



Anthony Khaiat
Candidates for Master of Business Analytics

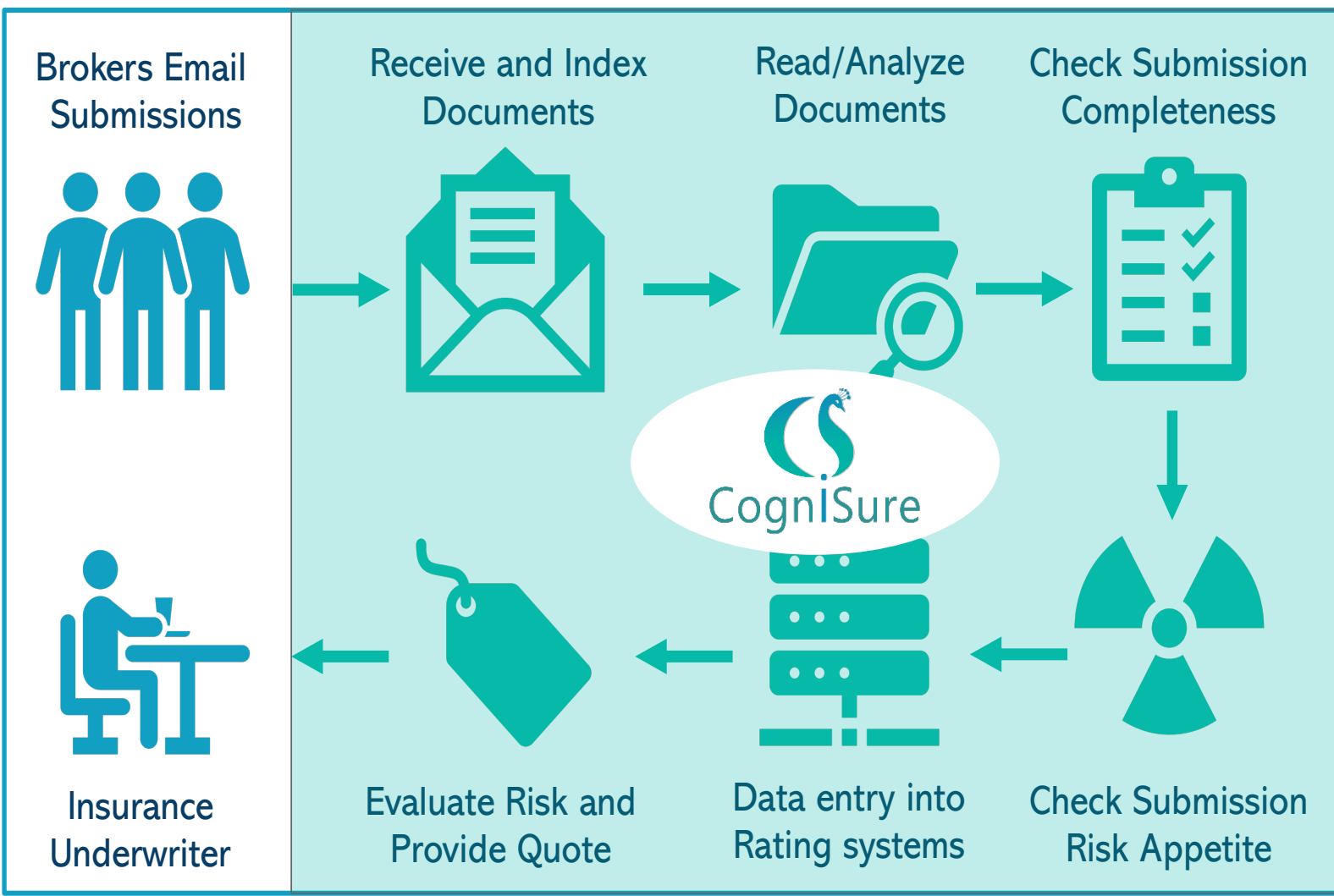
Jad Makki
Candidates for Master of Business Analytics

