Fast OCR & Fast EMR Platform

From PDF Documents to FHIR-first EMR Records

Client Overview and Solution Summary

Confidential — For Client Presentation

1 Industry Context and Operational Challenges

Healthcare and government programs increasingly rely on digital systems to capture structured information. However, in real-world operations, many critical workflows still depend on documents—especially PDFs—such as signed forms, historical records, externally submitted files, and appeal packages.

This creates a persistent gap between structured system data and the document reality used for review, approval, and audit.

1.1 Observed Challenges in Document-driven Workflows

- **High manual effort:** reviewers must read, interpret, and re-enter information from PDFs.

- **Inconsistent formats:** the same information appears in different layouts and document templates.

- **Audit complexity:** signed or scanned PDFs are authoritative, but difficult to analyze programmatically.

- **System fragmentation:** documents often originate outside the primary system of record.

These challenges slow down operations, increase error risk, and make audits and appeals more difficult.

2 Solution Approach and Conceptual Framework

Fast OCR & Fast EMR address these challenges by combining document intelligence with standards-based data modeling.

The solution is built on three guiding principles:

- **Document understanding:** use OCR to transform PDFs and scanned images into machine-readable content.

- **AI-assisted structuring:** apply constrained AI analysis to extract domain-specific information from documents.

- **Standards-based interoperability:** convert validated results into FHIR-aligned EMR records.

2.1 Conceptual End-to-End Flow

At a high level, the workflow proceeds as follows:

1. OCR reads text and layout from PDF documents.

2. AI analyzes document content and extracts structured fields.

3. Extraction results are stored with traceability and processing metadata.

4. Validated data is pushed to the server-side EMR layer.

5. A FHIR library converts the data into standardized EMR resources.

6. Structured records are persisted and displayed in the EMR interface.

2.2 Separation of Responsibilities

The platform intentionally separates document processing from enterprise data management:

- **Fast OCR** focuses on document-centric processing, review, and provenance.

- **Fast EMR** focuses on validated, long-lived records and interoperability.

  This separation improves scalability, reliability, and governance.

3 Implementation Strategy and Key Technologies

This section summarizes how the solution is implemented in practice, highlighting key technologies while keeping the focus on outcomes rather than low-level details.

3.1 Fast OCR: Document Intelligence and Review

Fast OCR provides a document-focused experience optimized for PDFs and scanned files:

- **In-browser PDF interaction:** PDF viewing and navigation using **pdf.js**.

- **Asynchronous processing:** background jobs managed via **arq** and **Redis** to ensure responsive user interaction.

- **AI-assisted extraction:** structured field extraction using **GPT**, orchestrated through **LangGraph** for a controlled, auditable workflow.

- **Document-layer storage: MongoDB** is used exclusively within Fast OCR to store extraction results, job state, and provenance metadata.

- **Deployment environment:** Fast OCR is deployed on **Fly.io** using container-based services.

3.2 Fast EMR: FHIR-first Data and Review Platform

Fast EMR provides the enterprise-facing system for validated data:

- **Web application:** an **Angular**-based interface for listing and reviewing structured records.

- **FHIR modeling:** a **.NET** service layer uses the **Microsoft FHIR SDK** to create and validate standardized healthcare resources.

- **Enterprise data store:** validated FHIR resources are persisted in **Azure Cosmos DB** for scalability and long-term retention.

- **Cloud hosting:** Fast EMR runs in **Azure** to align with enterprise security and governance requirements.

3.3 Document-to-EMR Data Handoff

Only validated and approved data crosses the boundary between Fast OCR and Fast EMR:

- Fast OCR produces structured outputs with document context.

- Approved data is transmitted to the EMR service.

- The EMR layer converts the payload into FHIR resources and persists them.

  This ensures a clean, auditable transition from documents to EMR records.

4 Delivered Outcomes and Future Expansion

4.1 What the Platform Delivers Today

The current implementation provides:

- A complete pipeline from PDF ingestion to FHIR-aligned EMR records.

- Human-in-the-loop validation with full document traceability.

- Clear separation between document processing and enterprise data storage.

- A scalable foundation suitable for audit, appeal, and review workflows.

4.2 Future Expansion and Scalability

The platform is designed to evolve as document volume and variety increase:

- Support for a wider range of PDF templates and document layouts.

- Multiple upload sources, including providers and external systems.

- A unified EMR-side presentation regardless of document origin.

- Enhanced evidence linking and explainability for reviewers and auditors.

5   Conclusion

Fast OCR & Fast EMR provide a production-ready solution for organizations operating at the intersection of documents and structured healthcare data. By combining OCR, AI-assisted extraction, asynchronous processing, and FHIR-first modeling, the platform reduces manual workload while improving consistency, transparency, and audit readiness.

End of Document