examples of real questions rewritten in different ways (formal phrasing, casual phrasing, and light dialect).

On the engineering side, a pragmatic stack keeps the scope realistic. A standard intent/entity model with a compact transformer encoder is enough for a first pass; dialogue flows handle the predictable parts (documents needed, list, link), while hand-off to a human remains possible for edge cases. Evaluation must look beyond accuracy: do students feel the answers are clear? do they finish tasks faster? A short usability check, alongside precision/recall/F1, gives a more honest picture. Finally, privacy is simple here—the bot uses public admissions

information and avoids personal data—so the prototype can be trialed safely and improved without touching student records.

If this background can be summed up in one line, it is this: a small, well-scoped Arabic chatbot, trained on real questions and measured by both accuracy and student experience, has a better chance to help during admission than a large, generic system.

Problem Statement

It has increasingly become popular in the past few years to consider chatbots as a tool to enhance student services in Arab universities, particularly for services like admissions and registration. But in practice when you dive into how these chat bots actually work in Arabic there are a lot of problems. I read six papers on distinct systems, and every single one of them presented something useful and, at the same time, something missing. For example, the Tayseer chatbot (Alabbas & Alomar, 2024) got really good accuracy with over 90% in tests. But it depended on users writing very structured inputs, which isn't realistic for everyone. Another chatbot used a neural network model (Alazzam et al., 2023), trained on a large Arabic dataset, and performed well in accuracy scores but it was never tested with real users, and it didn't support dialects at all. The Bashayer bot (Al-Abdullatif et al., 2023) was actually used by students on WhatsApp and helped improve motivation and access to learning. However, the study was short and didn't focus on admission tasks specifically. UFAQBot (Almurayh, 2021) showed that many Saudi universities don't even have working chatbot systems, and the one they tested didn't understand dialects well. Another bot called Nabiha (Al-Ghadhban & Al-Twairesh, 2020) was one of the earlier ones, but it couldn't learn or adapt it just gave fixed replies. Finally, a big review of 60+ studies (Ouali & El Garouani, 2024) pointed out that most Arabic chatbots still rely on rule-based logic and there's a huge lack of annotated

datasets. It also noted that very few systems actually focused on admission and registration most were just general support bots. So after looking at all these studies, it's clear that Arabic-language chatbots do have potential, but the way they're used for admission and registration is still limited and not fully developed. This research aims to bring together these findings and suggest better ways to design, test, and improve chatbot systems for real use in Arab universities.

Research Objectives

General Objective:

To explore how Arabic-language chatbots are used for admission and registration in Arab universities, and to understand what is still missing or needs improvement.

Specific Objectives:

- To look into different Arabic chatbot systems made for student support.
- To understand the common issues like tech problems or language barriers.
- To see how students actually interact with these bots (if they do).
- To compare a few chatbot examples based on what the studies show.
- To give simple suggestions for how to make these chatbots better.

Literary reviews

1.Tayseer: A Novel AI-Powered Arabic Chatbot Framework for Technical and Vocational Student Helpdesk Services

Objective:

Develop and evaluate taights, designed to help technical and professional students with an AI operated Arab Chatbot, and address linguistic and infrastructure challenges in Arabic Chatbot design, recording, planning and campus service.

Methods:

- Rasa for Arabic NLP was developed using frames and dietary transformer - based architecture.
- Interaction with two mode: free text and menu -based navigation.
- Data collected through pre -examination of 248 students and institutional documents.
- User reaction and system view were evaluated through calculations.

Central findings:

- Intention obtained more than 90% accuracy in classification and unit recognition.
- Better student satisfaction and administrative information access.
- Scalable and adaptable structure for Arab educational settings.

Critical analysis:

- Strength: Advanced NLP integration, double interaction mode, practical perfection, user-centered design.
- Weaknesses: Limited data set size, limited condensed depth, addiction to structured input.

