SMART INDIA HACKATHON 2025

- **Problem Statement ID –** SIH25080
- **Problem Statement Title-** Document overload at Kochi Metro Rail Limited (KMRL)- An automated solution
- **Theme-** Smart Automation
- **PS Category-** Software
- **Team ID-** 79677
- **Team Name-** Nityam

DocSetu

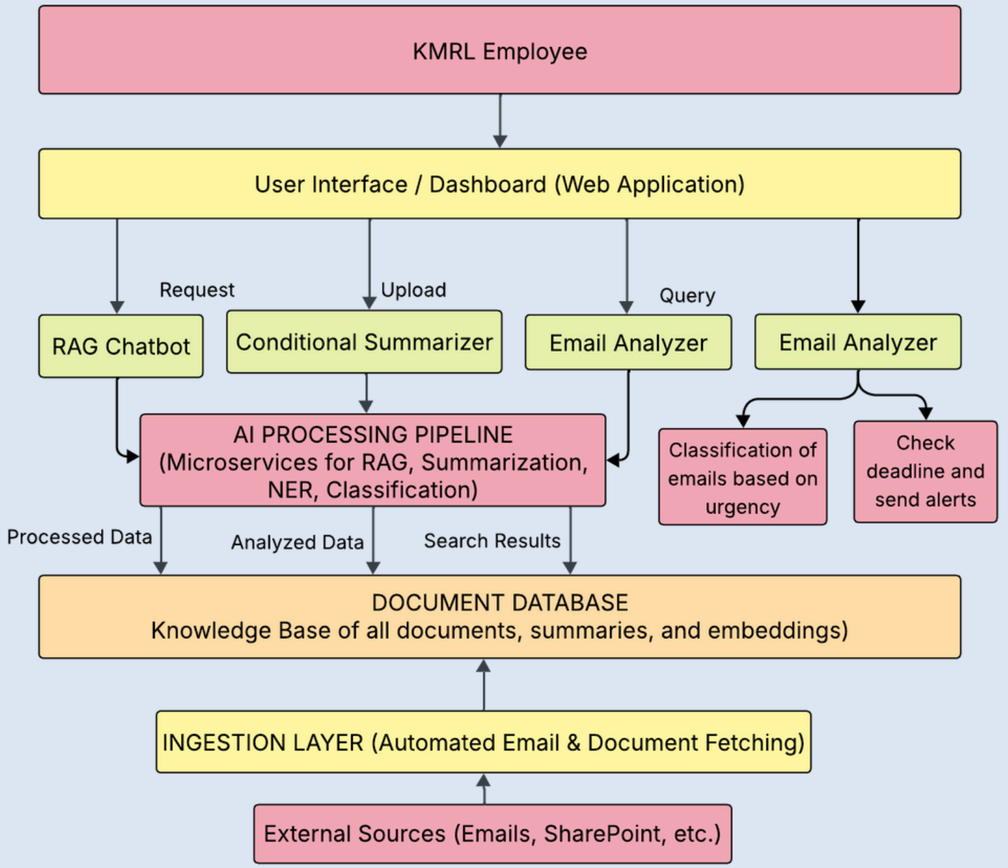
Proposed Solution:

- **Automated Summaries:** Automatically summarizes reports and extracts important data so leaders can make quicker decisions.
- **Conflict Detection:** Automatically flags documents for conflicts using a Knowledge Graph.
- **Automated Compliance:** Two AI agents auto-classify emails and documents to extract actions and deadlines, creating a sorted dashboard card.
- **Semantic Search:** Provides users with direct, synthesized answers to questions they are asking.
- **Perspective Summaries:** Generates summaries with user perspective from one document.

How AI-ML Solves the Problem:

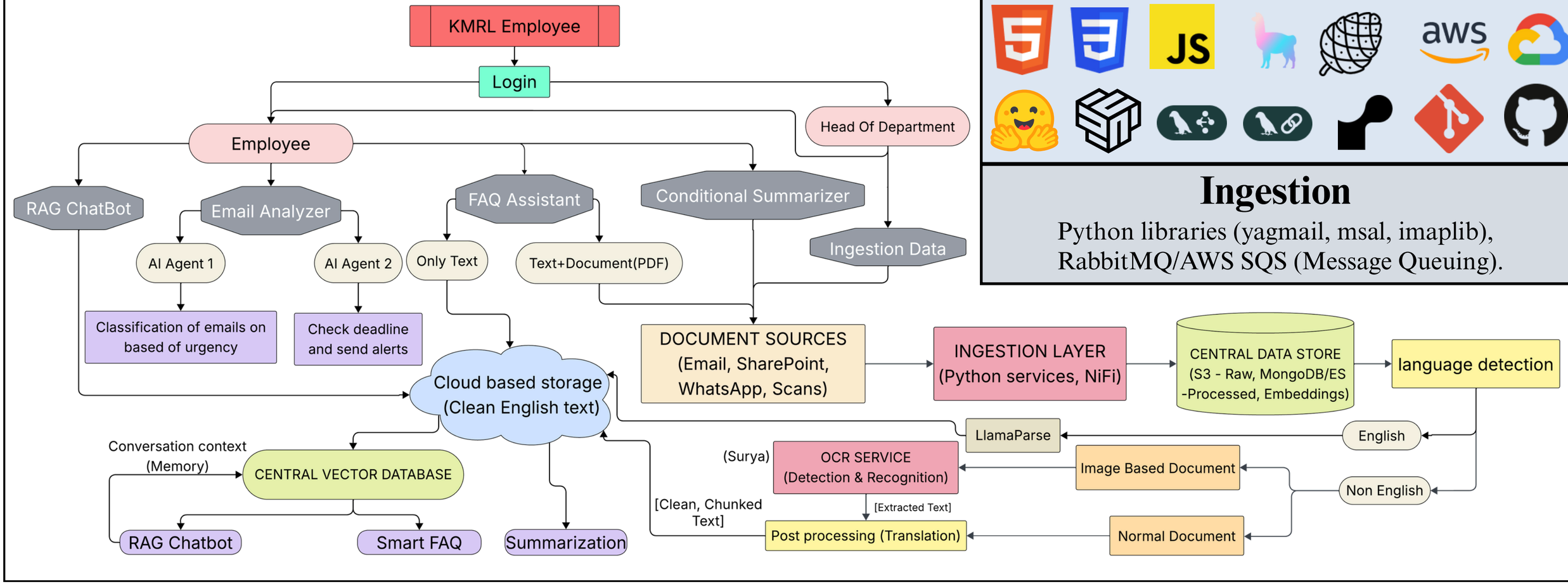
- **Smart Ingestion:** Parses emails and uses Surya OCR (for Malayalam) and LlamaParse (for English) to create structured, sorted data.
- **Conditional Summaries:** The Kimi model provides tailored, role-specific summaries to reduce information overload.
- **Unique Conflict Detection:** Automatically flags contradictions between documents using a Knowledge Graph, preventing costly errors.
- **Automated Compliance Workflow:** A dual-agent system that auto-classifies emails and documents, extracting actions and deadlines into a sorted dashboard card.
- **Semantic Search (Our USP: "Google4KMRL"):** A conversational search engine that has memory and provides direct, synthesized answers to preserve institutional knowledge.

