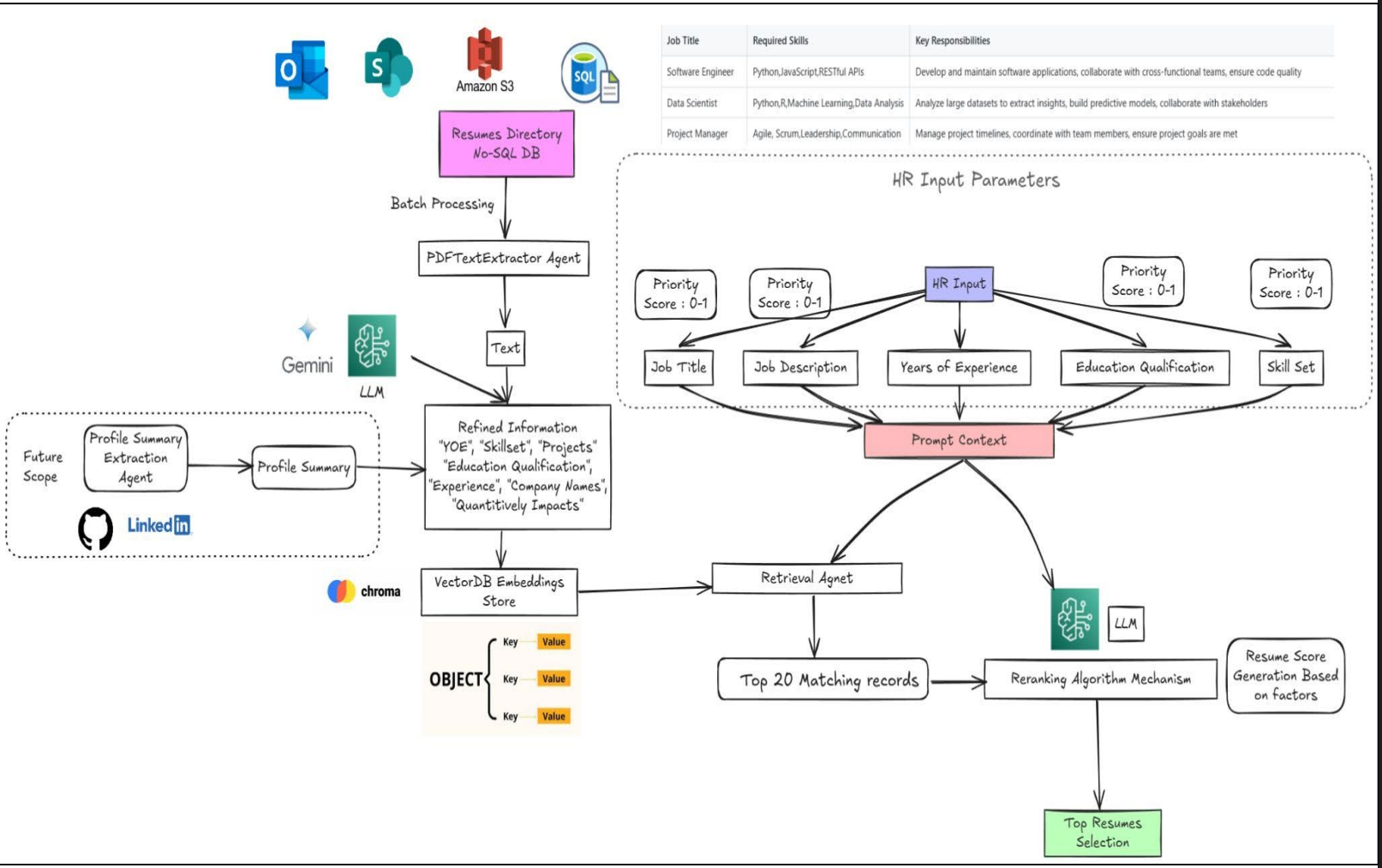
Image to Markdown

Extraction of Complex PPT Contents (Images) into Markdown




```
# Resume Screening Application Design

## Icons
- **O365**
- **Amazon S3**
- **SQL**

## Resumes Directory
- **No-SQL DB**

## Batch Processing
- **PDFTextExtractor Agent**

## HR Input Parameters
- **Priority Score: 0-1**
- **Job Title**
- **Job Description**
- **Years of Experience**
- **Education Qualification**
- **Skill Set**

## Prompt Context
- **Refined Information**
  - "YOE", "Skillset", "Projects"
  - "Education Qualification"
  - "Experience", "Company Names"
  - "Quantitively Impacts"

## VectorDB Embedding Store
- **OBJECT**
  - **Key** | **Value**
  - **Key** | **Value**
  - **Key** | **Value**

## Retrieval Agent
- **Top 20 Matching records**

## LLM
- **Reranking Algorithm Mechanism**
- **Resume Score Generation Based on factors**

## Top Resumes Selection

## Future Scope
- **Profile Summary Extraction Agent**
```

pip install doctomarkdown

```python
from doctomarkdown import DocToMarkdown
from openai import OpenAI
import os
from dotenv import load_dotenv
load_dotenv()

client = OpenAI(
    api_key=os.environ.get("OPENAI_API_KEY"),
)

app = DocToMarkdown(llm_client=client,
                    llm_model='gpt-4o')

result = app.convert_pptx_to_markdown(
    filepath="sample_docs/sample_ppt_2.pptx",
    extract_images=True,
    extract_tables=True,
    output_path="markdown_output"
)

for page in result.pages:
    print(f"Page Number: {page.page_number} | Page Content: {page.page_content}")
```