Agent-Specific Deontic Modality Detection in Legal Language

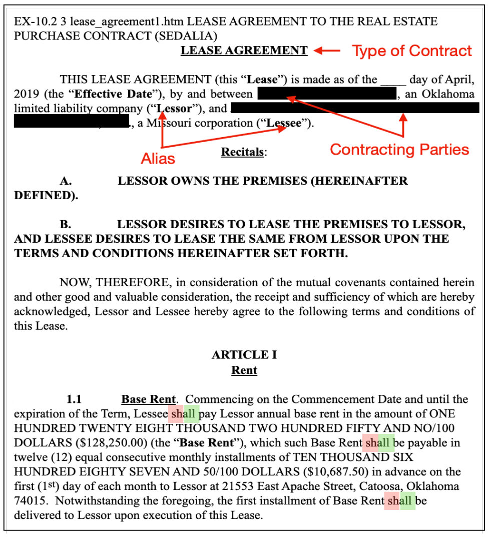
Abhilasha Sancheti[†‡], **Aparna Garimella**[‡], **Balaji Vasan Srinivasan**[‡], **Rachel Rudinger**[†]
[†]University of Maryland, College Park
[‡]Adobe Research
{sancheti, rudinger}@umd.edu
{garimell, balsrini}@adobe.com

Abstract

Legal documents are typically long and written in *legalese*, which makes it particularly difficult for laypeople to understand their rights and duties. While natural language understanding technologies can be valuable in supporting such understanding in the legal domain, the limited availability of datasets annotated for deontic modalities in the legal domain, due to the cost of hiring experts and privacy issues, is a bottleneck. To this end, we introduce, LEXDEMOD, a corpus of English contracts annotated with deontic modality expressed with respect to a contracting party or agent along with the modal triggers. We benchmark this dataset on two tasks: (i) agent-specific multi-label deontic modality classification, and (ii) agent-specific deontic modality and trigger span detection using Transformer-based (Vaswani et al., 2017) language models. Transfer learning experiments show that the linguistic diversity of modal expressions in LEXDEMOD generalizes reasonably from lease to employment and rental agreements. A small case study indicates that a model trained on LEXDEMOD can detect red flags with high recall. We believe our work offers a new research direction for deontic modality detection in the legal domain[1].

1 Introduction

A contract is a legal document executed by two or more parties. To sign a contract (*e.g.*, lease agreements, terms of services, privacy policies, EULA, etc.), it is important for these parties to precisely understand their obligations, entitlements, prohibitions, and permissions as described in the contract. However, for a layperson, understanding contracts can be difficult due to their length and the complexity of legalese used. Therefore, a layperson often signs agreements without even reading them (Cole, 2015; Obar and Oeldorf-Hirsch, 2020).



EX-10.2 3 lease_agreement1.htm LEASE AGREEMENT TO THE REAL ESTATE PURCHASE CONTRACT (SEDALIA)

LEASE AGREEMENT

THIS LEASE AGREEMENT (this "**Lease**") is made as of the ____ day of April, 2019 (the "**Effective Date**"), by and between [redacted], an Oklahoma limited liability company ("**Lessor**"), and [redacted], a Missouri corporation ("**Lessee**").

Recitals:

A. LESSOR OWNS THE PREMISES (HEREINAFTER DEFINED).

B. LESSOR DESIRES TO LEASE THE PREMISES TO LESSOR, AND LESSEE DESIRES TO LEASE THE SAME FROM LESSOR UPON THE TERMS AND CONDITIONS HEREINAFTER SET FORTH.

NOW, THEREFORE, in consideration of the mutual covenants contained herein and other good and valuable consideration, the receipt and sufficiency of which are hereby acknowledged, Lessor and Lessee hereby agree to the following terms and conditions of this Lease.

ARTICLE I
Rent

1.1 **Base Rent**. Commencing on the Commencement Date and until the expiration of the Term, Lessee shall pay Lessor annual base rent in the amount of ONE HUNDRED TWENTY EIGHT THOUSAND TWO HUNDRED FIFTY AND NO/100 DOLLARS ($128,250.00) (the "**Base Rent**"), which such Base Rent shall be payable in twelve (12) equal consecutive monthly installments of TEN THOUSAND SIX HUNDRED EIGHTY SEVEN AND 50/100 DOLLARS ($10,687.50) in advance on the first (1st) day of each month to Lessor at 21553 East Apache Street, Catoosa, Oklahoma 74015. Notwithstanding the foregoing, the first installment of Base Rent shall be delivered to Lessor upon execution of this Lease.

Figure 1: Sample contract[3] indicating the terminologies used to refer to the elements of a contract. 'shall' triggers obligation for Lessee and entitlement for Lessor. Contracting party or agent is referred to via an "alias" (such as Lessor or Lessee) throughout the contract.

Having a system which can provide an "at a glance" summary of obligations, entitlements, prohibitions, and permissions to a contracting party (henceforth, "agent"[2]), will be of great help not only to the agents but also to legal professionals for contract review. While existing language processing and understanding systems can be used for legal understanding, limited availability of annotated datasets in the legal domain due to the cost of hiring experts and privacy issues is a bottleneck. Furthermore, the highly specialized lexical and syntactic features of legalese make it difficult to directly apply systems trained on data from other linguistic domains (*e.g.*, news) to the legal domain.

For an "at a glance" summary of contracts, we first need to identify the obligations, entitlements, prohibitions, and permissions present in the contract for a given agent. Deontic modality is fre-

[1]The code and data are available at `https://github.com/abhilashasancheti/LexDeMod`

[2]Not related to semantic roles.

[3]Party names redacted for anonymity purpose.

Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 11563–11579
December 7-11, 2022

quently used in contracts to express such obligations, entitlements, permissions, and prohibitions of agents (Ballesteros-Lintao et al., 2016). For instance, 'shall', 'shall not', and 'may' is used to express 'obligation/entitlement', 'prohibition', and 'permission' respectively in example (1) below.

(1) a. Tenant **shall** pay the rent to the Landlord.
   b. Landlord **shall not** obtain financing or enter into any agreement affecting the Property.
   c. Landlord **may** continue this Lease in effect after Tenant's abandonment and recover Rent as it becomes due.

(2) a. Tenant **agrees** to pay the rent.
   b. Landlord **is responsible for** maintaining the structural soundness of the house.

However, existing works for identifying such deontic modality types (henceforth "deontic types") either use rule-based (Wyner and Peters, 2011; Peters and Wyner, 2016; Dragoni et al., 2016; Ash et al., 2020) or data-driven (Neill et al., 2017; Chalkidis et al., 2018) approaches, which cannot be directly used for our purpose. This is because rule-based approaches are not robust as they do not (in practice) capture the rich linguistic variety (*e.g.*, use of non-modal expressions in (2)) and ambiguity of modal expressions (*e.g.*, 'shall' in (1a)). Furthermore, annotated datasets used in the data-driven approaches do not consider multiple deontic types for a sentence and their association with the agent (*e.g.*, (1a) is an instance of 'obligation' for the Tenant and an 'entitlement' for the Landlord). Although, Funaki et al. (2020) introduced a corpus with annotations for rights, obligations, and associated agents, it does not cover all the deontic types. Moreover, different corpora consider different deontic types, lacking an accepted standard.

In this work, we address these issues through the following **contributions**: (a) we present a linguistically-informed taxonomy for annotating deontic types in the legal domain, and use the taxonomy to build a corpus (LEXDEMOD; §3) of English contracts with two types of annotations: (i) all deontic types expressed in a sentence with respect to an agent, and (ii) spans of modal *triggers*, *i.e.*, expressions (*e.g.*, bold-faced phrases in examples (1) and (2)) that evoke the modal meaning; (b) we benchmark the corpus on two tasks: (i) agent-specific multi-label deontic modality classification (§6), and (ii) agent-specific deontic modality and trigger span detection (§7) using state-of-the-art Transformer (Vaswani et al., 2017) models, and (c) we perform transfer learning experiments (§8) to investigate the generalizability of diverse modal expressions in LEXDEMOD and a case study to detect red flags (§9) in lease agreements.

2 Related Work

2.1 NLP in the Legal Domain

Prior works have investigated a number of tasks in legal NLP domain including legal judgement prediction (Aletras et al., 2016; Luo et al., 2017; Zhong et al., 2018; Chen et al., 2019; Chalkidis et al., 2019), legal entity recognition and classification (Cardellino et al., 2017; Chalkidis et al., 2017; Angelidis et al., 2018), legal question answering (Duan et al., 2019; Zhong et al., 2020), and legal summarization (Hachey and Grover, 2006; Bhattacharya et al., 2019; Manor and Li, 2019). While legal NLP covers a wide range of tasks, limited efforts have been made for contract review despite it being one of the most time-consuming and tedious tasks. Leivaditi et al. (2020) introduced a benchmark for lease contract review for detecting named entities and red flags. Hendrycks et al. (2021) introduced a large expert-annotated dataset and Tuggener et al. (2020) a large semi-automatically annotated dataset for provision type classification across a variety of contract types. However, these datasets do not contain deontic type annotations which the focus of this work.

2.2 Rights and Obligation Extraction

Existing works either propose rule-based methods (Wyner and Peters, 2011; Peters and Wyner, 2016) or use a combination of NLP approaches such as syntax and dependency parsing (Dragoni et al., 2016) for extracting rights and duties from legal documents such as Federal code regulations, European directives or customer protection codes. Another line of works (Bracewell et al., 2014; Neill et al., 2017; Chalkidis et al., 2018) use machine learning and deep learning approaches to predict deontic types with the help of small datasets. However, rule-based approaches are not robust due to the rich linguistic variety and ambiguity of modal expressions, and the annotated datasets do not consider multiple deontic types for a sentence and their association with the agents which is important for contract understanding. Matulewska (2010) analyzed contracts from different countries and types considering fine-grained deontic modalities covered in them but only considers obligation, permission and prohibition with temporal constraints. Ash et al. (2020) propose a rule-based unsupervised approach to identify deontic types with respect to an agent and compute statistics for rights and duties for an agent. However, rule-based approaches have limitations as mentioned above. Recently, Funaki et al. (2020) curate an annotated corpus of contracts for recognizing rights and obligations along with the agents using LegalRuleML (Athan et al., 2013). However, the corpus is not publicly available, does not annotate for modal triggers, and does not cover all the deontic types expressed in a contract.

