

Underwriting Copilot for Life Insurance

Integrating Multi-Model OCR and RAG for Medical Document Summarization and Context-Aware Conversational Assistance

Company Supervisor: Ryan Gaughan | Academic Supervisor: Prof. Zhao Rui
Harsh Sharma | MSBA Student, NUS

Problem & Motivation

- **Up to 3 days processing per complex case** - creating substantial operational bottlenecks
- **60-70% accuracy** on handwritten medical text with traditional OCR systems
- **Inconsistent decision-making** across underwriters leading to pricing disparities and regulatory compliance issues
- **Processing bottlenecks** severely limiting organizational scaling in competitive markets
- **Lack of integrated platforms** - underwriters struggle with fragmented tools and manual workflows
- **Critical Impact:** Organizations in the reinsurance industry process **thousands of medical documents monthly** - these operational challenges could significantly impact **customer satisfaction, policy issuance speed, and competitive market positioning** if left unaddressed

Our Solution: Underwriting Copilot

Full-Stack Enterprise Platform

- **React + TypeScript frontend** with intuitive underwriter interface
- **FastAPI backend** with asynchronous processing and comprehensive API endpoints
- **Multi-database architecture:** **SQLite** for transactional data + **ChromaDB** for vector embeddings

Multi-Model OCR Pipeline

- **MS OCR** as primary engine (**85%** of processing)
- **Vision Language Model (VLM)** for contextual understanding (**14%**)
- **Intelligent routing** based on document complexity and content type

Sequential AI Architecture

- LLM (**Llama 3.2**) for initial document analysis
- VLM (**Mistral 24B**) for complex texts and for **LangGraph** workflow (exploratory)

Key Results & Performance

Technology	Overall Accuracy	Handwriting Accuracy
VLM	85.7%	80.5%
MS OCR	78.6%	72.1%

Processing Time Cut: up to **50%** reduction

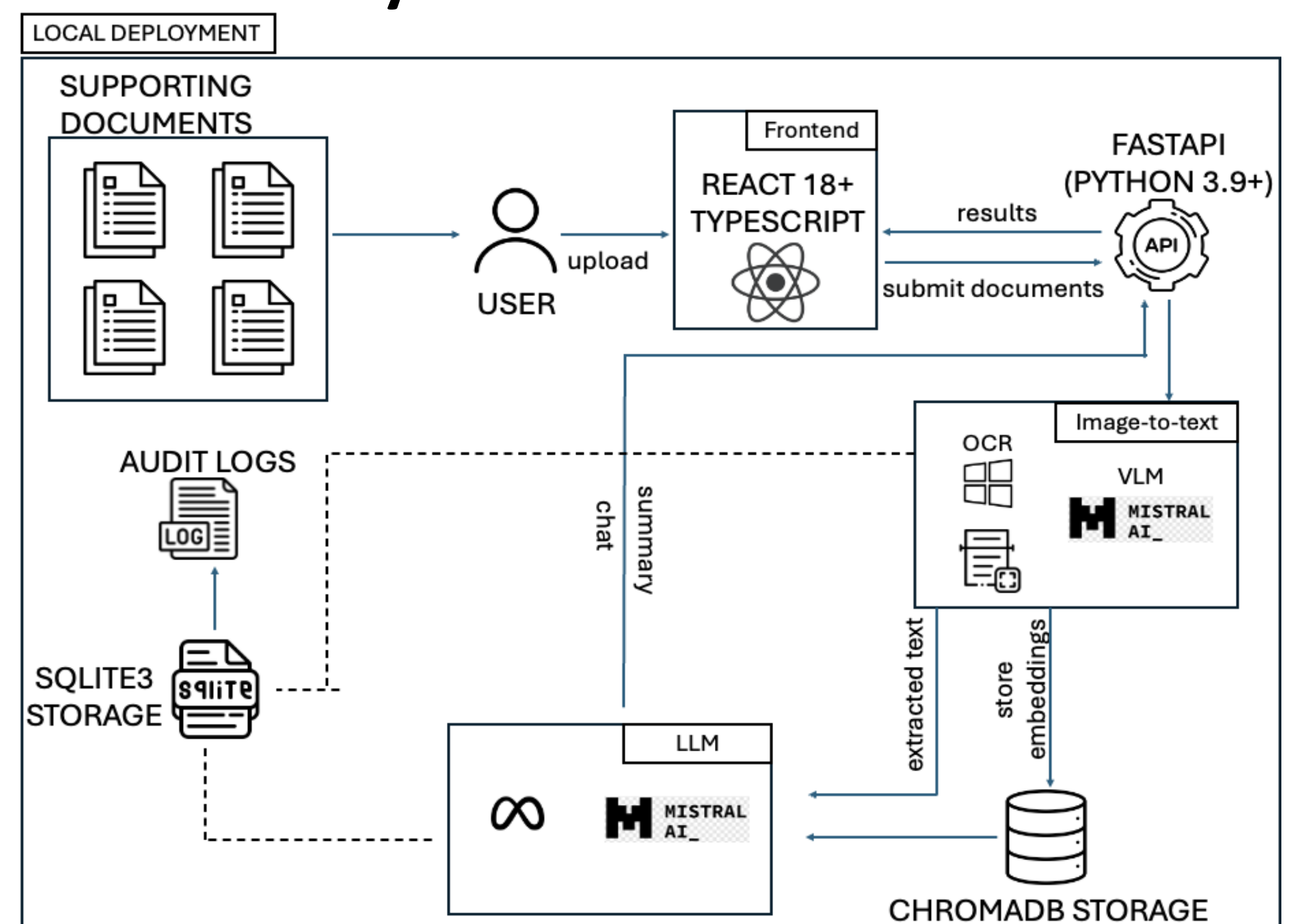
RAG Chat System Performance

- Document-Level Chat: **92.4%** accuracy
- Case-Level Chat: **89.7%** accuracy
- Mixed-Language Terms: **88.3%** accuracy
- Average Response Time: **24.1** seconds

