AI-Powered Redaction Tool

Introduction:

Organizations often need to redact sensitive information in PDF or image documents, such as personal identifiers, financial details, or proprietary data and HIPAA Compliance data to comply with privacy regulations and protect confidential information.

The manual redaction process is time-consuming, prone to errors, and unsuitable for large-scale operations. This case study highlights the development of an AI-powered redaction tool that automates this process efficiently.

Client Details:

Name: Confidential | **Industry:** Software, AI | **Location:** USA

Technologies:

Python, FastAPI, Docker, Azure, Kubernetes, Nginx, Linux, Ubuntu, RabbitMQ, GIT, OCR tools (Tesseract, Google Vision), NLP frameworks (spaCy, Hugging Face, LexNLP).

Project Description:

The platform is designed mostly to target Financial and Insurance firms with a focus on maintaining privacy. The platform can also be used by Medical Professionals doing research while hiding the personal and legal data in the file. It can also be used by any individual who wants to perform redactions on their documents.

It is an API that allows users to upload multiple documents in the form of PDF or image to a dashboard where they can store them and perform redactions on them anytime, anywhere.

There are two versions of the API: an on premise version and a cloud-based version. The on-premise version, which utilizes Tesseract for file processing, can be deployed on the client's infrastructure using a Docker server. This option is well-suited for industries with strict security requirements, such as banking and insurance. Meanwhile, the cloud version offers a flexible solution that can be integrated into any application.

AI-Powered Redaction Tool

The AI redaction tool has two basic end-user availability:

1. Client
2. Admin

Client:

Clients are the applications or websites. All users on a website share a single license registered to the application. The tool accepts pdf/Image files from client applications with redaction rules. The tool has option to select the OCR type, e.g. Google Vision or Tesseract. After the processing is completed, the redact file is delivered to the client application using a callback.

Users can purchase pages as subscription (yearly or monthly) or according to their need using the pay-as-you-go option.

Admin:

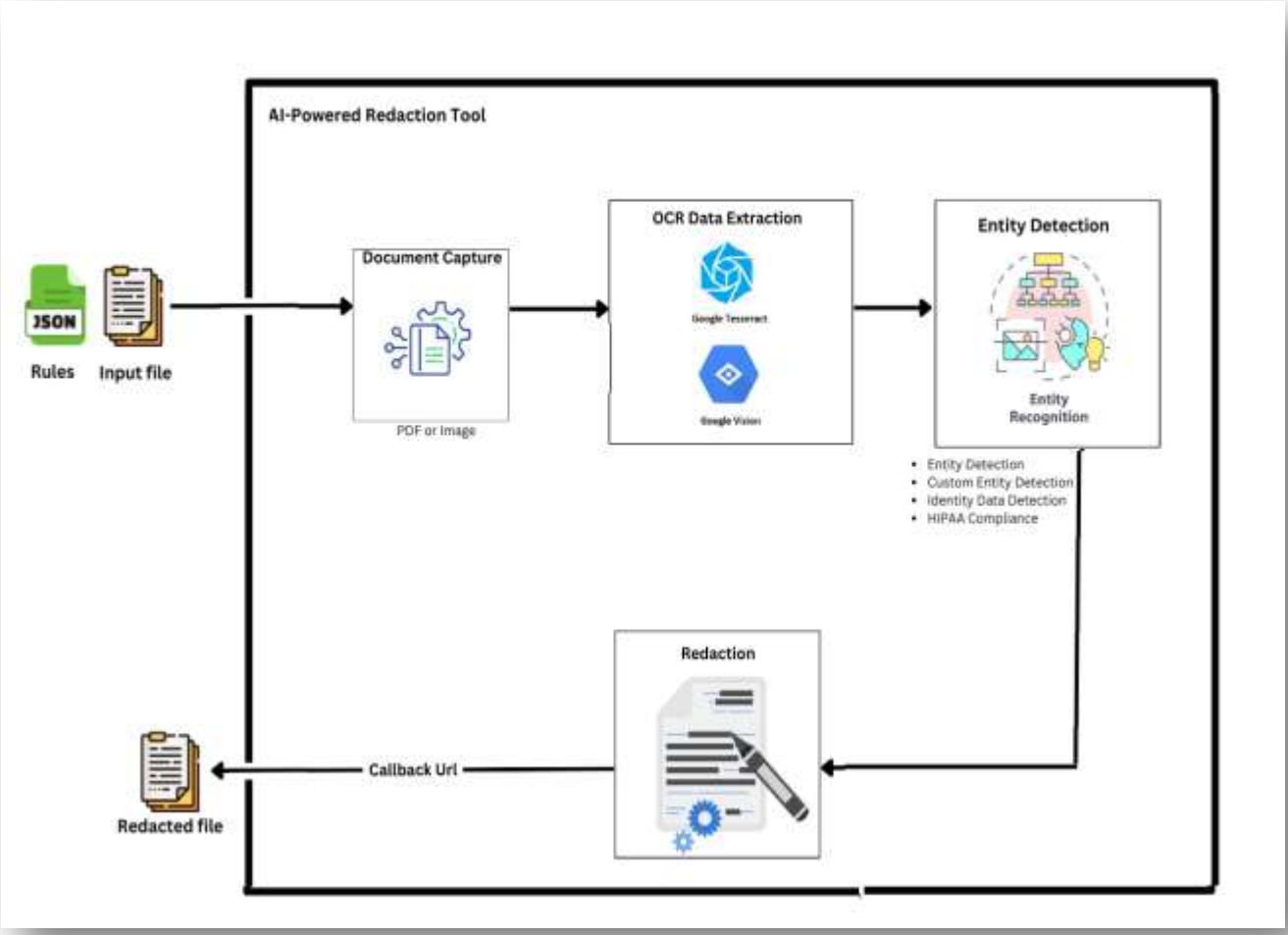
Admin is the owner of the platform with access to view all reports and usage analytics of the tool including signups, sessions, purchases, user feedbacks, storage, file upload logs, etc. The admin can add free pages to the accounts of users based on need. The admin can send automatic weekly reports to his team in email format, just by adding emails to his admin panel. The admin can also send promotional emails to users.

Methodology followed:

- Development of the tool is powered by Python Fast API Framework. The database used is PostgreSQL.
- Azure is used as the cloud platform.
- Docker is used for container management with Kubernetes. We followed Agile Scrum with Daily Standups and Sprint-based deliveries.

AI-Powered Redaction Tool

Architecture Design:



AI-Powered Redaction Tool

Screenshots: