Detecting Purpose Language in State Regulations

Chandan Aggarwal[1,†], Royce Koh[1,†], Cindy Zhang[1,†], Matthew Carey[1] and Sylvia Kwakye[1,*]

[1]Cornell University, Ithaca, New York, USA

Abstract

The legal effect of US state regulations often depends not only on explicit statements of rules, but also on purposes stated in the regulations. In this paper we demonstrate the use of a model for classifying regulations according to the presence or absence of 'purpose' language. Classifying 'purpose' sections is a foundational step toward deeper insights into the structure and intent of legal regulations.

Keywords

natural language processing, legal document analysis, binary classification

1. Introduction

State regulations, issued by executive agencies, are essential for clarifying the interpretation and implementation of statutes. These regulations often lack consistency in structure across different states. This inconsistency makes it challenging for non-expert individuals and organizations to access, understand, analyze, and compare the legal texts. Our motivation for this research was to reduce the financial and interpretative barriers to accessing this legal information.

The area of focus of our project is the 'purpose' sections within state regulations. Regulations are rarely arbitrary. These sections communicate the underlying rationale, the intended outcomes, and the broader context of particular regulations. In other words, such sections focus on the 'why' and not just the 'what' of regulations.

We were interested in finding out to what extent the drafters of state regulations communicate the value and necessity of the regulations. If they did, was such language easy to find, understand, and compare across jurisdictions?

Unsurprisingly, we found significant differences between states. Some states were consistent in providing explicitly named 'Purpose' sections, others had purpose sections with vague names like 'General', while some had purpose statements embedded within other sections. The diversity of approaches made purpose content difficult to identify with just heuristics. To overcome these difficulties, we present a machine learning model to classify regulatory texts into purpose and non-purpose categories. We aim to generalize the model to handle the diverse regulations across states.

The implications of this research are both practical and broad. 'Purpose' annotations can support open access to law by providing relevant context to readers of published regulations. For policymakers and researchers, a system capable of identifying purpose sections enables streamlined legal analysis and cross-jurisdictional comparisons of regulatory intent. This can support efforts to standardize regulatory frameworks, enhancing transparency and governance. In addition, the project aims to lay a groundwork for advanced analytics of regulatory content.

2. Related Work

AI and Law researchers have long sought to extract the purposes of legal rules for use in automated legal reasoning. In [1], Berman and Hafner propose that knowledge engineers should include the purposes of judicial doctrines in their representations of legal opinions for use in case-based reasoning. [2] suggests representing the 'purpose' of a ruling by discovering the decision maker's preference for a set of values. [3] and [4] use such values to automate the work of generating legal arguments and predicting case outcomes. [5] presents a Value-Based Reasoning Framework, in which a set of possible arguments is filtered to include only the arguments matching an agent's value preferences. [6] explores a method for predicting the outcomes of judicial cases by assigning weights to the values supported by the arguments in the case.

Prior work has explored the use of NLP to categorize legal texts, including regulations. [7] uses a classifier based on regular expressions to classify sentences found in statutes into 13 categories, but 'purpose' or 'value' sentences are not among the included categories. LLMs are used to discover factors in caselaw in [8], and to sort judicial opinions into thematic categories in [9]. [10] uses BERT [11] for a binary classification task, to determine whether statements by financial services providers can be considered 'promissory' under US regulations. [12] evaluates the accuracy of both BERT and GPT in assigning regulations to specified categories based on their purposes.

Our work also follows several other papers that use SetFit [?] to categorize legal text. [13] uses SetFit to detect Hohfeldian rights and privileges in UK legislative text, and [14] uses SetFit to categorize the rhetorical roles of sentences in legal judgment documents from Indian courts.

3. Data Exploration

The dataset we used was the Public.Resource.Org quarterly dump of US state regulations in XML for 2024 Q3 [15]. This open dataset attempts to include all the regulations currently in effect in the 50 states, which adds up to about 1.5 million provisions, including supersections. Our programmatic analysis of the XML found 1,170,452 section-level regulations with text content. Most of these regulations were currently in effect, because most of our data sources were state regulatory codes, and states remove outdated regulations from these codes on a regular basis. One significant exception is Arkansas, which publishes current and prior versions of its regulations together and does not clearly distinguish between them. In some states, instead of removing an outdated regulation, the state would remove the regulation's text content and then replace either the regulation's content or its name with a brief notation like 'Repealed' or 'Renumbered.'