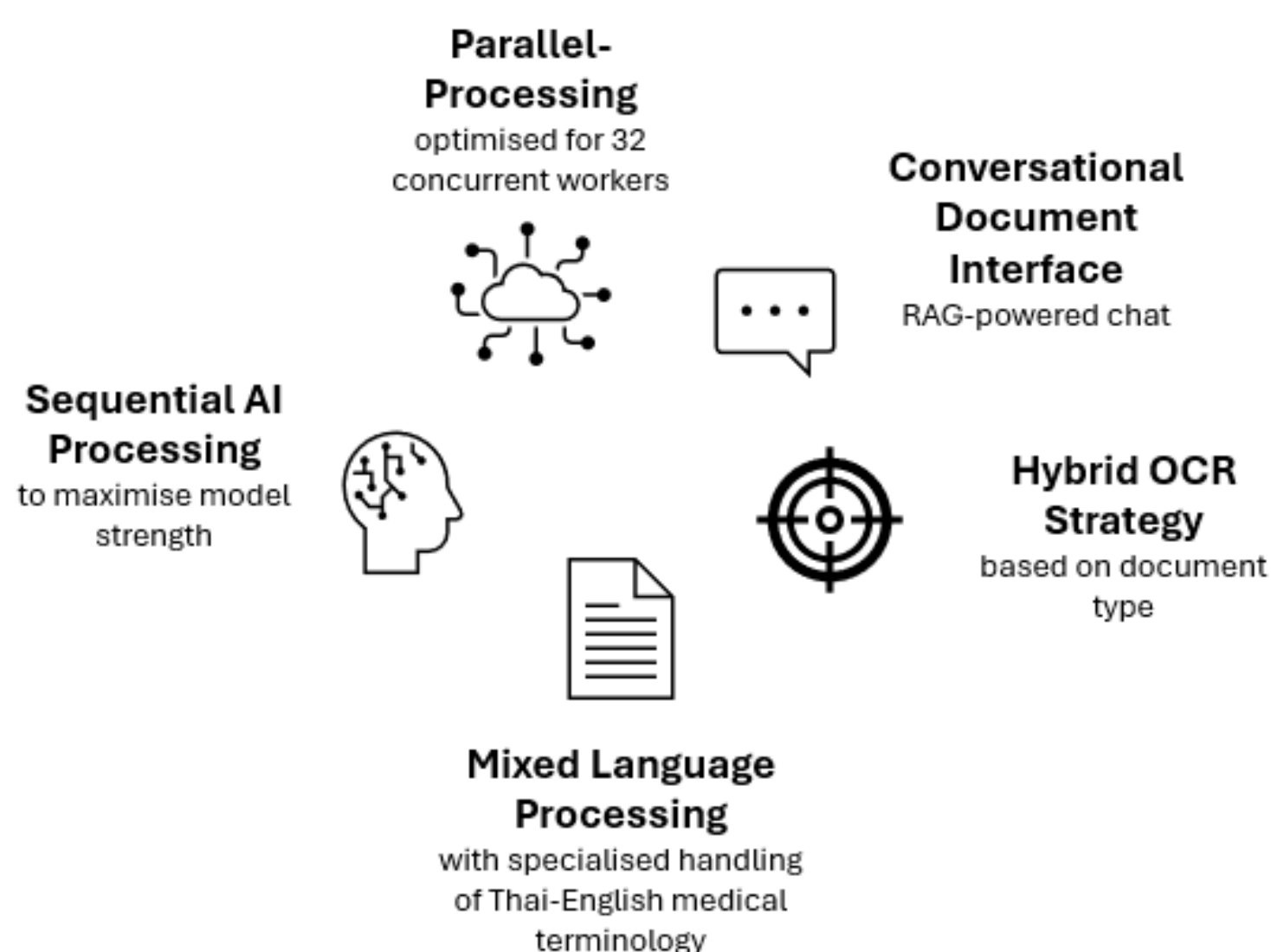
Medical Information Extraction

- Vital Signs: **98.4%** accuracy
- Medications: **94.7%** accuracy
- Diagnoses: **92.3%** accuracy
- Risk Factors: **89.6%** accuracy

System Architecture



Key Innovation



Business Impact & Achievements

Operational Benefits

- **Up to 50%** faster case processing
- Improved decision consistency across underwriters
- Production-ready system with comprehensive audit capabilities

Cost Optimization

- **90.7%** cost reduction with VLM vs MS OCR
- Strategic resource allocation based on document complexity

User Experience

- **4.6/5.0** user satisfaction from 3-underwriter evaluation
- **100%** adoption of conversational AI features
- Advanced conversational AI with semantic document search

Future Roadmap