2.3 Modality Annotation and Detection

Modality refers to the linguistic ability to describe alternative ways the world could be and is commonly expressed by modal auxiliaries such as *shall*, *will*, *must*, *can*, and *may*. Existing studies have proposed various modality annotation schemas for Portuguese (Hendrickx et al., 2012; Avila et al., 2015) and applied (Quaresma et al., 2014) it to build machine learning models to identify the deontic types. However, it does not cover all the deontic types and restrict the identification to three modal auxiliaries. While Athan et al. (2013) and Nazarenko et al. (2018) propose XML-based annotation schema to formally represent legal text in English and highlight the various interpretive issues that arose during the annotation, it does not consider trigger annotation. Although Rubinstein et al. (2013) and Pyatkin et al. (2021) consider trigger and modality type (not restricted to modal auxiliaries) annotations at different levels of granularity, fine-grained deontic types as well as association with the agent is not considered. As different studies consider different deontic types lacking an accepted standard, we present a linguistically-informed taxonomy for annotating deontic types and their triggers.

3 LexDeMod Dataset Curation

We first describe the dataset source (§3.1) followed by pre-processing (§3.2), annotation protocol (§3.3), and the quantitative and qualitative analysis (§3.4) of the collected dataset.

3.1 Dataset Source

We use the contracts available in the LEDGAR corpus (Tuggener et al., 2020) which comprises material contracts (Exhibit-10), such as agreements (*e.g.*, shareholder/employment/lease/non-disclosure), crawled from Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. EDGAR is maintained by the U.S. Securities and Exchange Commission (SEC[4]). The documents filed on SEC are public information and can be redistributed without a further consent.[5]

3.2 Contract Pre-processing

The raw contracts in LEDGAR are available in html format. We extract all the paragraphs (henceforth, "provisions") from the html (identified by <p> or <div> tags) of a contract, and heuristically filter the provisions defining any terminologies (identified by presence of phrases such as 'shall mean', 'means', 'shall have the meaning', 'has the meaning', etc.). As contracts have a hierarchical structure (*e.g.*, bullets and sub-bullets), we prepend (see A.1) the higher level context with the lower level (*e.g.*, combining sub-bullets with its context in the main bullet). After this, we heuristically extract the type of the contract (*e.g.*, lease or employment contract) and the alias (*e.g.*, "Lessee" in Figure 1) used to refer the contracting parties from the content of the contracts.

Contract Type Extraction. We heuristically scan the first 20[6] provisions to identify the type of the contract using regular expressions (all uppercase characters and presence of 'AGREEMENT').

Agent Alias Extraction. Agent in a contract can be either a person or a company. Therefore, we scan the first 20 provisions of a contract to find company mentions using lexnlp (Bommarito II et al., 2021)[7] and named entities with 'person' tag using spaCy (Honnibal et al., 2020) library. We then use regular expression (alias is mentioned in parenthesis (see Figure 1) following the agent mention) to extract the alias used to refer to the found agents in the provisions. For each type of contract, we manually select the most frequently occurring aliases extracted after using the regular expression.

We collect all the sentences of provisions belonging to a contract wherein alias for an agent is found. We posit that if a sentence does not contain an alias, then deontic type is not expressed for an agent. For instance, '*Any such month-to-month tenancy shall be subject to every other term, covenant and*

[4] https://www.sec.gov/

[5] https://www.sec.gov/privacy.htm#dissemination

[6] We found that the structure of contracts is not fixed and table of contents sometimes precedes the actual contract.

[7] lexnlp was better at extracting companies than spaCy.

Deontic Type	Description
Obligation (Obl)	Agent is required to have or do something
Entitlement (Ent)	Agent has the right to have or do something
Prohibition (Pro)	Agent is forbidden to have or do something
Permission (Per)	Agent is allowed to have or do something
No Obligation (Nobl)	Agent is not required to have or do something
No Entitlement (Nent)	Agent has no right to have or do something

Table 1: Taxonomy[8] for deontic type.

agreement contained herein.' is a rule and does not specifically mention any deontic type for an agent.

3.3 Annotation Protocol

Annotation task description. We propose agent-specific deontic modality detection tasks that address the following issues: (i) non-robustness of rule-based extraction of rights and duties as it cannot capture the rich linguistic variety and ambiguity of modal expressions; (ii) lack of standard taxonomy for annotating fine-grained deontic types; (iii) non-association of deontic type with the agent during annotation, and (iv) considering deontic type detection as a single class classification task. Consider, for instance, the following:

(3) a. [*Tenant*] Tenant **shall**$_{obl}$ pay the rent to the Landlord and **may**$_{per}$ use the parking space.
 b. [*Landlord*] Tenant **shall**$_{ent}$ pay the rent to the Landlord and may use the parking space.

In these examples, the words in bold evoke the modal expression, which we call a *trigger*. For Tenant as the [*Agent*], an obligation (obl) and a permission (per) are expressed in the sentence (3a), and an entitlement (ent) for the Landlord (3b).

Our data collection is performed via crowdsourcing on Amazon Mechanical Turk (AMT). We ask the workers to provide two types of annotations for each sentence with respect to an agent (referred to via an alias): (i) select all the deontic types expressed, and (ii) select trigger word(s) (as span) for each selected deontic type. If a sentence contains more than one agent, we duplicate it to get separate annotations with respect to each agent so that the workers focus their understanding with respect to one agent at a time. This task design choice helps in better estimation of the time taken to do each HIT (Human Intelligence Task) as the number of agent mentions in a sentence can vary. This also simplifies the custom annotation interface (see Figure 6) built to get the annotations. Detailed guidelines for annotation are provided in A.2 (Figure 5).

Taxonomy for deontic type. We base our taxonomy for deontic type annotation (Table 1) on the deontic logic theory of Von Wright (1951). Von Wright's categorical modals are best suited for legal contracts which talk about rights and duties of contracting parties (Ballesteros et al., 2020; Matulewska, 2010) than other categorizations (Chung, 1985; Palmer, 2001; Jespersen, 2013) which are not found in contracts (*e.g.*, desiderative, hortative). We also include no-obligation and no-entitlement categories to cover all possible modalities which were found on manual inspection of contracts.

Annotation process and requirements. As legalese is syntactically complex and difficult to understand, the annotation task is quite intricate in nature. To ensure that the workers properly understand the task, we first conduct a qualification test which explains the task with the help of right and wrong annotation examples along with explanations and contains 10 questions. The qualification test is open to workers with $\geq 95\%$ approval rate and $\geq 1,000$ approved HITs. Finally, we select 25 workers who answered all the qualification questions (details in A.3) correctly.

The main annotation task consists of 12 sentences per HIT, including 2 quality check questions to ensure workers provide good annotations. We publish 3 pilot HITs, with revised guidelines in each of them. We also manually check the annotations (selected randomly) to ensure quality and provide feedback to the workers. We observe a learning curve for the task and considerable variation in the time taken per HIT (7.5 ± 1.5 mins). After the pilots, the annotations were majorly performed by 3 workers. We publish a batch of 50 HITs with 3 annotations for each HIT from the 3 workers. As we found good inter-annotator agreement between the 3 workers (see §3.4), we collect only one annotation per HIT for the remaining HITs to get more sentences annotated within reasonable time.

3.4 Annotated Dataset Statistics and Analysis

Each contract contains $202.6(\pm 162.4)$ provisions on average (standard deviation in parentheses), with $2.2(\pm 1.7)$ sentences per provision; each contract consists of $306.4(\pm 235.8)$ sentences on an average. Among these, $75.8(\pm 14.4)\%$ of sentences per contract have atleast one agent mentioned in them, with an average length of $65(\pm 47)$.

We collect a total of $8,230$ trigger span annota-

[8]We also provide an additional option 'None' in case none of the deontic types is expressed or it is a rule.

Split	#Sent.	#Spans	Obl	Ent	Pro	Per	Nobl	Nent	None
Train	4282	5279	1841	1231	343	289	265	239	1071
Dev	330	421	176	86	20	18	21	22	78
Test	1777	1952	575	418	64	167	101	88	539

Table 2: Dataset Statistics.

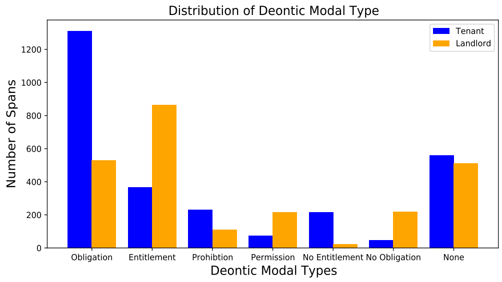


Figure 2: Distribution of deontic type with respect to Tenant and Landlord for lease agreements.

tions for 7,092 sentences from 23 lease contracts after considering HITs for which both the quality questions are correctly answered. For duplicate sentences, we retain those annotations that are in-line with one of the authors (and discard 14.1% of duplicated ones; a few examples are provided in in A.4). The test set comprises of sentences from 5 contracts including those for which we have 3 annotations per sentence, and rest of the sentences are divided into train and development sets such that the sentences from the same contract belong to the same set. We combine the 3 annotations for a subset of sentences in the test set using majority voting[9] for deontic type and by taking a union[10] of annotated trigger spans for the majority deontic types. The average inter-annotator agreement for each deontic type computed with Krippendorff's α (Krippendorff, 2018) is substantial ($\alpha = 0.65$) given the complexity of the task (see Table 7 for type-wise agreement). For trigger span annotation, the token-level inter-annotator agreement for the majority deontic types for a sentence is also substantial ($\alpha = 0.71$). The fine-grained dataset statistics after filtering and resolving disagreements are presented in Table 2.

