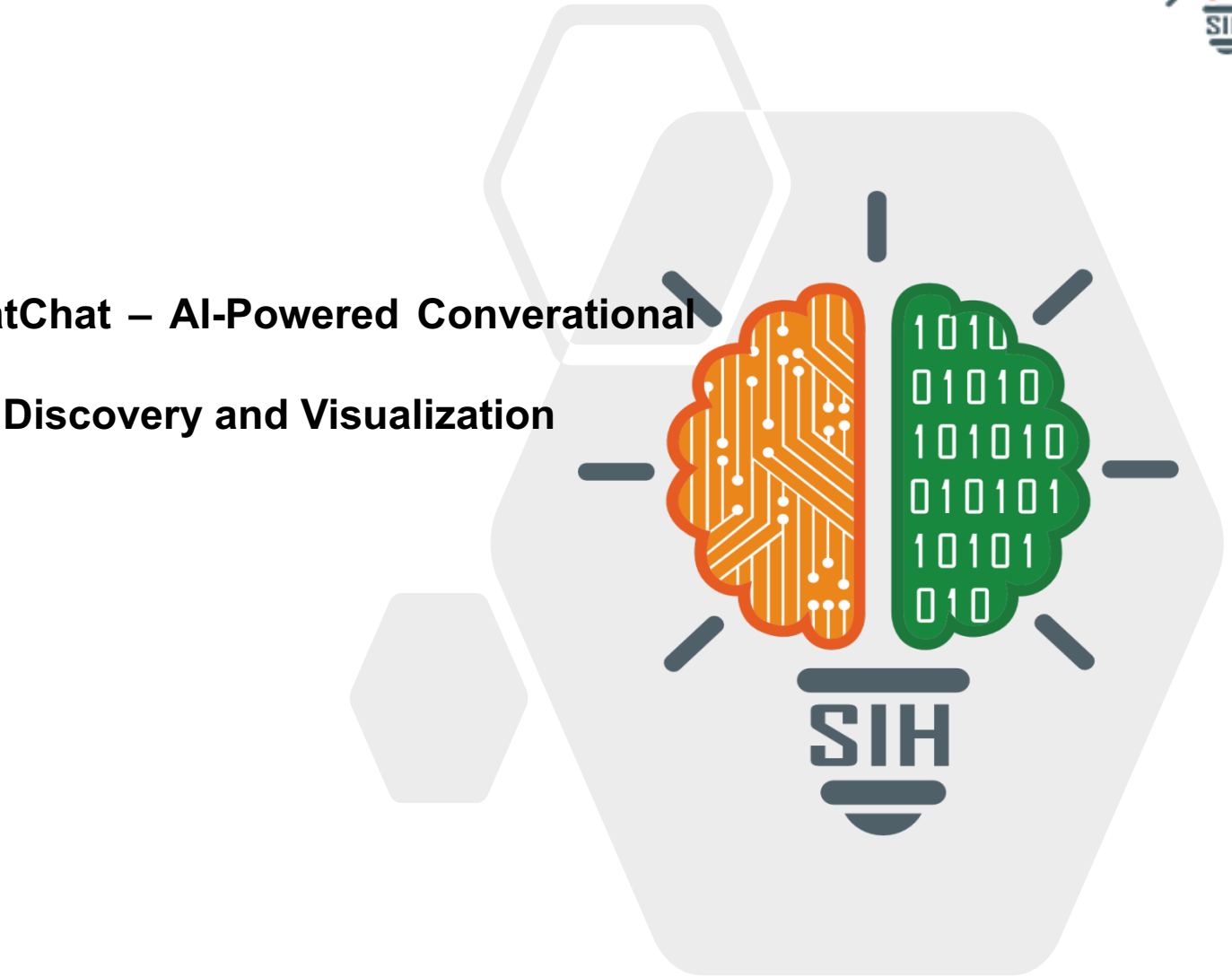
SMART INDIA HACKATHON 2025

SMART INDIA HACKATHON 2025

- **Problem Statement ID – 25040**
- **Problem Statement Title – FloatChat – AI-Powered Conversational Interface for ARGO Ocean Data Discovery and Visualization**
- **Theme - Miscellaneous**
- **PS Category – Software**
- **Team ID -**
- **Team Name – Vision Flux**



Vision Flux

OceanAI:Potential challenges and their solutions

SMART INDIA HACKATHON 2025

- **Challenge:** Ocean data in NetCDF files is often large and complex, making it difficult to query and retrieve efficiently. AI-based natural language queries can produce inaccurate results, and real-time visualization of multiple parameters like depth, salinity, and temperature remains a major challenge.
- **Solution:**
  - **Structure Data Access:** Convert NetCDF files into SQL(Parquet) for fast, easy retrieval of oceanographic information.
  - **Intelligent Query Engine:** Use RAG-powered AI to interpret natural language queries and return accurate, context-aware results.
  - **Rapid Data Search:** Implement vector database (CROMA DB) for instant retrieval of float metadata and measurements.
  - **Dynamic Visualization Hub:** Modular interactive plots and maps using Plotly and Leaflet for depth-time profiles, float trajectories, and parameter comparisons..
- **Unique Features:**
  - **Conversational Data Access:** Interact with complex ARGO ocean data using natural language queries, making insights accessible to both experts and non-technical users.
  - **Real-Time Interactive Visualization:** Instant depth-time plots, float trajectories, and parameter comparisons, enabling users to explore oceanographic trends dynamically.

