



LawLLM: Law Large Language Model for the US Legal System

Dong Shu
Northwestern University
Evanston, IL, United States
dongshu2024@u.northwestern.edu

Haoran Zhao
Northwestern University
Evanston, IL, United States
haoranzhao2024@u.northwestern.edu

Xukun Liu
Northwestern University
Evanston, IL, United States
xukunliu2025@u.northwestern.edu

David Demeter
Northwestern University
Evanston, IL, United States
ddemeter@u.northwestern.edu

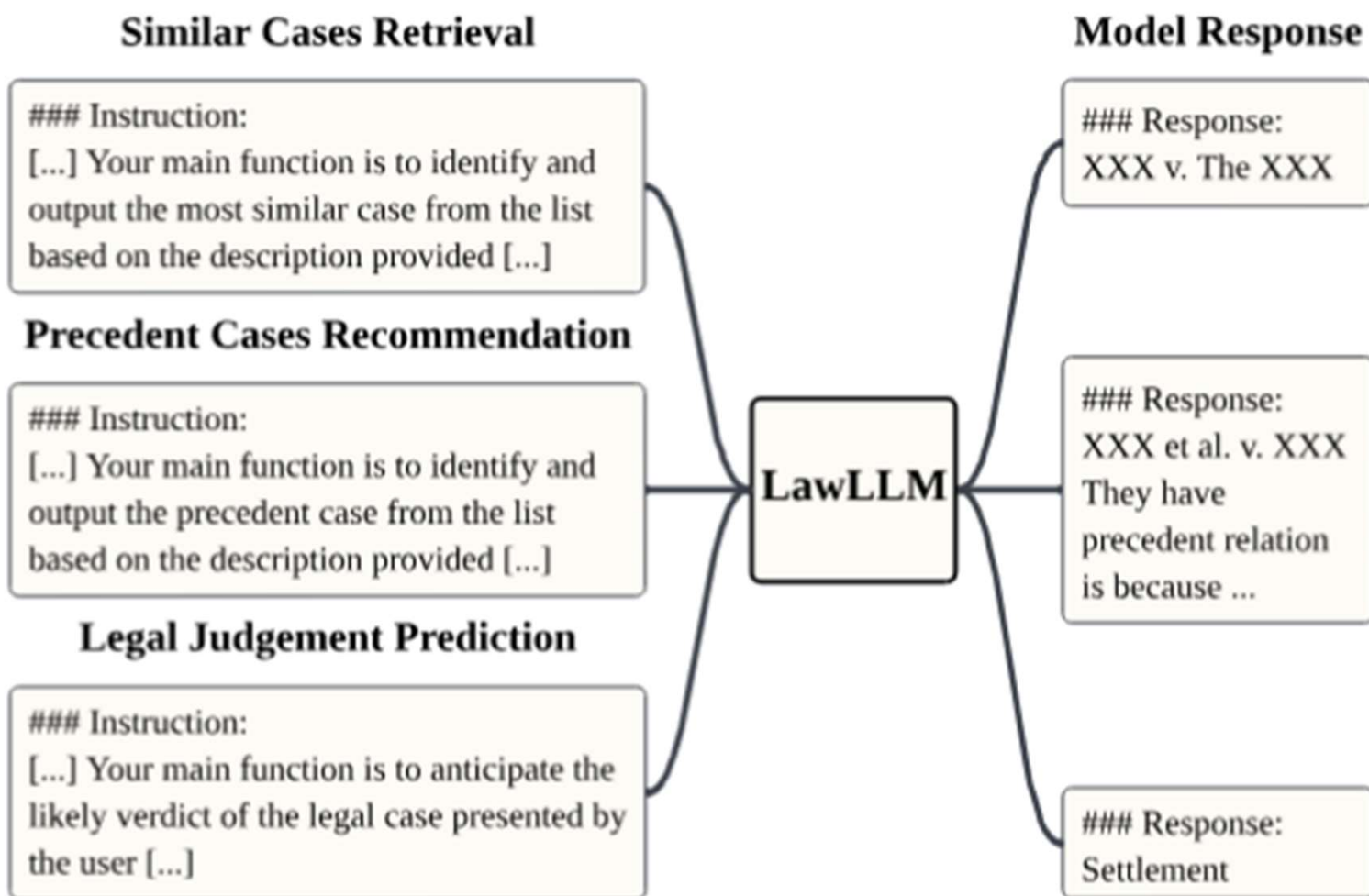
Mengnan Du
New Jersey Institute of Technology
Newark, NJ, United States
mengnan.du@njit.edu

