🚀 AI-Powered Email Classification & OCR for Loan Servicing

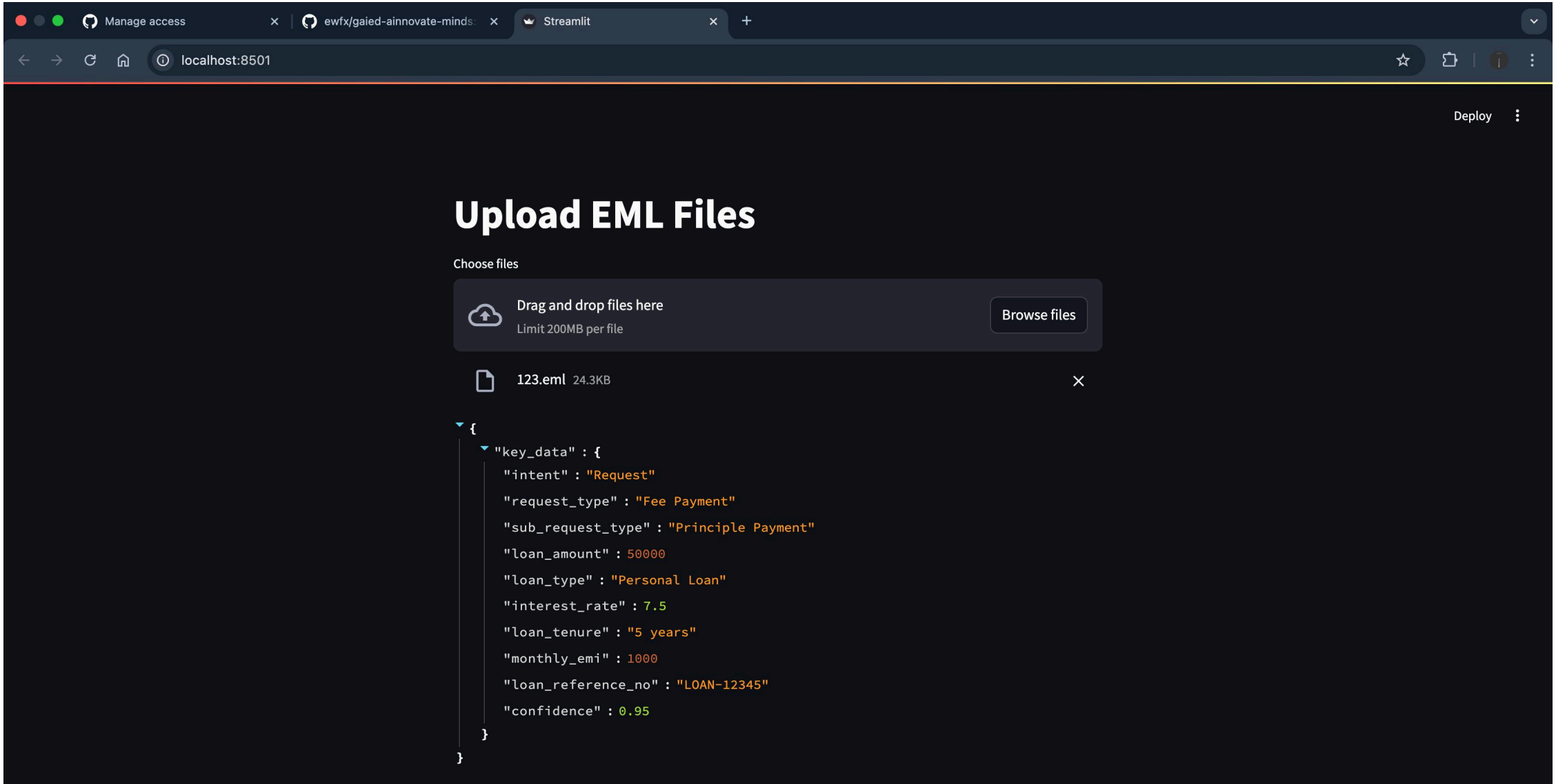
Team : AInnovate Minds

📌 Table of Contents

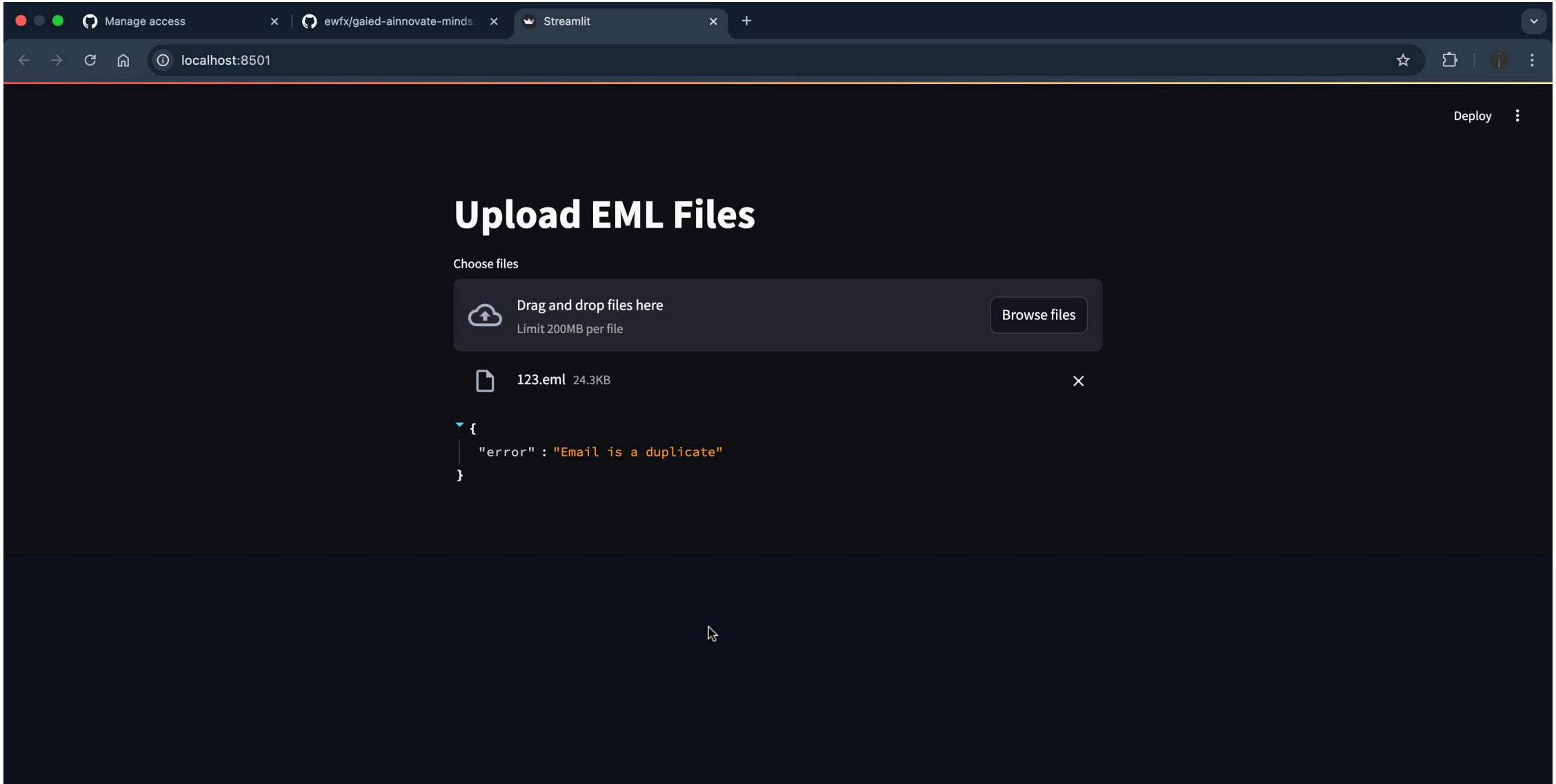
🎯 Introduction

- Loan servicing teams process **thousands of emails daily**
- Manual processing is **slow, error-prone, and resource-intensive**
- This project automates classification & processing using **AI, NLP, and OCR**
- Key Features:
  - ✓ **Email classification**
  - ✓ **Data extraction**
  - ✓ **Multi-request handling**
  - ✓ **Duplicate email detection**

Screenshots



Screenshots



💡 Inspiration

- Loan servicing teams struggle with **high email volumes** & **unstructured data**
- Aim: Build a **scalable AI automation system**
- Key Goals:
  - **Understand email context**
  - **Extract key data**
  - **Identify primary request intent**
  - **Detect duplicate emails**

⚙️ What It Does

- ✅ Email Classification
- ✅ Context-Based Data Extraction
- ✅ Multi-Request Handling
- ✅ Priority-Based Extraction
- ✅ Duplicate Email Detection

🛠️ How We Built It

- **AI Models:** Google Gemini, langchain_google_genai
- **OCR & Data Extraction:** Pytesseract, pdf2image, python-docx
- **Email Processing:** `email.parser`, MIME handling libraries
- **Duplicate Detection:** eml.parser, hashing
- **Data Processing:** pydantic, langchain, json
