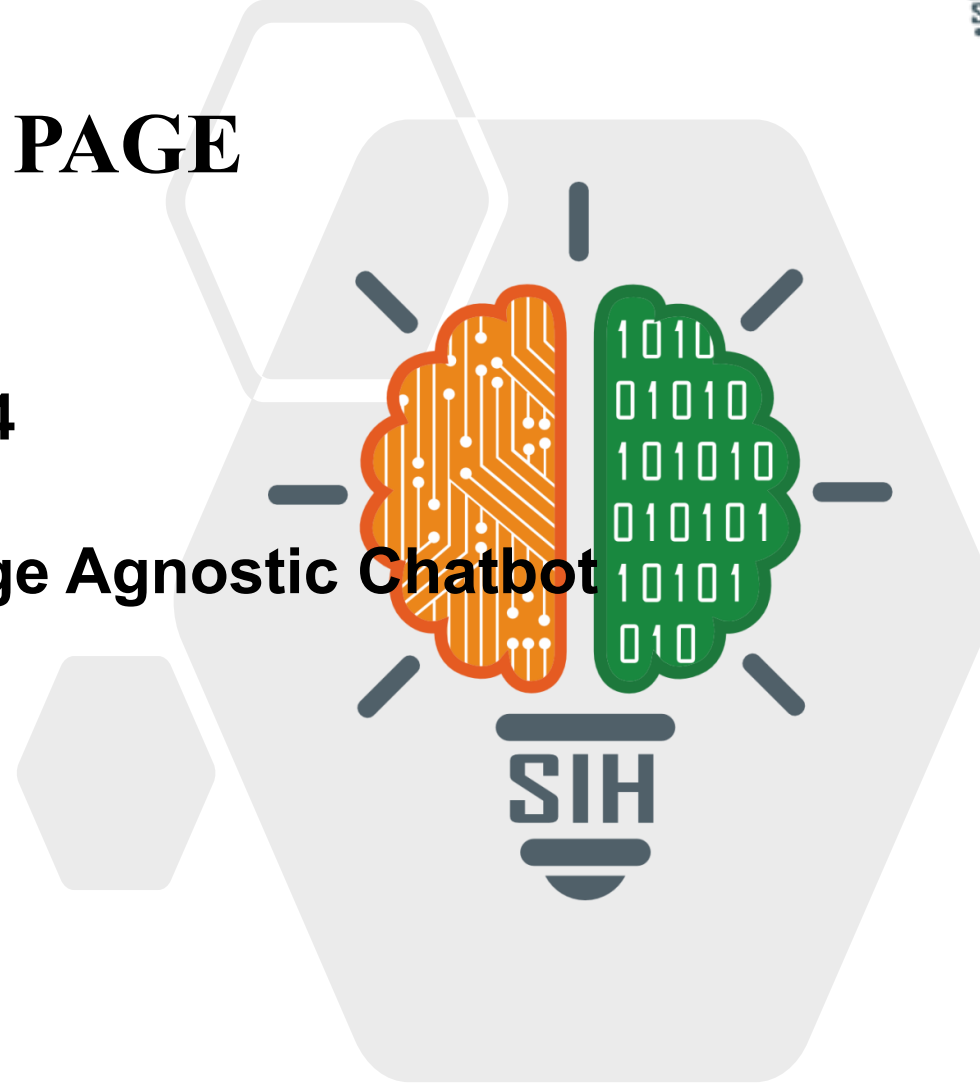


## TITLE PAGE

- Problem Statement ID – SIH25104
- Problem Statement Title- Language Agnostic Chatbot
- Theme- AI/ML
- PS Category- Software
- Team ID-
- Team Name - AXORA





AXORA

# SAARTHI



## ❖ Proposed Solution

- Our chatbot will ingest institutional FAQs, circulars, and PDFs to create a robust knowledge base. It will use **Large Language Models (LLMs) with Retrieval-Augmented Generation (RAG)** for advanced intent recognition, multilingual query handling, and context management, providing accurate and personalized answers to student queries in a conversational manner.

