

# Bridging the Gap: Mapping Layperson Narratives to Legal Issues with Language Models

Hannes Westermann<sup>1,\*</sup>, Sébastien Meeùs<sup>1</sup>, Mia Godet<sup>1</sup>, Aurore Troussel<sup>1</sup>, Jinzhe Tan<sup>1</sup>, Jaromir Savelka<sup>2</sup> and Karim Benyekhlef<sup>1</sup>

<sup>1</sup>Cyberjustice Laboratory, Faculté de droit, Université de Montréal, Québec, Canada

<sup>2</sup>School of Computer Science, Carnegie Mellon University, Pittsburgh, USA

## Abstract

Individuals without legal training (i.e., laypeople) typically tend to perceive their situation through facts, i.e., events that occur. Understanding which legal opportunities or remedies are available to them requires an analysis of which legal issues are raised by these facts, which may be difficult for laypeople to assess. This “gap” can cause laypeople to miss out on benefits or be unable to resolve their disputes. In this paper, we propose an approach to automatically analyze a factual description provided by a layperson in order to map it to potentially relevant legal issues. The system then suggests the issues to the user who may decide if and how to explore them. We demonstrate how this approach could be integrated in legal decision support tools, such as the JusticeBot, to guide users to the relevant guided pathways, while giving the user the possibility to verify the results. This has the potential to further increase the impact on access to justice of such tools. We evaluated the approach on real-world data collected in the JusticeBot project, and found that the system was able to identify the relevant legal issue in 93.5% of selected cases. Our findings can be leveraged by legal professionals and developers of legal decision support systems to alleviate the challenges related to bridging the gap between layperson language and legal issues.

## Keywords

JusticeBot, Access to Justice, Augmented Intelligence, Language Models, Sentence Embeddings, Human-computer interaction

## 1. Introduction

Many individuals have issues resolving their legal disputes. Most laypeople (i.e. individuals without legal training) will face a legal dispute at some point in their lives. These may include, e.g., issues related to debts, employment or consumer rights [1]. Many individuals do not know how to effectively resolve such disputes. They end up not doing anything at all, or trying to solve the issue on their own without professional support. Surveys show that many individuals believe they could have received a better result with more legal information [2].

Only a minority of people use the court system to resolve their issues. Using the court can be expensive, leading many individuals to self-represent. It can also have significant temporal and emotional costs [3]. Overall, a survey conducted in 2021 showed that only 21% of legal disputes that had come up over the past few years

had been resolved at the point when the survey was conducted [4]. Globally, 1.5 billion people are estimated to have legal issues they cannot resolve, and 4.5 billion people are estimated to be excluded from the opportunities the law provides [5].

A significant issue in providing legal information to laypeople is the gap between layperson language and legal language [6]. Laypeople often tend to think of their situation in terms of what has happened (such as “There is a water leak in my apartment”). In order to obtain legal relief, however, the layperson needs to link this factual occurrence to legal issues that can lead to remedies. Each factual situation may give rise to different legal criteria being fulfilled, and thus different remedies.

Lawyers are trained to recognize which facts fulfill a certain legal criterion, e.g., that a water leak can represent a failure of a landlord to fulfill their duties. Laypeople, however, may not be able to establish this link, which could lead them to struggle to understand which rights they have, or even that they have any legal rights at all. Knowing there is a legal right is, of course, an important precondition to commencing the enforcement of these rights. A survey conducted in 2009 showed that many of the individuals that did not act at all in response to a legal problem were not aware that their problem had a legal solution [2]. Further, laypeople may struggle to know which forms to employ when filing a claim in court [7], or which facts are relevant and need to be proved when in court [6, 8].

This gap between a layperson understanding of a situ-

*Proceedings of the Sixth Workshop on Automated Semantic Analysis of Information in Legal Text (ASAIL 2023), June 23, 2023, Braga, Portugal.*

\*Corresponding author.

✉ hannes.westermann@umontreal.ca (H. Westermann);

sebastien.meeus@umontreal.ca (S. Meeùs);

mia.godet@umontreal.ca (M. Godet);

aurore.troussel@umontreal.ca (A. Troussel);

jinzhe.tan@umontreal.ca (J. Tan); jsavelka@cs.cmu.edu

(J. Savelka); karim.benyekhlef@umontreal.ca (K. Benyekhlef)

☎ 0000-0002-4527-7316 (H. Westermann); 0000-0002-8259-2630

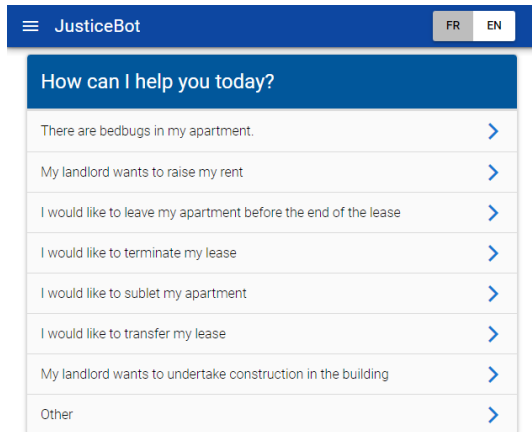
(J. Tan); 0000-0002-3674-5456 (J. Savelka); 0000-0001-9390-556X

(K. Benyekhlef)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)



**Figure 1:** After indicating that they are a tenant, the user is presented a list of legal issues that the system can handle. Translated from French.

ation and a legal understanding of a situation can further affect the usefulness of legal self-help tools [6], a powerful way of increasing access to justice [9, 10, 11]. We have noticed this in the user-feedback of the JusticeBot, a legal decision support tool focused on landlord-tenant disputes, built at the Cyberjustice Laboratory. The system can ask questions of the user, analyze their responses, and then provide them with legal information regarding their situation and the potential next steps that they could consider undertaking to resolve their dispute. The first version of the JusticeBot, built in collaboration with the tribunal administratif du logement (Housing tribunal of Quebec), has been used by over 20,000 users (over 140k page views). For a more complete description, see [12].

After accessing the JusticeBot at <https://justicebot.ca>, users first select whether they are a landlord or a tenant. If they choose, e.g., the tenant option, they are given a list of legal issues that the system can handle. Figure 1 shows the issues available for the tenant, such as “There are bedbugs in my apartment”, “I want to terminate my lease” or “Other”.

