



## Legal Document Analysis & Chronology Builder



### Project Overview

#### AI-Powered Legal Document Analysis Platform with Advanced Chronology Building

- **Purpose:** Automate legal document analysis and extract chronological events
- **Target Users:** Legal professionals, compliance teams, business analysts, researchers
- **Core Innovation:** Direct PDF-to-analysis workflow with AI-powered insights



### System Architecture

#### Frontend Layer

- **Streamlit Web Interface** (Port 8501)
- Interactive PDF upload and analysis selection
- Real-time configuration options
- Visual results display with charts and metrics

#### Backend Layer

- **FastAPI REST API** (Port 8000)
- 8 specialized endpoints for different analysis types
- PDF text extraction with OCR fallback
- AI processing with Groq LLM (Llama 3.3 70B)

#### AI Processing Layer

- **LLM Processor:** Advanced prompt engineering for legal analysis
- **Document Tagging:** Automatic section identification and chunking

- **Temporal Processing:** Date normalization and relative date resolution

## **Core Features & Capabilities**

### **Document Processing**

- **PDF Text Extraction:** Handles digital and scanned PDFs with OCR
- **Document Structure Analysis:** Automatic section identification
- **First-Page Processing:** Optimized for legal document headers

### **AI Analysis Modules**

- **Clause Analysis:** Identify and classify legal clauses (8 categories)
- **Chronology Builder:** Extract and organize chronological events
- **Document Chat:** Interactive Q&A with document content
- **Risk Assessment:** Identify legal risks with severity levels
- **Document Classification:** Categorize by type, subject, importance
- **Document Summarization:** Generate customizable summaries

### **Chronology Features**

- **Event Extraction:** 10 event types (Legal Action, Contract Event, Payment, etc.)
- **Temporal Resolution:** Normalizes dates and resolves relative expressions
- **Timeline Construction:** Builds ordered chronological sequences
- **Conflict Detection:** Identifies temporal inconsistencies

### **Technical Stack**

#### **Backend Technologies**

- **Python 3.11+:** Core programming language
- **FastAPI:** Modern, fast web framework for APIs

- **Pydantic**: Data validation and serialization
- **PyPDF2/pypdfium2**: PDF text extraction
- **Tesseract OCR**: Scanned document processing

### **AI & ML Technologies**

- **Groq API**: High-performance LLM inference
- **Llama 3.3 70B**: Large language model for analysis
- **LangChain**: LLM application framework
- **spaCy**: Natural language processing
- **Weaviate**: Vector database for search (optional)

### **Frontend Technologies**

- **Streamlit**: Interactive web application framework
- **Plotly**: Interactive data visualizations
- **Pandas**: Data manipulation and analysis

## **API Endpoints**

### **System Endpoints**

- GET / - API information and feature list
- GET /health - System health check

### **Document Processing**

- POST /extract-text/ - PDF text extraction and tagging

### **AI Analysis Endpoints**

- POST /analyze-clauses/ - Legal clause analysis (PDF upload)
- POST /extract-chronology/ - Chronology extraction (PDF upload)
- POST /chat-with-document/ - Interactive document chat (PDF upload)

- POST /summarize-documents/ - Document summarization (PDF upload)
- POST /assess-risks/ - Legal risk assessment (PDF upload)
- POST /classify-document/ - Document classification (PDF upload)

## Key Innovations

### Simplified Workflow

- **One-Step Process:** Upload PDF directly to any analysis mode
- **No Pre-processing:** Eliminates need for text extraction step
- **Real-time Configuration:** Options available in each analysis mode

### Advanced AI Capabilities

- **Plain-English Output:** Converts legal jargon to understandable language
- **Context-Aware Analysis:** Maintains document context throughout processing
- **Confidence Scoring:** Provides reliability metrics for all analyses

### Temporal Intelligence

- **Smart Date Resolution:** Handles relative dates like "two weeks later"
- **Event Classification:** 10 specialized event types for legal documents
- **Timeline Validation:** Detects and reports temporal conflicts

## Business Value

### Efficiency Gains

- **90% Time Reduction:** From hours to minutes for document analysis
- **Automated Processing:** Eliminates manual document review
- **Scalable Analysis:** Process multiple documents simultaneously

## Quality Improvements

- **Consistent Analysis:** Standardized AI-powered insights
- **Comprehensive Coverage:** No missed clauses or events
- **Risk Identification:** Proactive legal risk detection

## Cost Benefits

- **Reduced Legal Fees:** Faster document review process
- **Lower Compliance Costs:** Automated risk assessment
- **Improved Decision Making:** Data-driven legal insights

## Deployment & Usage

### System Requirements

- **Python 3.11+:** Minimum version requirement
- **4GB RAM:** Minimum (8GB recommended)
- **Internet Connection:** Required for AI processing
- **API Keys:** Groq API key (Weaviate optional)

### Quick Start

1. **Install Dependencies:** `pip install -r requirements.txt`
1. **Set Environment:** Configure `.env` file with API keys
1. **Start Backend:** `python run_backend.py` (Port 8000)
1. **Start Frontend:** `python run_frontend.py` (Port 8501)
1. **Access Platform:** Open <http://localhost:8501>

### Usage Workflow

1. **Upload PDF** → Select analysis mode → Configure options → Get results
1. **View Analysis** → Interactive charts, metrics, and detailed insights
1. **Export Results** → JSON, CSV, or summary reports

## **Target Applications**

### **Legal Practice**

- Contract analysis and clause identification
- Case timeline building and event tracking
- Legal risk assessment and compliance checking

### **Business Operations**

- Document classification and priority management
- Executive summaries and key point extraction
- Interactive document Q&A for quick insights

### **Research & Analysis**

- Document pattern analysis and comparison
- Chronological event sequencing
- Legal document structure analysis

## **Future Enhancements**

### **Planned Features**

- **Document Comparison:** Side-by-side analysis of multiple documents
- **Advanced Search:** Semantic and temporal search capabilities
- **Batch Processing:** Multiple document analysis workflows
- **API Integration:** Third-party legal system integration

### **Scalability**

- **Cloud Deployment:** AWS/Azure containerized deployment
- **Multi-tenant Support:** Organization-specific document processing
- **Advanced Analytics:** Machine learning insights and trends

Caution: Tested only with contract document attached in code base

Code Base link: [Demos](#)

Document links:

Total Research work document: [True-Private AI Workspace for Legal Document Analysis.docx](#)

Clause classification and analysis: [Clause Summarization Research work.docx](#)

One page document for clause classification+ Summarisation: [Clause Classification + Summarisation.docx](#)

Chronology Builder: [Chronology Builder\(Automatic Event & Timeline Generation from Documents\).docx](#)

Document Classification: [True-Private AI Workspace for Legal Document Analysis.docx](#)

Plugins Research work: [Word Add-ins or Plugins.docx](#)

Written by

Raju

AI Engineer

CognitBotz Solutions