

# Structuring Legal Text as Preparation for Chat-Bot Use

Lothar Hotz<sup>1,\*</sup>, Karen Wilke<sup>1,†</sup> and Mario Krebs<sup>2,†</sup>

<sup>1</sup>Hamburger Informatik Technologie-Center e.V., Vogt-Kölln-Straße 30, 22527 Hamburg

<sup>2</sup>Universität Bielefeld, Fakultät für Rechtswissenschaft, Forschungsstelle für Immobilienrecht (FIR), Universitätsstraße 25, 33615 Bielefeld

## Abstract

The increasing reliance on generative artificial intelligence (GenAI) in public administration is a pressing need due to the sector's prevalent textual and image-based activities. These activities include document generation, summarization, speech text creation, and explanation of large collections of documents, such as legal documents regarding laws, their interpretations, and judgments. One approach to harnessing domain-specific documents is Retrieval Augmented Generation (RAG), which enhances an Large Language Model's query with preselected text from a prepared dataset. However, the effectiveness of RAG depends heavily on the application and the task at hand. This paper presents a detailed approach to structuring legal text to answer building modernization-related questions. The proposed method will be integrated into an Intelligent Modernization, which aims to provide a tool for tenants and landlords to initiate modernization negotiations based on building functions such as stability, moisture protection, thermal insulation, sound insulation, fire safety, daylight access, aesthetics, and cost-effectiveness.

## Keywords

Generative Artificial Intelligence, Large Language Models, Retrieval Augmented Generation, Legal Text Structuring, Building Modernization,

## 1. Introduction

Making Generative Artificial Intelligence (GenAI), such as Large Language Models (LLMs) or Diffusion Models, accessible to public administration is an important task, because of the often textual and image-based activities in this sector. Such activities include the generation of documents, summaries, speech text, or explanations of large collections of documents, such as legal documents about laws, their interpretations, and judgments. One approach for the targeted use of domain-specific documents, e.g., documents of a collection, is Retrieval Augmented Generation (RAG). Simply speaking, for a prompt given to an LLM as a query, RAG augments it with selected text taken from a previously prepared and provided data set. However, this preparation greatly depends on the application and task to be performed. Especially for legal text, a query could simply be "Which law deals with building modernization", which would result in less structuring, a query such as "In which cases does a tenant have to pay modernization costs?" would result in the need of higher structuring legal text.

In this paper, we take a detailed structuring approach for enabling answers about legal text related to building modernization questions which come up when, e.g., an external thermal insulation system shall be installed on a front side of a building.

This research is part of the Intelligent Modernization Platform (IntelMOD) project, which aims to offer a tool to tenants and landlords to initiate modernization negotiations based on building functions: stability, moisture protection, thermal insulation, sound insulation, fire protection, daylight access, aesthetics and cost effectiveness [1].

---

*Artificial Intelligence for Digital Public Services, June 11, 2025, Koblenz, Germany*

\*Corresponding author.

†These authors contributed equally.

✉ lothar.hotz@hitec-hamburg.de (L. Hotz); karen.wilke@hitec-hamburg.de (K. Wilke); mario.krebs@uni-bielefeld.de (M. Krebs)

🌐 <https://kogs-www.informatik.uni-hamburg.de/~hotz/> (L. Hotz); <https://hitec-hamburg.de> (K. Wilke); <https://ekvv.uni-bielefeld.de> (M. Krebs)

