

[DataVerse – An AI-Powered Data Science Platform]

Final Year Project Proposal

by

Abdul Rehman Amer BSDSF21A001

Muhammad Soban Anjum BSDSF21A007

Nauman Ishaq BSDSF21A044

Advisor: Dr. Muhammad Arif Butt

Department of Data Science

Faculty of Computing and Information Technology

University of the Punjab

Lahore, Pakistan

(2024)

Abstract of the proposal

DataVerse: Revolutionizing University Experience with AI-Powered Data Science Platform combines cutting-edge technologies like **Retrieval-Augmented Generation (RAG)**, **Natural Language Processing (NLP)**, **Optical Character Recognition (OCR)** and **predictive analytics** to create a platform that aggregates admissions with features such as an **Intelligent Chatbot** for instant support, **Dynamic Personalized Faculty Websites** that highlight achievements and research, and **Admission Prediction** tools that analyze historical data for informed decision-making. DataVerse transforms university operations into a highly efficient, connected, and personalized digital ecosystem. Join us in shaping the future of education!

Objectives

The platform's objectives are to:

1. Generic Website:

Dive deep into detailed, dynamic faculty details that include qualifications, research, publications, teaching experience, and ongoing projects. This will contain details like last year's merit, courses' details, prerequisites, previous FYPs, FYPs rules & regulations and road map. There will also be a portion for announcements.

2. Faculty Website Creator with AI-Personalization:

Empower your faculty with the ability to craft their personalized, professional websites using **customizable AI-driven templates**. Each template is designed with flexibility in mind, allowing faculty to highlight their unique strengths, portfolios, and academic achievements in a manner that resonates with students and peers.

3. AI-Powered Intelligent Chatbot:

Powered by state-of-the-art **Retrieval-Augmented Generation (RAG)**, the chatbot is your round-the-clock AI assistant. Whether it's about faculty, courses, or admission processes, the chatbot provides **instant, personalized responses** with unparalleled accuracy. Queries about course prerequisites, professor availability, or even administrative details are just a message away.

4. AI-Powered Admission Prediction:

Streamline the student admission process using an AI system that **validates documents** and calculates admission aggregates effortlessly. The AI cross-references historical data like Matric, Intermediate and entry test results and current performance metrics to **predict admission success**, offering actionable insights for both students and admissions officers.

5. Deployment of the Project:

The Project will be deployed on the cloud and will be accessible to all students, faculty members and others. They will be able to visit the site and interact with it. This project will be open sourced and all the code will be available on github.

Methodology

1. Generic Website:

- **Content Management System (CMS):** Use platforms like WordPress or a custom-built CMS to allow faculty and admins to update profiles, courses, FYPs, and announcements easily.
- **Database Integration:** Develop a relational database (MySQL/PostgreSQL) or NoSQL (MongoDB) to store faculty details, course information, previous FYPs, and roadmaps.
- **Dynamic Frontend Design:** Implement front-end technologies like React to create responsive, dynamic faculty profiles and course pages.

- **RESTful APIs:** Create backend APIs (in Python, Node.js, or Django) to fetch data dynamically for sections like faculty profiles, merit lists, course details, and announcements.

2. Faculty Website Creator with AI-Personalization:

- **Dynamic Templates:** Teachers can customize their websites using predefined, dynamic templates built with React, CSS, HTML, and JavaScript.
- **AI-Powered Content Curation:** Automatically populate website sections by analyzing faculty data, including publications (e.g., from Google Scholar), research fields, and teaching history.
- **Sub-Domains:** Each personalized website will be hosted on a sub-domain of the main portal, implemented using the Django framework.
- **Hyperlinks to Existing Websites:** Faculty members can link to their existing websites through easily accessible hyperlinks, also implemented using Django.

3. AI-Powered Intelligent Chatbot:

- **Retrieval-Augmented Generation (RAG):** Train a chatbot model with **Natural Language Processing (NLP)** frameworks like Hugging Face or OpenAI's GPT using a combination of knowledge base retrieval (database/FAQs) and text generation to answer queries.
- **Custom Query Pipeline:** Implement a pipeline that parses queries, retrieves relevant data from databases, and generates personalized responses.
- **Data Source Integration:** Link the chatbot to structured data like course information, professor schedules, and admission rules through APIs to ensure real-time updates and accurate responses.
- **Continuous Learning:** Use Reinforcement Learning (RL) to train the chatbot to improve response quality based on user feedback over time.
- **Voice Oriented Chatbot:** The chatbot can also communicate with voice. We will be using the text to speech libraries like tortoise etc.