By clicking any of these options, the user will be taken through a legal guided pathway that helps them assess their rights and understand the potential next steps they can undertake to address their situation. If none of the issues are relevant to the user, they can click the “Other” option, which will inform them that their issue is not yet covered by the JusticeBot. They further have the option to submit a form describing their situation, so that we can evaluate which pathways should be added to the JusticeBot.

When analyzing the submitted descriptions, we noticed that a significant number of the issues described by the users were, in fact, covered by existing pathways

in the JusticeBot. This may have been due to the phenomenon described above. Several of the pathways (such as wanting to leave the apartment before the lease term) are reliant on the user knowing which legal remedy they want to explore. If they instead think of their situation in terms of facts, they may not know which remedies are even available to them, meaning that they miss out on the useful information contained in the JusticeBot. Since an enormous amount of factual situations could lead to certain consequences, adding all of the possible factual situations to this list would make it unwieldy and difficult to use.

In this paper, we propose a system that can analyze layperson descriptions of an event, and detect possible legal issues that may be relevant to their factual situation. This system can be used, for example, to suggest potentially relevant pathways to the user inside a legal decision support tool like the JusticeBot. Here, the user can be guided to the appropriate pathway, or even external sources if the issue is not covered. Thus, the user is more likely to find the pathway that is relevant to their situation, which can further increase the effectiveness of the decision support tools, contributing to an increase in access to justice. We describe the system, and evaluate its capability to correctly identify the relevant pathway based on real-world user data submitted to the JusticeBot.

## 2. Related and Prior work

Westermann et al. describe foundational principles of the JusticeBot design and reports initial proof of concept experiments [13]. In a follow-up publication, Westermann and Benyekhlef proposed a generalized methodology for building augmented intelligence tools for laypeople to increase access to justice [12]. Here, we devise a method to extend this methodology, by suggesting relevant pathways to the user based on layperson descriptions of factual situations.

The plain language movement<sup>1</sup> has criticized the verbosity of legal language, complicated syntax employed in legal texts, as well as overuse of specialized terms. In the past, there have been studies exploring the possibilities, effects and limitations of communicating legal information in a way that is more accessible to laypeople [14, 15, 16]. There have been early attempts to create a legal information retrieval system for laypeople [17]. Garimella et al. experimented with general natural language processing (NLP) text simplification methods on legal documents [18]. Uijttenbroek et al. describe a system that analyzes laypeople’s input in terms of a layperson ontology, maps the entities to a legal ontol-

<sup>1</sup>Plain Language: Beyond a Movement. Available at: <https://www.plainlanguage.gov/resources/articles/beyond-a-movement/> [Accessed 2023-5-3]

ogy, retrieves relevant case law, and finally presents the results to the layperson in a comprehensible way [19]. Fernández-Barrera and Casanovas focused on mapping layperson queries to ontologies in the domain of consumer mediation [20]. [6] explored the difference between layperson language and judicial language, finding that layperson submissions are difficult to use to predict case outcomes. Spot is an API that can analyze non-lawyer descriptions and link it to a standardized list of legal issues.<sup>2</sup> [21] compared the JusticeBot approach to asking ChatGPT questions in layperson language, finding that the answers given by ChatGPT had some issues when it comes to accuracy and reliability. Here, we propose an approach to map layperson factual descriptions to legal issues using language models, describe its use in the context of a decision support system, and evaluate its performance on real-world data.

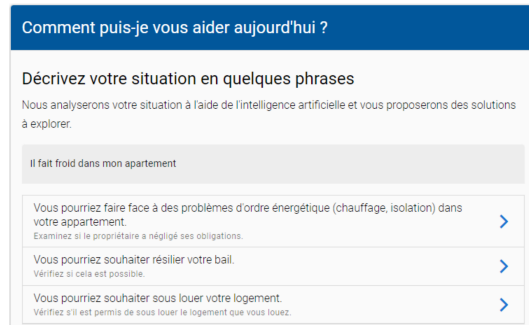
In NLP, the success of word embeddings (e.g. [22, 23]) was followed by an increasing interest in learning continuous vector representations of longer linguistic units such as sentences. This trend that has been reflected in AI & Law research as well [24, 25, 26]. Cer et al. [27] utilized the transformer architecture [28] and Deep Averaging Network [29] trained on the SNLI dataset. Reimers et al. build on top of BERT [30] and RoBERTa [31], which have been shown to be remarkably effective on a number of NLP tasks. Specifically, they used siamese and triplet network structures to derive semantically meaningful sentence embeddings [32]. Conneau et al. demonstrated the effectiveness of models trained on a natural language inference task (SNLI dataset [33]). They proposed a BiLSTM network with max pooling trained with fastText word embeddings [34, 35] as the best universal sentence encoding method [36]. While most of the earlier work was limited to a one or few languages, several approaches to obtain general-purpose massively multi-lingual sentence representations were proposed [37, 38, 39]. Such representations were utilized in many downstream applications, such as document classification [40], machine translation [41], question answering [42], hate speech detection [43], or information retrieval (IR) in the legal domain [44]. In this work, we utilize a multilingual sentence encoder [39] to embed and compare factual descriptions written by laypeople.

### 3. Proposed system

#### 3.1. Interface

To demonstrate the usefulness of the approach, we show how our method, linking layperson factual narratives to legal issues, can be used to support users of the JusticeBot. As discussed, users of the system frequently

<sup>2</sup><https://spot.suffolktilab.org/>



**Figure 2:** After the user described their factual situation, they are given suggestions of which pathways may be relevant to them.

struggle to identify the pathway that is relevant to their situation. Hence, we implemented a feature to automatically suggest a potentially relevant pathway based on a description of an issue provided by the user.

