# Structured Legal Argumentation with LLMs: A Study in Landlord-Tenant Law

<sup>1</sup>Department of Mathematics, Statistics, and Computer Science, University of Illinois Chicago, USA

<sup>2</sup>Pritzker School of Law & McCormick School of Engineering, Northwestern University, USA

# Interpretability in the legal context

- **Trust and Transparency**: For legal practitioners and other users to be able to understand the model's reasoning process, verify and trust it.
- Accountability: Judicial decisions must be clear and justifiable.
- **Expanding access to justice**: To make it easier for laypeople to engage with and assess AI-generated legal content.

# Pitfalls of Large Language Models (LLMs)

- False or misleading legal information: an LLM may invent laws, precedents, events ...
- Lack of interpretability: What is the LLM answer based on?
- **Hard to verify**: What makes the output correct?

### Research Question

Can LLMs with Context Augmentation and Chain of **Thought** prompting generate *accurate*, *factual*, *relevant*, and comprehensive legal reasoning for Landlord-Tenant problems?

**Input Exposition**: *Provide a detailed description of the legal scenario*, including relevant facts, context, and specific issues at stake.

**Task**: Generate structured legal arguments based on the exposition provided.

The Tenant and Landlord Laws considered are the Residential Landlord and Tenant Ordinance of the City of Chicago provided below:

[Text of the RLTO.]

**Output Format**: Argument 1 **Exposition**: Summarize relevant facts from the input that relate to this argument. **Specific Law**: Identify a specific law or statute that applies to the scenario.

Why This Law Applies: Explain how and why this law is relevant to the facts presented.

**Conclusion**: State the conclusion derived from the application of this law to the facts provided.

[Continue generating arguments as necessary, each focusing on a different applicable law.]

Figure 1: The prompt we used in GPT-40 to generate legal reasoning.

Gregoire Fournier<sup>1</sup> Daniel W. Linna Jr.<sup>2</sup>

# Landlord-Tenant Scenarios

We collaborated with the Law Center for Better Housing (LCBH), a non-profit law firm that aims to protect renters' rights in the Chicago area.

**Rentervention**, one of its programs, provides free resources to help tenants enforce their rights, including a chatbot that helps renters find the correct information for their issue.

Table 1:Scenarios and their Sources.

#	Description	Source
1-4	Heating and mold, rent increase, privacy, property damage	Claude
5–9	Cockroaches, eviction, landlord actions, leaking roof	Rentervention
10	Wear and tear	Authors

# Metrics for Legal Reasoning

The output was evaluated by a lawyer at the **LCBH**, who is an expert in landlord-tenant law, using the metrics:

- Accuracy for a given legal scenario measures how closely the set of generated arguments aligns with the true or expected answer.
- **Comprehensiveness** measures how well one argument coherently and concisely addresses the relevant aspects of the input legal scenario regarding the legal requirement cited.
- **Factuality** assesses whether an argument originates from the text of the RLTO.
- **Relevance** evaluates whether the argument logically relates to the legal scenario.

### Example

An example of a **relevant** and **factual** argument which is **not** accurate nor comprehensive.

- Exposition: The landlord provided a revised offer to reduce the rent increase to 7% if the tenant signed a two-year lease.
- Specific Law: Chicago Residential Landlord and Tenant Ordinance (RLTO) Section 5-12-140(a).
- Why This Law Applies: Section 5-12-140(a) states that no rental agreement may provide that the tenant agrees to waive or forgo rights, remedies, or obligations provided under this chapter.
- **Conclusion:** The tenant could argue that the landlord's offer to reduce the rent increase contingent on signing a two-year lease may be seen as an attempt to make the tenant waive their right to proper notice of rent increase, thus violating Section 5-12-140(a).

Figure 2: An example of an argument generated in Scenario # 2 about a rent increase.

# Analysis of the generated Legal Reasoning

- of an RLTO section.
- correct conclusion.

Figure 3: Comprehensiveness boxplot of the arguments across the ten scenarios.

### **Identified Limitations:**

- context.

- outputs.



• Factuality: 54 out of 55 generated arguments were factual. This single mistake resulted from the model's poor reformulation

• Accuracy: In eight of the ten scenarios, the generated reasoning was **accurate**. The LLM was able to identify the legal

requirement to apply, justify why it was applicable, and reach the

• **Relevance**: We identified two scenario types: those (e.g., #2, 6, 7) where most arguments lacked **relevance** to the correct legal reasoning, and others where they were largely relevant.

• **Comprehensiveness**: In all scenarios except #7 and #9, the model included the **correct legal reasoning** in at least one argument. It struggled to generate only high-quality arguments.



• Inability to identify legal issues **beyond the provided** 

• **Gaps** between the logical steps of the Chain of Thought output. • Challenges in assessing the **relevance** of generated arguments.

### **Future Work**

• Enhance the legal-reasoning process by querying users for additional information, ensuring more **context-aware** 

• Develop capabilities to generate **reasoning** from **multimodal inputs**, including text, images, and audio, for richer analysis.