4. AI-Enhanced Admission Prediction:

- **Document Scanning:** Implement Optical Character Recognition (OCR) systems (Google Vision API, Tesseract) to scan documents such as certificates and transcripts.
- **Predictive Analytics:** Use Machine Learning (ML) models (Random Forest, SVM, or Neural Networks) trained on historical admission data (Matric, Intermediate, and test scores) to predict the likelihood of admission based on defined criteria.
- **AI Decision System:** Integrate an AI-based decision system that analyzes aggregate scores and ranks applicants based on university-specific weightings and performance metrics.

Expected Outcomes

1. Dynamic Faculty Profiles:

- **Unlock a wealth of knowledge:** Faculty profiles are more than just bios; they're interactive portals showcasing their research, achievements, and expertise. Students will gain deeper insights, enabling them to connect with the **intellectual leaders of tomorrow**.
- **Transparent course details:** Students will love the **clarity and transparency** provided through detailed course prerequisites, merit lists, and FYP guidelines. It's a one-stop hub for academic planning and decision-making.

- **Instant updates:** Real-time announcements ensure that students never miss out on critical news. From **last-minute schedule changes** to **exclusive academic opportunities**, everything will be at their fingertips.

2. Faculty Website Creator with AI-Personalization:

- **InstaYour professional brand, amplified:** Faculty can now create **stunning, professional websites** in minutes, with AI guiding them every step of the way. These sites are more than just pages—they're **digital portfolios** showcasing research, publications, and achievements to a global audience.
- **Instant AI-curated content:** The AI will auto-generate recommendations, ensuring each website stands out. Faculty members will have **powerful, sleek websites** that reflect their academic journey and innovations.
- **Fully customizable:** With dynamic, AI-enhanced templates, faculty can create **one-of-a-kind websites** that adapt to their teaching style, research focus, and student engagement strategies.

3. AI-Powered Intelligent Chatbot:

- **Personalized guidance, 24/7:** Imagine an AI assistant that's always on! From answering intricate admission questions to suggesting courses based on interests, this chatbot is like having a **personal university advisor** on-call.
- **Next-level responsiveness:** Powered by cutting-edge AI, the chatbot will make interactions feel almost human—**instant, intelligent, and super engaging**. Say goodbye to endless searching, and hello to **immediate, precise answers**.
- **Adaptive learning:** The more it interacts, the smarter it becomes! Over time, it will offer **even more accurate, personalized recommendations**, making it a must-have for students navigating university life.

4. AI-Enhanced Admission Prediction:

- **Seamless, intelligent admissions:** No more manual checks! The AI will handle everything from **document verification** to **aggregate calculation**, ensuring **lightning-fast, error-free admissions**.
- **Admission insights in real time:** The AI not only processes but also predicts admission outcomes with stunning accuracy, using cutting-edge machine learning algorithms. It gives **students actionable feedback** on their chances, making the process transparent and stress-free.
- **Data-driven decisions:** With **intelligent, data-backed admissions**, universities will be able to **streamline the entire process**, saving time, reducing errors, and enhancing the overall experience for applicants.

1. INTRODUCTION

Background

First thing first, there is no Data Science website for the PUCIT. In today's educational landscape, universities face increasing demands to manage a vast amount of data, engage effectively with students and faculty, and streamline administrative processes. Traditional methods of handling faculty information, student queries, and admissions are often manual and time-consuming, leading to inefficiencies and potential inaccuracies. As universities continue to expand and adopt digital tools, there is a pressing need for a more integrated, intelligent approach to managing academic and administrative tasks.

Artificial Intelligence (AI) has emerged as a transformative technology capable of addressing these challenges by automating repetitive tasks, providing data-driven insights, and enhancing user interactions

through intelligent systems. The integration of AI into university systems can significantly improve efficiency, accuracy, and user experience.

