LexLead Legal Advisory System

Sameera Boppana, Jialong (Mark) Li, Yilin Shan, Jiayue Tian

Tools in the box

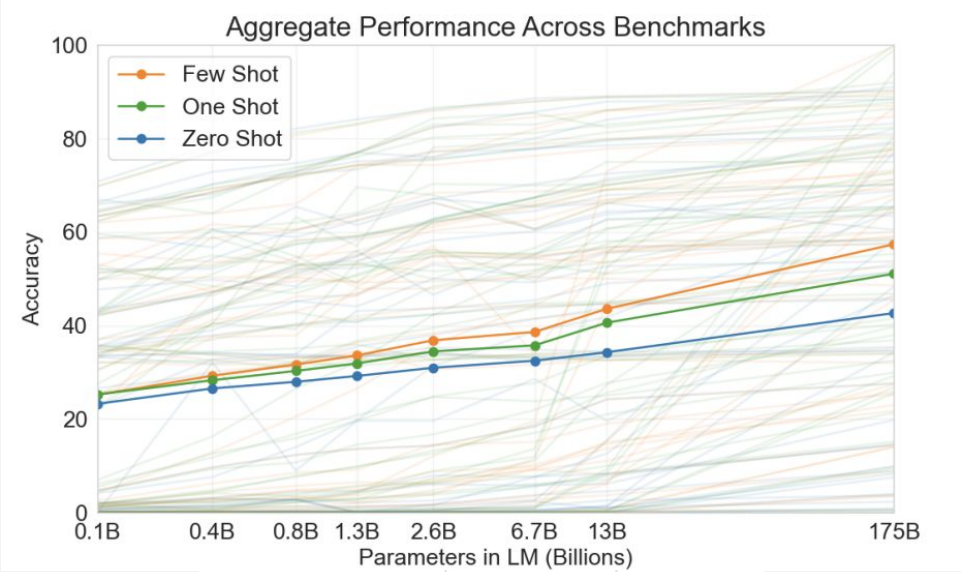
— Key technologies that we used.

> Ideas are cheap, and there are lots of them in the air. Your skill comes in when you decide which one to work on, and how well you execute on it.
>
> — John Schulman (former cofound of OpenAI, Anthropic Research)

- Large Models are **Few-Shot** Learners (NeurIPS 2020, OpenAI)
- **RAPTOR**: Recursive Abstractive Processing for Tree-Organized Retrieval (ICLR 2024, Stanford)
- **ReAct**: Synergizing Reasoning and Acting in Language Models (ICLR 2023, Princeton + Google)
- **Adaptive-RAG**: Learning to Adapt Retrieval-Augmented Large Language Models through Question Complexity (ACL 2024, KAIST)
- **LegalBench:** A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models (NeurIPS 2023, 20 institutions)

LLM are Few shot Learner

— go to the office hour! (to get more hints)



	0-shot	few-shot
privacy_policy_qa	12/50	24/50
contract_qa	44/50	48/50
hearsay	33/50	45/50
citation prediction	1/50	48/50

Few-shot is a **prompt engineering** technique, it works when you have a task-specific question to ask.

In plain words, you give a **demo** to LLM, like:

Question {Claim construction requires determining how a skilled artisan would understand a claim term in the context of the entire patent, including the specification.}
Correct Answer {Phillips v. AWH Corp}

Then, you answer more questions on this topic (citation prediction).

By feeding examples to LLM, just like a human, LLM can better understand the question and providing better answer.

ReAct (Reason-Action)

— Split hard problems into pieces and try different approaches.

The question: “What are the differences in firearm control laws across the various states in the U.S.”

Toolbox: {**RAG Search, Google Search, Self-reflection**}

Reason_1: {I need to show the opinion from both side, and quote the right sources properly}

Action_1: {Search pro-gun control laws in our database, and retrieve relevant law successfully!