Automating Risk-Clearance Scoring for Insurers with Large Language Models

Faculty Advisor: Dr. Retsef Levi
CogniSure Advisors: Sai Raman, Punitha Geetha

Overview

Insurance submissions cause operational inefficiencies



Challenges (3 V's)

Manual Document Field Extraction

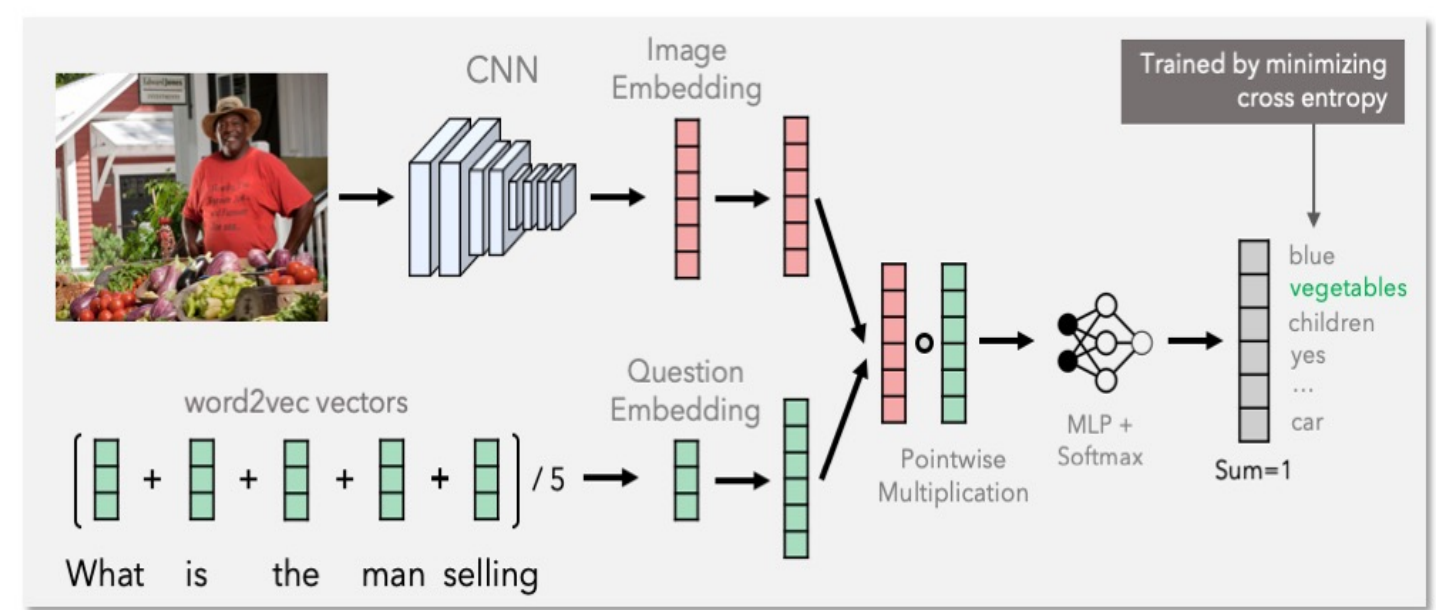
- Volume**
Process 1000s of PDFs and Emails
- Variety**
Extract from a multitude of field types
- Velocity**
Accurate, Fast, and an easy-to-use UI

End Goal

Insurance submissions cause operational inefficiencies
Create an automated application which can take any **document-type** or **email** and extract multiple fields (names, addresses, etc...) with high **speed** and **accuracy**