- **UI:** Streamlit

Workflow

1. Parse emails & extract attachments
2. Classify request types using AI models
3. Extract relevant fields
4. Detect multiple requests & primary intent
5. Check for duplicate emails
6. Return structured output

🚧 Challenges We Faced

- Handling unstructured email formats
- Ensuring high classification accuracy
- Extracting data contextually
- Multi-request processing
- Duplicate detection efficiency

🏗️ Tech Stack

- **AI Models:** Google Gemini, langchain_google_genai
- **OCR & Data Extraction:** Pytesseract, pdf2image, python-docx
- **Email Processing:** `email.parser`, MIME handling libraries
- **Duplicate Detection:** eml.parser, hashing
- **Data Processing:** pydantic, langchain, json
- **UI:** Streamlit

🔮 Future Integrations

- **CRM Integration:** Sync with loan servicing CRMs for seamless workflow automation
- **Cloud Storage:** Automate email & attachment storage in AWS S3 / Google Drive
- **API Endpoints:** Enable integration with banking systems via secure APIs
- **Advanced NLP:** Improve accuracy with fine-tuned models for domain-specific data
- **Multilingual Support:** Expand capabilities for global banking clients
- **Voice-to-Text Support:** Process audio-based loan servicing requests

👥 Team

🔗 **GitHub Repository:** [Project Link]

🚀 **Built for AI-Powered Loan Servicing Automation!**