Figure 2 shows the new interface related to the new feature in the JusticeBot. Instead of just a list of possible pathway options (as can be seen in Figure 1), users are now also shown a text box, and are invited to describe their factual situation. While they are typing, the system will retrieve suggestions of relevant pathways and display them to the user. Figure 2 shows the result of entering "I am cold" into the text box. As we can see, the system suggests three possible pathways that may be relevant to the user. Each suggestion consists of the following elements:

- **Factual explanation:** An explanation of what the system understood from the user's description. This can help the user verify that the system has correctly understood their situation. In our example, the first entry in the list states: "You may have issues with heating or insulation.", based on the factual situation described by the user.
- **Suggested action:** An explanation of what will be accomplished by clicking the link, i.e. the legal issue the user may want to explore. For example, the first entry in Figure 2 indicates that the user may wish to explore whether their landlord has not fulfilled their duties, and the consequences of such a situation. The suggested action is also important where the same factual situation (e.g. heating not working) can lead to different legal remedies becoming relevant (e.g. rent reduction and/or lease termination).
- **Link:** Once users click the suggestion, they will be taken:
  - To the relevant pathway, if the issue is covered by the JusticeBot. It can also take the

user to specific locations in the pathway, if the answers to certain questions are already evident from the factual description of the user.

- To an external site that has more information about the legal situation, if the issue is not covered. While providing the user with verified information in the JusticeBot is preferable, linking to trusted external sources can be helpful where the corresponding JusticeBot pathway has not yet been created.

As we can see, just like the JusticeBot itself, the pathway suggestion system acts as an augmented intelligence system, by suggesting pathways or external sources to the user, but never telling them what to do. By reading the factual explanation, the users are able to verify that the system has correctly understood their situation. We will explore this feature more in depth in Section 6.

### 3.2. Methodology

Next, let us take a look at the technical stack that enables the functionality of finding relevant pathways based on a layperson's factual description of a situation. The key idea underlying the approach is that the user query is not compared directly to the legal issue, but rather to a database of example descriptions of situations that would be covered by a certain legal issue.

The process consists of four steps:

- Section 3.3 - The creation of example descriptions of situations that would lead to certain legal issues becoming relevant.
- Section 3.4 - The creation of sentence embeddings for each example description.
- Section 3.5 - The indexing of these sentence embeddings, in order to be able to quickly retrieve example descriptions similar to a new description.
- Section 3.6 - In order to suggest relevant pathways to the user, we create an embedding of the factual description provided by the user and use the index to retrieve similar example descriptions, that are then used to suggest the relevant legal issue.

The methodology is similar to that presented in [26] and [25].

### 3.3. Creation of example situations

The first step is the creation of a database of content that can be used to match the user's description of a situation. We started by creating such descriptions ourselves, considering the legal issues that are currently covered by

the JusticeBot. These pathways describe different legal issues that a user may wish to explore. For each of these pathways, we formulated multiple factual situations that could give rise to this pathway being relevant for a user. Then, we put ourselves in the shoes of a user that faces this factual issue, and imagined how a user may describe their situation, in their own words. For each suggestion, we thus end up with the following elements:

1. A number of example descriptions how a layperson may express their situation. For example:
  - I need to wear a jacket indoors.
  - When I wake up, I have bites on my face.
  - I would like to go on vacation, what can I do?
  - I received a letter from my landlord, informing me of a rent increase.
2. A **factual explanation** of the situation, specifying what the model understood from the user description (e.g. "You may have heating issues", "You may have a bedbug infestation" or "You seem to want to leave your apartment for a while", see Section 3.1)
3. A **suggested action** that the user will undertake by pressing the suggestion (e.g., "Explore whether the landlord has breached their obligation to keep the apartment warm", or "Read more about bedbugs on an external site", see Section 3.1)
4. A **link** to the relevant legal issue, as covered in the JusticeBot. This can be the beginning of a pathway, or a deep link into the pathway. It can also be a link to external sources.

For example, linking the situation of a tenant being cold in their apartment to a pathway exploring whether the tenant can receive a rent reduction, these elements would be as follows:

1. Example descriptions of heating issues:
  - I need to wear a jacket indoors.
  - There is frost on the inside of my window.
  - I feel cold indoors every day
  - etc.
2. "You seem to have heating issues"
3. "Explore whether you can receive a rent reduction due to the landlord not fulfilling their obligations."
4. A reference to a relevant JusticeBot pathway.

The more of the example descriptions are present, the more likely it is that the description of the user will be similar to a previously created description (see section 3.6). To develop the database of example user descriptions, we use two sources:

**Seed example descriptions:** As described above, we write our own seed examples that describe how a user

might express their situation. We tried to be as varied as possible in formulating these prompts, to be able to provide relevant suggestions for as many user descriptions as possible. Many different types of descriptions can lead to the same pathway becoming relevant. For example, both “I would like to sublet my apartment” and “I am going on vacation” and “My friend wants to live in my apartment for a few months” could guide the user to explore the legal situation of subletting. Due to the power of language models, as we will see, it does not matter if the way the user describes their situation perfectly matches the example formulation—as long as the meaning is similar, the suggestion that is the most similar will often be correct.

**User-submitted example descriptions:** Second, as described in section 1, users of the JusticeBot tool are invited to submit a description of their situation if their case is not covered by the JusticeBot. These are yet another high-quality source of data, since they represent the real-world users’ descriptions of their situations. Thus, by using these descriptions, we are able to match the way laypeople describe their situations.

Beyond adding the user-submitted examples as training data to benefit the matching methodology, they can also be used to evaluate the performance of how well the suggestion feature works. Each submitted missing question description represents a user that was not able to identify the pathway that is relevant to them in the JusticeBot, even though some of these situations are already covered in the JusticeBot. Thus, by annotating these user-submitted descriptions, we can evaluate whether the system is able to surface a pathway that is relevant to them. The experimental setup is described below in Section 3.2.

### 3.4. Embedding of example situations using language model

To match user descriptions to the stored example descriptions, we embed the stored examples using a sentence encoder into a vector format, and then use an approximate nearest neighborhood model to retrieve sentences that are semantically similar to the user description.

To convert each sentence into an embedding, we use a multilingual universal sentence encoder. This model is a multi-task, dual-encoded [45] convolutional neural network, that has been pre-trained to embed texts from 16 languages into the same semantic embedding space [39]. We use a pre-trained version of this model available on tensorflow-hub.<sup>3</sup> The model takes a sentence as input, and produces a vector of 512 dimensions, capturing the semantic content of the text.

