



Invoice Processing Tech Stack

Component	Technology	Why This Tech? (Function & Purpose)
Messaging Broker	RabbitMQ	Enables decoupling, parallelism, and async communication between microservices; easy to set up & scale.
OCR Engine	Donut / LayoutLMv3	
Fallback OCR	Tesseract / EasyOCR	Transformer-based models trained for document layout + text extraction; robust across varied invoice formats.
Preprocessing	OpenCV / PIL / PyMuPDF	
Field Matching / Validation	FuzzyWuzzy / RapidFuzz / custom rules	Performs approximate string or numeric matching to validate fields like totals, UPCs, quantities.
Data Persistence (Relational)	MySQL / MariaDB	
Data Persistence (NoSQL)	MongoDB	Stores structured and validated invoice data; supports foreign keys for product/vendor matching.
API Layer	FastAPI	
Dashboard	Streamlit / Dash	Stores raw JSON outputs, full OCR snapshots, logs, and audit trails for traceability.
Anomaly Detection	Rule Engine / Light ML Model	
Human-in-the-Loop (HITL)	Custom React UI / Prodi.gy	Lightweight, async-friendly Python web framework to expose APIs for dashboards and internal systems.
Orchestration	Docker + Shell Scripts	

Monitoring & Alerting	Prometheus + Grafana	Monitors service uptime, message throughput, model performance, and latency against SLA targets.
Logging & Tracing	OpenTelemetry / ELK Stack	Captures logs, traces, and metrics to help with debugging and performance analysis.
Semantic Search (Optional)	Weaviate / FAISS / Pinecone	Used to match invoice text or product codes against vectorized catalog entries for better enrichment.
Model Retraining Pipeline	Hugging Face + MLflow	Manages fine-tuning, evaluation, and deployment of OCR/IDP models based on feedback from HITL interface.
Deployment Environment	AWS EC2 / ECS / EKS	Scalable infrastructure for containerized services, GPU inference, and secure database hosting.
Secrets Management	AWS Secrets Manager / Vault	Secures API keys, DB credentials, and model tokens used across microservices.
ETL / Scheduling	Apache Airflow / Prefect	Automates batch invoice processing pipelines, retries, and data exports on schedule.