Problem Statement

1. Inefficient Management of Faculty Information:

- **Unmet Need:** Universities need an effective way to manage and present faculty information, such as qualifications, research publications, and teaching schedules, in a centralized and accessible manner. Traditional methods involve manual updates and scattered data sources, which can lead to outdated information and increased administrative workload.
- **Significance:** Inefficiencies in managing faculty information can result in inaccuracies, missed collaboration opportunities, and increased time and effort for administrative staff. In large institutions, managing hundreds to thousands of faculty profiles manually can consume significant resources, leading to potential data discrepancies and administrative bottlenecks.

2. Fragmented Student-Faculty Communication:

- **Unmet Need:** There is a need for a streamlined communication system that facilitates real-time interactions between students and faculty. Traditional communication methods such as email and face-to-face meetings are often slow and can lead to delays in obtaining crucial information.
- **Significance:** Delays in communication can affect student satisfaction and academic performance. For instance, if response times average 24 to 48 hours, students may face delays in receiving important information or support, impacting their overall academic experience.

3. Error-Prone Admissions Prediction:

- **Unmet Need:** There is a need for an AI-driven system that can predict admission outcomes based on the academic documents students provide, helping students understand their chances of admission more accurately and efficiently.
- **Significance:** Relying on manual assessment for admission predictions can slow down the process and lead to inconsistent results. By using AI to predict admissions based on the provided documents, institutions can improve transparency and provide prospective students with quick and reliable insights. For example, AI predictions can significantly reduce the time it takes to provide feedback, potentially reducing wait times by 50% or more.

In Scope:

- Development of the generic website.
- Creation and integration of an AI chatbot for answering queries.
- Development of a centralized database for faculty profiles, Development of a tool for faculty to create personalized websites.
- Automation of the admissions process, including document verification and admission prediction.

Out of scope:

- AI Avatar that will be shown upon a query.
- Emotion recognition from voice inputs or real-time processing of audio or video.
- Handling or integration with external third-party educational databases beyond basic document verification.
- A comprehensive CGPA maintainer for every student.
- A state of the art plagiarism detection.
- Facial recognition system for security.

Limitations

- The platform will primarily support English-language content, with potential limitations in multi-language support based on resources.
- Real-time community features will be developed with basic moderation and may require further refinement for large-scale deployment.

2. LITERATURE REVIEW

Related Work: The integration of AI-powered chatbots in education has become increasingly popular, with several platforms offering innovative solutions. Two notable chatbot platforms, **Futr.ai** and **Verge AI**, have been instrumental in transforming how educational institutions interact with students.

- <https://futr.ai/#features>
- <https://verge-ai.com/chatbot-for-higher-education/>
- <https://gamma.app/> is kinda like having a dynamic website template building.

Gap Analysis:

Generic Website

- **Current Gaps:**
 - Many educational websites provide static faculty profiles, lacking detailed, dynamic information on qualifications, ongoing research, and real-time updates.
 - Information on past FYPs, merit requirements, and course roadmaps is often scattered across different pages, making it difficult for students to find comprehensive data in one place.
 - Announcements are generally scattered and not integrated into the same platform, leading to missed important updates.
- **Proposed Solution:**
 - A centralized platform that consolidates all academic and administrative information (faculty profiles, FYP rules, merit details, etc.) into a single, easily accessible hub.

AI-Powered Intelligent Chatbot

- **Current Gaps:**

- Most websites either do not offer a chatbot, or if they do, it's a simple, rule-based chatbot that fails to provide meaningful, personalized responses.
- The lack of an AI-powered chatbot means users often face delays in getting their queries answered, particularly outside working hours or for administrative questions.
- Current bots do not integrate advanced AI such as RAG, which leads to less precise and context-aware responses.

- **Proposed Solution:**

- Implement a Retrieval-Augmented Generation (RAG) chatbot for round-the-clock, highly accurate and personalized assistance related to admissions, courses, and faculty availability.

AI-Powered Admission Prediction

- **Current Gaps:**

- Universities do not have a module of admission prediction in their universities portal or website.
- Prospective students often lack personalized guidance based on their academic history, which makes it difficult for them to assess their chances of being admitted to a university.
- Current systems do not take advantage of historical data or predictive AI models to offer insights into potential admission outcomes.

- **Proposed Solution:**

- Develop an AI-powered system that predicts admission outcomes based on student-provided documents, calculating admission aggregates and performing predictive analysis using historical data. This would provide real-time insights on a student's likelihood of admission, giving them more transparency and helping them make informed decisions without replacing the manual admissions process.