<sup>3</sup><https://tfhub.dev/google/universal-sentence-encoder-multilingual/3>

### 3.5. Indexing of example embeddings

To retrieve semantically similar sentences from the database, we use the Annoy similarity search library released by Spotify<sup>4</sup>. We opted for this solution for its ease of use and minimal system requirements. The annoy library is a very quick and light implementation of an Approximate Nearest Neighbors algorithm proposed in [46]. The library enables us to build an index for the sentence embeddings created in the previous step. When supplied with a vector, the index can surface the N most similar vectors, in a fraction of a second.

### 3.6. Analysis of user query

When the user accesses the front-end for a block that has the NLP feature enabled, they are shown the screen in Figure 2. They are then able to type a factual description of their situation.

Once they pause writing, the text entered by the user is sent to the server. This means that the user can get relevant explanations even before they have completed the typing of their description, which saves time since they may already see the relevant suggestion after starting to write, making it unnecessary to complete their writing.

On the server side, once the factual description is received, the description is vectorized with the embedding model described above in 3.4. Then, the search index is used to retrieve the previously embedded examples, as described in section 3.5. This results in a list of the example descriptions that are the most similar to the user factual description. Since these example descriptions are linked to a relevant legal issue (see 3.3), we can retrieve the legal issue that was seen as relevant for an example description. The top three legal issues that have an example description linked to them that is the most similar to the user factual description are shown to the user.

The user is shown the factual explanation and suggested action (see Section 3.1). They can verify that the factual explanation corresponds to their situation, and click on the suggestion to be taken to the relevant page in the JusticeBot, or to an external source that provides useful information to them.

An advantage of the technique used in this approach is that it is very fast. The entire analysis performed on the server (i.e. embedding the user factual description, retrieving the most similar example descriptions and selecting the relevant suggestions) can be done in a few milliseconds. This means that the user can obtain suggestions very quickly.

<sup>4</sup>[github.com/spotify/annoy](https://github.com/spotify/annoy)

### 3.7. User feedback

Users interacting with the system provide valuable additional training data for the model, that can be verified by expert annotators. If a user writes the description of their facts into the text box that can be seen in Figure 2, they receive suggestions from the system. While some of these suggestions will be relevant, some will not. How the user acts in response to the suggestion can be a strong indicator of whether the suggestions are useful or not. When given suggestions, users can take three actions:

- They can click one of the suggestions. This indicates that their factual description is likely relevant to the link that they clicked. Thus, the description can be saved and potentially be included in the system as an example description.
- They may not click any of the suggestions, and instead click one of the links in list of pathways shown below. This means that the suggestions surfaced by the system were not relevant in this case, and that the user opted to use the standard list for selecting their pathway. In this case, their factual description can still be used, and added to the suggestion of the pathway that the user selected.
- If the user enters a factual description, and then clicks on the “Other” heading, this may be an indication that none of the options are satisfactory to them. Thus, the description of the user can be added to the “missing questions” database, as described in 1.

Thus, the system can collect data and improve over time, as users interact with it. Of course, the data would need to be anonymized, and the user be made aware that their input can be used in this way.

We have now seen the approach we developed to link users’ factual descriptions to legal issues. Next, we will describe how we evaluated this approach.

## 4. Experimental design

In order to evaluate the methodology described in Section 3.2, we analyze three research questions:

1. RQ1 - Can the proposed methodology achieve adequate performance in pointing individuals toward the correct legal issues using the seed examples only?
2. RQ2 - Can the performance of the methodology in guiding individuals toward the correct pathway be further increased by augmenting the database with real-life user factual descriptions?
3. RQ3 - Can the use of a language model and seed examples overcome the cold-start problem to rapidly achieve usable performance?

**Table 1**

Dataset, composed of prepared seed examples, and examples drawn from user submissions

| Legal issue               | N seed    | N user     | Total      |
|---------------------------|-----------|------------|------------|
| Bedbug infestation        | 6         | 12         | 18         |
| Rent increase             | 5         | 68         | 73         |
| Lease termination         | 9         | 13         | 22         |
| Sublease                  | 6         | 8          | 14         |
| Lease transfer            | 5         | 10         | 15         |
| Renoviction               | 6         | 18         | 24         |
| Eviction to increase size | 3         | 6          | 9          |
| Heating                   | 4         | 22         | 26         |
| Nuisance                  | 4         | 250        | 254        |
| Repossession              | 5         | 154        | 159        |
| Animals                   | 5         | 39         | 44         |
| <b>Total</b>              | <b>58</b> | <b>600</b> | <b>658</b> |

### 4.1. Dataset

To evaluate our research questions, we focused on the perspective of an individual who has indicated that they are a tenant, as shown in Figure 1. First, for each of the available pathways, we created seed example descriptions, by imagining how users might express themselves regarding situations that would benefit from the pathway. Further, we created some suggestions that take the user deeper into the pathway—for example, the issue of heating is linked to a pathway that allows the user to explore a rent reduction or lease termination. These example descriptions are referred to as “seed” in Table 1. We created them in french, but note that due to the multilingual nature of the embedding model used, even seed examples in other languages should be usable by the system.

Next, we analyzed the factual descriptions submitted by users as not covered by the JusticeBot (see Section 1). For the issues that are covered by the JusticeBot, we noted the corresponding pathway and added the examples to the dataset. In total, we annotated 3,250 such submitted feedback examples. These range from a single word to multiple sentences. This part of the dataset is referred to as “user” in Table 1. As can be seen in Table 1, we identified a substantial amount of situations where the JusticeBot contains relevant information, but the user was not able to find the relevant pathway.

In analyzing the submitted feedback, we also encountered many situations that were not yet covered by the Justicebot. As discussed in [12], this serves as an excellent basis for prioritizing the addition of new pathways. Further, links to external content can be introduced as a stop-gap measure that can already help, even if the JusticeBot pathway is not yet available. While this will need to be further explored, we already added three such pathways that seem to frequently re-occur, namely ques-

tions regarding animals, repossession and nuisances (see Table 1). For each of these, we created seed examples and assigned user examples.

