

Design Document: AI Legal Document Summarization Tool

Team Members

Natascha Banda

Morgan-Lee Blake

Michelle Ike

Unaswi Mudanga

Michelle Seth

1. Requirements Document

A Gen AI solution called the Legal Document Summarization Tool was created to address the dangers and inefficiencies associated with manual legal document analysis. The program is capable of summarizing long contracts, extracting important provisions, and identifying any compliance issues by utilizing sophisticated language models and subject modeling. In law firms and corporate legal departments, this increases the effectiveness of legal teams, lowers human error, and improves compliance readiness.

The core functionality of the system will revolve around natural language understanding and intelligent document parsing. The system must be able to summarize legal documents using BERT-based language models and extract categorized key clauses such as termination conditions, indemnity clauses, and confidentiality agreements. It must also identify and flag potentially risky terms or clauses that may raise compliance issues, using contextual understanding from retrieval-augmented generation (RAG) and classification logic. Additionally, BERTopic will be used for topic modeling to provide deeper document insights. The tool should support human-in-the-loop (HITL) validation, allowing legal professionals to review and edit AI-generated outputs before final use. It must enable document uploads in common formats like PDF and DOCX, and support API integrations for seamless connectivity with legal platforms such as Clio and iManage.

The system must maintain strong performance standards, including the ability to respond to document uploads and queries within two seconds for smaller files, and under ten seconds for contracts of up to fifty pages. Security is a top priority; therefore, all user data and document content must be protected using end-to-end encryption. The system must comply with GDPR and other relevant data privacy regulations, especially for clients handling sensitive legal information. It should be scalable, capable of serving individual legal practitioners as well as large enterprise teams. Clients should have the option for on-premise deployment to maintain data sovereignty. Additionally, the platform must ensure high availability, with a target of 99.9%

uptime, and incorporate role-based access control (RBAC) to safeguard user permissions and data access levels.