We first approached the dataset by using Python scripts for data exploration. We attempted to identify patterns and common keywords in purpose sections. We used the 'ydata-profiling' [16] Python library for some of this exploratory analysis.

The unit of analysis for our study was a single regulatory 'section.' In states that don't refer to regulations as 'sections,' we identified 'section-level' regulations that might be called "rule", "part," etc., depending on the state. Our expectation was that two to five percent of the regulations in the corpus would include 'purpose' statements. However, the average length of 'section-level' regulations varied widely from state to state. In states like Vermont and Arkansas, with very long section-level regulations, each individual regulation was more likely to contain at least some discussion of purpose, and thus the 'purpose' label was applied to a greater percentage of those states' regulations.

4. Training Classification Models

We trained classification models for state regulations using three approaches: LegalBERT [17], SetFit [18], and GliClass [19]. We chose text classification with LegalBERT, a transformer-based model designed specifically for legal text, because of its suitability for the legal nature of the task. We also

decided to use the few-shot SetFit framework for optimized fine-tuning, and GLiClass as a zero-shot classifier as another model for comparison.

After initial data exploration, we manually curated 100 samples of 'purpose' sections from the states of Alabama, California, Kansas, and New York. We considered those states to be representative of the overall corpus because they include the two states with the most regulatory text currently in effect (California and New York) and the state with the least (Kansas), based on the sizes of the XML archives at [15]. Our initial samples included only the text content of each regulation, not the headings. To quickly collect samples, we began with a keyword search of regulations containing the word 'purpose' and its synonyms. After a preliminary analysis of this initial sample set, we identified four broad subcategories of 'purpose' regulations, and described them as follows:

- `purpose-regulatory` (REG): 'The language of the section explicitly states that the purpose of the law is to regulate something. This is the catch-all type of regulatory purpose if it cannot be classified as one of the other types.'
- `purpose-with-scope` (SCOPE): 'In addition to being a regulatory purpose, this has language that indicates the scope of the regulation. The scope here is generally a reference to a specific type of activity or applicability. It could also be a reference to some specific chunk of regulation.'
- `purpose-with-authority` (AUTH): 'In addition to being a regulatory purpose, this states the authority under which the regulation is issued.'
- `purpose-administrative` (AD): 'In addition to being a regulatory purpose, this has language to indicate that the purpose of the regulation is administrative, such as to establish a commission or to set a fee.'

To create samples for each category, we manually assigned labels until we hit 25 examples for each label, disregarding any additional regulations for any given category. As expected, this process was time consuming. Subsequently, we generated extra candidates programmatically by classifying to one or zero of the four purpose categories with a keyword search in the heading, and the first 140 characters of the text of each regulation. The algorithm we used for the keyword search was as follows:

1. If the regulation does not include the keyword 'purpose,' exclude it from the sample.
2. If the regulation includes 'determination,' 'policy,' 'regulation,' 'rule,' or 'law,' classify it as 'purpose-regulatory' (REG).
3. If the regulation includes 'applicability,' 'scope,' or 'severability,' classify it as 'purpose-with-scope' (SCOPE).
4. If the regulation includes 'authority' or 'authorization', classify it as 'purpose-with-authority' (AUTH).
5. If the regulation includes 'administrative,' 'administration,' 'commission,' 'committee,' 'fee,' or 'mission,' classify it as 'purpose-administrative' (AD).
6. If none of the above keywords match, exclude the regulation from the sample.

We reviewed the results of this keyword classifier and selected another 23 samples to combine with each of our our hand-labeled samples. Examples of these classifications are shown in Table 1.

The goal of the keyword classifier was to prefilter samples and speed up manual labelling. However, out of curiosity, we applied the keyword-based classifier to 13,019 sample regulation sections from Alabama, California, Kansas, and New York. We found that 5,348 were classified as "non-purpose," 4,454 as 'purpose-regulatory' (REG), 1,126 as 'purpose-with-authority' (AUTH), 1,196 as 'purpose-administrative' (AD), and 755 as 'purpose-with-scope' (SCOPE). We did not expect this keyword classifier to be very accurate, but the ratio of purpose sections far exceeded what was expected. This result alerted us that we had significantly undersampled non-purpose sections in the training data. With an equal number of samples for each label, there were four times as many purpose as non-purpose sections.