3. Project Overview

Project Title: DataVerse – An AI-Powered Data Science Portal			
Group Leader: M. Soban Anjum			
Project Members:			
Name	Registration #	Email Address	Signature
Abdul Rehman Amer	BSDSF21A001	bsdsf21a001@pucit.edu.pk	
Muhamamd Soban Anjum	BSDSF21A007	bsdsf21a007@pucit.edu.pk	
Nauman Ishaq	BSDSF21A044	bsdsf21a044@pucit.edu.pk	
Project Goal: <ul style="list-style-type: none">• Functional Generic Website• AI Enabled Chat Bot with voice support• Faculty’s personalized website• AI-Driven Admission Prediction• Scalable and Modular			
Objective:			
Functional Generic Website	Create an AI tool for faculty to build personalized academic websites with templates, featuring merit lists, course details, prerequisites, previous FYPs, and announcements.		
Implement an AI-Powered Chatbot	Develop a voice-enabled chatbot using Retrieval-Augmented Generation (RAG) to provide accurate, context-rich responses to queries about faculty information, university policies, and complex inquiries on the academic roadmap and courses.		
Provide Faculty Website Creation Tools	Develop a tool to enable faculty members to create and maintain personalized academic websites using customizable templates.		
Predicting the Admission Process	Implement AI-based systems for document verification, aggregate calculation, and admission prediction.		

Deployment of the Website & AI solutions	Design and deploy AI-powered systems with seamless integration into university infrastructure, ensuring scalability, security, and user accessibility. Deploy the website using scalable cloud infrastructure for high availability, real-time updates, and user-friendly access to AI tools.
--	---

Project Success Criteria:
Assumptions: <ul style="list-style-type: none"> • The data provided (faculty and student) will be accurate and up-to-date. • We will have comprehensive data on faculty members, including their qualifications and expertise. • Previous students' academic and admission data will be accessible for analysis. • Detailed admission information, including weightage criteria and evaluation methods, will be available. • All faculty members possess basic knowledge of website operation and maintenance. • The Internet is required.
Risks & Obstacles: <ul style="list-style-type: none"> • Modular • Incompletion in case of no data • Degraded performance in case of no admission data • Scalability • Limited financial resources may impact the quality of the tools and technologies
Organization Address:
Target End Users: <ul style="list-style-type: none"> • Prospective Students • Faculty Members • Students of data science • University Administrators • Layman
Suggested Project Supervisor: <ul style="list-style-type: none"> • Dr. Muhammad Arif Butt
Approved By:
Date:

4. Tools and technologies used with reasoning

Frontend Development

1. React.js

Reasoning: React.js is a popular library for building user interfaces, particularly single-page applications where dynamic updates are needed without full page reloads. Its component-based architecture allows for reusable UI components, which is essential for a complex application like a university portal. React's strong community support and extensive ecosystem also offer numerous libraries and tools that can be integrated to enhance functionality.

2. Tailwind CSS or Bootstrap

Reasoning: Tailwind CSS/ Bootstrap CSS offers a utility-first approach to styling, which allows for rapid design and customization while maintaining consistency.

3. WebSockets (via Socket.io)

Reasoning: For implementing real-time features like chat or notifications, WebSockets are essential. Socket.io provides an easy-to-use API for real-time, bidirectional communication between clients and servers, which is ideal for community engagement and live interactions.

Backend Development

1. Django

Reasoning: Django is a high-level Python web framework that promotes rapid development and clean, pragmatic design. It includes a robust ORM (Object-Relational Mapping) system for database interactions, built-in admin interface for managing data, and strong security features. Django is well-suited for handling the backend logic of a university portal, including user management, content management, and integration with AI services.

2. GraphQL or REST API

Reasoning: GraphQL allows for more flexible and efficient querying of data compared to traditional REST APIs, particularly useful when the frontend requires complex data interactions. REST APIs, however, are simpler and widely used. The choice depends on the complexity of the data interactions and the team's familiarity with the technologies.

AI and Machine Learning

1. NLP Models (BERT, GPT, Lama, Gemini)

Reasoning: For the AI-powered chatbot, using advanced NLP models such as Lama or GPT can enhance the chatbot's ability to understand and respond to queries with contextually accurate answers. These models provide high-quality text generation and comprehension capabilities.