Qualitative analysis. Figure 2 shows the distribution of annotated spans[11] over deontic types with respect to each agent (Landlord and Tenant for lease agreements[12]) in the train set. Interestingly, tenants have more obligations and prohibitions, and fewer entitlements or permissions than landlords. 17.3% of the sentences have multiple trigger annotations, 48.6% of these sentences express multiple deontic types. 24.8% of the sentences do not express any deontic type with respect to a given agent. The dataset contains 383 unique triggers across all the deontic types. Table 3 lists the top 10 triggers for each deontic type, and Figure 3 shows the frequency-based wordcloud of the annotated triggers. 'Shall' constitutes 44.6% of the annotated triggers used to express not only obligation but entitlement, no-obligation, and no-entitlement as well. Prohibitions may be expressed using negation words (14.9%) between the context (*e.g.*, 'neither lessor nor lessee may') of a sentence. While modal auxiliaries (*e.g.*, *shall*, *will*, *may*) are more frequently used, 45.2% of the total unique triggers are non-modal expressions (*e.g.*, *agrees*, *rep-*

[9]We discard the sentence in case no majority is found.

[10]Union was performed in 15.54% of sentences and we manually corrected 9.4% of these sentences where union lead to incorrect triggers.

[11]For 'None' type, no span is annotated so the bar denotes the number of sentences labeled as none.

Type	Top 10 triggers
Obl	shall, will, agrees, agree, acknowledges, acknowledge, represents and warrants, shall be responsible for, undertakes, will be responsible for
Ent	shall, will, agrees, shall have the right to, shall be entitled to, represents and warrants, acknowledges, waives no rights, shall not, retains all other rights, will be entitled to
Pro	shall not, will not, may not, nor shall, not to be, neither lessor nor lessee may, in no event shall, nor will, will not allow, nor may
Per	may, is permitted, will allow, has the right, shall, or at landlord's option, shall be permitted to, shall be allowed
Nobl	shall not be liable, shall not be obligated to, shall not be required to, shall, shall have no obligation to, in no event shall landlord be obligated to, waives, shall not, shall have no liability
Nent	shall, shall have no right to, waives no rights, shall not, shall have no obligation to, waives, shall not be required, shall not be obligated, waive the right, shall not have the right to

Table 3: Top 10 triggers for each deontic type in decreasing order of frequency.

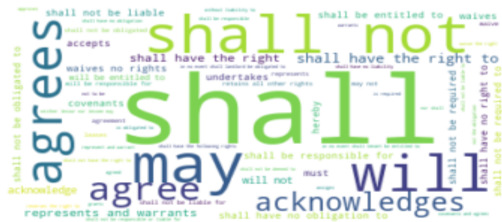


Figure 3: Frequency-based wordcloud of all the triggers.

resents

[12]We add the statistics for tenant, subtenant & lessee under Tenant, and landlord, sublandlord & lessor under Landlord.

resents) covering 20.3% of the annotated trigger spans. This shows that LEXDEMOD covers a wide variety of linguistic expressions of deontic modality in legalese, not restricted to modal auxiliaries. Annotated samples from the dataset are provided in Table 15 in A.9.

4 Proposed Benchmarking Tasks

Having established the rich variety and coverage of linguistic expressions for deontic modality in LEXDEMOD, we benchmark the corpus on the proposed two tasks defined below:
(i) Agent-specific multi-label deontic modality classification. This task aims at predicting all the deontic types expressed in a sentence with respect to an agent. We pose this as a multi-label classification task conditioned on a sentence and an agent.
(ii) Agent-specific deontic modality and trigger span detection. This task aims at identifying both the deontic type and corresponding triggers. We pose this as a token classification task. Every token in the corpus is assigned a BIOS tag if it belongs to a modal trigger, which is appended with a suffix indicating its deontic type. For instance, Tenant$_{O}$ *is*$_{B-OBL}$ *responsible*$_{I-OBL}$ *for*$_{I-OBL}$ paying$_{O}$ the$_{O}$ rent$_{O}$, where subscripts denote the BIOS tags.

For both the tasks, agent is conditioned using special tokens added at the beginning of a sentence. This simple strategy has been successfully used previously for controlled text generation tasks (Sennrich et al., 2016; Johnson et al., 2017; Rudinger et al., 2020; Sancheti et al., 2022).

5 Benchmarking Setup

We experiment with various pre-trained language models (PLMs) (Devlin et al., 2019; Liu et al., 2019), which have shown state-of-the-art performance on natural language understanding tasks, to study their performance on our proposed tasks. We fine-tune these models for both the tasks on binary cross-entropy loss for 20 epochs each with a batch size of 8, and maximum sequence length of 256 using HuggingFace's Transformers library (Wolf et al., 2020). The model(s) with the best macro-F1 score on the dev set is used to report results on the test set. Further implementation details are in A.6. We also partition the data according to the agent being conditioned to assess the performance of the trained models with respect to each agent.

6 Benchmarking Multi-label Classification

Comparison models. We experiment with three kinds of approaches for the agent-specific multi-label deontic modality classification task.
(1) Majority class predicted for each agent.
(2) Rule-based. We implement a rule-based approach similar to the one described in Ash et al. (2020) with additional conditioning on the agent. It searches for the presence of pre-defined modal triggers for a deontic type and associates it with the agent using dependency tags (*e.g.*, *nsubj*, *aux* or *agent*). We use spacy to tokenize each sentence and obtain a dependency parse. More details in A.5.
(3) Fine-tuning PLMs. We fine-tune the following PLMs differing in size and domain of data used for pre-training: (i) BERT-base-uncased (**BERT-BU**); (ii) RoBERTa-base (**RoBERTa-B**); (iii) RoBERTa-large (**RoBERTa-L**), and (iv) recently introduced Contract-BERT-base-uncased (**C-BERT-BU**) model (Chalkidis et al., 2020) which has been pre-trained on US contracts from the EDGAR library.

All the above models are trained assuming trigger span information is not available and full context (*i.e.*, sentence) is used. To understand the importance of **A**gent conditioning, **C**ontext, and **T**rigger for this task, we additionally train the following models: (i) **No-agent** where special token for agent is not used during training; (ii) **ACT-Masked** wherein everything in the context except the trigger span is masked using `[MASK]` token to hide the context but retain the positional information of the trigger; (iii) **AT** wherein only the tokens belonging to a trigger are used and multiple triggers are separated using a special token (*e.g.*, `[SEP]` or `</s>`), and (iv) **ACT** wherein all the triggers are appended at the end of the context separated by a special token (*e.g.*, `[SEP]` or `</s>`).
Evaluation measures. We report macro-averaged Precision, Recall, and F1 scores across all the types, calculated using Sklearn library (Pedregosa et al., 2011). We also report the Accuracy of predicting all the classes correctly for a sentence.
Results and analysis. We report the results for multi-label classification task in Table 4. While rule-based approach has better F1 score than majority type prediction for each agent, Transformer-based models outperform these baselines indicating their ability to better capture the linguistic diversity of expressing deontic modals. As expected, Rule-

Model	Accuracy	Precision	Recall	F1
Majority	39.53/28.66/34.38	6.49/5.23/11.72	14.29/14.29/21.09	8.92/7.66/15.03
Rule-based	61.32/50.54/56.22	**81.81**/75.21/**80.04**	46.66/45.54/46.64	50.13/46.16/48.76
BERT-BU	74.07/70.79/72.52	73.68/74.44/75.48	75.84/71.02/77.17	**78.81**/71.18/75.61
RoBERTa-B	75.53/71.42/73.59	73.54/72.17/74.48	78.39/72.88/78.31	74.90/71.91/75.66
C-BERT-BU	77.50/73.25/75.48	76.63/76.22/77.52	**80.47**/71.54/78.81	77.95/72.34/77.67
RoBERTa-L	**78.28**/**75.03**/**76.74**	75.05/**77.69**/77.30	79.59/**75.21**/**79.11**	76.71/**76.00**/**77.88**
RoBERTa-L-No-agent	51.28/47.45/49.46	57.09/53.79/58.32	65.01/55.52/60.08	55.75/52.56/57.53
RoBERTa-L-ACT-Masked	81.52/72.02/77.02	76.39/71.31/81.46	71.22/65.74/75.90	84.25/76.29/80.42
RoBERTa-L-AT	84.72/76.22/80.70	79.84/73.60/82.29	78.00/72.96/82.47	87.58/80.58/84.20
RoBERTa-L-ACT	**91.66**/**88.62**/**90.23**	**88.44**/**85.40**/**89.48**	**88.10**/**84.43**/**89.21**	**93.38**/**91.38**/**92.42**

Table 4: Evaluation results for agent-specific multi-label deontic modality classification task. Scores for **Tenant/-Landlord/Both** are averaged over 3 different seeds. BU, B, L, A, C, and T denote base-uncased, base, large, agent, context (sentence), and trigger, respectively. Development set results are provided in Table 9 in Appendix.

based approach has the highest overall precision but low recall due to the impossibility of enumerating all the rules. While C-BERT-BU, which is pre-trained on contracts, performs better than BERT-BU and RoBERTa-B, interestingly it achieves comparable F1 score to RoBERTa-L. This indicates that improvements from domain-specific pre-training may also be achieved with larger model size and more training data.

As RoBERTa-L performs the best on this task, we report the results for variants of this model to understand the importance of agent conditioning, context, and trigger, in the last block of Table 4. The performance of RoBERTa-L-No-agent, trained without agent conditioning, significantly drops as compared to RoBERTa-L, indicating the importance of agent conditioning during training and association of agent with the modality expressed in a sentence. Using trigger information during training (RoBERTa-L-ACT) significantly improves the performance over RoBERTa-L across all the measures, showing that triggers are indicative of specific deontic type. Higher scores for RoBERTa-L-AT than RoBERTa-L-ACT-Masked show that positional information of trigger span adds noise to the representations learned by the model. Further, context is also important for identifying deontic type, as all the metric scores drop when context is masked (RoBERTa-L-ACT-Masked) or not used (RoBERTa-L-AT) during training as compared to using all the information (RoBERTa-L-ACT).