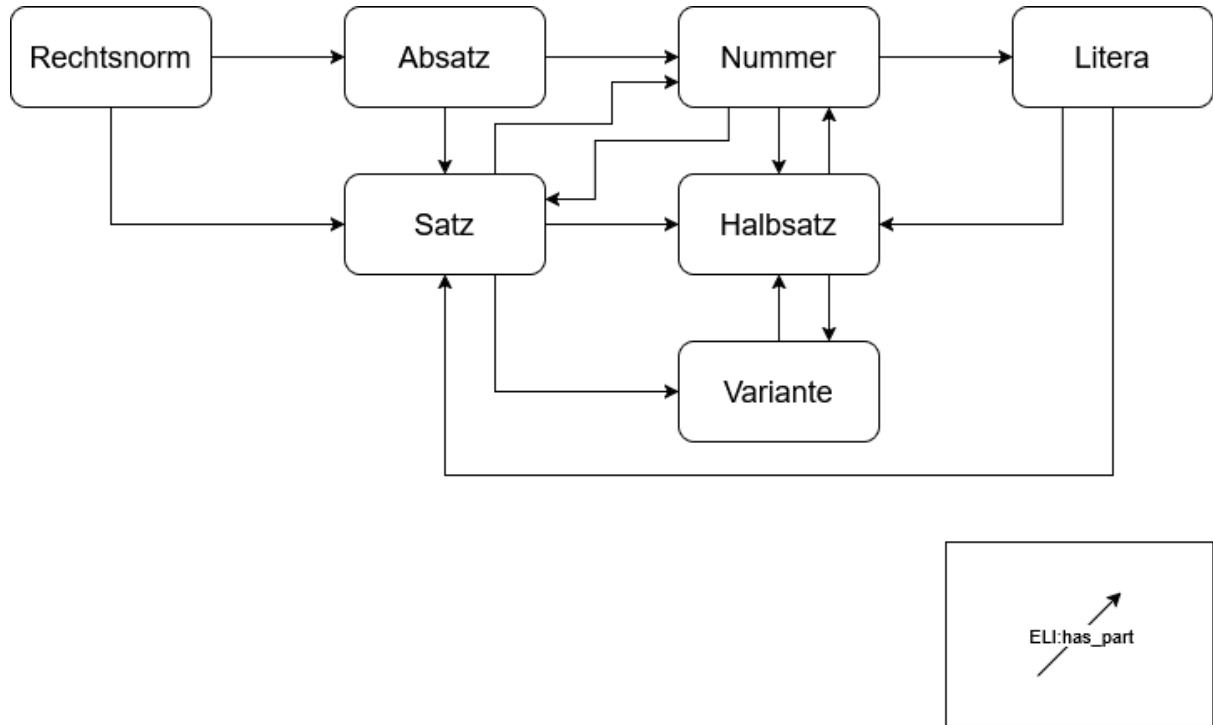
🆔 0000-0001-7370-7726 (L. Hotz)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## 2. Approach

As a first step, we identify the laws that are related to the building modernization tasks, an example is the *German Civil Code (BGB, Bürgerliches Gesetzbuch) Section 555a Maintenance measures*. With deep legal knowledge and on the basis of existing legal ontologies such as the European Legislation Identifier (ELI) [2], we intensively structure the identified legal norms. Fig. 1 presents a small part of the structure of a legal norm.



**Figure 1:** Structure of a legal norm

We are currently developing a new ontology based on this structured legal text with the Web Ontology Language (OWL)<sup>1</sup>. The final goal is to use this structure as a data set of a RAG architecture which augments a query with then essential parts of this structure, which enables a user to write a prompt that leads to our ontology being searched and an LLM providing an answer using the knowledge displayed in the ontology. Fig. 2 shows our enhancement of ELI. We identified seven different components of structuring statute books and seven different components of legal norms.