2. TensorFlow or PyTorch

Reasoning: Both TensorFlow and PyTorch are leading frameworks for building and training machine learning models. TensorFlow is known for its robust production tools and deployment options, while PyTorch is favored for its ease of use and dynamic computation graph. Either framework can be used to develop models for document verification and admission prediction.

3. OCR Technology (Tesseract or AWS Textract)

Reasoning: OCR technology is crucial for extracting text from uploaded documents. Tesseract is an open-source OCR engine, while AWS Textract offers a more advanced and scalable solution for extracting and analyzing text from documents.

4. MLlib (from Apache Spark)

Reasoning: For handling large-scale data processing and predictive analytics related to admissions, MLlib provides scalable machine learning algorithms and data processing capabilities that can handle big data efficiently.

- **Scikit-Learn:** Ideal for building predictive models to calculate admission likelihood based on student data.
- **RPA (Robotic Process Automation) with UiPath or BluePrism:** These can be employed to automate repetitive tasks like data entry, document verification, and calculating aggregates.

Database and Storage

1. PostgreSQL or MongoDB or MySQL or Sqlite or Microsoft SQL

Reasoning: PostgreSQL is a powerful, open-source relational database that provides advanced features like transactional integrity and complex queries. MongoDB is a NoSQL database suitable for storing unstructured data. The choice depends on the data structure and requirements of the project.

2. Cloud and DevOps

- Azure
- AWS
- Google Cloud

Authentication and Security

1. AI-Enhanced Security:

Reasoning: Implementing AI-based security measures, such as anomaly detection and fraud prevention algorithms, can enhance the overall security of the platform by identifying and mitigating potential threats.

2. No SQL injection:

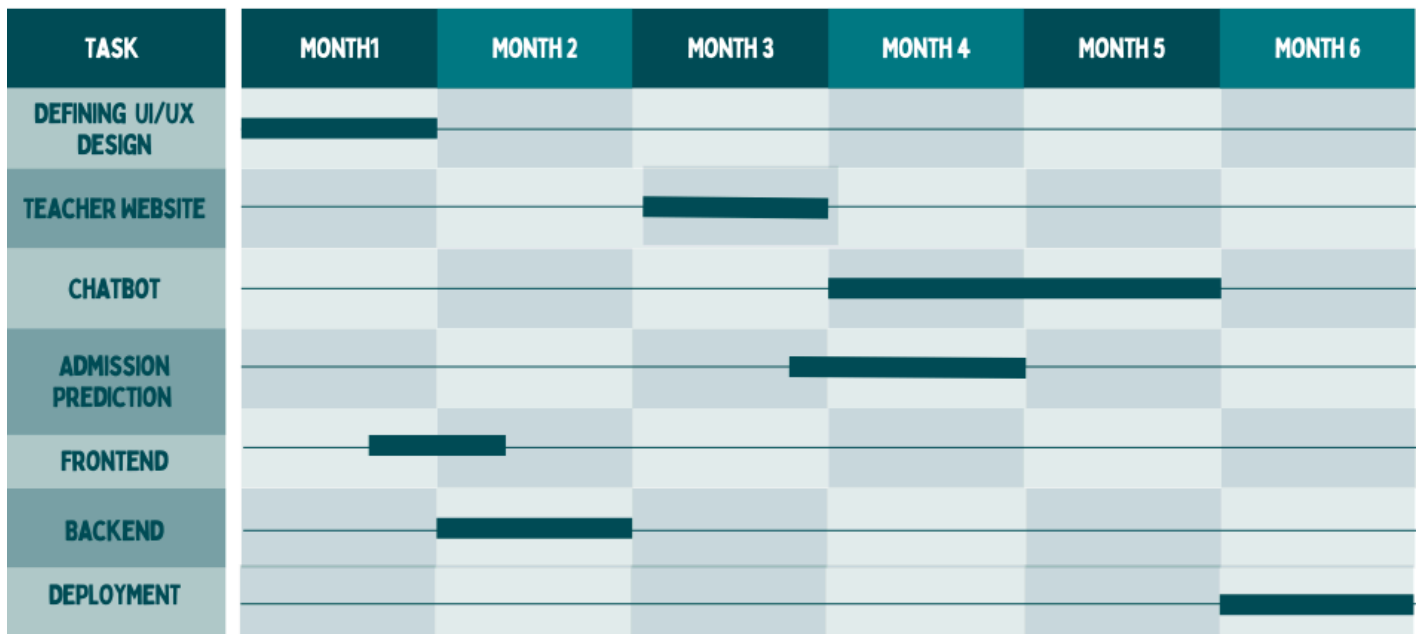
Reasoning: Our website will handle any sql injection and will not crack with these strategies.

