Software Requirements Specification (SRS) Document

for

LACPD1 AWS

Prepared by Niyusha Zarnegar, Dominick Daito Jr, Tuan Khai Tran, Averii Bell, Steven Partida, Daniel Hernandez, Viyoka Lim, Matthew Sanchez, Luis Rosas

Sponsored by LA County Public Defender's Office

Version 2.5

[12/06/24]

Revision History

Version	Date	Author	Change Description
0.1	9/30/24	Niyusha Zarnegar	Initial
0.2	10/01/24	Tuan Khai Tran	Introduction
0.3	10/02/24	Tuan Khai Tran	General Description
0.4	10/03/24	Tuan Khai Tran	Requirements
0.5	10/04/24	Daniel Hernandez	Requirements
0.6	10/04/24	Dominick Daito Jr	General Description
0.7	10/05/24	Averii Bell	General Description
0.8	10/05/24	Steven Partida	User Interfaces
0.9	10/6/24	Niyusha Zarnegar	Introduction & Formatting
0.10	10/09/24	Steven Partida	User Interfaces
0.11	10/09/24	Steven Partida	Performance Requirements
0.12	10/09/24	Dominick Daito Jr	General Description
1.1	11/08/24	Tuan Khai Tran	Introduction: Purpose
1.2	11/12/24	Daniel Hernandez	General Description
1.3	11/13/24	Viyoka Lim	Requirements
2.0	11/28/24	Viyoka Lim	1.2 and 2.1
2.1	12/02/24	Dominick Daito Jr	General Description
2.2	12/02/24	Dominick Daito Jr	General Description
2.3	12/02/24	Dominick Daito Jr	Requirements
2.4	12/05/24	Tuan Khai Tran	6. Legal and Ethical Consideration
2.4	12/05/24	Tuan Khai Tran	4. Requirement Specification
2.4	12/05/24	Tuan Khai Tran	5. Other Nonfunctional Requirements
2.5	12/05/24	Niyusha Zarnegar	User Characteristics

TABLE OF CONTENTS

1.
INTRODUCTION

1. Introduction

1.1 Purpose

The Software Requirements Specification (SRS) document describes a transcript analysis and case management system developed for the Los Angeles County Public Defender's Office (LACPD). This system is specifically designed to analyze court transcripts and flag police misconduct patterns. The system will integrate AWS cloud services and advanced data science techniques, particularly Natural Language Processing (NLP) and generative AI, to detect potential misconduct, identify key entities (such as officer names and badge numbers), and flag inconsistencies in testimonies for legal use.

The SRS aims to ensure all stakeholders, including project sponsors, developers, testers, and legal personnel, have a shared understanding of the system's capabilities and its role in supporting legal workflows.

1.2 Scope

The Transcript Analysis and Case Management System is a software solution created for the Los Angeles County Public Defender's Office (LACPD). Its main goal is to improve the efficiency and accuracy of legal workflows by enhancing the analysis of court transcripts. By utilizing AWS cloud services, Natural Language Processing (NLP), and generative AI models, the system provides real-time processing of transcripts, helping legal teams identify critical insights.

Key Benefits

- **Streamlined Workflows:** Automates tasks like document review and analysis.
- **Accurate Insights:** Highlights critical information such as misconduct patterns and inconsistencies.
- **Scalable Design:** Adapts to increasing case data without compromising performance.

Objectives and Capabilities

1. **Misconduct Detection:**
    - Automatically identify instances of police misconduct based on flagged patterns or keywords in transcripts and evidence documents.
    - Highlight repeated behaviors or systemic issues to strengthen arguments.

2. **Entity Extraction:**
    - Extract and categorize important entities such as officer names, badge numbers, dates, and locations.
    - Present these entities in a structured, searchable format for quick reference.
3. **Testimony Analysis:**
    - Analyze witness statements and cross-reference them to identify contradictions or inconsistencies.
    - Highlight tone shifts or emotional language that may impact case outcomes.
4. **Scalability and Performance:**
    - Leverage cloud-based infrastructure to process and store large volumes of legal documents.
    - Maintain low latency for real-time document retrieval and analysis, even with complex cases.

Applications and Benefits

Applications:

1. **Legal Case Preparation:**
    - Streamlines the review of transcripts and documents, allowing attorneys to focus on building arguments.
2. **Legal Research:**
    - Assists in identifying patterns or precedents relevant to ongoing cases.