End-to-End System Architecture



TECHNICAL APPROACH

Flow of construction activities & ML model used



Ingestion

Python libraries (yagmail, msal, imaplib), RabbitMQ/AWS SQS (Message Queuing).

Nityam

FEASIBILITY AND VIABILITY

SMART INDIA HACKATHON 2025

Operational

- Integrates with SharePoint, Email, Dashboards
- Automated Workflows: Auto-creates dashboard cards.
- Benefits multiple teams: Maintenance, Safety, Finance

Technical

- Multilingual Support: Works across English, Malayalam & more.
- Document Processing: Surya OCR + LlamaParse + moonshotai/kimi handle PDFs, images & emails.
- Conflict Detection: Graph DB auto-detects contradictions.

Economic

- Lower Upfront Costs: Open-source + cloud reduce initial investment.
- Higher Efficiency & Accuracy: Automations cut ~65% manual work and minimize rework/errors.
- Faster ROI: Lower costs + efficiency deliver quicker returns.

30% of Indian firms maximize AI value - higher than global avg. of 26%.

Potential risks and Challenges

- **Data privacy & security concerns.**
- **Handling multilingual like (English + Malayalam) documents.**
- **Knowledge Graph accuracy and validation**
- **AI limitations and staff adoption challenges**

Strategies for overcoming these challenges

- **Secure design:** Encryption & access control.
- **Localized training:** Fine-tune models on KMRL bilingual corpus.
- **Human-in-the-loop:** Validation to ensure accuracy.
- **Change management:** Staff training & awareness sessions.
- **Graph Validation:** Rule-based + human validation for knowledge graph accuracy.

Nityam

IMPACT AND BENEFITS

SMART INDIA HACKATHON 2025

POTENTIAL IMPACTS

- **Efficiency & Productivity:** Reduces manual work for faster, more accurate decisions.
- **Compliance & Safety:** Real-time alerts and tracking ensure adherence and reduce risks.
- **Knowledge & Scalability:** Preserves institutional knowledge and adapts easily to new data/workflows.
- **Collaboration & Future-Readiness:** Stronger coordination, better service delivery, and Smart City alignment.

BENEFITS

Workforce Impact: Quick access to critical info boosts efficiency, morale, and accountability.

Efficiency Gain: ~65% faster processing cuts costs and optimizes resources.

Sustainability: Reduced paper use through digitization, supporting eco-friendly Smart City goals.

Nityam

RESEARCH AND REFERENCES

SMART INDIA HACKATHON 2025

Reference from Google Scholar about our reasearch

- An Efficient Text Summarization Using Term and Inverse Frequency With Key Phrase Identification in Malayalam Language https://ieeexplore.ieee.org/document/10505438
- Build an AI-powered document processing platform with open source NER model and LLM on Amazon SageMaker | Artificial Intelligence https://aws.amazon.com/blogs/machine-learning/build-an-ai-powered-document-processing-platform-with-open-source-ner-model-and-llm-on-amazon-sagemaker/
- Learn by Interaction: Advancing Agentic AI for Web Automation with LangGraph By Jialin Wang ,Zhihua Duan https://www.cambridge.org/engage/coe/article-details/67b4455381d2151a02f5bdf3
- Surya – OCR, layout analysis, reading order & table recognition in 90+ languages https://github.com/datalab-to/surya

Online Resources

- Kaggle. Machine Learning for Education: Datasets and Models. https://www.kaggle.com/
- Generative AI: Introduction and Applications | Coursera https://shorturl.at/Dmi17

Documentation

- What is RAG?- Retrieval-Augmented Generation AI Explained - AWS https://aws.amazon.com/what-is/retrieval-augmented-generation/
- Langchain Official Document https://python.langchain.com/docs/introduction/
- Backend Using Flask - Flask Official Document https://flask.palletsprojects.com/en/stable/
- What is a Vector Database & How Does it Work? Use Cases +Examples | Pinecone
Agentic AI Using Langgrapg - Langgrapgh Official Document
https://langchain-ai.github.io/langgraph/concepts/why-langgraph/