Feature #6: Smart Document Processing & Compliance

Overview

Automate document handling, data extraction, and compliance tracking for CareLinkAI. This feature will save hours of manual work by automatically processing uploaded documents, extracting key information, and ensuring compliance requirements are met.

Business Value

Problems Solved

1. **Manual Data Entry** - Hours spent typing information from documents
2. **Document Disorganization** - Hard to find and track documents
3. **Compliance Gaps** - Missing or expired documents go unnoticed
4. **Slow Onboarding** - Document processing delays resident admission
5. **Human Errors** - Typos and missed information in manual entry

Benefits

- ⏱️ **Save 5-10 hours/week** per operator on data entry
- 📊 **95%+ accuracy** in data extraction
- ✅ **100% compliance tracking** - never miss required documents
- 🚀 **50% faster onboarding** - automated document processing
- 💰 **ROI in 2-3 months** - significant time savings

Core Features

1. Document Upload & Storage

- Drag-and-drop file upload
- Support for PDF, images (JPG, PNG), scanned documents
- Cloudinary integration for secure storage
- File size limits and validation
- Progress indicators

2. OCR & Text Extraction

- Automatic text extraction from images and PDFs
- Tesseract.js for client-side OCR
- Google Cloud Vision API as fallback
- Handles handwritten text
- Multi-page document support

3. AI-Powered Field Extraction

- OpenAI GPT-4 extracts structured data
- Identifies: names, dates, addresses, phone numbers, medical info, insurance details

- Context-aware extraction
- Confidence scores for each field
- Manual review for low-confidence extractions

4. Form Auto-Population

- Automatically fills resident/inquiry forms
- Maps extracted data to form fields
- Validation and error checking
- Review before saving
- Edit extracted data if needed

5. Document Classification

- Automatically categorizes documents:
- Medical Records
- Insurance Cards
- ID Documents (Driver's License, Passport)
- Contracts & Agreements
- Financial Documents
- Care Plans
- Emergency Contacts
- Other
- Machine learning-based classification
- Manual override option

6. Compliance Checking

- Tracks required documents per resident
- Expiration date monitoring
- Alerts for missing documents
- Compliance dashboard
- Automated reminders

7. Document Generation

- Template-based document creation
- Auto-fill with resident data
- Professional PDF generation
- Common templates:
- Admission Contracts
- Care Agreements
- Financial Agreements
- Emergency Contact Forms
- Medical Release Forms

8. Search & Organization

- Full-text search across all documents
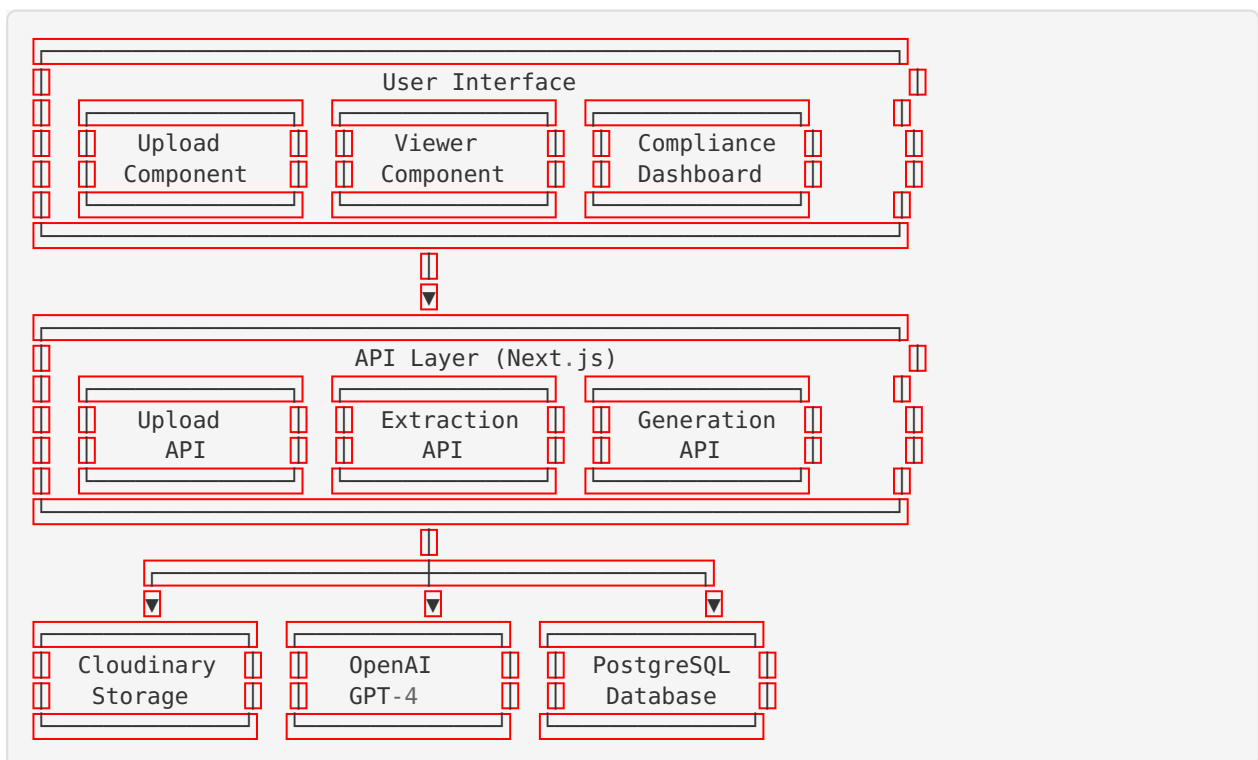- Filter by type, date, resident, status
- Tag system for custom organization