Benefits:

- **Time Efficiency:** Automates repetitive tasks, reducing manual effort.
- **Improved Accuracy:** Minimizes human error by flagging critical issues.
- **Enhanced Insights:** Provides actionable information that strengthens case strategies.
- **Scalability:** Accommodates growing data demands and user bases seamlessly.

Exclusions:

1. **Judgment Prediction:**
    - The software does not predict case outcomes or court decisions.
2. **Complex Legal Interpretation:**
    - It does not replace attorney expertise in interpreting legal arguments or evidence.
3. **Manual Data Input:**
    - Requires users to upload documents and provide metadata manually; it cannot fetch data from external sources.
4. **Cross-Language Analysis:**
    - Currently limited to documents in English; non-English transcripts require pre-translation.

1.3 Overview

This SRS document is organized as follows:

- Section 1: Provides an introduction, including the purpose, scope, and structure of the document.
- Section 2: Describes the general factors affecting the software, including the product perspective, product functions, and user characteristics.
- Section 3: Details the specific software requirements, including functional requirements, external interfaces, and performance criteria.
- Section 4: Lists any references to external documents or standards used in this specification.

1.4 Definitions, Acronyms, and Abbreviations

There will be more in Appendix A: Glossary

Term/Acronym	Definition
LAPD	Los Angeles Police Department
LACPD	Los Angeles County Public Defender
AWS	Amazon Web Services
SRS	Software Requirements Specification

ML	Machine Learning
NLP	Natural Language Processing
CMS	Case Management System
IEEE	Institute of Electrical and Electronics Engineers

1.5 References

a) *Provide a complete list of all documents referenced elsewhere in the SRS;*

b) Identify each document by title, report number (if applicable), date, and publishing organization

(The below is not official)

1. AWS Security and Compliance Whitepaper.
2. IEEE Std 830-1998: IEEE Recommended Practice for Software Requirements Specifications.

2.
GENERAL DESCRIPTION

2. General Description

2.1 Product Perspective

The LACPD Transcript Analysis System is a cloud-based software solution developed as a standalone application that integrates into LACPD's existing case management workflow. The system will utilize Box.com for data storage and AWS cloud services for computing power and machine learning model hosting. It will use Natural Language Processing (NLP) models to extract key information and classify different segments of the court transcripts. The product will also provide a web-based interface for querying and visualizing the relationships between various cases.

- **System Interfaces**
    - The system interacts with LACPD's case management processes by enabling the manual upload and export of relevant case data, such as court documents and transcripts. While direct integration through APIs with LACPD's existing systems is currently dependent on their approval and infrastructure capabilities, the system is designed to be integration-ready. This allows for future API-based synchronization, should LACPD permit direct data exchange. Exported insights and flagged data can be provided in widely compatible formats (e.g., JSON, XML, CSV) for manual or automated re-integration into external tools. Key data such as case details, court record hearings, transcripts, and other relevant documents will be exchanged, reducing the need for manual input.
- **User Interfaces**
    - The user interface is designed to be minimalistic, clean, and easy to use.
    - Being intuitive, it allows LACPD personnel to easily utilize and navigate data.
    - The UI will allow users to query court transcripts, view analysis results, and explore visualized data relationships through web-based dashboards.
- **Hardware Interfaces**
    - Requires minimal hardware. Users will access the system through standard internet-connected devices.
    - Functionality will occur on all screen sizes and device types while being optimized for low resources.

- **Software Interfaces**
    - Utilizes standard libraries for NLP tasks and may integrate with other LACPD software tools.
    - We may include libraries such as spaCy or NLTK for text processing and TensorFlow or PyTorch for machine learning tasks.
    - The software will also handle basic data formats like PDFs, plain texts, and structured data for importing and exporting transcripts.
- **Communications Interfaces**
    - Data will be transmitted securely over the Internet using HTTPS encryption. The system will comply with LACPD's communications protocols for data integrity and confidentiality.
- **Memory**
    - Efficiently manage memory usage, particularly during data processing phrases. Dynamically allocate resources based on system demands.
- **Operations**
    - Operated by the LACPD IT department with guidelines in place for updates, security checks, and user support. Regular maintenance will be held to ensure proper functionality and the latest support.
- **Site Adaptation Requirements**
    - The deployment will be tailored to the LACPD operational environment such as a variety of cases, workflow, training management, and data backup and recovery.
    - As the department continues to grow, further updates will be pushed to adhere to different changes.