Vision Flux

TECHNICAL APPROACH

SMART INDIA HACKATHON 2025

Technology Stack

- **Front End:** Streamlit
- **Backend:** Flask, MCP sever, Python
- **Database:** Open Source Data Lake(Iceberg), CROMA DB
- **Gen AI:** LangChain, LangGraph, RAG pipeline, Azure AI foundry
- **Cloud & Deployment tools:** Docker, Azure Fabric/Databricks EMR
- **Miscellaneous:** xarray, netCDF4, (Reading and processing ARGO NetCDF files), Pandas, Numpy(Data manipulation), GeoPandas (Geospatial data handling for maps), PySpark
- **ETL:** Azure Data Factory , Azure Databricks

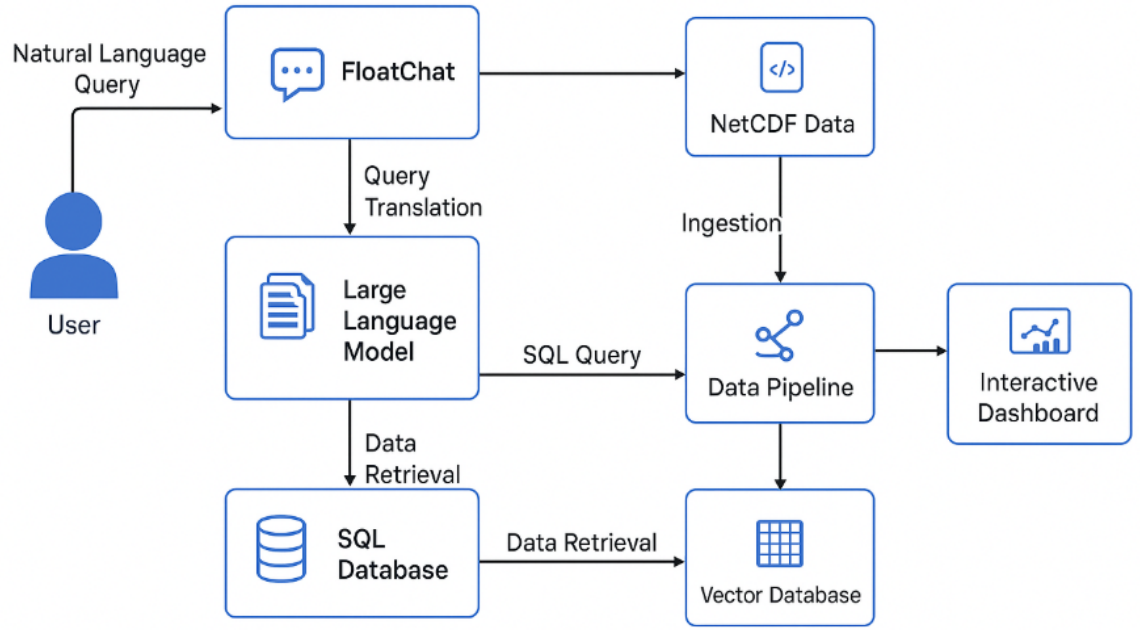


Fig2: Architecture diagram of the proposed solution

Vision Flux

FEASIBILITY AND VIABILITY

SMART INDIA HACKATHON 2025

Feasibility

- **Technical**
  - Can be built quickly using Python and open-source tools.
  - Easy to expand later to include more ocean data.
  - AI ensures accurate answers from natural language questions.
  - Cloud and Docker make it easy to run anywhere.
- **Economic**
  - Low cost to build using free tools.
  - Helps scientists and policymakers make better decisions.
  - Can be adopted by research institutes and government agencies in the future.

Potential Challenges and Strategic Solutions

- **Risk:** Users may struggle to interact with complex ARGO ocean data.
- **Strategy:** Implement a simple chatbot interface with interactive visualizations to make data easy to access and understand.

- **Risk:** AI may provide incorrect or irrelevant responses.
- **Strategy:** Use Retrieval-Augmented Generation (RAG) to ensure answers are grounded in actual database records.

- **Risk:** Slow performance due to large datasets.
- **Strategy:** Optimize queries and use vector databases (FAISS) for fast data retrieval.

- **Risk:** Difficulties in scaling to more datasets like satellite or glider data.
- **Strategy:** Design a modular and cloud-ready architecture for easy expansion

Vision Flux

IMPACT AND BENEFITS

SMART INDIA HACKATHON 2025

Vision Flux

RESEARCH AND REFERENCES

SMART INDIA HACKATHON 2025

1. https://argo.ucsd.edu/data/data-visualizations/
2. https://www.digitalocean.com/resources/articles
3. https://www.capellasolutions.com/blog/faiss-vs-chroma-lets-settle-the-vector-database-debate.in/
4. https://www.ukargo.net/data/view_data
5. https://biogeochemical-argo.org/data-tools.php
6. https://www.researchgate.net/publication/33549908_The_Argo_Program_Observing_the_Global_Ocean_with_Profiling_Floats