- Quick access to recent documents
- Document history and versions

Technical Architecture

Technology Stack

- **Frontend**: React, TypeScript, TailwindCSS
- **Backend**: Next.js API Routes, TypeScript
- **Database**: PostgreSQL with Prisma ORM
- **File Storage**: Cloudinary
- **OCR**: Tesseract.js (client-side), Google Cloud Vision API (server-side)
- **AI**: OpenAI GPT-4 for field extraction
- **PDF Generation**: pdf-lib
- **PDF Viewing**: react-pdf

System Architecture

```
┌────────────────────────────────────────────────────────┐
│                    User Interface                        │
│  ┌──────────────┐  ┌──────────────┐  ┌──────────────┐   │
│  │   Upload     │  │   Viewer     │  │  Compliance  │   │
│  │  Component   │  │  Component   │  │  Dashboard   │   │
│  └──────────────┘  └──────────────┘  └──────────────┘   │
└────────────────────────────────────────────────────────┘
                           │
                           ▼
┌────────────────────────────────────────────────────────┐
│                 API Layer (Next.js)                      │
│  ┌──────────────┐  ┌──────────────┐  ┌──────────────┐   │
│  │   Upload     │  │  Extraction  │  │  Generation  │   │
│  │    API       │  │    API       │  │    API       │   │
│  └──────────────┘  └──────────────┘  └──────────────┘   │
└────────────────────────────────────────────────────────┘
                           │
          ┌────────────────┼────────────────┐
          ▼                ▼                ▼
  ┌──────────────┐  ┌──────────────┐  ┌──────────────┐
  │  Cloudinary  │  │    OpenAI    │  │  PostgreSQL  │
  │   Storage    │  │    GPT-4     │  │   Database   │
  └──────────────┘  └──────────────┘  └──────────────┘
```

Database Schema

New Models

Document

```
model Document {
  id                String          @id @default(cuid())
  type              DocumentType
  category          String?         // Auto-classified category
  fileName          String
  fileUrl           String          // Cloudinary URL
  fileSize          Int
  mimeType          String

  // OCR & Extraction
  extractedText     String?         @db.Text
  extractedData     Json?           // Structured data from AI
  extractionStatus  ExtractionStatus @default(PENDING)
  extractionError   String?

  // Compliance
  isRequired        Boolean         @default(false)
  expirationDate    DateTime?
  complianceStatus  ComplianceStatus @default(PENDING)

  // Relationships
  residentId        String?
  resident          Resident?       @relation(fields: [residentId], references: [id])
  inquiryId         String?
  inquiry           Inquiry?        @relation(fields: [inquiryId], references: [id])
  uploadedById      String
  uploadedBy        User            @relation(fields: [uploadedById], references: [id])

  // Metadata
  tags              String[]
  notes             String?
  version           Int             @default(1)

  createdAt         DateTime        @default(now())
  updatedAt         DateTime        @updatedAt

  @@index([residentId])
  @@index([inquiryId])
  @@index([type])
  @@index([complianceStatus])
}

enum DocumentType {
  MEDICAL_RECORD
  INSURANCE_CARD
  ID_DOCUMENT
  CONTRACT
  FINANCIAL
  CARE_PLAN
  EMERGENCY_CONTACT
  OTHER
}

enum ExtractionStatus {
  PENDING
  PROCESSING
  COMPLETED
  FAILED
}

enum ComplianceStatus {
```

```
    PENDING
    COMPLIANT
    MISSING
    EXPIRED
    EXPIRING_SOON
  }
```