Manual inspection of deontic type-wise (Table 11) performance reveals that permission is the easiest, while no-entitlement and prohibition are the hardest to identify. This can be due to the use of limited variety in expressing permissions (majorly 'may'), while use of negation within context for expressing prohibitions which makes it harder to identify. For tenant, obligation is identified more accurately than entitlements (vice-versa for landlord) as expected due to higher frequency of obligations for tenant and entitlements for landlord.

Figure 4a shows the trend for RoBERTA-L's F1 score as train data size varies indicating that the rate of increase in F1 decreases with additional data.

7 Benchmarking Trigger Span Detection

Comparison models. We experiment with three kinds of approaches for the agent-specific deontic modality and trigger span detection task.

(1) Majority. 'Shall' is the most used trigger as shown in Figure 3 and is used to express obligations for Tenant while entitlements for Landlord. This baseline tags each occurrence of 'shall' with S-OBL for tenant or S-ENT for landlord as agent.

(2) Rule-based. We tag occurrences of pre-defined modal triggers in a sentence with the deontic type predicted using the rule-based approach (§6).

(3) Fine-tuning PLMs. We fine-tune the same models as described in §6 on a token classification task to predict the BIOS tags. Additionally, we train a 'No-agent' model to verify the importance of agent conditioning.

Evaluation measures. We report macro-averaged Precision, Recall, and F1 scores, calculated using seqeval library (Nakayama, 2018). We also report the Accuracy of predicting the BIOS tags for a sentence. Following (Pyatkin et al., 2021), we report these metrics in labeled (both deontic type and trigger span considered) and unlabeled (only trigger span without deontic type is considered) settings.

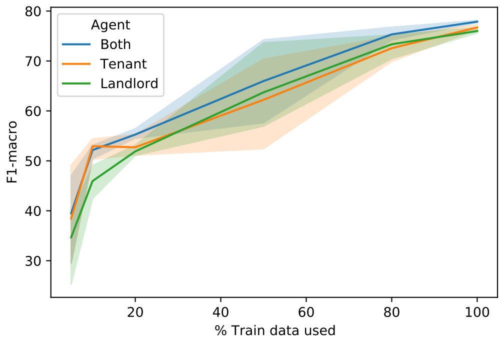
Results and Analysis. Labeled and unlabeled metric scores for trigger span detection task are reported in Table 5. RoBERTa-L has the best labeled F1 score which evaluates for both trigger detection and its correct deontic type identification. Similar

Model	Labeled Accuracy	Labeled Precision	Labeled Recall	Labeled F1	Unlabeled Accuracy	Unlabeled Precision	Unlabeled Recall	Unlabeled F1
Majority	97.16/96.85/97.01	5.58/4.40/9.98	10.89/10.31/16.00	7.38/6.17/12.27	97.28/97.04/97.17	41.30/39.59/40.51	50.61/42.86/46.76	45.48/41.16/43.41
Rule-based	97.85/97.67/97.76	**77.42/76.68/79.66**	32.42/33.24/33.61	40.00/39.30/40.58	97.89/97.73/97.81	**72.59/73.97/73.22**	40.07/35.34/37.73	51.64/47.83/49.80
BERT-BU	98.45/98.36/98.41	53.04/56.11/56.48	58.49/59.05/61.97	55.11/**57.01**/58.80	**98.55/98.52/98.53**	68.87/69.92/69.38	76.07/75.22/75.65	72.29/72.46/72.38
RoBERTa-B	98.40/98.24/98.32	53.03/52.43/55.57	63.65/**59.63/64.00**	57.08/55.31/53.91	98.49/98.41/98.46	68.99/67.44/68.22	78.18/**75.36/76.78**	73.27/71.14/72.22
C-BERT-BU	98.44/**98.39/98.42**	53.46/54.70/57.08	60.76/57.37/62.42	56.45/55.68/59.31	98.52/**98.52**/98.52	69.49/70.85/70.15	76.89/74.26/75.59	72.99/**72.52**/72.76
RoBERTa-L	**98.45**/98.27/98.37	54.99/55.58/57.37	**65.55**/58.88/63.74	**59.19**/56.71/**60.04**	98.54/98.39/98.47	69.78/69.56/69.68	**79.16**/74.35/**76.78**	**74.18**/71.88/**73.06**
RoBERTa-L-NA	97.64/97.75/97.69	32.45/36.71/36.36	48.92/43.93/46.79	34.68/38.72/39.45	98.26/98.22/98.24	61.73/64.63/63.12	75.21/72.84/74.03	67.79/68.46/68.11

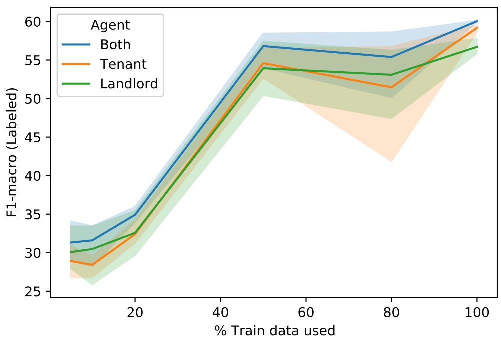
Table 5: Evaluation results for agent-specific modal trigger span detection task. Macro-averaged Precision, Recall and F1 scores are presented for **Tenant/Landlord/Both**. Scores are averaged over 3 different seeds. BU, B, L, and NA denote base-uncased, base, large, and no-agent respectively. Dev set results are shown in Table 10 in Appendix.

to the classification task, Rule-based approach outperforms other models on precision however, lags behind in recall for the same reason. Size of the model (RoBERta-L) is instrumental than domain knowledge of C-BERT-BU. Consistently higher unlabeled scores, indicate that models are able to identify the trigger words. However, associating triggers with the correct deontic type is a harder task, owing to the multiple deontic types that a trigger can be used to express (*e.g.*, shall in Table 3) them. Similar to the classification task, importance of agent conditioning is evident from the last row with significant drop in F1 scores (even lower than Rule-based approach in Labeled score). Higher accuracy scores are due to the majority tokens being labeled as 'O'. Trends with dataset size variation are shown in Figure 4b. Manual analysis of deontic type-wise span detection (Table 11) reveals that prohibition, no-entitlement, and no-obligation are hard to identify. Similar trends were observed for tenant and landlord as in §6.

These results show that identification of both triggers and associating it with the deontic type is a difficult task owing to the linguistic variety of expressions used in legal language.



(a) Classification



(b) Trigger Span Detection

Figure 4: RoBERTA-L's performance with varying train dataset size for the two tasks.

8 Beyond Lease Contracts

To investigate if the diverse linguistic expressions used for expressing deontic modality is specific to a contract type, we collect annotations via AMT using the same annotation protocol (§3.3) for: (1) 470 sentences from 3 employment contracts in the LEDGAR corpus, and (2) 154 sentences from 4 rental agreement templates freely available at PandaDoc.[13] We evaluate the performance of the best model (RoBERTa-L) for both the tasks on these sentences and report the results in Table 6. We observe a performance drop (more for employment contracts than rental agreements) when compared to model's performance on lease agreements, although it is significantly better than the rule-based approach demonstrating the non-robustness of rule-based approach towards diverse linguistic expressions. This drop is more prominent for employment contracts due to the lease-specific agent conditioning (*e.g.*, tenant) used during training while commonly occurring agents in employment contracts are employee, employer, etc.

To account for this, we additionally train models with anonymizing the agent mentions in the dataset in two ways: (i) **RoBERTa-L-AR**– all occurrences of an agent are replaced with the same token (*e.g.*,

[13] https://www.pandadoc.com/ We chose rental agreements as they are commonly used by layperson than lease contracts from SEC.

Model	Accuracy	Precision	Recall	F1
Multi-label Classification (Rental/Employment)				
Majority	36.36/27.45	11.87/8.80	19.10/15.15	14.46/11.11
Rule-based	41.56/47.45	53.77/64.63	34.54/35.00	33.27/37.22
RoBERTa-L	73.16/48.72	83.08/52.87	63.42/48.90	68.90/48.32
RoBERTa-L-AR	55.19/42.55	56.87/59.29	52.38/46.48	50.66/50.30
RoBERTa-L-ARR	70.35/64.68	76.79/70.05	63.14/64.62	65.89/65.36
Trigger Span Detection (Labeled) (Rental/Employment)				
Majority	96.09/97.37	18.33/4.23	1.90/7.08	3.42/5.30
Rule-based	96.40/97.83	56.25/59.66	23.69/19.65	29.62/27.45
RoBERTa-L	97.48/97.78	49.74/36.80	45.87/37.84	45.58/34.87
RoBERTa-L-AR	97.22/98.15	49.97/48.86	44.43/42.99	44.22/43.42
RoBERTa-L-ARR	97.60/98.38	59.42/53.14	47.83/43.84	49.61/45.47

Table 6: Results for rental/employment contracts.[14]

'a1' for tenant), and (ii) **RoBERTa-L-ARR** –agent is randomly replaced with a token consistent within a sentence. Replacing agent mentions leads to significant improvements for employment contracts in both the tasks, although evaluating these models on rental agreements (see Table 6) and lease data shows (see Table 12) an expected drop in the performance. These experiments show that the linguistic expressions captured by LEXDEMOD are also generalizable to other types of contracts.

9 Case Study: Red flag Detection

To investigate if our agent-specific deontic modality classifier is capable of identifying the red flags annotated by Leivaditi et al. (2020) for lease agreements, we compare the predictions on the dev set from ALeaseBERT, proposed by Leivaditi et al. (2020), and RoBERTa-L model trained on LEXDEMOD dataset. For each sentence in the red flags dataset, we predict the deontic modality with respect to each of the agent alias mentioned in that sentence. If any one of the deontic types is expressed for any of the agents then we consider the prediction as positive otherwise negative. We find that (see Table 14 in A.8) the model trained on LEXDEMOD has high recall and low precision while ALeaseBERT has high precision but low recall for the positive class. Our model was able to predict all the red flags predicted by ALeaseBERT and some additional red flags. This is expected as many permissions or entitlements may not be red flags but may belong to a deontic type. We also found that there were payments related obligations which were predicted as red flags by our model but were not annotated as red flags in the dataset. Therefore, our model could also be used to filter important sentences which could indicate some red flags due to high recall.