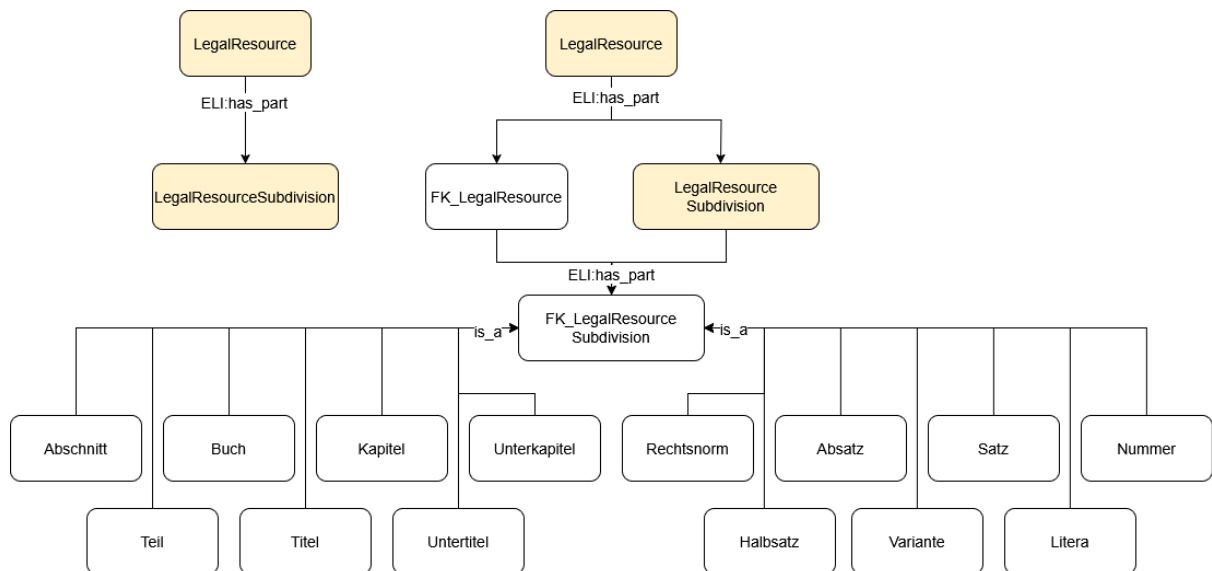
## 3. Experiments

### 3.1. Experimental Setup

In order to examine the state of the art, to test our approach and to subsequently plan further research, we conducted two experiments with ChatGPT-4o after defining quality criteria for the subsequent responses.

The first experiment was conducted to examine the current state of the art by prompting two specific legal questions about building modernization. Two questions about possible financial consequences were asked. One of them was a general question, the follow-up question was an example containing specific information about the measure, its cost, the apartment, and the rent increase. The quality criteria for the general question for this experiment are the general correctness of the conclusion on whether or not the landlord is allowed to increase the rent and the limitations that influence the rent.

<sup>1</sup><https://www.w3.org/OWL/>



**Figure 2:** Specialization structure of parts of a legal text as extension of ELI

The quality criteria for the follow-up question are the correct calculation of the rent, the underlying juridical restrictions of the specific case and the general correctness of the conclusion on whether or not the landlord is allowed to increase the rent to the planned extent. The second experiment was conducted to determine the impact of our ontology on the model's solution. We provided ChatGPT with our current ontology and asked to use it to answer a specific legal question with the very same information that we provided in the second question of the first experiment. The quality criteria for the follow-up question of the first experiment were applied, as well.

### 3.2. First Experiment

As a first question, we created the following prompt:

*"We are in Germany. My landlord insulated the outer wall and increased the rent. Are they allowed to do that?"*

In this case, we expected to receive general information about the modernization law which ChatGPT provided, even though some of the information was missing or misleading.

Table 1 shows the criteria for the quality of possible responses and if ChatGPT met them:

Quality criterium	Evaluation
General conclusion correctness (yes or no)	yes
<i>Restrictions and special cases</i>	partly
Distinguishing between modernization cost and repair cost	no
Showing maximum rent increase in % of modernization cost	yes
Showing maximum rent increase in €	no
Show cases in which the rent cannot be increased	yes
Show cases in which the maximum rent is lower	no

**Table 1**

Results: Experiment 1, Prompt 1

We then refined our prompt so that it showed more specifications of our use case:

*"The total cost of the measure for my apartment is 10,000€, my apartment's size is 30 square meters and my rent used to be 207€. The new rent is supposed to be 273.66€. Is that allowed?"*

In this case, we expected to receive an answer referring to the two most important points: Firstly, the system must distinguish between modernization costs and maintenance costs. Maintenance costs

must be deducted from the total costs (§ 559 Abs. 2 S.1 Hs. 1 BGB). Secondly, the cap in § 559 Abs. 3a S. 2 BGB demands a maximum increase of 2€ per sqm.

In our use case, the legal basis entitling to a rent increase is § 559 Abs. 1 S. 1 BGB. However, § 559 Abs. 2 S. 1 Hs. 1 BGB states that maintenance costs must not be taken into account when calculating the rent increase. Furthermore, it is necessary to consider whether a rent cap is applicable here. First, it must be examined whether the rent cap exception under § 559 Abs. 3a S.1 BGB applies. If it does not apply, the rent cap according to § 559 Abs. 3a S. 2 must also be examined. As shown in 3, in the present case, the system failed to consider the deduction required under § 559 Abs. 2 S. 1 Hs. 1 BGB. Furthermore, the rent cap was either not taken into account at all or was only incorrectly considered (limited to § 559 Abs. 3a S. 1 BGB instead of § 559 Abs. 3a S. 2 BGB).