Yongfeng Zhang
Rutgers University
New Brunswick, NJ, United States
yongfeng.zhang@rutgers.edu

The 33rd ACM International Conference on Information and Knowledge Management
Boise, Idaho, USA | 21-25 October, 2024

Background

- 1) Different countries have different law system, it is impossible to develop a model that suitable for all law system
 - 1) What is possible is that we can develop a model that can perform multi legal tasks under the same law system.
- 2) US operates under a common law system
 - 1) A common law system is based on judicial precedent (i.e. precedent case)
- 3) Lots of researchers believe that similar case are precedent case, but is not.

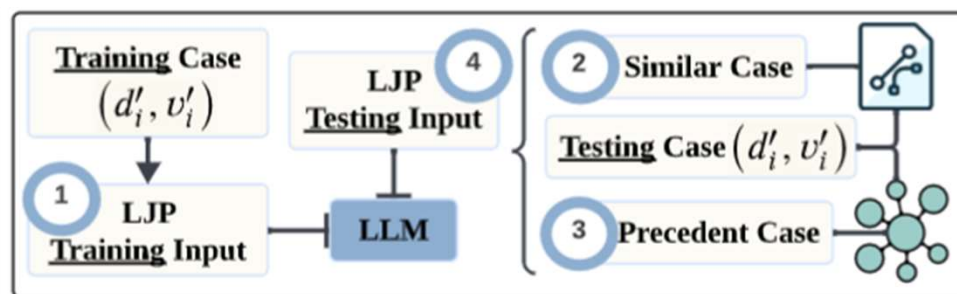
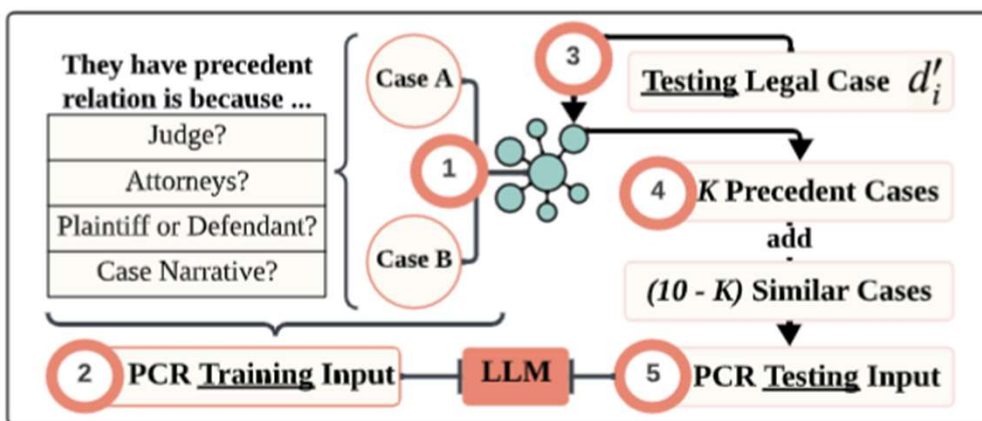
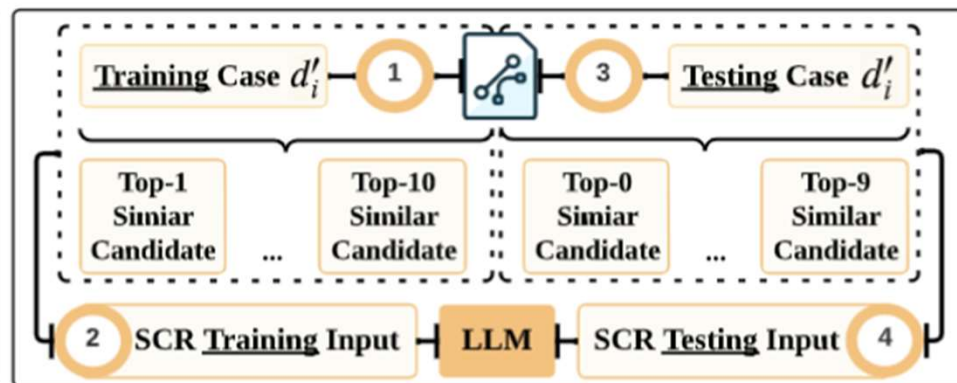
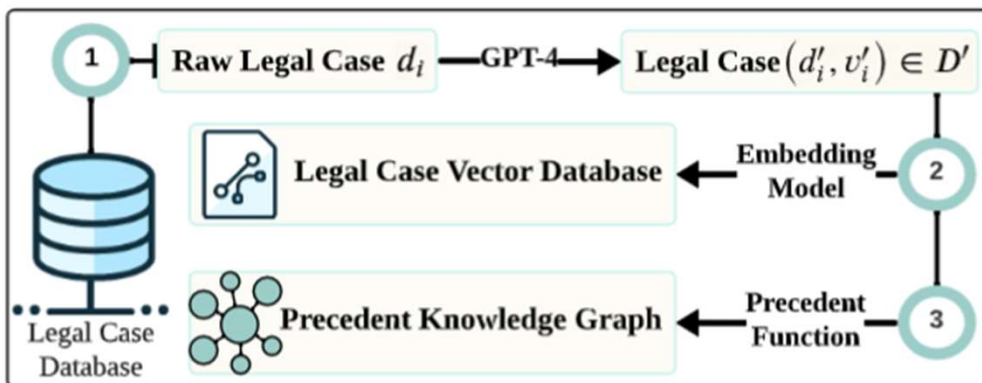