[14]Unlabeled scores are provided in Appendix in Table 13.

10 Conclusion and Future Work

We introduced LEXDEMOD for deontic modality detection in the legal domain which consists of diverse linguistic expressions of deontic modality. We propose and benchmark two tasks namely, agent-specific multi-label deontic modality classification, and agent-specific deontic modality and trigger span detection using transformer-based models. While evaluation results are promising, there is substantial room for improvement. We demonstrated the generalizability of the diverse linguistic expressions captured in LEXDEMOD via transfer learning experiments to employment and rental lease agreements. The small case study on red flag detection using our data showed the usability of our dataset. We leave joint-modeling of the two tasks and using these identification models for generating "at a glance" summary of contracts for future work.

11 Limitations

We note a few limitations: (1) Although we demonstrate reasonable generalization to employment agreements, our dataset is limited to lease agreements which may not cover all the linguistic expressions for deontic modality occurring in legal domain. (2) The custom interface built for collecting annotations does not support non-contiguous trigger-span selection which may result in some contract type specific triggers (only for triggers with negation). Future work may consider handling non-contiguous spans and other challenges associated with it (*e.g.*, representing non-contiguous trigger spans for a category in the BIO span). (3) As we focus on identifying agent-specific deontic modalities, we only consider sentences where the agent alias is explicitly mentioned. This helped in simplifying the annotation process and cost efficiency. Therefore, our models may not work well when no agent alias is mentioned in the given sentence. We leave the collection of annotations for sentences not explicitly mentioning agent alias for future work. (4) Our data collection and modeling assume that agent alias is known apriori (for which we perform agent alias extraction) as we focus on the identification task. Extending this work to any other type of agreement will require similar alias extraction method (*e.g.*, employee, employer for employment agreement) or a more sophisticated model to identify the agent implicitly.

12 Ethical Considerations

We are committed to ethical practices and protecting the anonymity and privacy of individuals who have contributed. We ensure that the privacy of the annotators is protected. For annotations, \$7.5/hr was paid per task.

Societal Impact. We recognize and acknowledge that our work carries a possibility of misuse including malicious adulteration of summaries generated by extracting sentences identified by our model and adversarial use of the model to mislead users. Such kind of misuse is common to any prediction model therefore, we strongly recommend coupling any such technology with external expert validation. The purpose of this work is to provide aid to the legal personnel or layperson dealing with legal contracts for a better understanding of the legal documents, and not to replace any experts. As contracts are long documents, identification of sentences that express deontic types can help in significantly reducing the number of sentences to read or highlighting the important parts of the contract which may need more attention.

13 Acknowledgements

We would like to thank Ani Nenkova and the anonymous reviewers for their useful feedback and comments on this work. We acknowledge the support from Adobe Research unrestricted gift funding for this work. The views contained in this article are those of the authors and not of the funding agency.

References

Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. 2016. Predicting judicial decisions of the european court of human rights: A natural language processing perspective. *PeerJ Computer Science*, 2:e93.

Iosif Angelidis, Ilias Chalkidis, and Manolis Koubarakis. 2018. Named entity recognition, linking and generation for greek legislation. In *JURIX*, pages 1–10.

Elliott Ash, Jeff Jacobs, Bentley MacLeod, Suresh Naidu, and Dominik Stammbach. 2020. Unsupervised extraction of workplace rights and duties from collective bargaining agreements. In *2020 International Conference on Data Mining Workshops (ICDMW)*, pages 766–774. IEEE.

Tara Athan, Harold Boley, Guido Governatori, Monica Palmirani, Adrian Paschke, and Adam Wyner. 2013. Oasis legalruleml. In *proceedings of the fourteenth international conference on artificial intelligence and law*, pages 3–12.

Luciana Beatriz Avila, Amália Mendes, and Iris Hendrickx. 2015. Towards a unified approach to modality annotation in portuguese. In *Proceedings of the Workshop on Models for Modality Annotation*.

Miguel Ballesteros, Rishita Anubhai, Shuai Wang, Nima Pourdamghani, Yogarshi Vyas, Jie Ma, Parminder Bhatia, Kathleen McKeown, and Yaser Al-Onaizan. 2020. Severing the edge between before and after: Neural architectures for temporal ordering of events. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5412–5417, Online. Association for Computational Linguistics.

Rachelle Ballesteros-Lintao, Maria Regina P Arriero, Judith Ma Angelica S Claustro, Kristina Isabelle U Dichoso, Selenne Anne S Leynes, Maria Rosario R Aranda, and Jean Reintegrado-Celino. 2016. Deontic meanings in philippine contracts. *International Journal of Legal Discourse*, 1(2):421–454.

Paheli Bhattacharya, Kaustubh Hiware, Subham Rajgaria, Nilay Pochhi, Kripabandhu Ghosh, and Saptarshi Ghosh. 2019. A comparative study of summarization algorithms applied to legal case judgments. In *European Conference on Information Retrieval*, pages 413–428. Springer.

Michael J Bommarito II, Daniel Martin Katz, and Eric M Detterman. 2021. Lexnlp: Natural language processing and information extraction for legal and regulatory texts. In *Research Handbook on Big Data Law*. Edward Elgar Publishing.

David Bracewell, David Hinote, and Sean Monahan. 2014. The author perspective model for classifying deontic modality in events. In *The Twenty-Seventh International Flairs Conference*.

Cristian Cardellino, Milagro Teruel, Laura Alonso Alemany, and Serena Villata. 2017. Legal NERC with ontologies, Wikipedia and curriculum learning. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 254–259, Valencia, Spain. Association for Computational Linguistics.

Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. 2019. Neural legal judgment prediction in English. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4317–4323, Florence, Italy. Association for Computational Linguistics.

Ilias Chalkidis, Ion Androutsopoulos, and Achilleas Michos. 2017. Extracting contract elements. In *Proceedings of the 16th edition of the International Conference on Articial Intelligence and Law*, pages 19–28.

Ilias Chalkidis, Ion Androutsopoulos, and Achilleas Michos. 2018. Obligation and prohibition extraction using hierarchical RNNs. In *Proceedings of the 56th*

Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 254–259, Melbourne, Australia. Association for Computational Linguistics.

Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. LEGAL-BERT: The muppets straight out of law school. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2898–2904, Online. Association for Computational Linguistics.

Huajie Chen, Deng Cai, Wei Dai, Zehui Dai, and Yadong Ding. 2019. Charge-based prison term prediction with deep gating network. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6362–6367, Hong Kong, China. Association for Computational Linguistics.

Sandra Chung. 1985. Tense, aspect and mood. *Language typology and syntactic description*, pages 202–258.

G Marcus Cole. 2015. Rational consumer ignorance: When and why consumers should agree to form contracts without even reading them. *JL Econ. & Pol'y*, 11:413.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Mauro Dragoni, Serena Villata, Williams Rizzi, and Guido Governatori. 2016. Combining nlp approaches for rule extraction from legal documents. In *1st Workshop on MIning and REasoning with Legal texts (MIREL 2016)*.

Xingyi Duan, Baoxin Wang, Ziyue Wang, Wentao Ma, Yiming Cui, Dayong Wu, Shijin Wang, Ting Liu, Tianxiang Huo, Zhen Hu, et al. 2019. Cjrc: A reliable human-annotated benchmark dataset for chinese judicial reading comprehension. In *China National Conference on Chinese Computational Linguistics*, pages 439–451. Springer.

Ruka Funaki, Yusuke Nagata, Kohei Suenaga, and Shinsuke Mori. 2020. A contract corpus for recognizing rights and obligations. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 2045–2053, Marseille, France. European Language Resources Association.

Ben Hachey and Claire Grover. 2006. Extractive summarisation of legal texts. *Artificial Intelligence and Law*, 14(4):305–345.

Iris Hendrickx, Amália Mendes, and Silvia Mencarelli. 2012. Modality in text: a proposal for corpus annotation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 1805–1812, Istanbul, Turkey. European Language Resources Association (ELRA).

Dan Hendrycks, Collin Burns, Anya Chen, and Spencer Ball. 2021. Cuad: An expert-annotated nlp dataset for legal contract review. *arXiv preprint arXiv:2103.06268*.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spacy: Industrial-strength natural language processing in python.

Otto Jespersen. 2013. *The philosophy of grammar*. Routledge.

Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, et al. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Klaus Krippendorff. 2018. *Content analysis: An introduction to its methodology*. Sage publications.

Spyretta Leivaditi, Julien Rossi, and Evangelos Kanoulas. 2020. A benchmark for lease contract review. *arXiv preprint arXiv:2010.10386*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Bingfeng Luo, Yansong Feng, Jianbo Xu, Xiang Zhang, and Dongyan Zhao. 2017. Learning to predict charges for criminal cases with legal basis. *arXiv preprint arXiv:1707.09168*.

Laura Manor and Junyi Jessy Li. 2019. Plain English summarization of contracts. In *Proceedings of the Natural Legal Language Processing Workshop 2019*, pages 1–11, Minneapolis, Minnesota. Association for Computational Linguistics.

Aleksandra Matulewska. 2010. Deontic modality and modals in the language of contracts.

Hiroki Nakayama. 2018. seqeval: A python framework for sequence labeling evaluation. Software available from https://github.com/chakki-works/seqeval.

Adeline Nazarenko, François Levy, and Adam Wyner. 2018. An annotation language for semantic search of legal sources. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

James O' Neill, Paul Buitelaar, Cecile Robin, and Leona O' Brien. 2017. Classifying sentential modality in legal language: a use case in financial regulations, acts and directives. In *Proceedings of the 16th edition of the International Conference on Articial Intelligence and Law*, pages 159–168.

Jonathan A Obar and Anne Oeldorf-Hirsch. 2020. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, 23(1):128–147.

Frank Robert Palmer. 2001. *Mood and modality*. Cambridge university press.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.