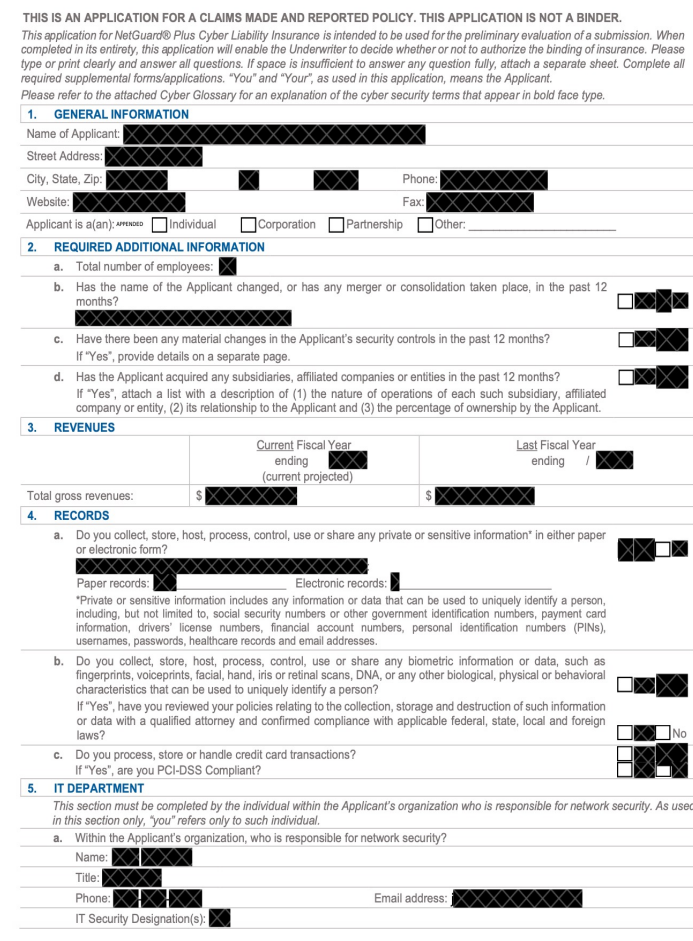
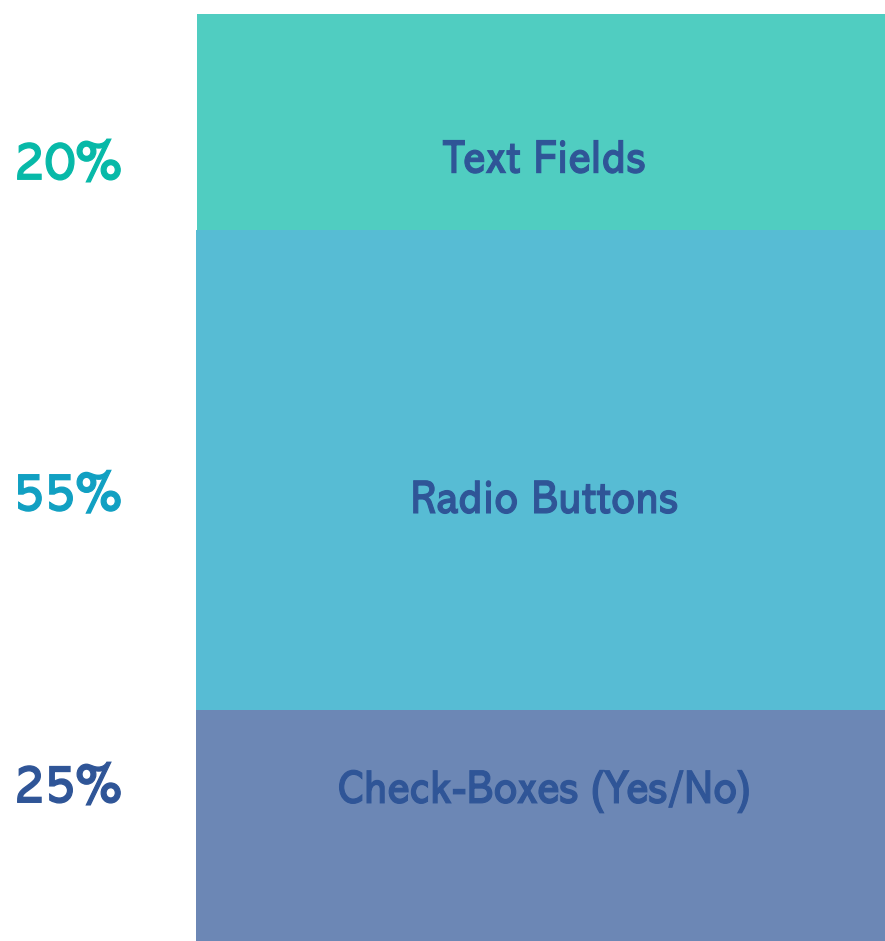
Solution