Contribution

- 1) We propose LawLLM, which is adept at handling a range of legal tasks, including Legal Judgement Prediction (LJP), Precedent Case Recommendation (PCR), and Similar Case Retrieval (SCR).
- 2) LawLLM distinguishes between precedent cases and similar cases, providing clarity on the objectives of each task. This clarification enables the future research to develop tailored strategies for those tasks.
- 3) Experimental results indicate that LawLLM outperformed all baseline models, including the GPT-4 model, across all three tasks.

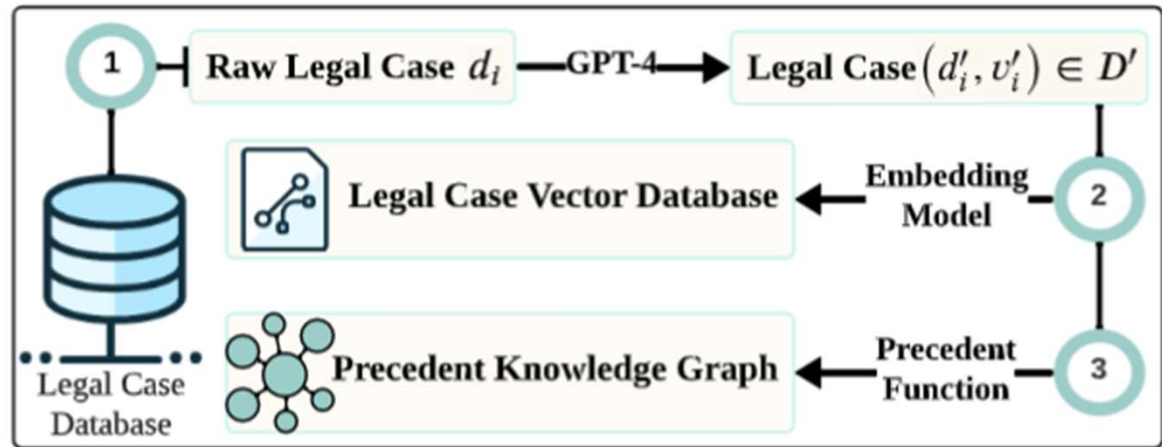
Difference between precedent case & similar case

- 1) A precedent case must have been closed before the input legal case.
- 2) Precedent cases are those that were actually considered by judges in making their decisions, unlike similar cases which might not have been taken into account.
- 3) Similar cases share textual and thematic similarities in the case narrative or might fall into similar case categories, while precedent cases might seem unrelated at face value.
- 4) There might be jurisdiction restrictions in precedent cases.
- 5) While a legal case's precedent case can sometimes be the same as a similar case, this is not always the case.