## ❖ Addressing the problem

- **Eliminate queues:**  
Student gets immediate answer to repetitive queries instead of waiting a long time.
- **Breaks Language Barriers:**  
Support queries in Hindi, English and at least five regional languages.
- **Frees up staff:**  
Deflects routine inquiries, allowing university staff to focus on complex, high-priority tasks.

## ❖ Uniqueness of the solution

- **Document-grounded answers:** no hallucination because all responses are pulled from circulars/FAQs.
- **Multilingual Context Management:** The ability to maintain conversation flow in various languages is a key differentiator.
  - **Scalable multilingual support:** Future-proof for adding new regional languages.
- **Lightweight deployment** (volunteers can update FAQs by just uploading a new PDF).
- **Holistic Integration:** It will integrate seamlessly with the college site and messaging platform, maximizing its reach.

### FAQ Chatbot

Ask questions about admissions / reporting. (Uses fuzzy matching against the extracted FAQ JSON)

Your question

E.g. Where do I report at RTU?

Ask

> Show full FAQ list

Prototype link

<https://sih-chatbot.streamlit.app/>



AXORA

# TECHNICAL APPROACH



SMART INDIA  
HACKATHON  
2025

## ❖ Technology stack



### LLM\_Engine:

Open-source LLMs (e.g., LLaMA 3, Mistral, GPT-compatible APIs) with RAG pipeline for intent recognition and accurate responses.



### Programming:

Python for backend logic with Flask/Django.



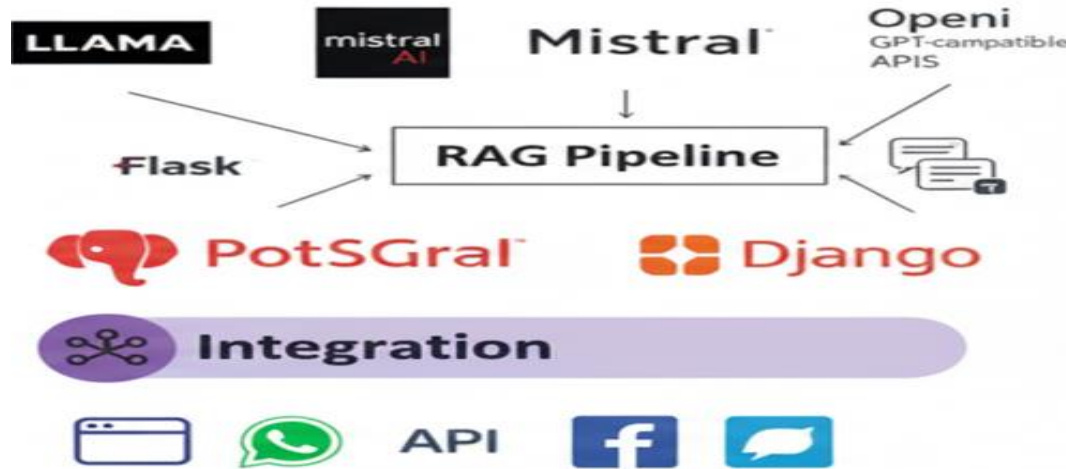
### Databases:

PostgreSQL or MongoDB for logging conversations and storing the knowledge base.