However, **LLM fail to find laws that against gun control**, because we only stores the ILLINOIS law in our database}

Reason_2: {Since we can’t get against gun control laws in our RAG database, let’s do **Google Search**!}

Action_2: {Search against-gun control laws on Google.}

Final Answer {based on all reason-action chains.}

Smart LLM can do this implicitly (OpenAI o1), but it requires millions of training data via Reinforcement Learning with Human Feedback (RLHF), the way DeepMind researchers train AlphaGO.

ReAct achieves this explicitly by introducing multi-agent, and each agent is a small, fine-tuned LLM.

Adaptive RAG

— RAG is not always useful

- Perplexity always use RAG because their user will ask anything.
- But sometimes, RAG is not useful, or even harm the LLM performance.
  - It is time-consuming, and sometimes it can only include noise.

Why, say our question is, “what is the time the contract being signed”.

Then, RAG is harmful.

So, techniques like Adaptive RAG is like a Dam’s water valve, to control the water/information flow.

Dataset Overview: Illinois Statutes

Description:
The dataset used to train and test our Legal Retrieval Augmented Generation (RAG) system includes all statutes in Illinois law.

- Statutes are codified laws passed by legislative bodies, providing a structured and factual foundation for legal applications.

Why Statutes?

- **Factual Nature**:Compared to case law or regulations, statutes offer direct, unambiguous rules and principles.
- **Comprehensive Coverage**: Addresses various domains of public interest and governance.

Key Content Areas in the Dataset:

- **Government**: State institutions and public administration.
- **Education**: Policies for educational systems.
- **Regulation**: Rules for societal compliance.
- **Human Needs**: Essential services and support.
- **Health and Safety**: Public health and safety standards.
- **Agriculture and Conservation**: Farming and resource management.
- **Transportation**: Vehicles and transit laws.
- **Rights and Remedies**: Civil liberties and legal recourse.
- **Business and Employment**: Corporate and labor regulations.

Reference:https://www.ilga.gov/legislation/ilcs/ilcs.asp

Benchmark Questions

To evaluate our Illinois Statutes RAG system, we created custom benchmark questions due to the absence of pre-existing, Illinois-specific questions.The questions primarily focus areas include education, health, and regulation, addressing policies, public safety, and compliance rules.

Question Design:

- **Total Questions**: 45
  - **15 True/False Statements**
    - Assess whether a given statement aligns with the statutes.
    - Example Question: Can a licensed currency exchange operate at multiple locations under the same license?
    - Example Answer: No.
  - **30 Short-Answer Questions**
    - Fact-based questions requiring specific answers about acts or provisions.
    - Example Question: What is required for a 16-year-old to donate blood?
    - Example Answer: Written permission or authorization from their parent or guardian.

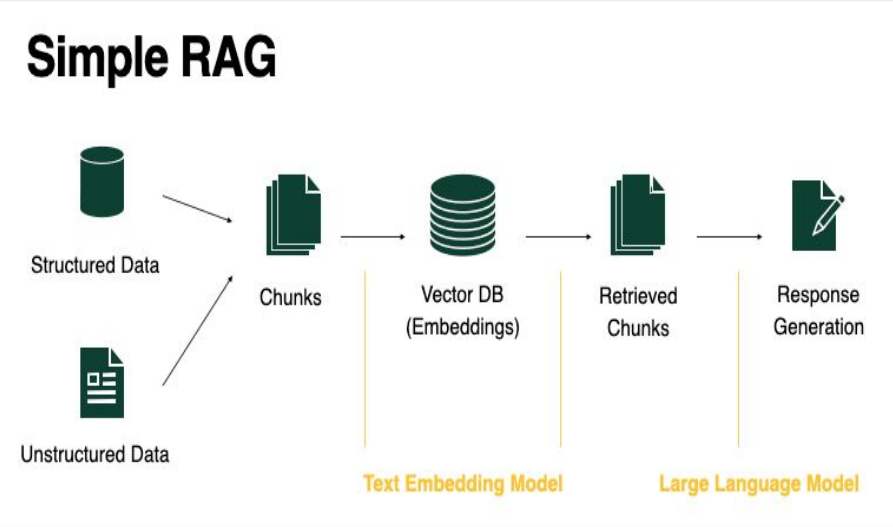
Simple RAG

To evaluate and improve performance, we implemented two different models for comparison:

Simple RAG Workflow:

- A straightforward implementation of the common RAG framework.
- **Key Steps**:
  - Load data into documents.
  - Use the **Recursive Text Splitter** to split data into smaller chunks.
  - Generate embeddings for the chunked text using **Vertex AI Embedding**.
  - Store the embedded vectors in the **GCP BigQuery Vector Store**.
  - Use a retriever to find relevant answers based on the user's question.
  - Use **Gemini Pro LLM** to refine and generate the final response.

Advanced RAPTOR Model:A more sophisticated approach for enhanced retrieval and generation performance.

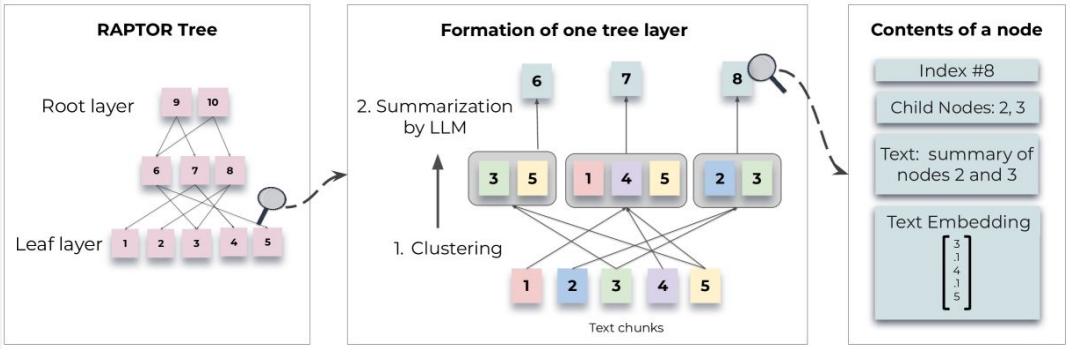


RAPTOR

RAPTOR is good at answering questions that require cross-sectional information or contextual understanding.

For example: (Apple Inc. v. Samsung Electronics Co. (2011))

“Through this case, what arguments and legal principles can be leveraged to protect a new technological invention or to challenge an existing patent?”



RAPTOR (example)

Q: Through this case, what arguments and legal principles can be leveraged to protect a new technological invention or to challenge an existing patent?

A: Several legal principles and arguments can be leveraged to protect a new technological invention or challenge an existing patent, as outlined through the Apple v. Samsung case:

1. Claim Construction: In patent law, claim construction is a critical step where the court determines the meaning of the key terms. In a dispute, each party may have differing interpretations of a term. For example, Samsung and Apple disagreed on the term "display," with Apple's broader interpretation of the term ultimately winning out.
2. Validity: An existing patent's validity can be challenged on grounds of functionality (if the features claimed are dictated by the product's use) or obviousness (if the invention seems an evident step based on previous innovations).
3. Infringement: A patent holder can argue about an infringement and the potential harm it may cause to its brand's goodwill or market share if a competing company's product appears substantially similar to the patented design, thereby creating consumer confusion.
4. Equity Rules: Non-automatic application of categorical rules and presumptions when seeking equitable relief like injunctions. Each situation calls for unique considerations.
5. Patent as a Property Right: This principle emphasizes a patent owner's exclusive right to make, use, and sell their invention. It reinforces the notion of protecting new technological inventions.
6. Licensing History: Previous instances where a patentee has licensed the patent can indicate if a reasonable royalty compensates for the infringement or not.
7. Necessity Core technology: Whether the patented invention is necessary to the product's functionality or not can play a pivotal role in establishing infringement claims.

These principles help secure rights over new technological inventions and provide a legal framework to challenge an existing patent.