- Data Preprocessing
- Similar Case Retrieval
- Precedent Case Recommendation
- Legal Judgment Prediction

Data Preprocess



A legal case has {Title, Date, Judge, Plaintiff(s), Plaintiff's Attorney(s), Defendant(s), Defendant's Attorney(s), Case Detail, Precedent Relationship}.

1. From raw legal case data to summarized case detail & verdict
2. Convert processed data into vector database
3. Convert processed data into precedent knowledge graph

Prompt for Similar Case Retrieval , Precedent Case Recommendation

SCR

Instruction:

You are a legal expert who specializes in comparing user-supplied legal cases to a list of candidate legal cases, which includes titles and content. Your main function is to identify and output the title of the most similar case from the list based on the description provided.

You should only output the case title and not any other information.

Consider the following choices:

Choice 1:

[Case 1...]

Choice 2:

...

Choice 10:

[Case 10...]

Input:

[Input Case...]

PCR

Instruction:

You are a legal expert who specializes in comparing user-supplied legal cases to a list of candidate legal cases, which includes titles and content. Your main function is to identify and output the precedent case from the list based on the description provided.

You should only output the reasoning process and case title.

Consider the following choices:

Choice 1:

[Case 1...]

Choice 2:

...

Choice 10:

[Case 10...]

Input:

[Input Case...]

Metric

- PCR and SCR:

1. Top-1

2. Top-3

3. Top-5

4. Not found

Prompt for Legal Judgment Prediction

Instruction:

You are a legal expert who specializes in predicting outcomes for legal cases. Utilize your internal knowledge base to predict verdict. Your main function is to anticipate the likely verdict of the legal case presented by the user.

You should only output the verdict and not any other information.

Consider the following choices:

- 1. Defendant Wins*
- 2. Plaintiff Wins*
- 3. Settlement*
- 4. Case Dismissal*

Input:

[Input Case...]