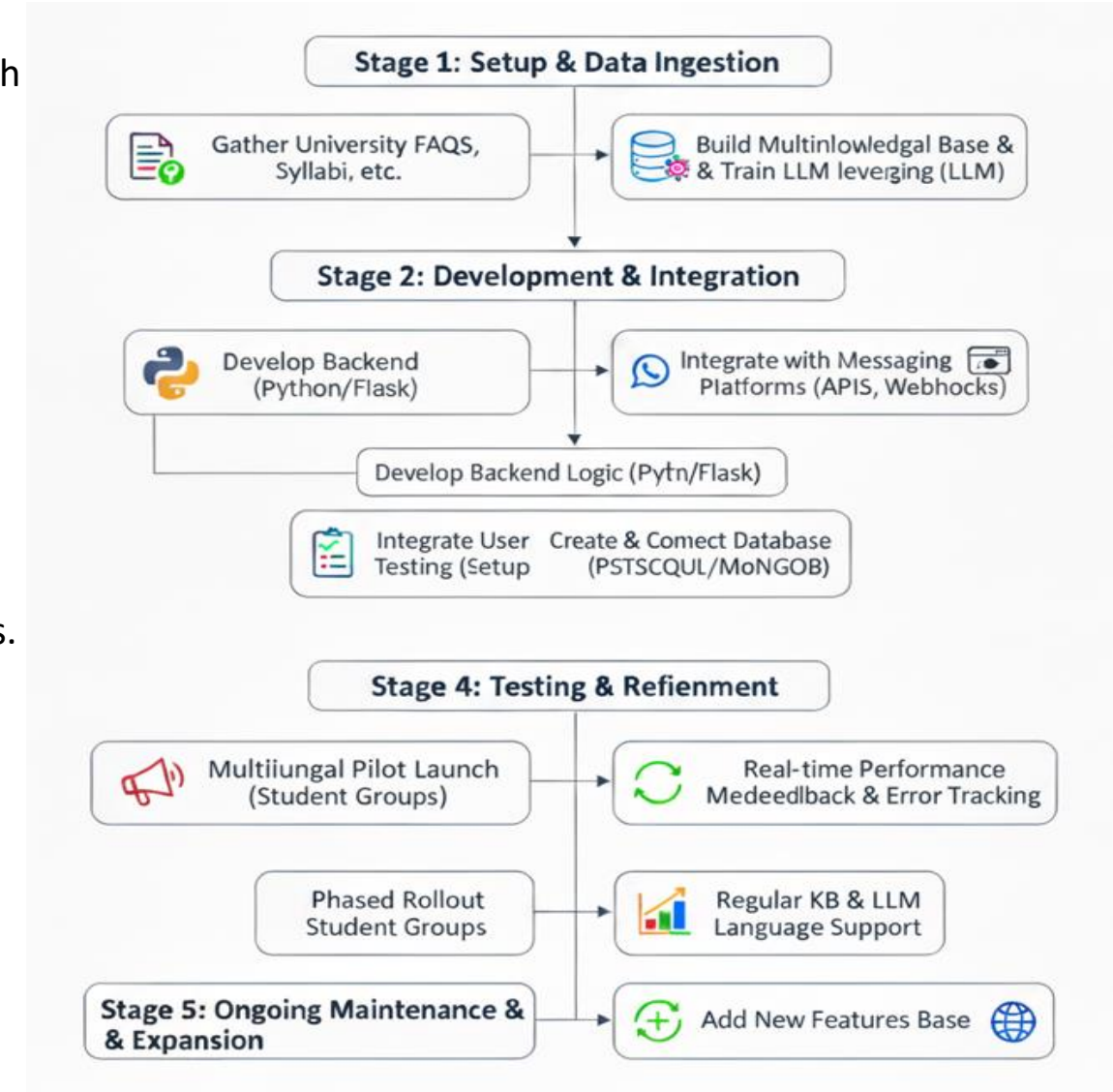


### Integration:

Web hooks and APIs for connecting to web and messaging platforms.



## ❖ Implementation & Methodology





AXORA

# FEASIBILITY AND VIABILITY



## Feasibility

- The core technology (LLMs) is available open-source.
- Data is readily available within educational institutions.
- The solution is highly scalable to accommodate a large number of users and queries.



## Challenges and Risks

- **Data Accuracy:** Ensuring the knowledge base is always up-to-date and accurate to prevent misinformation.
- **Regional Language Nuances:** Accurately recognizing colloquialisms and slang in different regional languages can be a challenge.
- **Privacy:** Ensuring all interactions respect user privacy and data security.



## Solution

- **Continuous Improvement:** Regularly review conversations and update the knowledge base.
- **Advanced AI for Intent Recognition:** Utilize pre-trained multilingual LLMs and RAG
- **Data Security & Privacy:** Implement robust protocols and encryption.