2.2 Product Functions

The primary functions of the system include:

- Real-time Transcript Ingestion and optional OCR Processing: Allow upload of court transcripts in various formats (PDF, DOCX), converting non-searchable text into searchable format where needed. OCR Processing will be used as a failsafe.
    - If the file includes non-machine-readable text (e.g., scanned transcript), OCR is applied to extract content.

- Text Analysis, Entity Extraction, and Misconduct Detection: Utilize NLP and generative AI models to identify and extract key entities (e.g., officer names, badge numbers, locations, dates, charges). The system will analyze transcripts to detect potential patterns of misconduct or inconsistencies within and across transcripts.
- Relationship Linking and Case Correlation: Use machine learning algorithms to identify links between cases based on shared attributes and contextual similarities, presenting attorneys with visual case connections.
- Multi-Level Flagging System: Generate flags on transcripts and highlight key findings. Flagging will incorporate color-coded indicators to convey levels of confidence in misconduct detection (e.g., red for high-confidence flags, yellow for moderate).

2.3 User Characteristics

The intended users of the system include:

- Attornies, Paralegals, Legal Assistants, Detectives, and Investigators: Use the system to review case transcripts and identify key information.
    - Attorney:
        - Varying levels of experience from recent law school graduates to years of practice.
        - Familiar with databases and case management software.
    - Paralegals and Legal Assistants:
        - Education level often includes an associate's degree or bachelor's degree. A certificate in paralegal studies would be a bonus.
        - Experience typically possesses some years in legal environments, such as document management, case preparation, and client interaction.
        - Technical expertise may be familiar with legal documentation and basic software tools.
        - They may have specialized training in investigative techniques and are comfortable with technological tools.
- LACPD IT Administrators: Utilize the management portal to ensure the proper state and function of the software system.
    - LACPD IT Administrators:

    - Education level often carries a degree in information technology or computer science.
    - Experience in managing software systems and security performance.
    - Technical Expertise includes a high level of technical knowledge with proficiency in software management, system integration, and data security.

2.4 General Constraints

- The system must comply with LACPD's data privacy and security regulations.
- AWS cloud services will be the primary platform for all data processing and storage.
- The system must support multi-user access with role-based permissions.
- The system should have a 24 hour RPO and a <= 6 hour RTO

2.5 Assumptions and Dependencies

Availability of Training Data (Transcripts)

- The development and effectiveness of machine learning models for this project rely on continuous access to accurate, high-quality transcripts. Court transcripts, as the primary data source, must be:
    - Timely and accurately transcribed, capturing recent cases and maintaining consistent language patterns.
    - Formatted uniformly to support reliable processing across models.
    - Extensive enough to enable effective model training for accurate insights.
- Any issues, such as delays or formatting inconsistencies, could significantly impact the model's performance. Therefore, uninterrupted access to high-quality transcripts is critical to project success.

Integration with LACPD's Case Management System (CMS) and Other Databases

- Successful implementation requires seamless integration with LACPD's existing CMS and related databases, allowing the system to retrieve relevant data, process it, and store results. Key integrations include:

- Bi-directional data exchange with LACPD's CMS to enable real-time updates, facilitating immediate availability of flagged information.
- Secure APIs for data transfers to maintain data integrity and confidentiality across systems.

- Any interruptions or updates to LACPD's existing systems may impact the functionality and performance of the machine learning models. Ensuring compatibility and establishing reliable communication protocols with these systems is essential to maintain a seamless workflow for the legal team.

AWS Infrastructure and Service Availability

- AWS infrastructure will support scalable, flexible, and secure deployment for model training, inference, and data storage. The system will rely on specific AWS services, such as:
  - **SageMaker**: For model training, inference, and deployment.
  - **Box.com:** For secure storage of transcript data and associated metadata.
  - **Comprehend**: For Natural Language Processing and key entity extraction.
- The assumption is that AWS services will remain accessible, reliable, and cost-effective throughout the project lifecycle. Any changes or interruptions to these services, such as unexpected downtimes or cost increases, may impact project timelines, budget, and outcomes.

