Project Title:	**Document Automation System (DASH)**
Role	**Data Scientist**
Description:	● Built a semantic search and document QnA system for retrieving A large-scale document automation platform used by multiple enterprise clients, capable of processing over 1 million documents per day. **Key Contributions:** ● Built asynchronous, multi-threaded services using **Python multiprocessing** to handle massive file queues and maintain low latency. ● Engineered scalable pipelines using **AWS services** (S3, SQS, EC2) integrated with **IBM MQ** for message-driven processing. ● Designed modular pipelines to handle **OCR (PaddleOCR, Textract)** and **LLM-driven extraction** using models like **DONUT, LayoutLM, Pix2Struct**. ● Fine-tuned transformer models using **LoRA, QLoRA, and PEFT** to achieve domain-specific extraction of names, amounts, dates, and checkboxes. ● Implemented **custom rule engines** for validating extracted fields against business rules (e.g., amount mismatch, missing signature). ● Built dashboards using **Grafana and Redash** to monitor system metrics like job failure rate, throughput, and accuracy. ● Reduced document processing turnaround time by 45% and improved extraction accuracy by 38% through iterative testing and model enhancement. ● Coordinated with DevOps to set up **CI/CD** for auto-deployment of microservices, ensuring rapid feature rollout.

Project Title:	**Document Automation System (DASH)**
Role	**Data Scientist**
Description:	● Built a semantic search and document QnA system for retrieving A large-scale document automation platform used by multiple enterprise clients, capable of processing over 1 million documents per day. **Key Contributions:** ● Built asynchronous, multi-threaded services using **Python multiprocessing** to handle massive file queues and maintain low latency. ● Engineered scalable pipelines using **AWS services** (S3, SQS, EC2) integrated with **IBM MQ** for message-driven processing. ● Designed modular pipelines to handle **OCR (PaddleOCR, Textract)** and **LLM-driven extraction** using models like **DONUT, LayoutLM, Pix2Struct**. ● Fine-tuned transformer models using **LoRA, QLoRA, and PEFT** to achieve domain-specific extraction of names, amounts, dates, and checkboxes. ● Implemented **custom rule engines** for validating extracted fields against business rules (e.g., amount mismatch, missing signature). ● Built dashboards using **Grafana and Redash** to monitor system metrics like job failure rate, throughput, and accuracy. ● Reduced document processing turnaround time by 45% and improved extraction accuracy by 38% through iterative testing and model enhancement. ● Coordinated with DevOps to set up **CI/CD** for auto-deployment of microservices, ensuring rapid feature rollout.

Project Title:	**Document Automation System (DASH)**
Role	**Data Scientist**
Description:	● Built a semantic search and document QnA system for retrieving A large-scale document automation platform used by multiple enterprise clients, capable of processing over 1 million documents per day. **Key Contributions:** ● Built asynchronous, multi-threaded services using **Python multiprocessing** to handle massive file queues and maintain low latency. ● Engineered scalable pipelines using **AWS services** (S3, SQS, EC2) integrated with **IBM MQ** for message-driven processing. ● Designed modular pipelines to handle **OCR (PaddleOCR, Textract)** and **LLM-driven extraction** using models like **DONUT, LayoutLM, Pix2Struct**. ● Fine-tuned transformer models using **LoRA, QLoRA, and PEFT** to achieve domain-specific extraction of names, amounts, dates, and checkboxes. ● Implemented **custom rule engines** for validating extracted fields against business rules (e.g., amount mismatch, missing signature). ● Built dashboards using **Grafana and Redash** to monitor system metrics like job failure rate, throughput, and accuracy. ● Reduced document processing turnaround time by 45% and improved extraction accuracy by 38% through iterative testing and model enhancement. ● Coordinated with DevOps to set up **CI/CD** for auto-deployment of microservices, ensuring rapid feature rollout.