3. Regular Checks:

Reasoning: We will be doing regular security checks to test our security and to see if there is any vulnerability in the code or the website. And this is an open source project, the community will help it to make it stronger and stronger day by day.

4. WORK DIVISION

GANTT CHART



- **Frontend:** This will be led by Abdul Rehman Amer - BSDSF21A001.
- **Backend:** This will be led by Nauman Ishaq - BSDSF21A044.
- **Machine Learning and AI:** This will be managed and pipelined by M. Soban Anjum - BSDSF21A007.

5. REFERENCES

- [1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*, Long Beach, CA, USA, 2017, pp. 6000-6010.
- [2] F. Rosenblatt, *Perceptrons: An Introduction to Computational Geometry*. Washington, DC, USA: U.S. Government Printing Office, 1958.
- [3] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533-536, Oct. 1986. DOI: 10.1038/323533a0.
- [4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778.
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS)*, Lake Tahoe, NV, USA, 2012, pp. 1097-1105.
- [6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS)*, Montreal, Canada, 2014, pp. 2672-2680.
- [7] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484-489, Jan. 2016. DOI: 10.1038/nature16961.
- [8] React, "Introduction to React," [Online]. Available: <https://react.dev/>. [Accessed: 29-July-2024].
- [9] Tailwind CSS, "Installation," [Online]. Available: <https://tailwindcss.com/docs/installation>. [Accessed: 02-Aug-2024].
- [10] Bootstrap, "Introduction," [Online]. Available: <https://getbootstrap.com/docs/5.3/getting-started/introduction/>. [Accessed: 06-Aug-2024].
- [11] MongoDB, "MongoDB," [Online]. Available: <https://www.mongodb.com/>. [Accessed: 24-Aug-2024].
- [12] Django, "The Web framework for perfectionists with deadlines," [Online]. Available: <https://www.djangoproject.com/>. [Accessed: 26-Aug-2024].
- [13] Django Documentation, "Django 5.1 documentation," [Online]. Available: <https://docs.djangoproject.com/en/5.1/>. [Accessed: 03-Sep-2024].

- [14] Django REST Framework, "Django REST framework," [Online]. Available: <https://www.django-rest-framework.org/>. [Accessed: 14-Sep-2024].
- [15] MySQL, "MySQL," [Online]. Available: <https://www.mysql.com/>. [Accessed: 20-Sep-2024].
- [16] PostgreSQL, "PostgreSQL," [Online]. Available: <https://www.postgresql.org/>. [Accessed: 20-Sep-2024].
- [17] Python, "Python," [Online]. Available: <https://www.python.org/>. [Accessed: 20-Sep-2024].
- [18] Comm100, "Higher Education Chatbots: A Complete Guide," [Online]. Available: <https://www.comm100.com/blog/higher-education-chatbots/>. [Accessed: 18-August-2024].
- [19] Tidio, "How to Create a Chatbot for a Website," [Online]. Available: <https://www.tidio.com/blog/how-to-create-a-chatbot-for-a-website/>. [Accessed: 29-Aug-2024].
- [20] ChatBot, "How to Build Your Chatbot," [Online]. Available: <https://www.chatbot.com/help/build-your-chatbot/how-to-build-your-chatbot/>. [Accessed: 01-Sep-2024].
- [21] Topflight Apps, "How to Build a Chatbot," [Online]. Available: <https://topflightapps.com/ideas/how-to-build-a-chatbot/>. [Accessed: 04-Sep-2024].
- [22] Docsbot.ai, "Docsbot," [Online]. Available: <https://docsbot.ai/>. [Accessed: 20-Sep-2024].
- [23] ChatBot, "Documentation," [Online]. Available: <https://www.chatbot.com/docs/>. [Accessed: 21-Sep-2024].
- [24] Sendbird, "AI Chatbot Overview," [Online]. Available: <https://sendbird.com/docs/ai-chatbot/guide/v1/overview>. [Accessed: 22-Sep-2024].