Data Security and Compliance Requirements

- The system must comply with LACPD's data privacy and security policies, as well as applicable legal and regulatory standards. This includes:
  - Adherence to secure data handling practices within AWS, such as encryption at rest and in transit, to protect sensitive case information.
  - Role-based access control to ensure that only authorized personnel can access the system's data and functions.
- Maintaining this security and compliance is critical to building a trusted system for legal use, and any changes in policy may require updates to the system's design or architecture.

3.
REQUIREMENTS

3. Requirements

3.1 Functional Requirements

3.1.1 Transcript Upload and Storage

- **Description**: The system shall allow users to upload court transcripts in supported formats (PDF, DOCX). OCR will convert any non-searchable document text into a searchable format.
- **Requirement**: The system shall store all uploaded and processed transcripts in Box.com with encryption, ensuring secure access for authorized personnel.

3.1.2 Text Analysis, Key Entity Extraction, and Misconduct Detection

- **Description**: The system shall utilize NLP models to extract key entities within transcript text (e.g., officer names, badge numbers, locations, charges) and detect misconduct patterns or inconsistencies, automatically flagging critical sections for legal review.
- **Requirement**: AWS Comprehend or an equivalent NLP model will handle entity extraction and misconduct detection. Results will be stored in a structured, searchable format accessible to authorized users.

3.1.3 Case Linking and Relationship Detection

- **Description**: The system shall identify and link related cases by analyzing shared attributes or contextual similarities, providing visual connections for easier analysis and tracking.
- **Requirement**: The system shall employ clustering or similarity algorithms to detect and present case relationships to users through an intuitive interface, facilitating cross-case analysis.

3.1.4 Multi-Level Flagging System

- **Description**: The system shall generate color-coded flags within transcripts, indicating varying confidence levels in detected issues (e.g., red for high-confidence flags, and yellow for moderate confidence).
- **Requirement**: Flags shall be generated and applied automatically by the system based on NLP analysis results, with options for users to export flagged sections or entire transcripts in PDF or CSV formats for further review and case reporting.

3.2 External Interface Requirements

3.2.1 Dashboard

- Purpose: Central hub for accessing and managing case information.
- Key Features:
    - Overview of Active Cases: Display a list of ongoing cases with quick links to associated documents.
    - Recent Activity Feed: Show recently uploaded, reviewed, or flagged documents.
    - Notifications Panel: Highlight critical updates, such as flagged issues, pending uploads, or incomplete tasks.
    - Search Bar: Provide a global search tool for locating cases, transcripts, or flagged information.

3.2.2 Case Explorer

- Purpose: Allow users to manage and organize case-related files.
- Key Features:
    - File Directory: Organized folder structure by case ID, client name, or metadata tags.
    - Filter Options: Enable sorting by date, document type, flagged issues, or entity mentions.
    - Bulk Actions: Allow multiple files to be uploaded, moved, or annotated simultaneously.
- Note:
    - Enhancements tool
    - Case Timeline View: Introduce a timeline that displays key case milestones (e.g court date, filing deadlines)

3.2.3 Document Viewer

- Purpose: Provide a workspace for attorneys to analyze and interact with case documents.
- Key Features:
    - Split-Screen View:
        - Left Panel: Display the full transcript or case document.
        - Right Panel: Highlight flagged content (e.g., misconduct patterns, inconsistencies).
    - Search Within Document: Allow users to search for specific terms or phrases in the transcript.

- Annotations: Enable users to add notes, highlight key sections, or bookmark pages.
- Entity Insights: Present identified entities (e.g., officer names, badge numbers) in an interactive side panel.
- Export Options: Allow attorneys to download annotated or flagged versions of documents.

- Note:
  - Version Control: Include a feature to track changes and view previous versions of annotated documents

3.2.4 Advanced Search and Filter

- Purpose: Help users find specific information quickly.
- Key Features:
  - Global Search: Search across all cases for transcripts, flagged issues, or specific keywords.
  - Advanced Filters: Narrow down results based on:
    - Date range.
    - Document type (e.g., transcript, evidence).
    - Metadata (e.g., client name, case ID).
  - Saved Searches: Let users save frequent searches for future use.

3.2.5 Insights and Reports

- Purpose: Summarize critical findings from transcripts and flagged documents.
- Key Features:
  - Misconduct Analysis: Present flagged patterns, inconsistencies, and correlations.
  - Entity Report: List all identified entities (e.g., names, locations) with their frequency of occurrence.
  - Case Summary: Generate a high-level overview of the case, including flagged issues and annotations.