#### **4.2. RQ1 - Our methodology could achieve adequate performance in pointing individuals toward the correct legal issues, using only seed examples.**

Our first research question explores if the proposed methodology is able to leverage the power of the language model used to achieve strong performance when retrieving legal issues, even if only the seeded examples are used. This would be a strong indication that the NLP feature could be useful even if no user data has been collected, and thus that new user needs can be quickly addressed as they arise. Of course, for this to work, the methodology needs to be able to draw links between the comparatively clean and structured seed examples and the real-world descriptions written by laypeople, that are likely to be much more varied in terms of factual situation, content and tone.

To investigate this research question, we trained a model on only the seed example descriptions prepared by our team. For each user-submitted factual description, we test whether the correct suggestion is surfaced first (P@1) or in the top 3 suggestions shown to the user (P@3), i.e., whether the correct pathway is suggested to the user in the first place, or visible at all in the interface shown in Figure 2. As such, here the “training data” used consists of the seed example data, while the “test data” consists of the user-submitted example descriptions (see Table 1).

#### **4.3. RQ2 - The performance of the methodology to guide individuals toward the correct pathway can be further increased by adding real-life user factual descriptions.**

The second research question analyzes if adding additional factual examples, provided by the users, to “train” the system will increase the amount of correct suggestions. This would indicate that adding additional, real-world data to the system increases its performance, i.e., that it improves over time.

To investigate this research question, we trained a model based on the seed data and the user-submitted factual descriptions. Then, for each user-submitted factual description, we use the methodology to retrieve the suggested pathway. Of course, we excluded the user-submitted pathway that is currently used in retrieval when evaluating the result, i.e. in a leave-one-out cross-

validation setting. As such, the “training data” consists of the seed data and the user-submitted data minus one sample, while the “test data” consists of the held-out sample. We report the scores of whether the correct suggestion is displayed first (P@1) and whether the correct suggested pathway is part of the top 3 results shown to the user (P@3).

#### **4.4. RQ3 - Using a language model and seed examples can overcome the cold-start problem to rapidly achieve usable performance**

The third research question analyzes whether the language models used in the methodology can bring the benefit of quickly adapting to the training data. Traditional machine learning often runs into the cold start problem, where it takes a while to be able to learn enough to generalize beyond the training data. RQ3 investigates whether our approach can help overcome this issue. This could be the case due to our use of language models, which have been trained on large corpora to absorb language patterns. Further, we use seed examples, that are synthetically created training examples aimed to teach the models some patterns before real data can be collected, which could also help in overcoming the cold-start problem.

To investigate this issue, we perform an experiment where we continually add training data and see how the performance of the model develops (compare [47]), when tested against test data (100 random samples of the user-submitted data that we withheld). We compare our language model-based approach to a more traditional strong baseline in the form of a support vector machine (SVM) trained on TF-IDF representations of the example descriptions. For both models, we run two variants of the test: one where the model is first trained on the seed examples and then on user-submitted examples, and one where it is only trained on the user-submitted examples (user\_only).

## **5. Results**

### **5.1. RQ1 - Our methodology could achieve adequate performance in pointing individuals toward the correct legal issues, using only seed examples.**

Table 2 shows the performance of the model when trained on the seed examples prepared by our team only. Overall, 54.8% of the users would be provided with a suggestion that is relevant to them by the system. The performance

**Table 2**

Precision values when only the seed example descriptions are used for training, evaluated on user examples.

| Legal issue               | P@1          | P@3          | Missed       |
|---------------------------|--------------|--------------|--------------|
| bedbug                    | 33.3%        | 33.3%        | 66.7%        |
| rent increase             | 69.1%        | 76.5%        | 23.5%        |
| lease termination         | 69.2%        | 76.9%        | 23.1%        |
| sublease                  | 25.0%        | 37.5%        | 62.5%        |
| lease transfer            | 60.0%        | 70.0%        | 30.0%        |
| renoviction               | 50.0%        | 77.8%        | 22.2%        |
| eviction to increase size | 50.0%        | 66.7%        | 33.3%        |
| heating                   | 72.7%        | 77.3%        | 22.7%        |
| nuisance                  | 25.2%        | 42.4%        | 57.6%        |
| reposession               | 40.3%        | 52.6%        | 47.4%        |
| animals                   | 74.4%        | 79.5%        | 20.5%        |
| <b>Total</b>              | <b>41.7%</b> | <b>54.8%</b> | <b>45.2%</b> |

**Table 3**

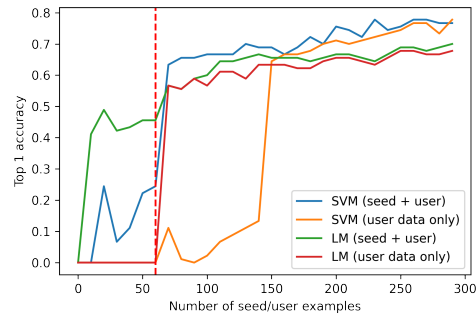
Precision values when only seed and user examples are used for training, evaluated on user examples.

| Legal issue               | P@1          | P@3          | Missed      |
|---------------------------|--------------|--------------|-------------|
| bedbug                    | 50.0%        | 75.0%        | 25.0%       |
| rent increase             | 83.8%        | 98.5%        | 1.5%        |
| lease termination         | 46.2%        | 84.6%        | 15.4%       |
| sublease                  | 25.0%        | 100.0%       | 0.0%        |
| lease transfer            | 50.0%        | 70.0%        | 30.0%       |
| renoviction               | 33.3%        | 88.9%        | 11.1%       |
| eviction to increase size | 33.3%        | 83.3%        | 16.7%       |
| heating                   | 54.5%        | 81.8%        | 18.2%       |
| nuisance                  | 78.4%        | 94.0%        | 6.0%        |
| reposession               | 79.9%        | 96.1%        | 3.9%        |
| animals                   | 82.1%        | 94.9%        | 5.1%        |
| <b>Total</b>              | <b>74.5%</b> | <b>93.5%</b> | <b>6.5%</b> |

varies widely between the different legal issues. For the animals category, inquiring about questions regarding animals in an apartment, almost 80% of the users would have received a relevant suggestion. For a bedbug infestation, on the other hand, only a third of the users would have received the relevant pathways. We will discuss this divergence in Section 6.

### 5.2. RQ2 - The performance of the methodology to guide individuals toward the correct pathway can be further increased by adding real-life user factual descriptions.