2. Arabic Educational Neural Network Chatbot

Objectives:

To design and evaluate a neural network-based Arabic chatbot that answers questions about educational regulations for students in the UAE, with a focus on Arabic NLP in education.

Methods:

- Data collected from the UAE Ministry, consisting of 300,000 Arabic words.
- Pre-processing using NLP techniques with a Five-team Custom Neural Network Model.
- Automatic evaluation metrics used include accuracy, precision, recall, and F1 score.

Key Findings:

- The chatbot achieved high accuracy and F1 scores for regulatory questions.
- This demonstrates the feasibility of using deep learning for Arabic educational chatbots.

Critical Analysis:

- Strengths: The neural network architecture is well-suited for Arabic morphological analysis, resulting in strong automated performance.
- Weaknesses: The system is limited to the domain of educational regulations, lacks real user testing, and does not account for dialectical variations in Arabic.

3. Implementing the Bashayer Chatbot in Saudi Higher Education

Objectives:

To assess the impact of Sansheer on learning strategies, including a WhatsApp-based chatbot, inspiration, registration, and educational support among students.

Methods:

- A semi-professional design involving 60 students divided into experimental and control groups.
- Motivations and learning strategies were measured using a revised version of the MSLQ questionnaire.
- The chatbot was hosted on WhatsApp and included various educational tasks.

Key Findings:

- There was a significant increase in inspiration and self-efficacy in the experimental group.
- The use of cognitive and metacognitive learning strategies also increased.
- Features for electronic registration and resource access were provided.

Critical Analysis:

- Strength: The study benefited from Real -world Glory, a strict quantitative assessment and integration with a popular platform.
- Weaknesses: The boundaries included a small sample size, only focused on students, short -term study periods and lack of qualitative insight.

4. The Challenges of Using Arabic Chatbot in Saudi Universities (UFAQBot)

Objectives:

To focus on the linguistic and technical challenges faced by Arab chatbots and to identify the obstacles to their adoption in Saudi universities.

Methods:

- Examining 44 websites of Saudi universities to check for the presence of chatbots.
- Conducting an experimental test of the Sara chatbot with 30 participants.

- Analyzing the linguistic and technical limitations of Arab chatbots.

Key Findings:

- No universities have implemented a chatbot, and live chat features are rarely available and limited.
- Sara Chatbot demonstrated a high frequency of failure due to varab in Arab input.
- Linguistic challenges include dialectical differences, morphological questions and orthographic ambiguity.

Critical Analysis:

- Strength: The Study provides a comprehensive evaluation of linguistic and technical aspects with a real world test of a chatbot.
- Weaknesses: No academic chatbot was directly tested; The study only provides ideologically

5. Nabiha: An Arabic Dialect Chatbot

Objectives:

Develop Nabiha, utilizing the Saudi Arabic dialect to assist students with educational inquiries across various platforms.

Methods:

- Collected data consisting of 248 questions and answers from student interactions.
- Implemented an AIML-based chatbot on Android, Twitter, and the web.
- Conducted thorough tests involving 13 IT students.

Key Findings:

- The chatbot was evaluated as somewhat satisfactory.
- Users appreciated the concise interface, though it was limited by consistent responses and a small dataset.
- Accessibility across multiple platforms significantly enriched the experience.

Critical Analysis:

- Strengths: Utilization of established methodologies, integration of various platforms, and reliance on real student data.
- Weaknesses: Limited dataset, lack of practical depth, and absence of AI learning or adaptation capabilities.

6. Arab Chatbot's challenges and solutions: a systematic literature review

Objective:

To systematically review challenges and solutions in Arab Chatbot development, especially in education.

Methods:

- A Prisma-based systematic review of 64 studies spread from 2000 to 2023.
- Analyzed linguistic, technical and practical obstacles.
- Assessment functionality, dataset and performance measurement.

Central findings:

- Large challenges identified include: Budhamfold, anotat data set deficiency, NLP deficiency (natural language treatment) equipment.