3.2.6 Upload Interface

- Purpose: Simplify the process of adding documents to the system.
- Key Features:
  - Drag-and-Drop Upload: Allows users to easily drag and drop files for upload.
  - File Validation: Immediately check file type, size, and format compatibility.
  - Progress Indicator: Show upload status with estimated time to completion.
  - Error Handling: Notify users of issues (e.g., file too large) with actionable guidance.

3.2.7 Notifications and Alerts

- Purpose: Keep users informed about system activity and issues.
- Key Features:
  - Success Notifications: Confirm successful actions (e.g., "File uploaded successfully").
  - Error Alerts: Notify users of system errors with clear instructions (e.g., "Connection lost. Please try again.").
  - Reminders: Prompt users about pending actions or deadlines (e.g., "Review flagged documents in Case #12345.").

3.2.8 User Settings and Preferences

- Purpose: Allow customization to enhance user experience.
- Key Features:
  - Profile Settings: Update user details, such as name or role.
  - Accessibility Options: Adjust font size, color contrast, or language preferences.
  - Notification Preferences: Choose how and when to receive alerts (e.g., email or in-app).

3.3 User Interfaces

- The User Interface of the LACPD Transcript Analysis System is crucial for ensuring that users, regardless of their technical expertise can easily utilize and interact with the system to perform their tasks efficiently and accurately. This section will specify the requirements for

the system's user interfaces, including configuration details, content organization, and guidelines for optimizing user experience.

- Requirements of Each Interface Between the Software Product and Its Users
  - Screen Layout
    - Dashboard: The system's primary screen should be a dashboard that provides users with an overview of their current tasks and cases. The dashboard will include:
      - A case search bar for navigation and retrieval of court transcripts.
      - Visualizations of case relations (graphs, charts, etc)
      - Summary of recently accessed cases and documents.
      - Quick access buttons for common tasks such as “Analyze Transcripts”, “Download Report”, and “View Case History”.
    - Search Results and Case View: When a user performs a search, the result should be displayed in an easy-to-read format, showing key metadata such as:
      - Case ID
      - Case Title
      - Case Date
      - A summary of the transcripts
      - The user can click on individual results for more detailed information including full analysis.
    - Analysis Results: When a transcript is processed, the results should be displayed in a structured format that includes:
      - Key entities (e.g. names, locations, charges, etc.) highlighted within the transcript
      - Categorized segments of the transcript (e.g. introduction, witness testimony, verdict, etc)
      - Confidence score that indicates the reliability of the model analysis.
      - Interactive graphs or charts displaying relationships between entities (e.g. the relationship between defendants, witness, and location.)

- Reports and Exports: The system will allow users to download reports summarizing the analysis results. Reports will include:
    - Full court transcripts with analysis annotations.
    - Summary tables of key information.
    - Charts for insights.
- Error or Confirmation Notification: When users interact with the system, feedback involving error and confirmation notifications will be given to help users navigate the user interface.
    - Error message (e.g. “Invalid case ID”)
    - Success message (e.g. “Report generated successfully!”)
- Navigation and Interaction: The interface should provide navigation menus with clear labels to access various sections of the application, such as:
    - Case Search
    - Transcript Analysis
    - Reports
    - Settings/Configuration (admin/IT staff)
- A back button will allow users to return to the previous screen or dashboard.
- Content Formatting
    - Reports should be organized in a readable manner and format. Key fields such as case numbers, dates, and extracted entities will be presented in a tabular format where filtering and sorting will be an option

- General Interface Characteristics
    - Consistency: The software shall maintain a consistent look and feel across all screens and modules, including uniform color schemes, fonts, and button styles.
    - Responsiveness: The interface shall be responsive and adapt to various screen sizes and resolutions.
- Optimizing User Interaction
    - Do’s

        - Provide Immediate Feedback: The system shall offer immediate visual or auditory feedback in response to user actions (e.g., button clicks, documentation download)
        - Use Familiar Icons and Symbols: Commonly recognized icons shall be used to represent standard actions
    - Don’t
        - Information Overload: The interface shall not display unnecessary information that may overwhelm the user.