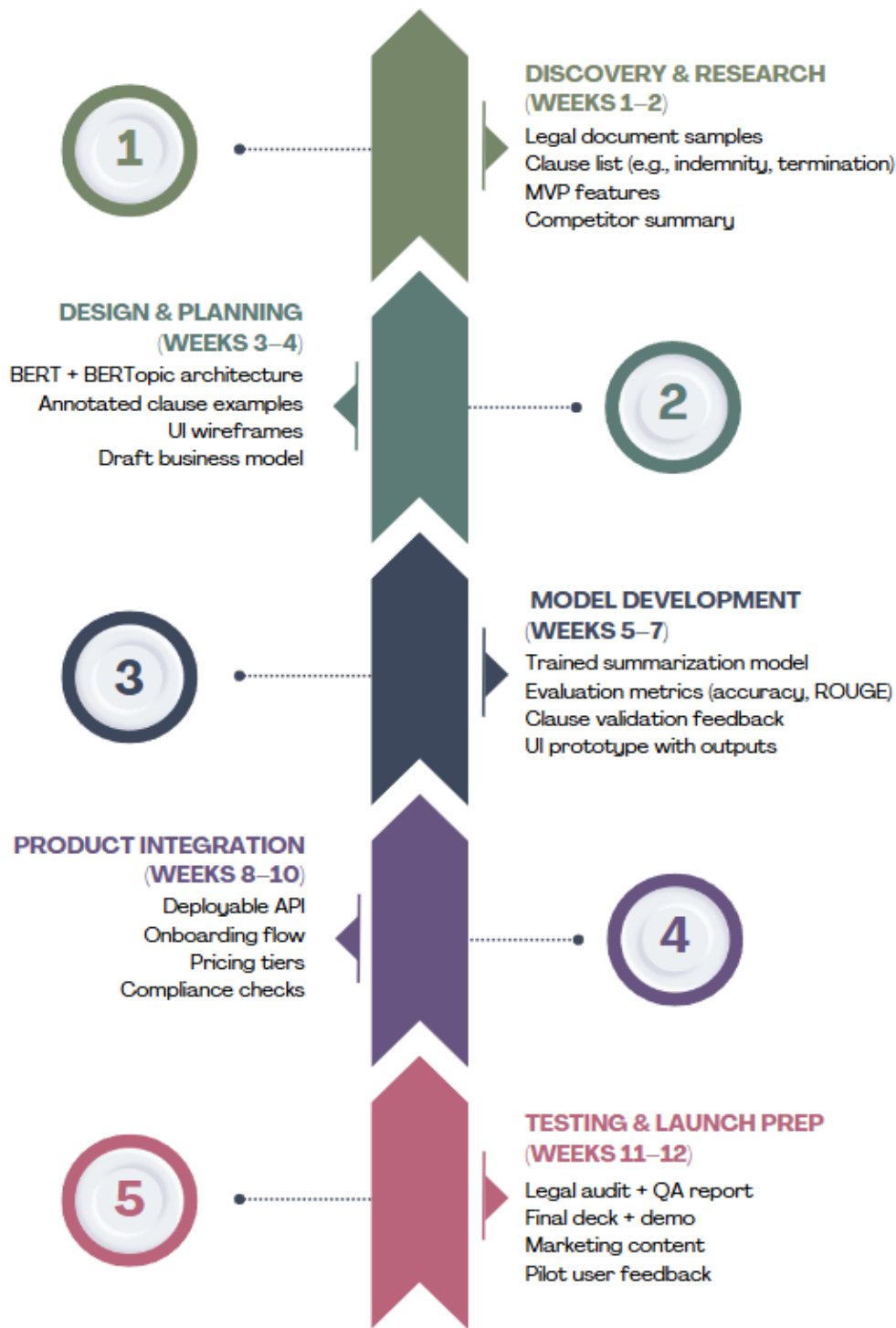
2. User Stories

To ensure the tool meets the needs of its intended users, we developed and prioritized user stories using the MoSCoW method. Core user needs focus on streamlining contract review through automated summarization and clause detection. High-priority (“Must-Have”) stories address the ability to upload legal documents, generate concise summaries, flag risk clauses, and support non-legal users with accessible insights. “Should-Have” and “Could-Have” features, such as document comparison, clause-location linking, and guided onboarding, enhance usability but are not essential for the MVP. Future features like multilingual support and custom AI fine-tuning were identified as out of scope for the current phase.

Priority	User Story
Must-Have	<ul style="list-style-type: none">• As a legal analyst, I want to receive a concise summary of a 50+ page legal contract so that I can review the document more efficiently.• As a compliance officer, I want the tool to flag clauses related to risk so that I can identify potential legal liabilities quickly.• As a non-legal stakeholder (e.g., project manager), I want to understand key obligations in a contract so that I can make informed decisions without relying on legal jargon.• As a user, I want to upload legal documents in common formats (PDF, DOCX) so that I can start analyzing them immediately.• As a team admin, I want to access all document summaries from a dashboard so that I can monitor

	contract analysis progress in one place.
Should-Have	<ul style="list-style-type: none"> • As a legal team lead, I want to compare multiple contract summaries side by side so that I can spot inconsistencies across vendors. • As a user, I want to highlight or click on summarized clauses to see their location in the full document so that I can validate their context. • As a user, I want the tool to integrate with our document management system so that I can avoid uploading manually.
Could-Have	<ul style="list-style-type: none"> • As a legal professional, I want the tool to provide jurisdiction-specific clause explanations so that I can ensure regional compliance. • As a new user, I want access to onboarding guides and tooltips so that I can quickly learn how to use the tool. • As a user, I want to provide feedback on summaries so that the model can improve over time (active learning loop).
Won't-Have	<ul style="list-style-type: none"> • As a legal firm, I want to fine-tune the AI on my internal documents so that it aligns with my writing style and risk preferences. • As a user, I want multilingual summarization support so that I can analyze international contracts.

Legal AI Roadmap



4. Tasks

This section outlines the key responsibilities undertaken by each team member across the five major phases of the AI Legal Document Summarization Tool project. Each phase was designed to support a structured, goal-driven workflow—from initial research through to final system testing and launch. The responsibilities listed below reflect a division of labor that emphasizes technical development, legal expertise, product strategy, and user experience design.

Phase 1: Discovery & Research (Weeks 1–2)

Validating the Problem and Understanding the Legal Landscape

Objective: Validate the problem, collect legal documents, and understand user needs.