DocumentTemplate

```
model DocumentTemplate {
  id          String    @id @default(cuid())
  name        String
  description String?
  type        DocumentType
  template    Json      // Template structure
  fields      Json      // Field mappings
  isActive    Boolean   @default(true)

  createdAt   DateTime  @default(now())
  updatedAt   DateTime  @updatedAt
}
```

Implementation Timeline

Week 1: Foundation (Days 1-7)

Phase 1A: Storage & Database
- Set up Cloudinary integration
- Create database schema
- Run Prisma migrations
- Create upload API endpoint

Phase 1B: Upload UI
- Build upload component
- Drag-and-drop functionality
- File validation
- Progress indicators
- Document list view

Week 2: Extraction (Days 8-14)

Phase 2: OCR
- Integrate Tesseract.js
- Google Cloud Vision API setup
- Text extraction API
- Process uploaded documents

Phase 3: AI Field Extraction
- OpenAI integration for extraction
- Field mapping system
- Form auto-population
- Validation and error handling

Week 3: Classification & Compliance (Days 15-21)

Phase 4: Classification
- Document type classification
- Category assignment
- Search functionality
- Filtering and organization

Phase 5: Compliance
- Compliance rules engine
- Required documents tracking
- Expiration monitoring
- Compliance dashboard

Week 4: Generation & Polish (Days 22-28)

Phase 6: Document Generation
- Template system
- Auto-fill functionality
- PDF generation
- Common templates

Phase 7: Testing & Deployment
- Comprehensive testing
- Bug fixes
- UI polish
- Performance optimization
- User documentation
- Deployment

Success Metrics

Technical Metrics

- ✅ Upload success rate > 99%
- ✅ OCR accuracy > 95%
- ✅ Field extraction accuracy > 90%
- ✅ Classification accuracy > 85%
- ✅ Processing time < 30 seconds per document

Business Metrics

- ⏱️ Time saved: 5-10 hours/week per operator
- 📊 Data entry errors reduced by 80%
- ✅ Compliance rate increased to 100%
- 🚀 Onboarding time reduced by 50%
- 😊 User satisfaction > 4.5/5

User Roles & Permissions

Operators

- Upload documents

- View all documents
- Extract and edit data
- Generate documents
- View compliance dashboard

Admins

- All operator permissions
- Manage templates
- Configure compliance rules
- View analytics
- Delete documents

Family Members

- View documents (limited)
- Upload documents for their resident
- No editing permissions

Security & Privacy

Data Protection

- Encrypted file storage (Cloudinary)
- Encrypted database fields for sensitive data
- HIPAA compliance considerations
- Access logging and audit trails

File Validation

- File type restrictions
- File size limits (10MB per file)
- Virus scanning (future enhancement)
- Content validation

Future Enhancements (Post-Launch)

1. **Batch Processing** - Upload and process multiple documents at once
2. **Mobile App** - Upload documents from mobile devices
3. **E-Signature Integration** - DocuSign/HelloSign integration
4. **Advanced Analytics** - Document processing insights
5. **AI Training** - Improve extraction accuracy over time
6. **Multi-Language Support** - OCR for non-English documents
7. **Version Control** - Track document changes over time
8. **Automated Workflows** - Trigger actions based on document uploads

---

Status: Planning Complete - Ready for Implementation
Start Date: December 19, 2025
Target Completion: January 16, 2026 (4 weeks)