Visual Language Answering Models (VQAs)



Exploratory Data Analysis

Beyond Traditional OCR; 3 different field types require a new approach

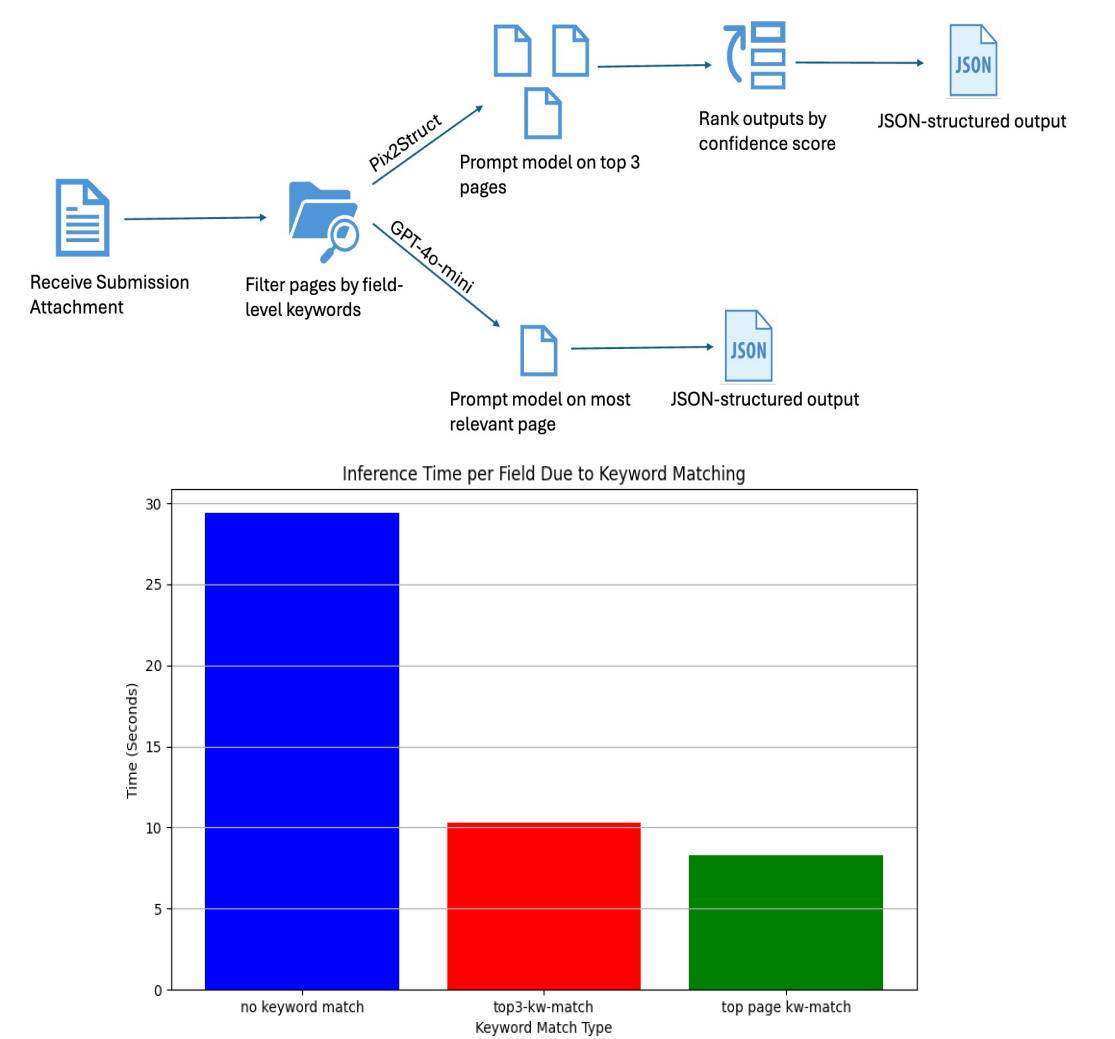


OCR

DONEGAL Policy Period: o1/0172022 0110172023,
(CLAIM: NONE Loss Date: oe (0250_26087
(Claimant Name: INoNe status: 'None Description:
NONE 'Adaitonal Description: NONE Policy Totals
Incurred Losses Total: \$0.00 Policy Period:
o1/0172021 0101/2022 Loss Date Loe:
0250_26067 Claimant Name: NONE Status None
Description NONE 'Adsjonal Description: NONE
Process Date: October 04,2022 Reported Date:
(Gamat Number: 'Type of Loss (Cass Code:
Expenses \$0.00 FRreported Date: 'Gtamant
Number: 'Type of Loss: (Cass Coe: Statement of
Losses Type of policy: WorkersComp 1000013558
Wedee inc Incurred Losses Expenses Subrogation
Total Loss