3.4 Hardware Interfaces

- The system shall support interaction with AWS cloud hardware for model training and deployment.

3.5 Software Interfaces

Amazon EC2

Name: Amazon Elastic Compute Cloud

Mnemonic: EC2

Specification Number: AWS-EC2-001

Version Number: Latest available at project initiation

Source: AWS

Purpose:

EC2 instances provide computational resources to support model inference, data processing, and other computationally intensive tasks as required by the system.

Interface Definition:

Input: Compute job requests (e.g., for data preprocessing or model inference).

Output: Processed data passed to storage services like S3 or directly to SageMaker for further analysis.

Format Reference: AWS EC2 API (AWS Developer Documentation) for starting, stopping, and monitoring instances.

- **AWS Services**: The system shall use AWS SageMaker for NLP model training and inference, S3 for secure storage of transcripts, and EC2 for computational needs. Data processed on AWS will be encrypted and accessible only to authenticated users.
- The system shall use AWS SDKs to interact with cloud services and APIs for external communication.
- The system shall use Box.com to store testimony transcripts and other related documents securely.
- The system shall integrate with Box.com for secure cloud storage and management of court transcripts and related documents.

3.6 Communications Interfaces

- The system shall use secure HTTPS protocols for all data transfers and communications between the client and the web server
- The system shall use AWS's encryption and associated security services to ensure the data stays secure throughout the data pipeline including storage

4.
REQUIREMENTS SPECIFICATION

4. Requirements Specification

4.1 Functional Requirements

4.1.1 Transcript Upload and Storage

- The system shall allow users to upload court transcripts in supported formats (PDF, DOCX).
- The system may apply OCR to convert non-machine-readable transcripts into searchable text.
- The system shall store all uploaded transcripts securely in Box.com, with encryption at rest and in transit.
- The system shall associate uploaded transcripts with case metadata, such as case ID, date, and document type, for efficient organization and retrieval.

4.1.2 Text Analysis, Entity Extraction, and Misconduct Detection

- The system shall utilize NLP models to identify key entities (e.g., officer names, badge numbers, locations) within transcripts.
- The system shall detect and flag potential misconduct patterns, including inconsistencies within transcripts and related cases.
- The system shall generate structured, searchable outputs for all analysis results, including entity lists and flagged sections.
- The system shall calculate and display confidence scores for flagged findings.

4.1.3 Case Linking and Relationship Detection

- The system shall analyze shared attributes or contextual similarities between cases to identify relationships.
- The system shall present case linkages visually, using graphs or relationship diagrams, to highlight connections and shared entities.
- The system shall allow users to filter relationship results by criteria such as case date, involved parties, or flagged patterns.

4.1.4 Multi-Level Flagging System

- The system shall assign color-coded flags to indicate levels of confidence in detected misconduct patterns (e.g., red for high confidence, yellow for moderate).

- The system shall allow users to view and manage flagged sections within the transcript viewer.
- The system shall enable the export of flagged findings and associated metadata in formats such as PDF or CSV.

4.2 External Interface Requirements

4.2.1 Dashboard

- The system shall provide a dashboard displaying active cases, recent activity, and notifications.
- The system shall include a global search bar to locate cases, transcripts, or flagged findings.

4.2.2 Case Explorer

- The system shall allow users to organize and manage case-related documents using folder structures.
- The system shall support filtering and sorting by attributes such as document type, upload date, and flagged status.

4.2.3 Document Viewer

- The system shall provide a split-screen view with the transcript on one side and flagged content on the other.
- The system shall support annotations, bookmarking, and interactive exploration of entities.

4.2.4 Advanced Search and Filter

- The system shall allow users to perform global searches across all cases using keywords or metadata.
- The system shall support advanced filters such as date range, document type, or flagged content status.

4.2.5 Insights and Reports

- The system shall generate summary reports of flagged misconduct, identified entities, and key case insights.
- The system shall allow export of reports in multiple formats, including PDF and Excel.

4.2.6 Upload Interface

- The system shall support drag-and-drop file uploads and manual file selection.
- The system shall provide real-time validation feedback for file type and size.

4.2.7 Notifications and Alerts

- The system shall notify users of key events, such as flagged document readiness or pending tasks.
- The system shall provide error messages with actionable guidance for troubleshooting.