Wim Peters and Adam Wyner. 2016. Legal text interpretation: Identifying hohfeldian relations from text. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 379–384, Portorož, Slovenia. European Language Resources Association (ELRA).

Valentina Pyatkin, Shoval Sadde, Aynat Rubinstein, Paul Portner, and Reut Tsarfaty. 2021. The possible, the plausible, and the desirable: Event-based modality detection for language processing. *arXiv preprint arXiv:2106.08037*.

Paulo Quaresma, Amália Mendes, Iris Hendrickx, and Teresa Gonçalves. 2014. Automatic tagging of modality: identifying triggers and modal values. In *Proceedings 10th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation*, pages 95–101. European Language Resources Association.

Aynat Rubinstein, Hillary Harner, Elizabeth Krawczyk, Dan Simonson, Graham Katz, and Paul Portner. 2013. Toward fine-grained annotation of modality in text. In *Proceedings of the IWCS 2013 workshop on annotation of modal meanings in natural language (WAMM)*, pages 38–46.

Rachel Rudinger, Vered Shwartz, Jena D. Hwang, Chandra Bhagavatula, Maxwell Forbes, Ronan Le Bras, Noah A. Smith, and Yejin Choi. 2020. Thinking like a skeptic: Defeasible inference in natural language. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4661–4675, Online. Association for Computational Linguistics.

Abhilasha Sancheti, Balaji Vasan Srinivasan, and Rachel Rudinger. 2022. Entailment relation aware paraphrase generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(10).

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Controlling politeness in neural machine translation via side constraints. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 35–40.

Don Tuggener, Pius von Däniken, Thomas Peetz, and Mark Cieliebak. 2020. LEDGAR: A large-scale multi-label corpus for text classification of legal provisions in contracts. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1235–1241, Marseille, France. European Language Resources Association.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Georg Henrik Von Wright. 1951. Deontic logic. *Mind*, 60(237):1–15.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Adam Wyner and Wim Peters. 2011. On rule extraction from regulations. In *Legal Knowledge and Information Systems*, pages 113–122. IOS Press.

Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. 2018. Legal judgment prediction via topological learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3540–3549.

Haoxi Zhong, Chaojun Xiao, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. 2020. Jecqa: A legal-domain question answering dataset. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9701–9708.

A Appendix

A.1 Combining Bullets

We combine the higher level context (bullets- “parent”) with the lower level context (sub-bullet- “child”) owing to the hierarchical nature of contracts by iterating over the provisions in a contract in sequential order and following the below rules. Combination can be done in two ways: (i) concatenating, and (ii) merging. We find a sub-bullet via regular expression (`^\([ivx]+|^\([a-zA-Z]+ |^[\d.\d]+ `) pattern matching.

- If the child is not a complete sentence (identified by the presence of S in root of constituency parse), parent is a complete sentence, and parent does not contain ‘follow’ or ‘below:’ then remove ‘:’ from the end of parent and append the child (we call this, merging).
- If child starts with a lower case and parent does not contain ‘follow’ or ‘below:’ then remove ‘:’ and append child irrespective of the root label of constituency.
- If parent ends with ‘the following:’ then remove ‘the following:’ and append the child if it is not a complete sentence else do not remove ‘the following:’ and just append the child (we call this, concatenating).
- If none of the above rule satisfies and the parent ends with a ‘:’ then just concatenate the child with the parent.

A.2 Annotation Guidelines

We present the instructions, and the correctly and incorrectly annotated examples with explanations provided to the annotators in Figure 5. The custom annotation inference built to collect the data is shown in Figure 6. We manually annotate 50 sentences and use them as quality check questions to ensure annotators are sincerely and correctly annotating each HIT. Type-wise inter-annotator agreement for the sentences in test split is shown in Table 7.

A.3 Qualification Questions

We ask 10 multiple choice questions in the pre-qualification task consisting of 5 questions to test the understanding of identifying the correct deontic type and 5 questions to test their understanding of trigger span selection for a deontic type.

Obl	**Ent**	**Pro**	**Per**	**Nobl**	**Nent**	**None**
0.82	0.68	0.44	0.82	0.76	0.41	0.65

Table 7: Deontic type-wise inter-annotator agreement (α) for the test set.

Type	**Heuristic triggers**
Obl	shall/will be required, shall be obligated, shall, must, will, have to, should, ought to have, will/shall be paid
Ent	shall/will be entitled, shall/will be paid, shall/will retain, shall/will receive, shall have the right to, shall be retained, shall be kept, shall be claimed, shall be accessible, shall be owned, shall be determined
Pro	shall/will/must/may not, cannot, shall have no right, can not, shall/will not be allowed, shall not assist, shall/will be prohibited
Per	shall be permitted, shall also be permitted, can, may, could, shall/will be allowed
Nobl	shall/will not be liable for, shall/will not be obligated to, shall/will not be obligated for, shall/will not be responsible for, shall/will not be required to
Nent	shall/will not entitled to, shall/will not have the right to, shall/will not be entitled for

Table 8: Triggers used to identify the deontic types.

A.4 Resolving Disagreements

Disagreement in the annotation for duplicate sentences is resolved by one of the authors. The disagreement could occur because of any missing modality in case of multiple modalities expressed in a sentence, incorrect interpretation of the sentence, or human error in terms of annotating with respect to a tenant or a landlord. Consider the below sentence: “[landlord] After final approval of the Final Plans by applicable governmental authorities, no further changes may be made thereto without the prior written approval of both Landlord and Tenant.", it was annotated as ‘prohibition’ for landlord by one of the annotators and ‘none’ by another annotator. As the prohibition mentioned in the sentence is not for the landlord, the correct annotation is ‘none’. Therefore, we retain the correct annotation for the example and discard the sentence with the incorrect annotation. Another example is “[landlord] All conditions and agreements under the Lease to be satisfied or performed by Landlord have been satisfied and performed." which was incorrectly annotated as an ‘obligation’.

A.5 Rule-based Approach

We first curate a pre-defined list of triggers (Table 8) used to express deontic types in legal domain following Ash et al. (2020). Then, tokenize and obtain

Model	Accuracy	Precision	Recall	F1
Majority	47.87/26.06/38.48	7.60/4.83/12.43	14.29/14.29/20.29	9.92/7.22/15.26
Rule-based	52.13/42.25/47.88	63.81/65.28/75.54	48.08/32.69/40.07	47.42/34.63/42.75
BERT-BU	75.71/67.61/72.22	73.80/68.20/72.50	76.14/61.52/72.56	74.16/62.42/72.13
RoBERTa-B	72.69/69.48/71.31	73.94/74.61/73.71	76.06/74.94/76.50	73.65/73.86/74.36
RoBERTa-L	77.13/67.37/72.93	76.93/68.95/74.29	78.01/69.14/75.73	76.79/67.76/74.32
C-BERT-BU	76.60/68.31/73.03	78.17/69.27/75.52	81.19/68.62/77.43	79.14/67.49/75.92

Table 9: Evaluation results for agent-specific multi-label deontic modality classification task on development set. Scores are averaged over 3 different seeds. BU, B, and L denote base-uncased, base, and large respectively.

Model	Labeled				Unlabeled			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Majority	97.26/96.93/97.11	7.23/4.73/11.96	10.43/10.42/15.52	8.54/6.50/13.36	97.41/97.20/97.32	52.91/46.81/50.30	51.81/44.00/48.40	52.36/45.36/49.33
Rule-based	97.72/97.57/97.65	58.27/58.99/72.43	35.61/17.53/25.83	40.13/25.52/34.36	97.76/97.62/97.70	80.22/66.67/75.35	37.82/22.67/31.20	51.41/33.83/44.12
BERT-BU	98.23/98.03/98.14	57.07/45.78/54.75	57.81/45.74/55.23	56.91/44.18/54.14	98.41/98.29/98.36	72.70/67.68/70.47	72.37/67.33/70.17	72.49/67.50/70.30
C-BERT-BU	98.18/97.95/98.08	54.09/60.30/56.96	55.31/53.18/57.01	54.12/52.86/55.73	98.35/98.20/98.29	69.58/68.08/68.90	70.99/68.67/69.97	70.26/68.37/69.43
RoBERTa-B	98.12/97.83/97.99	55.33/48.38/53.63	58.95/52.51/57.78	56.73/49.80/55.31	98.28/98.06/98.18	70.30/67.56/69.08	73.40/71.78/72.69	71.80/69.56/70.81
RoBERTa-L	98.09/97.83/97.97	56.81/47.91/54.88	60.63/51.78/59.10	57.69/49.29/56.33	98.23/98.05/98.15	70.67/66.11/68.64	73.58/69.56/71.82	72.09/67.78/70.19
RoBERTa-L-NA	97.57/97.76/97.65	35.59/44.35/40.60	46.54/48.02/48.67	36.81/45.7/43.16	98.27/98.19/98.23	69.75/69.00/69.43	74.61/72.89/73.86	72.08/70.86/71.55

Table 10: Evaluation results for agent-specific modal trigger span detection task on development set. Macro-averaged scores for Tenant/Landlord/All are presented for precision, recall and F1 measures. Scores are averaged over 3 different seeds. BU, B, and L denote base-uncased, base, and large respectively.

Deontic Type	Classification			Span Detection		
	Precision	Recall	F1	Precision	Recall	F1
Obl	84.87	87.83	86.32	76.93	80.80	78.82
Ent	79.20	85.65	82.30	66.96	77.89	72.02
Pro	60.76	75.00	67.13	48.94	68.66	57.14
Per	91.25	87.43	89.30	90.79	82.63	86.52
Nobl	74.58	87.13	80.37	31.88	38.94	35.06
Nent	67.95	60.23	63.86	29.73	32.67	31.13
None	82.60	73.10	77.56	–	–	–

Table 11: Deontic type-wise results for agent-specific multi-label classification and modal trigger span detection (labeled) task on test set from the best (out of 3 seeds) RoBERTa-L model.

the dependency parse and part of speech (POS) tags each each token in a sentence using spaCy python library. We describe the heuristic algorithm (by observing patterns in the train set) which searches for the presence of pre-defined triggers in a given sentence to extract its position (start index), each of the agents' mention, and its dependency tag for a sentence in Algorithm 1.