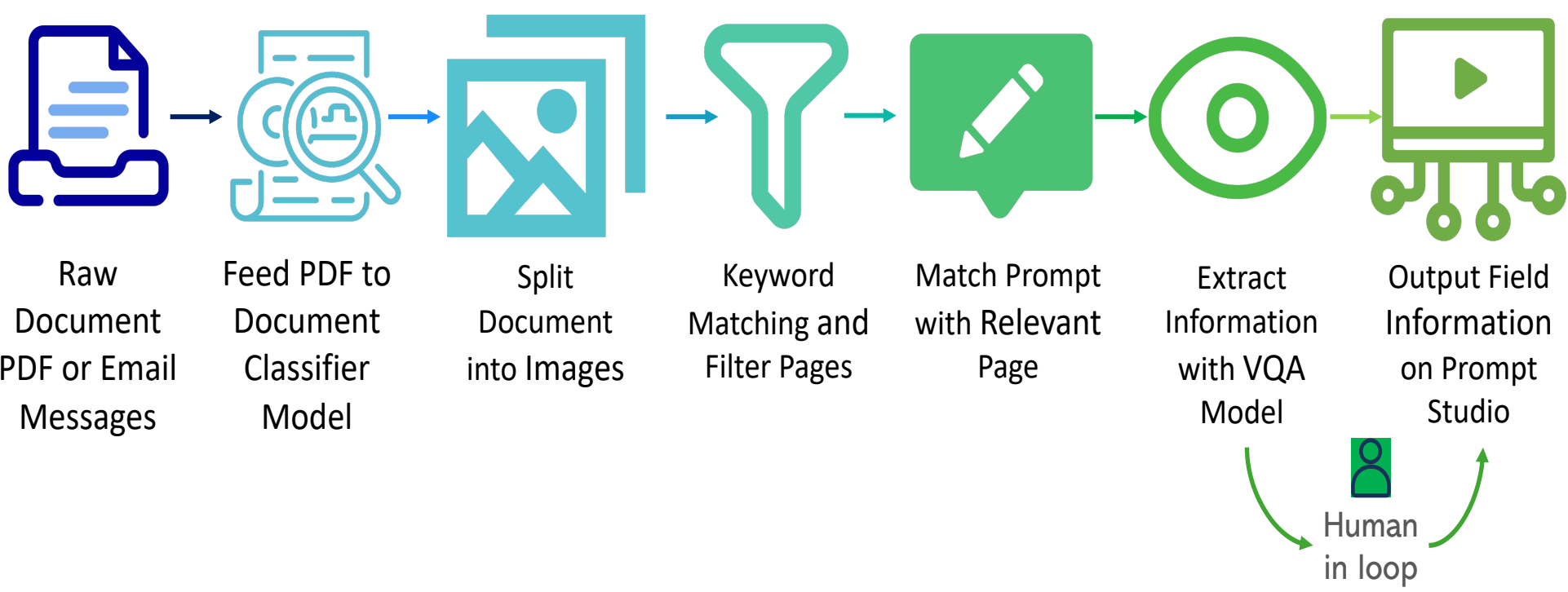
Data Processing

Keyword matching reduces page count, creating faster runtimes



Methodology