Table 3 shows the performance when the system is “trained” on the seed example description as well as user-submitted example descriptions, and evaluated on one of the user-submitted examples at a time. As we can see, the performance is much higher than when using the seed



**Figure 3:** A graph showing the top 1 precision for using a Support Vector Machine (SVM) model versus our approach (LM), when seed examples are used for training versus only user-submitted data. Left of red line: Seed data. Right of red line: User data.

examples only. Overall, 74.5% of user-submitted queries would have received the relevant suggestion at the top spot, while 93.5% of users with one of the captured situations would have received a relevant suggestion in the three suggestions that they were shown. Only 6.5% of users would not be shown a relevant suggestion. The results vary between the different issues—for lease transfer, about 30% of the users would not see the relevant suggestion, while for sublease, 100% of users would see the relevant suggestion in the 3 entries surfaced to them.

### 5.3. RQ3 - Using a language model and seed examples can overcome the cold-start problem to rapidly achieve usable performance

Figure 3 shows the performance impact of using our language model-based approach versus a support vector machine, with or without the seed examples. As we can see, our approach reaches reasonable performance much quicker than the SVM. For the SVM, the use of seed examples increases the performance and the speed at which the model can adapt to the real-world data.

## 6. Discussion

We have described the methodology to detect relevant legal issues from layperson factual descriptions, and presented the results of our evaluation. Let us discuss these results, and some aspects of the proposed methodology in increasing access to justice.



### **6.1. RQ1 - Our methodology can achieve adequate performance in pointing individuals toward the correct legal issues, using only seed examples.**

The results of exploring RQ1 are shown above in Table 2. As we can see, the proposed approach is able to surface the correct legal issue in the top 3 suggestions for 54.8% of the user descriptions.

This may seem like a relatively low number. However, it is important to consider the difficulty of the performed task. First, we only use the very few seed examples to “train” the model. For each suggestion, there are only 3–9 such seed examples. Further, the tasks consist of analyzing layperson language and linking it to structured legal issues. Layperson texts are notoriously difficult to deal with. For example, in [6], the authors found that submissions by unguided pro-se litigants could not be used to predict the outcome of cases, likely due to the gap between common parlance and legal language. Even though we attempted to create seed examples that are similar to how users are likely to describe their issues, it is possible that our one-sentence examples are very different from the way real-world users describe their issues. For example, the average length of the seeded examples ( 51 characters) is less than half the length of the user-submitted examples ( 109 characters). Further, the factual situation mentioned in the user-submitted examples are likely to be much more diverse than the limited number of situations captured in the seed examples.

From this viewpoint, the model being able to identify the correct legal issue in over half of the cases is a reasonable result. It shows the power of the language model used to absorb the semantic meaning of the sentences. Of course, there is room for improvement, which will be explored in future work.

### **6.2. RQ2 - The performance of the methodology to guide individuals toward the correct pathway can be further increased by adding real-life user factual descriptions.**

RQ2 investigates how well our approach works when previous user-submitted data is used to “train” the model. The overall performance is strong, with 93.5% of the user-submitted factual descriptions surfacing the correct legal issue (P@3). The performance varies somewhat between the different classes (between 70% and 100% at P@3), which may be due to the “semantic homogeneity” of the classes - perhaps, certain legal issues have much more divergent situations and descriptions that could give rise to them (compare [47]).

We observe that adding data from the user-submitted

queries definitely enhances the performance of the model, compared to when using seed explanations only. This shows that such a system built using real-world data could work well in practice, and contribute to increasing access to justice.

### **6.3. RQ3 - Using a language model and seed examples can overcome the cold-start problem to rapidly achieve usable performance**

In our third research question, we investigate whether language models and seed questions can overcome the cold-start problem. This seems to be the case, as the language models is much quicker to learn a pattern from the provided data than the SVM, both in the configuration with and without seed data. Likewise, when the seed data is used, the SVM is quicker to adapt to the real-world user-submitted data. Thus, it seems like both of our approaches paid off and show promise in overcoming the issue of a cold start. However, it also seems that potentially another approach (such as SVM or BERT-based classifiers) should be used once enough data has been collected, to guarantee the best performance. Future work will investigate how other models will perform at different stages, and when the switch should be made.

### **6.4. Augmented Intelligence**

The question remains whether the performance shown by these models is “adequate” for deployment in a real-world model. A factor that is important to consider is that the feature as described here is conceptualized as an “*augmented intelligence*” approach, that aims to support the user instead of replacing them. It does not automatically take the user to a pathway, but instead surfaces three suggestions that may be relevant to the situation described by the user. Each suggestion has a “factual explanation”, that allows the user to verify that the model correctly analyzed their factual situation. If the factual explanation of the suggestion matches the situation of the user, the legal issue is guaranteed to be relevant, as it has been encoded by legal experts. Thus, the user is given a way to meaningfully verify whether the system worked, without having to understand the legal particularities of their situation — they merely have to read the “factual explanation” and verify that the system correctly understood their situation.

Thus, the “cost” of a failure of the system is relatively low. If the user enters a description of their situation, and none of the suggestions seem relevant, they can simply ignore the feature and continue using the JusticeBot. However, if the system surfaces a relevant suggestion, a user can be directed to the appropriate pathway, that

they otherwise may have missed. Even if a fraction of the users benefits from using the suggestions described in this paper, access to justice could be improved.

While, of course, additional empirical evaluations need to be undertaken to make sure that the use of the system is safe, this framing may make the achieved performance “adequate” for an initial deployment. Further, once the system is live, the data collected can be a powerful way to enhance the system, as described in Section 3.7.

### 6.5. Potential benefits of the approach

As we have seen, the proposed system appears to perform well in identifying the legal issues from real-world, user-submitted factual descriptions. Integrating such system into a decision support tool, such as the JusticeBot, could play an important role in overcoming the gap between regular language and legal language, and give laypeople a better chance to understand how the laws apply to them or the legal remedies available to them. This could be an important step towards increasing access to justice.