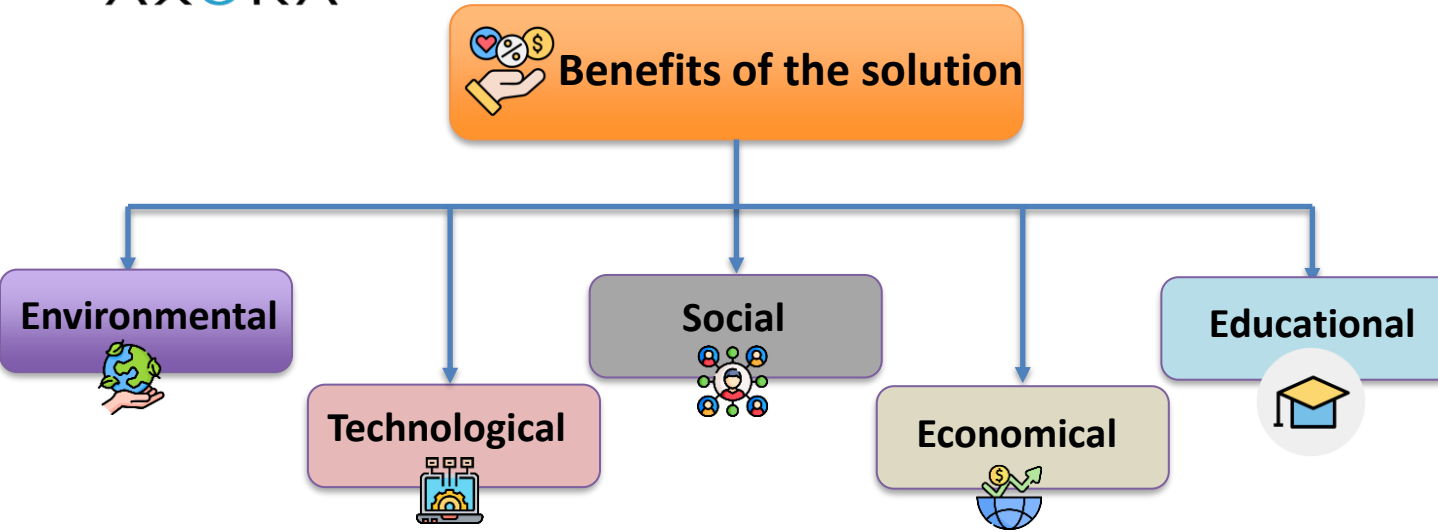


## Business Potential

- **Cost savings:** Reduces the workload of administrative staff, cutting HR costs.
- **SaaS Model** Charge institutions a monthly/annual fee per student or per campus.
- One chatbot can handle thousands of queries simultaneously **Subscription**

### *Factual Corner*

- **Large Target Market** – India has **1,000+ universities and 40,000+ colleges**; most face admission/query overload. Even a small adoption (1–2%) gives thousands of potential customers.
- **Accessibility Advantage** – Unlike traditional chatbots, our system supports **regional languages**, making it more inclusive for Tier-2/Tier-3 cities.
- **Plug-and-Play Feasibility** – No coding needed to update FAQs; just upload PDFs. This ensures adoption without needing a technical team.
- **Competitive Edge** – Most existing chatbots are either English-only or expensive SaaS products. Our solution is **cheaper, open-source friendly, and multilingual**.



## ❖ Potential impact on the target audience

- **Students:** Instant, 24/7 access to information, reducing stress and confusion.
- **College Staff:** Reduced workload on administrative staff, allowing them to focus on complex, human-centric tasks.

## ❖ Benefits of the solution

### Social:

Promotes **equitable access** to information for students from diverse linguistic backgrounds.

### Economical:

Reduces operational costs associated with manual query handling and staff time.

### Educational:

Provides a modern, user-friendly platform that improves the overall student experience.

### Environmental:

- Reduced Paper Consumption
- Lower Carbon Footprint (Travel & Commuting)
- Promoting Digital-First Mindset

### Technological :

- **24/7 Availability:** Students can get answers anytime
- **Multilingual Support:** Breaks down language barriers.
- **Instant Information Retrieval:** Students receive immediate, satisfaction.





# RESEARCH AND REFERENCES



- **Details / Links of the reference and research work**
  - Official SIH 2025 Problem Statement <https://www.sih.gov.in/sih2025PS>
  - LLM + RAG Documentation
    - Langchain <https://python.langchain.com/docs/tutorials/rag/>
    - Haystack [https://docs.haystack.deepset.ai/docs/get\\_started](https://docs.haystack.deepset.ai/docs/get_started)