Quality criterium	Evaluation
Correct calculation of rent increase	yes
Distinguishing between modernization cost and repair cost	no
Finding special case of previous rent	no
General conclusion correctness (yes or no)	no

**Table 2**

Results: Experiment 1, Prompt 2

We expect the structuring of legal norms with a closed-world ontology to produce a more exact output because the reasoning will occur in a predetermined environment and hallucinations will be prevented. Furthermore, we will have full control over the sources that are used to obtain the underlying information.

In order to conduct a rent increase in accordance to § 559 Abs. 1 BGB, we want to develop the ontology so that the relationships between the subdivisions of norms are considered and the system is able to reason according to the information structure provided. By interpreting the deduction of maintenance costs and the rent caps as "exceptions" within an ontology, it becomes possible to conduct a structured and interrelated examination based on the relevant legal provisions. In our example, we have only considered the maintenance cost deduction and the rent caps. Naturally, a legal provision is even more complex, meaning that not all interrelations have been taken into account so far and, therefore, have not yet been mentioned here.

### 3.3. Second Experiment

As a second experiment, we provided ChatGPT with our ontology in an .owl file and the following prompt:

*"You are a law expert that receives the following ontology about modernization of buildings. It contains concepts with the general structure of legal norms and instances. A user asks the following question: "We are in Germany. My landlord had a measure done to insulate the outer wall. The total cost of the measure for my apartment is 10,000€, my apartment's size is 30 square meters, and my rent used to be 207€. The new rent is supposed to be 273.66€. Is that allowed?" Answer the question according to all legal aspects that the ontology contains."*

In this case, we expected ChatGPT to find the content of § 555 Abs. 3a S. 2 BGB that is displayed in our ontology verbatim, and to interpret it correctly.

However, this was not the case. As shown in table 3, although the part of the answer consulting the legal aspect was shorter and contained more legal norms, this particular norm crucial for a correct answer was not consulted. Instead, ChatGPT came to the same flawed conclusion as in the first experiment and wrongly assessed that the procedure was legal. However, because of the indeterminism of LLMs, a reproduction of this result cannot be guaranteed.

Quality criterium	Evaluation
Correct calculation of rent increase	yes
Distinguishing between modernization cost and repair cost	no
Finding special case of previous rent	no
General conclusion correctness (yes or no)	no

**Table 3**

Results: Experiment 2

## 4. Discussion

In our first experiment, we discovered certain limits of ChatGPT when asking for legal advice.

In our second experiment, we showed that even a logical and compositional structure of legal text is not necessarily enough of an enhancement for ChatGPT to answer legal questions, even if the entities provide the legal answer verbatim.

We draw the conclusion that simply providing an LLM with an ontology containing legal norms is not enough. Instead, we plan to introduce constraints to the ontology, provide a reasoning system with an ontology and ask an LLM to use said reasoning system to generate an answer.

## Declaration on Generative AI

Generative AI (ChatGPT) was used exclusively in the writing of this paper to linguistically review the text provided in its entirety by the authors and to improve its comprehensibility.

## Acknowledgements

This research was funded by the Federal Ministry for Economic Affairs and Climate Action (Bundesministerium für Wirtschaft und Klimaschutz), Germany, within the “Joint project: EnOB: IntelMOD - Intelligent modernization platform based on functional cost splitting; sub-project: Infrastructure of the modernization platform“ (“Verbundvorhaben: EnOB: IntelMOD - Intelligente Modernisierungsplattform auf Basis des Funktionalen Kostensplittings; Teilprojekt: Infrastruktur der Modernisierungsplattform.“), FKZ 03EN1094D.

## References

- [1] K. David, Funktionales Kostensplitting zur Ermittlung von Mieterhöhungen nach energetischen Maßnahmen - Eine Handlungsempfehlung auf Basis theoretischer und empirischer Untersuchungen, doctoralthesis, HafenCity Universität Hamburg, 2019. URL: <https://repos.hcu-hamburg.de/handle/hcu/505>.
- [2] ELI Taskforce, ELI, the European Legislation Identifier, Technical Report, Publications Office of the European Union, 2024. URL: <https://op.europa.eu/en/publication-detail/-/publication/4fc75171-f616-11ee-8e14-01aa75ed71a1/language-en>. doi:10.2830/8537, accessed: 2025-04-27.