At the same time, it is important to keep in mind that the evaluation presented in this paper is limited to examples where the situation described by the user is already part of the JusticeBot. If the situation the user describes is not covered by a suggestion, they would be provided with a list of irrelevant pathways. However, as described above, they would be able to realize this, by seeing that the factual explanation does not apply to them. Even so, in expanding the coverage of legal decision support tools and the JusticeBot, it remains important to add new pathways for frequently recurring situations, or to add references to external content, to ensure that as many people as possible can be helped. The data collected through the use of the NLP feature could be an important tool in determining which pathways to add (see Section 3.7). We also plan to experiment with setting a threshold for sentence similarity that needs to be exceeded for suggestions to be returned at all, which should minimize providing obviously irrelevant suggestions to the users.

We anticipate that the feature described in this paper would also work well for other approaches, beyond the JusticeBot. Connecting layperson language to legal issues is important in many legal tasks. Being able to automate this process, even partially, could lead to many interesting projects e.g. in automated document drafting.

### 6.6. Unauthorized practice of law?

One also has to be cognizant of other risks of using the proposed approach, including the prohibition against the unauthorized practice of law in many jurisdictions. The distinction between providing legal information and legal advice may not always be clear [48]. The JusticeBot itself is specifically designed in a way to only provide legal

information, by merely supplying context to the user, and letting them make the relevant decisions. The feature described in this paper, however, goes a bit further than this, by providing the user possible legal pathways that they can explore, based purely on a factual description. An argument could be made that this could be seen as giving legal advice. However, looking at the information provided in Figure 2, it is important to note that the system is focused on augmenting the intelligence of the user—it is the user that makes the decisions. The user decides whether the surfaced suggestion makes sense and thus which pathway they want to explore. Further, clicking the suggestion leads back to the JusticeBot, which only provides legal information.

## 7. Future Work

This paper leaves ample space for future work. *First*, the dataset needs to be expanded, beyond the currently limited number of issues. Adding new suggestions concerning other frequently occurring situations could expand the usefulness of the tool and provide more context for evaluation. Large language models (LLMs), such as GPT-4 [49], have been used to generate legal example situations [21] and perform legal annotation tasks [50, 51], and may thus be an important way to make this task more efficient. The system should also be evaluated in the context of JusticeBot tools in other legal domains. *Second*, other embedding approaches or nearest neighbor search methods could be tried. Likewise, other machine learning models (including LLMs such as GPT-4) could be used. *Third*, applying the method described here to other tools would be an interesting way to explore how generalizable the approach is. *Fourth*, pilot studies with end-users could help us understand the real-world utility of the system and identify further areas of improvement.

## 8. Conclusion

We described an approach to map layperson factual descriptions to legal issues. We described and discussed the approach. The initial evaluations on real-world user data are promising, and could represent an important addition to the JusticeBot methodology and access to justice.

## Acknowledgments

We would like to thank the Cyberjustice Laboratory at Université de Montréal, the LexUM Chair on Legal Information and the Autonomy through Cyberjustice Technologies (ACT) project for their support of this research.

## References

- [1] T. C. Farrow, A. Currie, N. Aylwin, L. Jacobs, D. Northrup, L. Moore, *Everyday legal problems and the cost of justice in canada: Overview report*, Osgoode Legal Studies Research Paper (2016).
- [2] A. Currie, *The legal problems of everyday life*, in: *Access to justice*, Emerald Group Publishing Limited, 2009.
- [3] N. Semple, *The cost of seeking civil justice in canada*, *Can. B. Rev.* 93 (2015) 639.
- [4] L. Savage, S. McDonald, *Experiences of serious problems or disputes in the canadian provinces*, 2021, *Juristat: Canadian Centre for Justice Statistics* (2022) 1–28.
- [5] *Justice For All - Final Report*, Technical Report, The Task Force on Justice, New York: Center on International Cooperation, 2019.
- [6] K. Branting, C. Balhana, C. Pfeifer, J. S. Aberdeen, B. Brown, *Judges are from mars, pro se litigants are from venus: Predicting decisions from lay text.*, in: *JURIX*, 2020, pp. 215–218.
- [7] J. Macfarlane, *The national self-represented litigants project: Identifying and meeting the needs of self-represented litigants final report* (2013).
- [8] R. Susskind, *Online courts and the future of justice* (2019).
- [9] L. K. Branting, *Advisory systems for pro se litigants*, in: *Proceedings of the 8th international conference on Artificial intelligence and law*, 2001, pp. 139–146.
- [10] D. Thompson, *Creating new pathways to justice using simple artificial intelligence and online dispute resolution*, *IJODR* 2 (2015) 4.
- [11] J. Zeleznikow, *Using web-based legal decision support systems to improve access to justice*, *Information & Communications Technology Law* 11 (2002) 15–33.
- [12] H. Westermann, K. Benyekhlef, *Justicebot: A methodology for building augmented intelligence tools for laypeople to increase access to justice*, in: *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*, 2023.
- [13] H. Westermann, V. R. Walker, K. D. Ashley, K. Benyekhlef, *Using factors to predict and analyze landlord-tenant decisions to increase access to justice*, in: *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, 2019, pp. 133–142.
- [14] R. P. Charrow, V. R. Charrow, *Making legal language understandable: A psycholinguistic study of jury instructions*, *Columbia law review* 79 (1979) 1306–1374.
- [15] M. E. Masson, M. A. Waldron, *Comprehension of legal contracts by non-experts: Effectiveness of plain language redrafting*, *Applied cognitive psychology* 8 (1994) 67–85.
- [16] M. Myška, T. Smejkalová, J. Šavelka, M. Škop, *Creative commons and grand challenge to make legal language simple*, in: *AI Approaches to the Complexity of Legal Systems. International Workshop AICOL-III. Lecture Notes in Computer Science()*, vol 7639., Springer, 2012, pp. 271–285.
- [17] L.-C. Paquin, F. Blanchard, C. Thomasset, *Loge-expert: from a legal expert system to an information system for non-lawyers*, in: *Proceedings of the 3rd international conference on Artificial intelligence and law*, 1991, pp. 254–259.
- [18] A. Garimella, A. Sancheti, V. Aggarwal, A. Ganesh, N. Chhaya, N. Kambhatla, *Text simplification for legal domain: Insights and challenges*, in: *Proceedings of the Natural Legal Language Processing Workshop 2022*, 2022, pp. 296–304.
- [19] E. M. Uijttenbroek, A. R. Lodder, M. C. Klein, G. R. Wildeboer, W. Van Steenberg, R. L. Sie, P. E. Huygen, F. Van Harmelen, *Retrieval of case law to provide layman with information about liability: Preliminary results of the best-project*, Springer, 2008.
- [20] M. Fernández-Barrera, P. Casanovas, *From user needs to expert knowledge: mapping laymen queries with ontologies in the domain of consumer mediation*, in: *AI Approaches to the Complexity of Legal Systems. International Workshop AICOL-III. Lecture Notes in Computer Science()*, vol 7639., Springer, 2012, pp. 286–308.
- [21] J. Tan, H. Westermann, K. Benyekhlef, *Chatgpt as an artificial lawyer?*, in: *Artificial Intelligence for Access to Justice (AI4AJ 2023)*, 2023.
- [22] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, *Distributed representations of words and phrases and their compositionality*, *Advances in neural information processing systems* 26 (2013).
- [23] T. Mikolov, K. Chen, G. Corrado, J. Dean, *Efficient estimation of word representations in vector space*, *arXiv preprint arXiv:1301.3781* (2013).
- [24] L. Zhong, Z. Zhong, Z. Zhao, S. Wang, K. D. Ashley, M. Grabmair, *Automatic summarization of legal decisions using iterative masking of predictive sentences*, in: *ICAIL 2019*, 2019, pp. 163–172.
- [25] H. Westermann, J. Šavelka, K. Benyekhlef, *Paragraph similarity scoring and fine-tuned bert for legal information retrieval and entailment*, in: *New Frontiers in Artificial Intelligence, Lecture Notes in Computer Science*, Springer International Publishing, 2021.
- [26] H. Westermann, J. Šavelka, V. R. Walker, K. D. Ashley, K. Benyekhlef, *Sentence embeddings and high-speed similarity search for fast computer assisted annotation of legal documents*, in: *JURIX 2020*, volume 334, IOS Press, 2020, p. 164.
- [27] D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R. S.