A.6 Implementation Details

We run each model on 3 seed values. We use Adam optimizer with a linear scheduler for learning rate with an initial learning rate of $2e^{-5}$, and warm-up ratio set at 0.05. All the models are trained and tested on NVIDIA Tesla V100 SXM2 16GB GPU machine. We experiment with batch size $\in \{2, 4, 8\}$, number of epochs $\in \{3, 5, 10, 20, 30\}$, learning rate $\in \{1e^{-5}, 2e^{-5}, 3e^{-5}, 5e^{-5}\}$, and warm-up ratio $\in \{0.05, 0.10\}$. BERT-base (110M parameters) and Roberta-base (125M parameters)

Model	Accuracy	Precision	Recall	F1
Multi-label Classification				
RoBERTa-L	76.74	77.30	79.11	77.88
RoBERTa-L-AR	47.91	56.99	59.91	56.60
RoBERTa-L-ARR	66.18	69.75	74.14	71.22
Trigger Span Detection (Labeled)				
RoBERTa-L	98.37	57.37	63.74	60.04
RoBERTa-L-AR	98.12	50.11	59.54	53.63
RoBERTa-L-ARR	98.42	58.42	64.88	61.19
Trigger Span Detection (Unlabeled)				
RoBERTa-L	98.47	69.68	76.78	73.06
RoBERTa-L-AR	98.36	67.03	75.76	71.07
RoBERTa-L-ARR	98.54	70.65	76.76	73.58

Table 12: Evaluating RoBERTa-L-AR and RoBERTa-L-ARR on lease test set.

Model	Accuracy	Precision	Recall	F1
Majority	96.09/97.53	54.55/41.26	4.55/34.16	8.39/37.38
Rule-based	96.40/97.85	84.62/73.28	16.67/21.72	27.85/33.51
RoBERTa-L	97.78/98.09	69.85/54.32	68.44/60.76	69.14/57.36
RoBERTa-L-AR	97.81/98.31	68.77/59.64	69.44/57.05	69.09/58.30
RoBERTa-L-ARR	97.78/98.46	69.51/61.81	70.45/57.50	69.94/59.58

Table 13: Unlabeled Metric scores for trigger span detection task on rental/employment contracts.

models took 46mins, and RoBERTa-large (355M parameters) took 2hrs to train for each of the tasks.

A.7 Additional Results

Table 9, and 10 shows the results for the two tasks on the dev set. Table 11 shows the type-wise results from the best performing model. Table 12 shows the performance of models trained with anonymized agent on the test set of lease contracts.

Algorithm 1 `Rule-based Heuristic`

```
Inputs: List T of pre-defined triggers, List A of aliases for the type of contract to process.
Outputs: List L of tuples containing (Deontic type, trigger, agent, start index) for all the sentence in the contract.
L ← [], I ← []                                                        // Initialization
for each sentence in contract do
    // Initialize a list to keep account of visited trigger indices
    visited ← []
    for each t in T do
        if t in sentence then
            // Initialize a list of trigger indices
            indices ← []
            for each t in sentence do
                if start index of t ∉ visited then
                    indices ← start index
                    visited ← start index
                end if
            end for
            for word in sentence do
                if word.dependency is ROOT or word.pos ∈ [VERB, AUX] then
                    for child in word.children do        // Iterate over the children of word in the dependency tree
                        If a1 ∈ A is 'nsubj/nsubjpass' of word & child==t[0] & child.dependency is 'aux' & child.index in indices then L ← (Type(t), t, a1, child.index)    // Rule 1
                        If Rule 1 & a2 ∈ A is a 'conj' of a1 then L ← (Type(t), t, a2, child.index) // Rule 2
                        If child1.dependency is 'agent' & child2==t[0] & child2.dependency is 'aux' & a1 ∈ A in children(child1)=child3 then L ← (Type(t), t, a1, child2.index)    // Rule 3
                        If Rule 3 & a2 ∈ A in conjunction of child3 & VERB in conjunction of word & t1 is 'aux' of VERB then L ← (Type(t1), t1, a2, t1.index)    // Rule 4
                        If Rule 3 & not Rule 4 & a2 ∈ A in conjunction of child3 & child==t[0] & child.dependency is 'aux' & child.index in indices then L ← (Type(t), t, a2, child.index)    // Rule 5
                        If child.dependency in ['pobj', 'dobj'] & a1 ∈ A is in conjunction of children(child)=child1 & VERB in conjunction of word & t1 is 'aux' of VERB then L ← (Type(t1), t1, a1, t1.index)    // Rule 6
                        If child==t[0] & child.dependency is 'aux' & child.index in indices & VERB in conjunction of word & t1 is 'aux' of VERB & 'agent' in children(conjunction VERB)= child1.dependency & a2 ∈ A in children(child1) then L ← (Type(t), t, a2, child.index)    // Rule 7
                        If child==t[0] & child.dependency is 'aux' & child.index in indices & VERB in conjunction of word & t1 is 'aux' of VERB & not Rule 7 then L ← (Type(t1), t1, Agent(t), t1.index)    // Rule 8
                    end for
                end if
            end for
        end if
    end for
end for
```

Table 13 shows the unlabeled metric scores for generalizability to rental and employment contracts.

Model	Precision	Recall	F1
ALeaseBERT	82.35	8.09	14.74
Ours	8.53	87.28	15.54

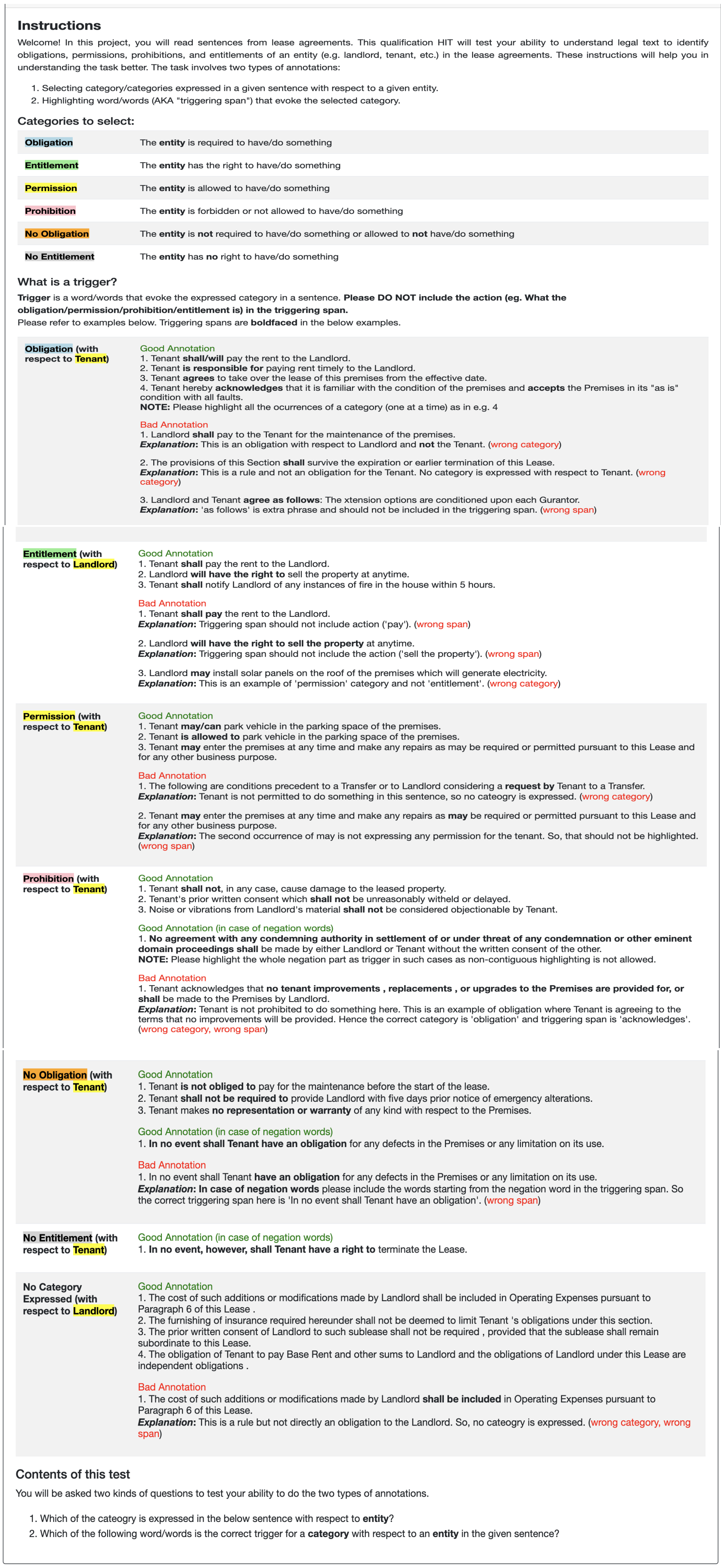
Table 14: Results from the red flag detection case study. Our (ALeaseBERT) denotes RoBERTa-L model trained on LexDeMod (Red flags dataset (Leivaditi et al., 2020)).

A.8 Case Study: Red flag Detection

Evaluation scores for the red flag detection case study are presented in Table 14.

A.9 Annotated Examples for Deontic Types

Samples annotations are provided in Table 15.



Instructions

Welcome! In this project, you will read sentences from lease agreements. This qualification HIT will test your ability to understand legal text to identify obligations, permissions, prohibitions, and entitlements of an entity (e.g. landlord, tenant, etc.) in the lease agreements. These instructions will help you in understanding the task better. The task involves two types of annotations:

1. Selecting category/categories expressed in a given sentence with respect to a given entity.
2. Highlighting word/words (AKA "triggering span") that evoke the selected category.

