

Project Proposal: Legal Assistant Agent

Hope McClew



What is the Problem?

Legal documents are difficult for non-lawyers to understand.

Statutes, court opinions, and regulations use technical, complex language.

Even professionals spend hours locating and interpreting relevant sources.

There is no simple way to ask a legal question and receive a clear, cited answer.

Goal: Build an AI agent that can retrieve, interpret, and summarize verified legal information.



Why Do I Care?

- 1) Law shapes everyday choices and business decisions.
- 2) What's legal in one place might be illegal somewhere else.
- 3) Making legal information easy to understand strengthens fairness and awareness.
- 4) When more people understand the law, they can make smarter, more informed decisions.
- 5) Broader public understanding can inspire new, informed solutions to legal and policy problems.



Technical Approaches & Methodology

Knowledge Base: Create a small corpus of open-license legal articles and summaries (e.g. simplified case law, Cornell Legal Information Institute pages)

Search Method: Implement TF-IDF retrieval (Milestone 1) to rank documents by relevance

Prompting Techniques: Use chain-of-thought prompting (Milestone 2) to let the model plan search and reasoning steps

Language Model: Use a Hugging Face pretrained model for text generation and reasoning (Milestone 3)

Agent Workflow: Combine iterative retrieval + reasoning + evidence summarization (Milestone 4)



Rationale

- 1) TF-IDF ensures interpretable, efficient search on small datasets.
- 2) Chain-of-thought prompts improve model transparency in reasoning.
- 3) Open-source LLMs (via Hugging Face) are explainable and ideal for reproducibility.



Timeline & Deliverables

Week 1 (Oct 27-31): Setup & Knowledge Base

Week 2 (Nov 3-7): Milestone 1

Week 3 (Nov 10-14): Milestone 2

Week 4 (Nov 17-21): Milestone 3

Week 5 (Nov 24-28): Milestone 4

Week 6 (Dec 1-5): Final Deliverables & Presentation

Deliverables: Create an AI agent that answers narrow legal questions and provides cited evidence.

Scope:

- 1) Statutory definitions
- 2) Filing deadlines
- 3) Elements of legal claims
- 4) Jurisdiction basics
- 5) Case outcomes

Out of Scope:

- 1) Offering legal advice
- 2) Handling ambiguous or novel questions



Final Data and Adjustments

Most data entries were concerned with general legal information in the U.S. This includes amendment rights and protections, immigration policy, tax code, contracts, and tenant rights.

I found I needed to increase the max tokens for my model, since legal documents often contain long and complex technical jargon, all of which constitute important tokens. Without an increase, the model often missed the key answer.

Testing

Question: What rights does the First Amendment protect?

Final Answer: The First Amendment protects freedom of speech, religion, press, assembly, and petition.

Trajectory:

Step 1

Thought: To find out what rights the First Amendment protects, I need to search for information about the First Amendment.

Action: search[query="First Amendment rights"]

Observation: {"tool": "search", "query": "First Amendment rights", "results": [{"id": "doc6", "title": "First Amendment Rights", "snippet": "The First Amendment protects freedom of speech, religion, press, assembly, and petition. These rights limit the government's ability to restrict expression or punish individuals for their beliefs."}, {"id": "doc7", "title": "Fourth Amendment Search and Seizure", "snippet": "The Fourth Amendment protects individuals from unreasonable searches and seizures. Police general..."}

Step 2

Thought: The First Amendment protects freedom of speech, religion, press, assembly, and petition.

Action: search[query="First Amendment rights"]

Observation: {"tool": "search", "query": "First Amendment rights", "results": [{"id": "doc6", "title": "First Amendment Rights", "snippet": "The First Amendment protects freedom of speech, religion, press, assembly, and petition. These rights limit the government's ability to restrict expression or punish individuals for their beliefs."}, {"id": "doc7", "title": "Fourth Amendment Search and Seizure", "snippet": "The Fourth Amendment protects individuals from unreasonable searches and seizures. Police general..."}

Step 3

Thought: The First Amendment protects freedom of speech, religion, press, assembly, and petition.

Action: finish[answer="The First Amendment protects freedom of speech, religion, press, assembly, and petition."]

Observation: done

You can see the agent search through multiple documents related to various amendments, and eventually successfully list all freedoms protected by the first amendment.



Testing

Question: When are police required to give Miranda warnings?
Final Answer: Police are required to give Miranda warnings when arresting a suspect.
Trajectory:
Step 1
Thought: I should search for key facts related to the question.
Action: search[query="(auto) refine the user question", k=3]
Observation: {"tool": "search", "query": "(auto) refine the user question", "results": [{"id": "doc4", "title": "Pythagorean theorem", "snippet": "In mathematics, the Pythagorean theorem is a fundamental relation in Euclidean geometry among the three sides of a right-angled triangle: the square of the hypotenuse equals the sum of the squares of the other two sides."}, {"id": "doc2", "title": "The Starry Night", "snippet": "The Starry Night is an oil-on-canvas painting by Dutch Post-Impressionist painter Vinc..."}]
Step 2
Thought: To find when police are required to give Miranda warnings, I'll need to search for relevant information.
Action: search[query="(auto) refine the user question", k=3]
Observation: {"tool": "search", "query": "(auto) refine the user question", "results": [{"id": "doc4", "title": "Pythagorean theorem", "snippet": "In mathematics, the Pythagorean theorem is a fundamental relation in Euclidean geometry among the three sides of a right-angled triangle: the square of the hypotenuse equals the sum of the squares of the other two sides."}, {"id": "doc2", "title": "The Starry Night", "snippet": "The Starry Night is an oil-on-canvas painting by Dutch Post-Impressionist painter Vinc..."}]
Step 3
Thought: I should search for key facts related to the question.
Action: search[query="(auto) refine the user question", k=3]
Observation: {"tool": "search", "query": "(auto) refine the user question", "results": [{"id": "doc4", "title": "Pythagorean theorem", "snippet": "In mathematics, the Pythagorean theorem is a fundamental relation in Euclidean geometry among the three sides of a right-angled triangle: the square of the hypotenuse equals the sum of the squares of the other two sides."}, {"id": "doc2", "title": "The Starry Night", "snippet": "The Starry Night is an oil-on-canvas painting by Dutch Post-Impressionist painter Vinc..."}]
Step 4
Thought: I should search for key facts related to the question.
Action: search[query="(auto) refine the user question", k=3]
Observation: {"tool": "search", "query": "(auto) refine the user question", "results": [{"id": "doc4", "title": "Pythagorean theorem", "snippet": "In mathematics, the Pythagorean theorem is a fundamental relation in Euclidean geometry among the three sides of a right-angled triangle: the square of the hypotenuse equals the sum of the squares of the other two sides."}, {"id": "doc2", "title": "The Starry Night", "snippet": "The Starry Night is an oil-on-canvas painting by Dutch Post-Impressionist painter Vinc..."}]
Step 5
Thought: Police are required to give Miranda warnings when arresting a suspect.
Action: finish[answer="Police are required to give Miranda warnings when arresting a suspect."]
Observation: done

A vaguer question prompted a vaguer and less useful result. You can see the model repeat steps.



Results

The model does well with straightforward questions that can be answered with a simple list. The model is less useful with vaguer questions about the application of different laws and policies.

Question: What are the basic elements required to form a valid contract?

Final Answer: The essential elements required to form a valid contract include: a clear and unambiguous offer, consideration, a legal obligation, mutual assent, and a material breach clause.