4.3 Logical Database Requirements

- The system shall maintain a database linking transcripts, flagged findings, and associated metadata.
- The system shall enforce referential integrity between cases, documents, flags, and metadata tables.
- The system shall store relationships between cases using graph structures for efficient retrieval and visualization.
- The system shall support indexing for search capabilities on attributes such as case ID, entity type, and flagged content.

4.4 Design Constraints

- The system shall comply with AWS's encryption standards for data security, including AES-256 for data at rest and TLS for data in transit.
- The system shall support real-time processing of large datasets, scaling horizontally using AWS services like SageMaker and EC2.
- The system shall adhere to role-based access control (RBAC) to restrict access based on user roles and permissions.
- The system shall ensure compatibility with existing LACPD tools, including Box.com and Salesforce.
- The system shall achieve a recovery time objective (RTO) of ≤6 hours and a recovery point objective (RPO) of 24 hours.

5.
OTHER NONFUNCTIONAL REQUIREMENTS

5. Other Nonfunctional Requirements

5.1 Performance Requirements

- **User Capacity**: The system is expected to support multiple simultaneous users, including attorneys, investigators, and administrative staff. The exact number of supported users under normal and peak conditions will be determined after the client's consultation and a detailed workload analysis.
- **Response Time**: The system should provide acceptable response times for user actions, including uploading documents, retrieving flagged findings, and performing searches. Specific response time thresholds will be defined based on user feedback and operational requirements gathered during implementation.
- **Data Processing**: The system shall efficiently handle data upload, processing, and retrieval tasks. Detailed processing speed and throughput benchmarks will be established after further testing with realistic transcript datasets.
- **Scalability**: The system shall be designed to scale horizontally to accommodate increasing user demand and data volumes. Specific scalability targets will be refined after the client's input and evaluation of operational needs.

5.2 Safety Requirements

- The system shall provide safeguards to ensure the integrity and security of sensitive legal data.
- In the event of a system failure, the system shall implement measures to preserve data integrity and enable recovery within the defined RTO of ≤6 hours.
- The system shall include fail-safe mechanisms to prevent accidental data loss or corruption during processing or user operations.
- The system shall adhere to industry standards for legal software to prevent unauthorized access, misuse, or data breaches that could harm clients or attorneys.

5.3 Security Requirements

- The system shall enforce role-based access control (RBAC) to ensure that only authorized personnel can access sensitive data and perform specific operations.
- All data at rest shall be encrypted using AES-256, and all data in transit shall be encrypted using TLS protocols.

- The system shall require multi-factor authentication (MFA) for all users with administrative roles.
- Logs of all user activities, including access, modifications, and uploads, shall be maintained for auditing purposes.
- The system shall comply with legal and regulatory standards, such as GDPR and CCPA, as well as the LACPD's internal data security policies.
- Regular security audits and penetration testing shall be conducted to identify and address vulnerabilities.

5.4 Software Quality Attributes

- **Adaptability**: The system shall be adaptable to changes in legal workflows and easily configurable to accommodate updates in legal standards or organizational needs.
- **Availability**: The system shall be operational 24/7 with an uptime of at least 99.5%, except during scheduled maintenance.
- **Reliability**: The system shall ensure accurate and consistent results across all transcript analyses, with error rates kept below a predefined threshold to be determined after initial testing.
- **Maintainability**: The system shall be modular to facilitate updates and maintenance, ensuring minimal downtime and impact on users during upgrades.
- **Interoperability**: The system shall integrate seamlessly with Box.com, AWS, and other LACPD tools, ensuring data exchange without compatibility issues.
- **Usability**: The user interface shall be intuitive and easy to navigate, ensuring minimal training requirements for end-users.

5.5 Business Rules

- Attorneys and investigators shall have access only to case files assigned to them, as per LACPD’s case management policies.
- Administrative staff shall have read-only access to case documents unless explicitly granted editing permissions.
- All flagged findings shall require manual review by authorized legal personnel before inclusion in court submissions.

- Data retention policies shall comply with LACPD's guidelines, ensuring that case files are archived or deleted based on predefined timelines.
- External analysts and consultants shall be granted temporary access to specific cases or reports, with access automatically revoked after a set duration.
- All system changes or updates shall require approval from LACPD IT administrators before deployment.