Metric for LJP

1. Accuracy

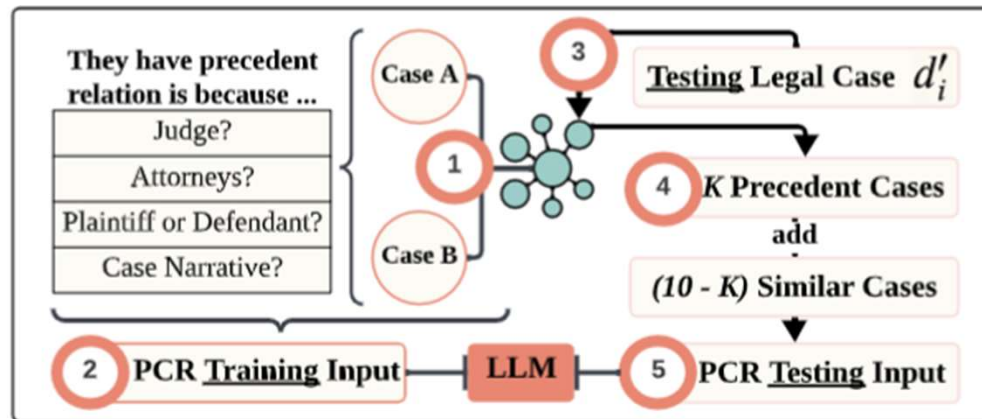
2. F1

Similar Case Retrieval



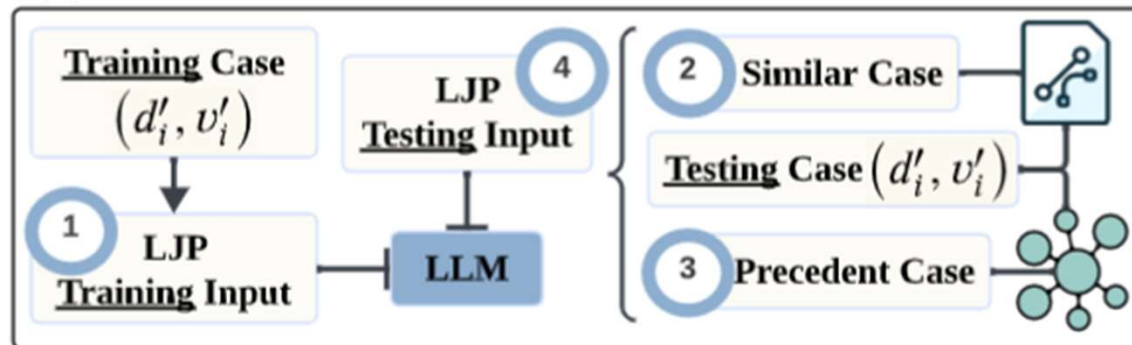
1. For each case, we retrieve top 10 similar cases from vector database.
2. Then use these cases to construct the training prompt
 1. The expected output will be the top 1 similar case
3. Testing stage is using the same process, but with top 0-9 similar cases

Precedent Case Recommendation



1. In top-1 setting, we extract case A and case B. Case A will be the input case, and Case B will be the precedent case of Case A. Rest of 9 cases are filled with similar case retrieved from vector database
 1. If there is a precedent relationship between them, we use BERT embedding to check the similarity between various case features (e.g. Judge, Case detail, etc.)
2. Note that, we also have top-3 and top-5 setting

Legal Judgment Prediction



1. We employ two-shot in-context learning during the testing phase.
2. For each testing case, we will retrieve a similar case from vector database and a precedent case from the knowledge graph.

Dataset

- CaseLaw Project

Table 1: Datasets Statistics

DATASETS	CaseLaw
Language	English
# State and Federal Totals	6,930,777
# Train case	1,00,000
# Test case	20,000
Avg. length per case (words)	2695.38

Table 2: SCR Test Results

Method	top-1 \uparrow	top-3 \uparrow	top-5 \uparrow	Not Found \downarrow
llama2-7b	0.083	0.197	0.309	0.406
gemma-7b	0.181	0.428	0.536	0.121
vicuna-13b	0.185	0.372	0.564	0.187
guanaco-13b	0.077	0.214	0.375	0.372
gpt3.5	0.219	0.579	0.691	0.148
gpt4	0.274	0.526	0.708	0.005
LawLLM	0.298	0.632	0.816	0.001

Table 3: PCR Test Results

Method	top-1 \uparrow	top-3 \uparrow	top-5 \uparrow	Not Found \downarrow
llama2-7b	0.069	0.148	0.343	0.479
gemma-7b	0.187	0.386	0.519	0.124
vicuna-13b	0.175	0.352	0.506	0.203
guanaco-13b	0.073	0.198	0.357	0.383
gpt3.5	0.154	0.325	0.504	0.165
gpt4	0.262	0.514	0.697	0.007
LawLLM	0.318	0.597	0.832	0.001

Table 5: LJP Test Results

Method	Accuracy \uparrow (Zero-shot)	F1 \uparrow (Zero-shot)	Accuracy \uparrow (Few-shot)	F1 \uparrow (Few-shot)
llama2-7b	0.235	0.239	0.473	0.455
gemma-7b	0.317	0.287	0.568	0.527
vicuna-13b	0.503	0.432	0.645	0.594
guanaco-13b	0.281	0.247	0.491	0.463
gpt3.5	0.558	0.546	0.679	0.647
gpt4	0.573	0.563	0.732	0.712
LawLLM	0.636	0.591	0.794	0.758



LawLLM: Law Large Language Model for the US Legal System

Dong Shu
Northwestern University
Evanston, IL, United States
dongshu2024@u.northwestern.edu

Haoran Zhao
Northwestern University
Evanston, IL, United States
haoranzhao2024@u.northwestern.edu

Xukun Liu
Northwestern University
Evanston, IL, United States
xukunliu2025@u.northwestern.edu

David Demeter
Northwestern University
Evanston, IL, United States
ddemeter@u.northwestern.edu

Mengnan Du
New Jersey Institute of Technology
Newark, NJ, United States
mengnan.du@njit.edu

Yongfeng Zhang
Rutgers University
New Brunswick, NJ, United States
yongfeng.zhang@rutgers.edu

Thank you

Q&A

dongshu2024@u.northwestern.edu