- Most existing chatbots are rule -based, while deep learning techniques are rarely used due to insufficient data.
- Recommendations for improvement include the development of hybrid models, construction of strong datasets and implementation of standardized assessment methods.

Critical analysis:

- Strength: Review provides a broad and practical observation with clear recommendations.
- Weaknesses: Input and registration processes were not sufficiently emphasized on focus, and the conclusions have a lack of new empirical data.

7. Intelligent Chatbot for Admission in Higher Education (Aloqayli & Abdelhafez, 2023)

Objective:

Build and test a chatbot to support admission tasks and reduce staff effort by giving students faster answers.

Methods:

- Developed on the Botsify platform.
- Trained mainly with structured FAQs from admission documents.
- Performance tested using accuracy measures and response tracking.

Central Findings:

- Accuracy was about 92% for clear, structured questions.
- Helped reduce repeated queries to staff.
- Worked well in providing quick responses.

Critical Analysis:

- Strengths: High accuracy with fixed inputs, practical design, quick answers.
- Weaknesses: Poor handling of free-text queries, no strategy for adapting to yearly admission changes, limited testing.

8. An Inquiry Smart Chatbot System for Al-Zaytoonah University (Al-Madi et al., 2024)

Objective:

In this work, the authors tried to build a smart chatbot to help students at Al-Zaytoonah University with their admission and registration queries.

Methods:

- Collected data from FAQs and admission documents specific to the university.
- Incorporated some use of the local dialect for better interaction.
- Tested with a small sample of students.

Central Findings:

- Responded well to standard admission questions.
- Students reacted positively to the use of local dialect.
- Improved speed in providing basic information.

Critical Analysis:

- Strengths: Context-specific design, language sensitivity, positive student engagement.
- Weaknesses: Limited to one university, no integration with student information systems, not tested in wider settings.

9. Bilingual AI-Driven Chatbot for Academic Advising (Bilquise et al., 2022)

Objective:

This paper tried to design a chatbot that can talk in both Arabic and English. The idea was to help students get academic advising more easily, instead of having to ask staff all the time.

Methods:

- The authors created the system using Python and added a simple AI-based dialogue handler. It was tested in a basic way to see how well it could respond to student questions in the two languages.
- Tested through small trials with students and functional evaluation.

Central Findings:

- Accuracy 80% in English, 75% in Arabic.
- Helpful for students who switch between the two languages.

Critical Analysis:

- Strengths: Innovative bilingual feature, relevant for multilingual student populations.
- Weaknesses: Limited dataset, no deep focus on dialects, small-scale testing without real-world validation.

10. College Enquiry Chatbot using Rasa Framework (Meshram et al., 2021/22)

Objective:

Create an enquiry chatbot using Rasa to answer student questions about admission and services.

Methods:

- Trained on a predefined dataset of admission FAQs.

- Used Rasa for dialogue management.
- Evaluated with system trials for accuracy and responsiveness.

Central Findings:

- Effective in answering the trained questions.
- Provided fast responses during trials.

Critical Analysis:

- Strengths: The study showed a clear structure and the use of Rasa made the design easy to follow. I also found the implementation simple, which makes it practical for small projects. Weaknesses: However, the system depended too much on its dataset.
- Whenever students asked new or unusual questions, it failed. Another issue is that testing was limited, so the results cannot be trusted for real admission use.

Research Methodology

This research will use a practical step-by-step approach to make sure the objectives can actually be achieved. The work is not about building a perfect chatbot from the start, but about testing what can be done in a real admission and registration setting divided into four main stages. Each stage is designed to move step by step toward achieving the project objectives while keeping the focus on practical testing.

Phase 1: Literature Review

The first thing I need to do is look at what other people have already done with chatbots in universities. I will focus more on the ones that tried to use Arabic or a mix of Arabic and English, because that is closer to my work. The point here is not just to list studies, but to see what actually worked for them and what problems they faced. I expect to find issues like how dialects are handled badly or how most

systems are not really linked to the official student platforms. To do this, I will read different types of materials, from research papers to case studies and even some technical reports.