6.
LEGAL
AND
ETHICAL
CONSIDERATIONS

6. Legal and Ethical Considerations

This project addresses key legal and ethical challenges, including user privacy, potential harm, intellectual property, and system security. Adhering to the ACM Code of Ethics and Professional Conduct, our approach ensures compliance with ethical standards and legal requirements.

User Privacy:

Handling sensitive court transcripts requires compliance with principles 1.6 (Respect Privacy) and 1.7 (Honor Confidentiality). We minimize data collection, anonymize records where feasible, and implement encryption and role-based access controls to secure sensitive information. The system complies with GDPR, CCPA, and LACPD-specific privacy standards, ensuring users' transparency and control over their data.

Avoiding Harm:

To mitigate risks like biased analyses or misuse, we follow principle 1.2 (Avoid Harm) by testing and validating NLP models to reduce errors and biases. Security protocols, including HTTPS and AWS services, protect against unauthorized access. Flagged results support attorneys' decisions without replacing their expertise.

Intellectual Property:

The intellectual property rights for the **LACPD Transcript Analysis System** will reside exclusively with the Los Angeles Public Defender's Office (LACPD). All software components, including the Natural Language Processing (NLP) models, machine learning pipelines, and user interface designs, are developed specifically for LACPD's use and are the sole property of the organization. Additionally, any case-related data, transcripts, or other sensitive information processed within the system remains under the ownership of LACPD, and the system is designed not to retain or use this data beyond its intended scope. To protect the integrity of this intellectual property, the system employs secure communication protocols, such as HTTPS, ensuring that data exchanged between LACPD systems and the application remains confidential and secure.

System Security:

Aligned with Principle 2.9, the system leverages AWS services for scalable, reliable, and secure deployment. Regular updates and thorough security audits are conducted to ensure compliance with evolving industry standards, safeguarding the system against potential vulnerabilities and maintaining the trust of stakeholders.

The system prioritizes privacy and minimizes potential harm by implementing robust data protection measures, respecting intellectual property rights, and adhering to legal and ethical guidelines. By integrating these practices, the system delivers actionable insights while ensuring data security, user confidentiality, and ethical compliance.

APPENDIX

Appendix A: Glossary

Term / Acronym	Definition
LACPD	Los Angeles County Public Defender, the primary stakeholder and user of the system.
NLP	Natural Language Processing, a branch of AI focused on understanding and interpreting text data.
OCR	Optical Character Recognition, a technology for converting non-machine-readable text into digital text.
AWS	Amazon Web Services, the cloud platform used for hosting, processing, and data storage.
SRS	Software Requirements Specification, a document outlining the software's goals, design, and requirements.
RBAC	Role-Based Access Control, a security mechanism that restricts system access based on user roles.
MFA	Multi-Factor Authentication, a security method requiring two or more verification factors for access.
RTO	Recovery Time Objective, the maximum acceptable time to restore a system after a failure.
RPO	Recovery Point Objective, the maximum acceptable amount of data loss measured in time before a recovery.
GDPR	General Data Protection Regulation, an EU regulation for data protection and privacy.
CCPA	California Consumer Privacy Act, a California law for protecting consumer data privacy.

Appendix B:

This appendix will contain the models that visually represent the system's design and its interactions. For this project, include:

- **Data Flow Diagram (DFD)**: Visual representation of how court transcripts flow through the system, including upload, OCR processing, analysis, and output.
- **Entity-Relationship Diagram (ERD)**: Show relationships between core database entities, such as:
  - Transcripts
  - Cases
  - Users
  - Flags
- **State-Transition Diagram**: Define how documents move between states (e.g., Uploaded -> Processed -> Flagged -> Reviewed).
- **System Architecture Diagram**: High-level diagram illustrating AWS services (e.g., S3, SageMaker, Comprehend) and how they interact with the system.

Each of these models should be created using tools like Lucidchart, Visio, or similar, and attached here for clarity.

Appendix C: To Be Determined List

Collect a numbered list of the TBD (to be determined) references that remain in the SRS so they can be tracked to closure.

Reference Number	Description
TBD-1	Define exact response times for user actions (e.g., search, upload).
TBD-2	Finalize the maximum number of simultaneous users the system will support.
TBD-3	Specify acceptable latency for large-scale transcript uploads.
TBD-4	Confirm exact integrations with external systems like Salesforce.
TBD-5	Establish detailed benchmarks for NLP model accuracy (e.g., acceptable error rate).