- John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, et al., Universal sentence encoder, arXiv preprint arXiv:1803.11175 (2018).
- [28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, *Advances in neural information processing systems* 30 (2017).
- [29] M. Iyyer, V. Manjunatha, J. Boyd-Graber, H. Daumé III, Deep unordered composition rivals syntactic methods for text classification, in: *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)*, 2015, pp. 1681–1691.
- [30] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
- [31] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, arXiv preprint arXiv:1907.11692 (2019).
- [32] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, arXiv preprint arXiv:1908.10084 (2019).
- [33] S. R. Bowman, G. Angeli, C. Potts, C. D. Manning, A large annotated corpus for learning natural language inference, arXiv preprint arXiv:1508.05326 (2015).
- [34] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching word vectors with subword information, *Transactions of the association for computational linguistics* 5 (2017) 135–146.
- [35] A. Joulin, E. Grave, P. Bojanowski, T. Mikolov, Bag of tricks for efficient text classification, arXiv preprint arXiv:1607.01759 (2016).
- [36] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, A. Bordes, Supervised learning of universal sentence representations from natural language inference data, arXiv preprint arXiv:1705.02364 (2017).
- [37] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: *NAACL 2019, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186.
- [38] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, É. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: *ACL 2020, 2020*, pp. 8440–8451.
- [39] Y. Yang, D. Cer, A. Ahmad, M. Guo, J. Law, N. Constant, G. H. Abrego, S. Yuan, C. Tar, Y.-H. Sung, B. Strope, R. Kurzweil, Multilingual universal sentence encoder for semantic retrieval, 2019. arXiv:1907.04307.
- [40] G. Lai, B. Oguz, Y. Yang, V. Stoyanov, Bridging the domain gap in cross-lingual document classification, arXiv preprint arXiv:1909.07009 (2019).
- [41] R. Aharoni, M. Johnson, O. Firat, Massively multilingual neural machine translation, in: *NAACL-HLT, Vol. 1 (Long and Short Papers)*, 2019, pp. 3874–3884.
- [42] P. Lewis, B. Oguz, R. Rinott, S. Riedel, H. Schwenk, Mlqa: Evaluating cross-lingual extractive question answering, in: *ACL 2020, 2020*, pp. 7315–7330.
- [43] S. S. Aluru, B. Mathew, P. Saha, A. Mukherjee, Deep learning models for multilingual hate speech detection, arXiv preprint arXiv:2004.06465 (2020).
- [44] V. Zhebel, D. Zubarev, I. Sochenkov, Different approaches in cross-language similar documents retrieval in the legal domain, in: *International Conference on Speech and Computer*, Springer, 2020, pp. 679–686.
- [45] M. Chidambaram, Y. Yang, D. Cer, S. Yuan, Y.-H. Sung, B. Strope, R. Kurzweil, Learning cross-lingual sentence representations via a multi-task dual-encoder model, 2019. arXiv:1810.12836.
- [46] P. Indyk, R. Motwani, Approximate nearest neighbors: towards removing the curse of dimensionality, in: *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, 1998, pp. 604–613.
- [47] H. Westermann, J. Šavelka, V. R. Walker, K. D. Ashley, K. Benyekhlef, Data-centric machine learning: Improving model performance and understanding through dataset analysis, in: *Legal Knowledge and Information Systems: JURIX 2021: The Thirty-fourth Annual Conference*, Vilnius, Lithuania, 8-10 December 2021, volume 346, IOS Press, 2022, p. 54.
- [48] J. E. Cabral, A. Chavan, T. M. Clarke, J. Green, Using technology to enhance access to justice, *Harvard Journal of Law & Technology* 26 (2012) 241–324.
- [49] OpenAI, Gpt-4 technical report, 2023. arXiv:2303.08774.
- [50] J. Savelka, K. Ashley, M. Gray, H. Westermann, H. Xu, Can gpt-4 support analysis of textual data in tasks requiring highly specialized domain expertise?, in: *ASAIL’23: 6th Workshop on Automated Semantic Analysis of Information in Legal Text*, 2023.
- [51] J. Savelka, Unlocking practical applications in legal domain: Evaluation of gpt for zero-shot semantic annotation of legal texts, arXiv preprint arXiv:2305.04417 (2023).