Phase 2: Data Collection and Preparation

The next step for me is to collect the data that the chatbot will actually learn from. Most of this will come from admission forms, registration rules, and the type of questions that students usually ask every semester. The problem with official documents is that they are often too formal and not written in the way students normally speak. So, I will take that information and rewrite it into simple, everyday questions. Some will be in Modern Standard Arabic, others will sound more casual, and in many cases students mix in English words, so I will include that too. After gathering everything, I will clean the dataset, remove repeated lines, and sort the questions into groups such as deadlines, required documents, or course registration. Finally, I will label each example so that the chatbot can learn what type of intent or category the question belongs to.

Phase 3: System Design and Development

When the dataset is ready, the next step is to start building the chatbot. The idea is to keep it simple but still useful. I will use Python and probably Rasa because it has many tools already built in and it saves time. The system does not need to be fancy, it just has to do a few main things. First, it should figure out what the student is asking about in general. Second, it has to notice the small but important details, like dates or course codes, because without that the answer won't be very helpful. Third, it should keep the conversation going in a way that feels natural, not just stop after one short reply. That happens a lot with weak chatbots, and I want to avoid it. Flexibility is also part of the plan. Some students will type questions in their own words, others will prefer clicking on menu choices, so the design will make room

for both ways. To make testing easier, I will put the chatbot on a simple website. This way, students can just try it directly in their browser without needing to install anything or deal with complicated steps.

Phase 4: Testing and Evaluation

The last stage is testing the chatbot. At first, it will be checked with a prepared list of questions to measure how accurate it is using metrics like precision, recall, and F1 score. After that, I will involve a small group of students who will use the chatbot in realistic admission and registration scenarios. Their opinions will be collected using short surveys to see how clear and helpful the system feels in practice. By putting together the technical results with the student feedback, the project should give a balanced view of how useful the chatbot really is and where improvements are still needed.

Ethical Considerations

No private data will be used in this project. Students who participate in the pilot will give informed consent, and their feedback will stay anonymous.

Limitations

It is expected that the chatbot will not handle all dialects or unexpected questions. The dataset will also be limited in size. These issues will be noted and suggested as areas for future improvement.

Expected Outcomes

Functional Prototype: A chatbot that can answer admission and registration queries in Arabic, with some English support.

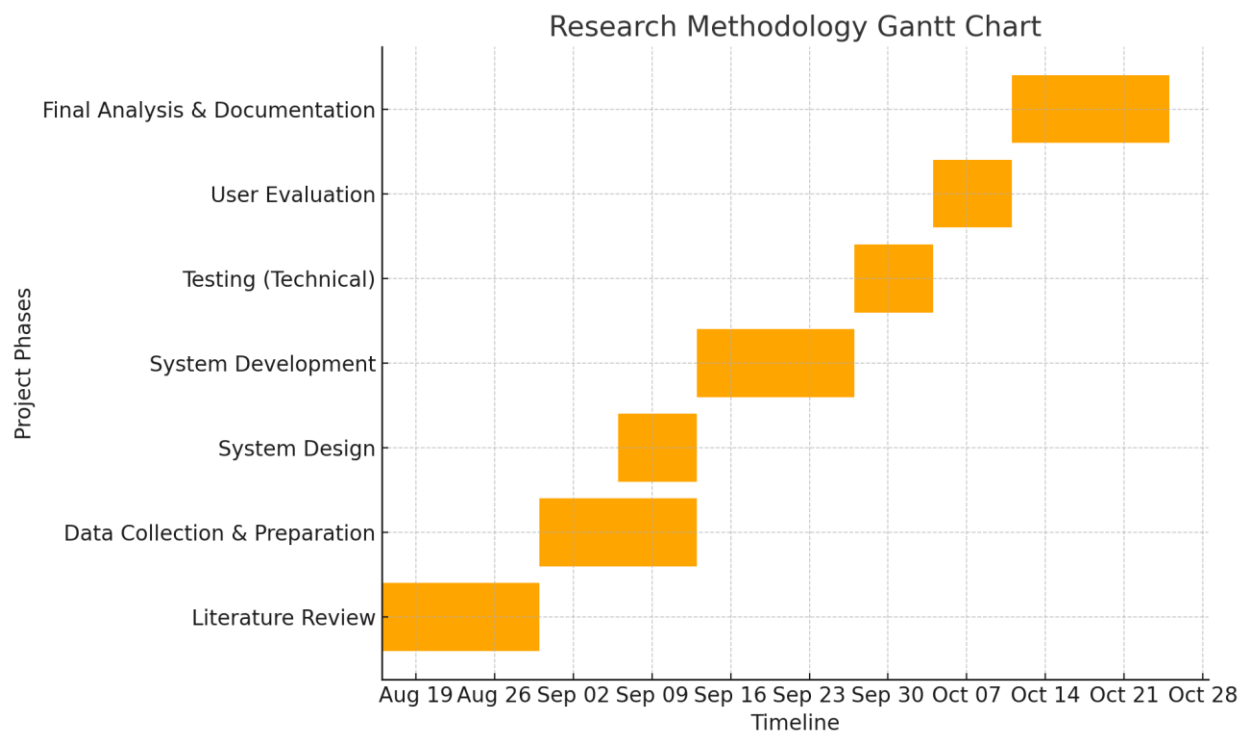
Identification of Challenges: Insights into the main barriers facing Arabic chatbots (dialects, ambiguity, limited datasets).