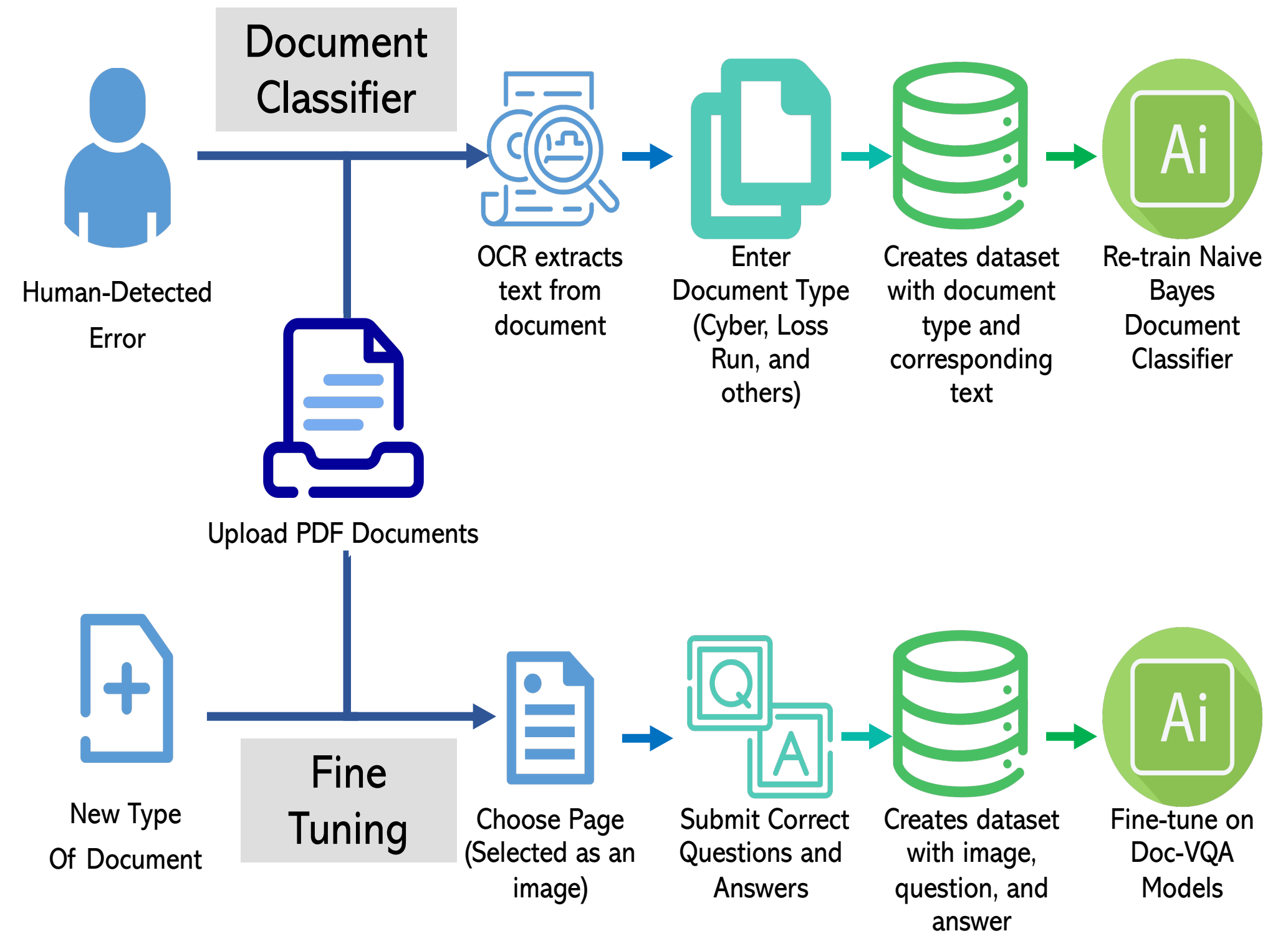
End-to-End system that receives PDFs/email with prompts and generates JSON with corresponding answers