Team Member	Tasks
Michelle Ike (Tech Lead)	<ul style="list-style-type: none"> - Research summarization benchmarks tailored to legal documents (e.g., LegalBERT) - Identify technical constraints such as input length limitations, interpretability requirements, and compute needs
Unaswi Mudanga (Data Scientist)	<ul style="list-style-type: none"> - Collect and analyzed a sample of legal documents to examine clause structure and frequency - Explore recurring clause patterns to inform future modeling inputs
Morgan-Lee Blake (Legal Analyst)	<ul style="list-style-type: none"> - Define high-priority clause categories and risk indicators based on real-world legal contracts - Identify common challenges in manual contract review through interviews with legal practitioners
Natascha Banda (Product Manager)	<ul style="list-style-type: none"> - Develop the initial product scope, MVP feature list, and success KPIs - Establish the team’s collaborative workspace and organize the discovery backlog
Michelle Seth (UX/Marketing)	<ul style="list-style-type: none"> - Conduct interviews with legal professionals to understand workflows and pain points - Compile a competitor landscape report highlighting strengths, weaknesses, and market gaps

Phase 2: Design & Planning (Weeks 3–4)

Defining the Solution Blueprint and User Experience

Objective: Finalize model plan, create UI wireframes, validate MVP features.

Team Member	Tasks
--------------------	--------------

Michelle Ike (Tech Lead)	<ul style="list-style-type: none"> - Design the model architecture combining BERT for summarization and BERTopic for topic modeling - Estimate compute and memory requirements based on document size and architecture complexity
Unaswi Mudanga (Data Scientist)	<ul style="list-style-type: none"> - Prototype clause-level topic modeling pipelines and evaluated multiple embedding libraries - Develop a preprocessing workflow for document standardization and tokenization
Morgan-Lee Blake (Legal Analyst)	<ul style="list-style-type: none"> - Provide annotated examples of legal clauses and flagged compliance-relevant content - Define legal rules and decision criteria for highlighting sensitive clauses
Natascha Banda (Product Manager)	<ul style="list-style-type: none"> - Refine the strategic roadmap and develop a business canvas with pricing and adoption considerations - Initiate research into strategic partnerships with legaltech vendors and firms
Michelle Seth (UX/Marketing)	<ul style="list-style-type: none"> - Design user interface wireframes for the contract summary and clause views - Create user personas to guide interaction flows and ensure usability alignment

Phase 3: Model Development (Weeks 5–7)

Building the Core Intelligence Engine

Objective: Build summarization + topic modeling engine.

Team Member	Tasks
Michelle Ike (Tech Lead)	<ul style="list-style-type: none"> - Fine-tune BERT for domain-specific summarization and integrate BERTopic for clause clustering - Direct training iterations and adjust model parameters to suit legal data
Unaswi Mudanga (Data Scientist)	<ul style="list-style-type: none"> - Implement the data cleaning, embedding generation, and evaluation pipeline - Assess topic and summary quality using metrics such as topic coherence and summary accuracy
Morgan-Lee Blake (Legal Analyst)	<ul style="list-style-type: none"> - Review extracted clauses and summaries for legal accuracy and omissions - Provide targeted feedback to improve clause classification fidelity

Natascha Banda (Product Manager)	<ul style="list-style-type: none"> - Ensure alignment of model capabilities with product objectives and legal stakeholder needs - Draft the API integration strategy to support downstream deployment
Michelle Seth (UX/Marketing)	<ul style="list-style-type: none"> - Develop user interface prototypes using model-generated summaries and clause outputs - Create summary card templates and contextual information overlays for flagged clauses

Phase 4: Product Integration (Weeks 8–10)

Connecting the Dots: Backend, Frontend, and User Workflows

Objective: Merge frontend and model, develop onboarding & workflow logic.