Benchmark Analysis

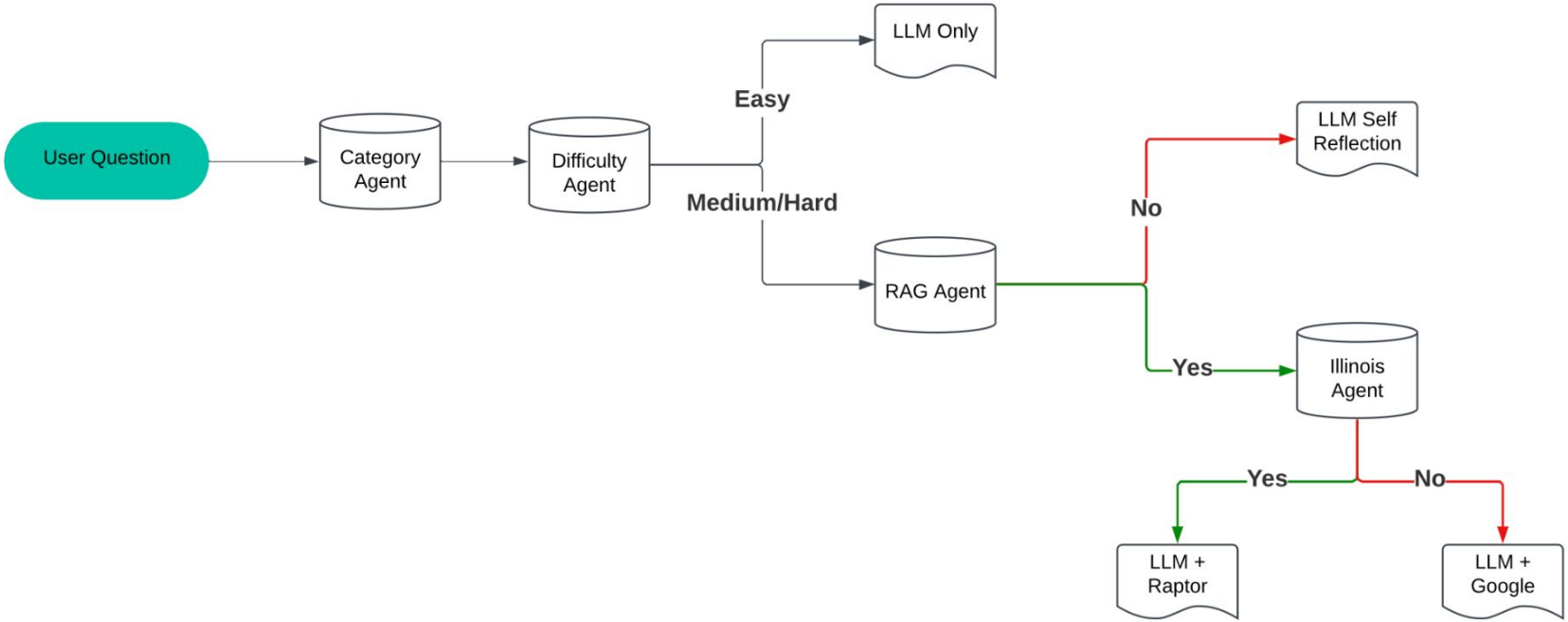
Binary Questions

	Average Time (per question)	Question Evaluated	Correctly Answered	Accuracy
Simple RAG	27.9 seconds	15	12	80.0%
RAPTOR	11.5 seconds	15	11	73.3%
LLM (base model)	18.0 seconds	15	11	73.3%

Short Answer Questions

	Average Time (per question)	Question Evaluated	Correctly Answered	Accuracy
Simple RAG	37.9 seconds	30	11	36.7%
RAPTOR	14.0 seconds	30	25	83.3%
LLM (base model)	14.4 seconds	30	3	10%

LLM Agent Workflow



Legal Advisory System Demo

https://lexlead-streamlit-demo-175227783063.us-central1.run.app/