- PDF-to-Image**
Each page in each PDF document is extracted as individual images with email messages
- Keyword Matching**
Match prompts with pages by finding the top 3 pages with the most keywords
- VQA Model**
VQA models fine-tuned on scanned documents + GPT4o

Methodology

Learn with Active Learning framework that includes Fine-Tuning and Document Classifier Modules



Results

Key metrics summary across different document types

Field Names	Doc-VQA Accuracy	GPT4o Accuracy
Applicant Name	70%	60%
Employee Count	83%	50%
Current Revenue	86%	100%
Previous Revenue	86%	100%
Multi-factor Authentication	100%	100%
Encryption	71%	100%
Private Information	75%	75%

Doc-VQA: 82%
GPT-4o: 84%

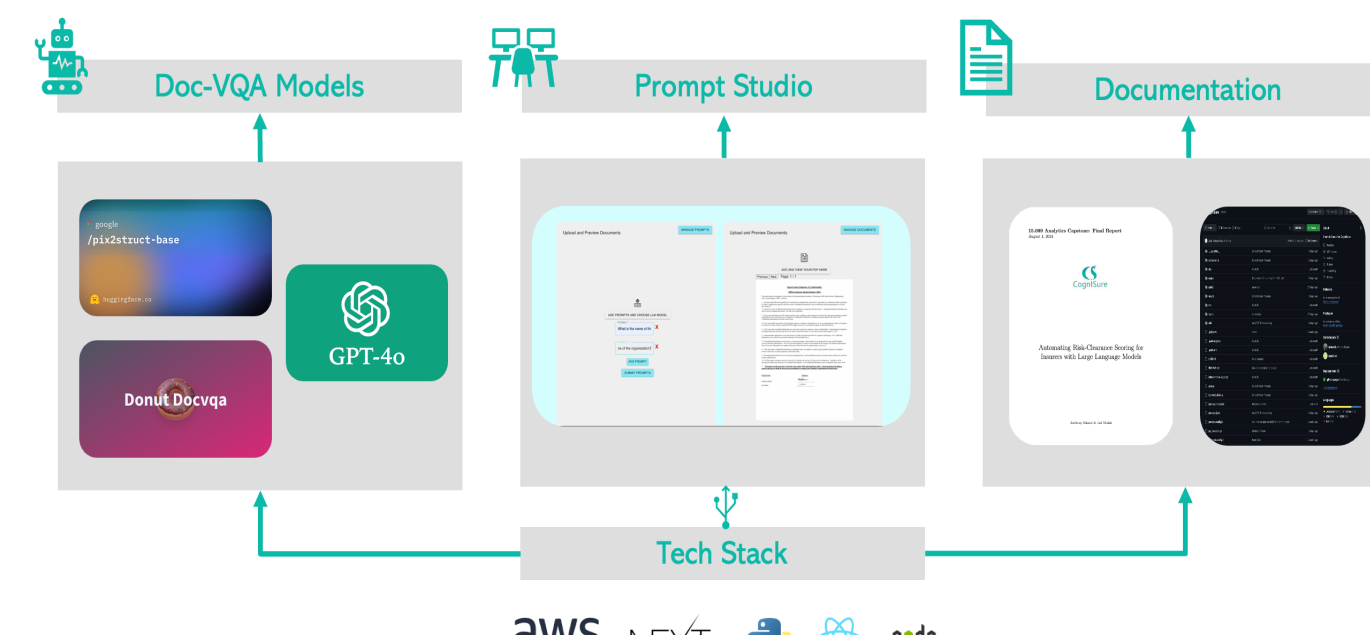
Doc-VQA Cost/Document ~\$0.005
GPT-4o Cost/Document ~\$0.05

Accuracy Increase 173% YoY

Model Speed (15 Fields)
Doc-VQA: ~150 seconds
GPT-4o: ~90 seconds
Donut: ~400 seconds

Deliverables

Doc-VQA Models + Prompt Studio with Scalable Pipeline



Business Value

- Customers**
 - +30% faster quote turnaround
 - +80% accuracy on documents
 - Increase in throughput ratio
- Shareholders**
 - 50% in submission costs
 - 40% operational costs
 - Higher client loyalty and retention