Team Member	Tasks
Michelle Ike (Tech Lead)	<ul style="list-style-type: none"> - Package the model into a deployable API and implement clause classification endpoints - Enable real-time access to model outputs within the application interface
Unaswi Mudanga (Data Scientist)	<ul style="list-style-type: none"> - Evaluate full-length document performance and adjusted the inference pipeline accordingly - Address performance bottlenecks and implement optimizations for production readiness
Morgan-Lee Blake (Legal Analyst)	<ul style="list-style-type: none"> - Conduct quality assurance on legal outputs by assessing clause accuracy and completeness - Identify edge cases for further rule-layer development or model refinement
Natascha Banda (Product Manager)	<ul style="list-style-type: none"> - Design the user onboarding workflow, including guided entry points for contract uploads - Develop wireframes for pricing interface and finalize tier testing plans
Michelle Seth (UX/Marketing)	<ul style="list-style-type: none"> - Finalize frontend components and ensure usability of the summary and risk flag views - Integrated clause-level interactivity and refined UI based on internal usability reviews

Phase 5: Testing & Launch Prep (Weeks 11–12)**Final Validation, User Feedback, and Go-to-Market Readiness***Objective: Test full system, document results, prep launch deck.*

Team Member	Tasks
Michelle Ike + Unaswi Mudanga	<ul style="list-style-type: none"> - Conduct end-to-end system evaluations using ROUGE scores, latency benchmarks, and scalability testing - Compile performance documentation to support go/no-go decision for launch
Morgan-Lee Blake (Legal Analyst)	<ul style="list-style-type: none"> - Complete a comprehensive legal audit of model output and validated clause identification accuracy - Develop a final legal compliance checklist and quality control protocol
Natascha Banda (Product Manager)	<ul style="list-style-type: none"> - Finalize the investor pitch deck outlining product capabilities, market positioning, and pricing strategy - Confirm launch deliverables and supported early-access user communications
Michelle Seth (UX/Marketing)	<ul style="list-style-type: none"> - Facilitate live walkthroughs with pilot users and collected post-session feedback - Refine UI components and prepare marketing assets for launch, including web copy and visuals

5. Responsibilities & Roles

Phase	Team Member	Role	Key Responsibilities
1: Discovery & Research (Weeks 1–2)	Natascha Banda	Product Manager	Define MVP scope, establish team tools, organize backlog
	Michelle Ike	Tech Lead	Research legal NLP models (e.g., LegalBERT), assess technical constraints
	Unaswi Mudanga	Data Scientist	Collect/analyze legal documents, identify clause patterns
	Morgan-Lee Blake	Legal Analyst	Identify legal risks, interview legal professionals

	Michelle Seth	UX/Marketing	Conduct user interviews, research competitors
2: Design & Planning (Weeks 3–4)	Natascha Banda	Product Manager	Refine roadmap, explore legaltech partnerships
	Michelle Ike	Tech Lead	Design model architecture, estimate compute needs
	Unaswi Mudanga	Data Scientist	Prototype topic modeling, design preprocessing workflow
	Morgan-Lee Blake	Legal Analyst	Annotate legal clauses, define rules for compliance
	Michelle Seth	UX/Marketing	Create wireframes, define user personas
3: Model Development (Weeks 5–7)	Natascha Banda	Product Manager	Align model to business goals, plan API integration
	Michelle Ike	Tech Lead	Train/fine-tune summarization model, monitor training
	Unaswi Mudanga	Data Scientist	Run evaluation pipeline, assess output quality
	Morgan-Lee Blake	Legal Analyst	QA model summaries, provide legal feedback
	Michelle Seth	UX/Marketing	Build UI with model outputs, design summary cards
4: Product Integration (Weeks 8–10)	Natascha Banda	Product Manager	Design onboarding experience, draft pricing wireframes
	Michelle Ike	Tech Lead	Deploy model API, integrate clause output to UI
	Unaswi Mudanga	Data Scientist	Optimize pipeline, fix latency issues
	Morgan-Lee Blake	Legal Analyst	Validate legal accuracy, flag edge cases
	Michelle Seth	UX/Marketing	Finalize dashboard, improve interactivity
5: Testing & Launch Prep (Weeks 11–12)	Natascha Banda	Product Manager	Prepare pitch deck, manage early access rollout
	Michelle Ike + Unaswi Mudanga	Tech + Data	System testing, benchmark performance, document results

	Morgan-Lee Blake	Legal Analyst	Final legal audit, QA protocols, compliance checklist
	Michelle Seth	UX/Marketing	Host user walkthroughs, finalize marketing assets