Categories to select:

Obligation	The **entity** is required to have/do something
Entitlement	The **entity** has the right to have/do something
Permission	The **entity** is allowed to have/do something
Prohibition	The **entity** is forbidden or not allowed to have/do something
No Obligation	The **entity** is **not** required to have/do something or allowed to **not** have/do something
No Entitlement	The **entity** has **no** right to have/do something

What is a trigger?

Trigger is a word/words that evoke the expressed category in a sentence. **Please DO NOT include the action (eg. What the obligation/permission/prohibition/entitlement is) in the triggering span.**
Please refer to examples below. Triggering spans are **boldfaced** in the below examples.

Obligation (with respect to Tenant)	Good Annotation 1. Tenant **shall/will** pay the rent to the Landlord. 2. Tenant **is responsible for** paying rent timely to the Landlord. 3. Tenant **agrees** to take over the lease of this premises from the effective date. 4. Tenant hereby **acknowledges** that it is familiar with the condition of the premises and **accepts** the Premises in its "as is" condition with all faults. **NOTE:** Please highlight all the ocurrences of a category (one at a time) as in e.g. 4 Bad Annotation 1. Landlord **shall** pay to the Tenant for the maintenance of the premises. ***Explanation*:** This is an obligation with respect to Landlord and **not** the Tenant. (wrong category) 2. The provisions of this Section **shall** survive the expiration or earlier termination of this Lease. ***Explanation*:** This is a rule and not an obligation for the Tenant. No category is expressed with respect to Tenant. (wrong category) 3. Landlord and Tenant **agree as follows**: The xtension options are conditioned upon each Gurantor. ***Explanation*:** 'as follows' is extra phrase and should not be included in the triggering span. (wrong span)
Entitlement (with respect to Landlord)	Good Annotation 1. Tenant **shall** pay the rent to the Landlord. 2. Landlord **will have the right to** sell the property at anytime. 3. Tenant **shall** notify Landlord of any instances of fire in the house within 5 hours. Bad Annotation 1. Tenant **shall pay** the rent to the Landlord. ***Explanation*:** Triggering span should not include action ('pay'). (wrong span) 2. Landlord **will have the right to sell the property** at anytime. ***Explanation*:** Triggering span should not include the action ('sell the property'). (wrong span) 3. Landlord **may** install solar panels on the roof of the premises which will generate electricity. ***Explanation*:** This is an example of 'permission' category and not 'entitlement'. (wrong category)
Permission (with respect to Tenant)	Good Annotation 1. Tenant **may/can** park vehicle in the parking space of the premises. 2. Tenant **is allowed to** park vehicle in the parking space of the premises. 3. Tenant **may** enter the premises at any time and make any repairs as may be required or permitted pursuant to this Lease and for any other business purpose. Bad Annotation 1. The following are conditions precedent to a Transfer or to Landlord considering a **request by** Tenant to a Transfer. ***Explanation*:** Tenant is not permitted to do something in this sentence, so no cateogry is expressed. (wrong category) 2. Tenant **may** enter the premises at any time and make any repairs as **may** be required or permitted pursuant to this Lease and for any other business purpose. ***Explanation*:** The second occurrence of may is not expressing any permission for the tenant. So, that should not be highlighted. (wrong span)
Prohibition (with respect to Tenant)	Good Annotation 1. Tenant **shall not**, in any case, cause damage to the leased property. 2. Tenant's prior written consent which **shall not** be unreasonably witheld or delayed. 3. Noise or vibrations from Landlord's material **shall not** be considered objectionable by Tenant. Good Annotation (in case of negation words) 1. **No agreement with any condemning authority in settlement of or under threat of any condemnation or other eminent domain proceedings shall** be made by either Landlord or Tenant without the written consent of the other. **NOTE:** Please highlight the whole negation part as trigger in such cases as non-contiguous highlighting is not allowed. Bad Annotation 1. Tenant acknowledges that **no tenant improvements , replacements , or upgrades to the Premises are provided for, or shall** be made to the Premises by Landlord. ***Explanation*:** Tenant is not prohibited to do something here. This is an example of obligation where Tenant is agreeing to the terms that no improvements will be provided. Hence the correct category is 'obligation' and triggering span is 'acknowledges'. (wrong category, wrong span)
No Obligation (with respect to Tenant)	Good Annotation 1. Tenant **is not obliged to** pay for the maintenance before the start of the lease. 2. Tenant **shall not be required to** provide Landlord with five days prior notice of emergency alterations. 3. Tenant makes **no representation or warranty** of any kind with respect to the Premises. Good Annotation (in case of negation words) 1. **In no event shall Tenant have an obligation** for any defects in the Premises or any limitation on its use. Bad Annotation 1. In no event shall Tenant **have an obligation** for any defects in the Premises or any limitation on its use. ***Explanation*: In case of negation words** please include the words starting from the negation word in the triggering span. So the correct triggering span here is 'In no event shall Tenant have an obligation'. (wrong span)
No Entitlement (with respect to Tenant)	Good Annotation (in case of negation words) 1. **In no event, however, shall Tenant have a right to** terminate the Lease.
No Category Expressed (with respect to Landlord)	Good Annotation 1. The cost of such additions or modifications made by Landlord shall be included in Operating Expenses pursuant to Paragraph 6 of this Lease . 2. The furnishing of insurance required hereunder shall not be deemed to limit Tenant 's obligations under this section. 3. The prior written consent of Landlord to such sublease shall not be required , provided that the sublease shall remain subordinate to this Lease. 4. The obligation of Tenant to pay Base Rent and other sums to Landlord and the obligations of Landlord under this Lease are independent obligations . Bad Annotation 1. The cost of such additions or modifications made by Landlord **shall be included** in Operating Expenses pursuant to Paragraph 6 of this Lease. ***Explanation*:** This is a rule but not directly an obligation to the Landlord. So, no cateogry is expressed. (wrong category, wrong span)

Contents of this test

You will be asked two kinds of questions to test your ability to do the two types of annotations.

1. Which of the cateogry is expressed in the below sentence with respect to **entity**?
2. Which of the following word/words is the correct trigger for a **category** with respect to an **entity** in the given sentence?

Figure 5: Instructions and examples provided to the annotators.

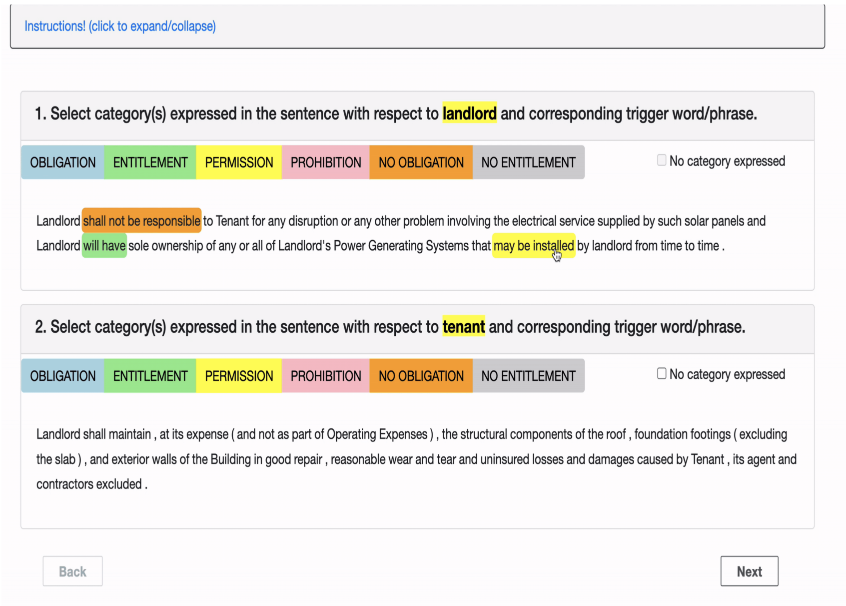


Figure 6: Annotation Interface.

Type	Examples
Obl	[tenant] Tenant **shall** repair any damage resulting from such removal and **shall** restore the Property to good order and condition. [tenant] Tenant **acknowledges** and **agrees** that Landlord shall have the right to adopt reasonable rules and regulations for the use and/or occupancy of the Leased Premises and Tenant **agrees** that it shall at all times observe and comply with such rules and regulations.
Ent	[tenant] Tenant **shall also have the right to** use the roof riser space of the Building. [lessor] Rent **shall** be payable at Lessor 's place of business , or such other place as Lessor may direct from time to time. [landlord] Landlord **reserves the right to** modify Common Areas, provided that such modifications do not materially adversely affect Tenant's access to or use of the Premises for the Permitted Use.
Pro	[lessee] Lessee **shall not** commit or allow waste to be committed on the Premises, and Lessee **shall not** allow any hazardous activity to be engaged in upon the Premises. [lessor] **Neither Lessor nor Lessee may** record this Lease nor a short - form memorandum thereof.
Per	[tenant] Tenant **may**, without Landlord's consent, before delinquency occurs, contest any such taxes related to the Personal Property. [lessor] Additional keys **may** be furnished at a charge by Lessor.
Nobl	[tenant] For the avoidance of doubt, to the extent there is a bank vault in the Premises, Tenant **shall have no obligation to** remove such vault on surrendering the Premises. [lessor] Further, **in no event shall Lessor have any obligation to** repair any damage to, or replace any of Lessee's furniture, trade fixtures, equipment or other personal property.
Nent	[landlord] Landlord hereby **waives the right to** any revenue that may be generated as a result of the use of the roof by Tenant or any other third - parties pursuant to the terms of the Lease during the Term. [lessee] The Lessee **will not be entitled to** a reimbursement of any part of the Rent, even if in practice the Building Capacity for which it has paid has not been utilized.
None	[lessor] For the avoidance of doubt, it is hereby clarified that wherever the word Lessor is written this means: "the Lessor and/or anyone acting on its behalf". [landlord] Other than the Purchase Agreement, this Lease represents the entire agreement and understanding between Landlord and Tenant with respect to the subject matter herein, and there are no representations, understandings, stipulations, agreements or promises not incorporated in writing herein.

Table 15: Sample annotated sentences for each deontic type with respect to an [Agent] and trigger annotations in **bold-face**.