Practical Value: Evidence of reduced staff workload and quicker access to admission information for students.

Contribution to Research: A case study that highlights both the potential and the current limits of chatbot applications in Arabic higher education.

Foundation for Future Work: A roadmap for improving dataset quality, scaling to multiple universities, and integrating with student information systems.

Gantt chart



References

1. Alabbas and M. Alomar, "Tayseer: AI-powered Arabic chatbot framework," *International Journal of Advanced Computer Science*, vol. 15, no. 2, pp. 45–56, 2024.
2. M. Alazzam, M. Alkhatib, and K. Shaalan, "Arabic educational neural network chatbot," *Journal of Computer Science Research*, vol. 12, no. 4, pp. 220–229, 2023.
3. Al-Abdullatif, F. Al-Shehri, and R. Al-Mutairi, "Bashayer chatbot in Saudi higher education," *Journal of Educational Technology*, vol. 19, no. 3, pp. 87–95, 2023.
4. Almuryah, "Challenges of using Arabic chatbots in Saudi universities," *Arab World Journal of Information Technology*, vol. 8, no. 2, pp. 65–72, 2021.
5. D. Al-Ghadhban and N. Al-Twaireesh, "Nabiha: An Arabic dialect chatbot," *Journal of Natural Language Engineering*, vol. 26, no. 5, pp. 403–420, 2020.
6. S. Ouali and A. El Garouani, "Arabic chatbots challenges and solutions: A systematic review," *Applied Sciences*, vol. 14, no. 6, p. 2547, 2024.
7. Aloqayli and H. Abdelhafez, "Intelligent chatbot for admission in higher education," *Journal of Artificial Intelligence Applications*, vol. 11, no. 1, pp. 33–42, 2023.
8. N. A. Al-Madi, F. Ahmad, and M. Hasan, "An inquiry smart chatbot system for Al-Zaytoonah University of Jordan," *International Journal of Computer Applications*, vol. 48, no. 1, pp. 21–29, 2024.
9. G. Bilquise, S. Ibrahim, and K. Shaalan, "Bilingual AI-driven chatbot for academic advising," *International Journal of Computational Linguistics*, vol. 8, no. 3, pp. 112–121, 2022.
10. S. Meshram, A. Patil, and R. Khan, "College enquiry chatbot using Rasa framework," *International Journal of Computer Science and Mobile Computing*, vol. 10, no. 2, pp. 14–20, 2021.