Separately, a legal subject matter expert on our team manually labeled samples of 50 randomly selected regulations for each of four states: Alabama, Alaska, California, and Texas. The California and

Table 1

Examples of keyword-based 'purpose' regulation classifications used for training

Heading Text	Body Text	Citation	Assigned Class
Purpose Of The Rules	The purpose of this subchapter is to assure that elevators and other automated conveyances are correctly and safely installed and operated within the state by authorizing and enforcing rules for the design, installation, operation and maintenance of automated people conveyances, and by licensing mechanics and inspectors who work on these conveyances.	Ala. Admin. Code r. 480-8-1-.01	REG
Purpose and Scope	This article sets forth rules to be observed when Department employees conduct eyewitness identification procedures. This article does not apply to field show-ups.	Cal. Code Regs. Tit. 10, § 2698.22	SCOPE
Scope and Purpose	The purpose of this article is to implement and make specific the provisions of Public Resources Code, Division 3, Chapter 1, Article 4.6 (commencing with section 3280), to accomplish the purposes of Article 4.6 as declared in Statutes of 2022, chapter 365, section 1 (SB 1137).	Cal. Code Regs. Tit. 14, § 1765	SCOPE
Purpose and Statutory Authority	The rules and regulations contained in this Subchapter are for the purpose of implementing provisions of the Unclaimed Property Law and are authorized by Code of Civil Procedure Section 1580.	Cal. Code Regs. Tit. 2, § 1150	AUTH
Authority and Purpose	The Commissioner promulgates these regulations pursuant to the implied authority granted by California Insurance Code Sections 791 et seq. and 15 U.S.C. Sections 6801(b) and 6805(b) to implement California Insurance Code and Gramm-Leach-Bliley privacy provisions consistent with providing individuals the maximum privacy protections permitted by those laws.	Cal. Code Regs. Tit. 10, § 2689.1	AUTH
Purpose	The purpose of this Chapter is to provide a fee schedule to be charged for analysis run on certain products, animals or fowl when the request for analysis originates from private citizens or agencies other than public agencies.	Ala. Admin. Code r. 80-1-12-.01	AD

Table 2

Count of sample regulations human-annotated as containing purpose language

State	Purpose of regulation			Purpose of statute		
	Yes	No	Maybe	Yes	No	Maybe
AL	5	44	1	0	49	1
AK	1	48	1	0	50	0
CA	0	50	0	0	50	0
TX	3	47	0	0	50	0

Texas samples included the sections' heading text, but the Alabama and Alaska samples did not. The subject matter expert applied two tags per regulation. One tag indicated whether the regulation stated the purpose of any statute, and the other tag indicated whether the regulation stated the purpose of any regulation. (If the regulation stated its own purpose, this was marked 'Yes'.) If determining the correct tag would have required reading the headings of super-sections or other material not available to the classifier, the subject matter expert marked 'Maybe'. The distribution of tags is shown in Table 2. We created a 'non-purpose' (NON) category in our classifiers' training data from the regulations with two 'No' tags from the California and Texas files, and added a few false positive sections we observed during data exploration.

At the end of this exercise, we had at least 50 samples for each type of purpose section.

4.1. Training and Finetuning

As previously mentioned, we used three classifiers: LegalBERT, SetFit (initially with paraphrase-MiniLM-L6-v2), and GliClass.

We started out with a LegalBERT classifier, without any finetuning, but the results were not satisfactory. The model often misclassified non-purpose sections as purpose sections. Also disappointing were preliminary setfit results with a generic sentence transformer (paraphrase-MiniLM-L6-v2). We suspected that this may be due to the models' lack of domain-specific training on regulatory texts. Even though LegalBERT was designed for legal text, its original training data did not include regulatory content, leaving gaps in its understanding of domain-specific nuances critical for distinguishing between purpose and non-purpose sections.

Subsequently, we adapted LegalBERT to the domain with continued pretraining on the representative set of regulations from Alabama, California, Kansas, and New York. We created a new classifier that used the updated LegalBERT as the transformer embedding model within a SetFit classifier.

We trained our classifiers with 8 samples per label to evaluate the model's initial performance. Then we increased the sample size to 16, then 32 per label, ensuring the text of each section was appropriately labeled for its category.

Table 3
Combined classification report for 8, 16, and 32 samples

Samples	Label	Precision	Recall	F1-score
8	NON	0.60	0.50	0.55
8	AD	0.78	0.58	0.67
8	REG	0.45	0.56	0.50
8	AUTH	0.27	0.30	0.29
8	SCOPE	0.57	0.62	0.59
16	NON	1.00	0.67	0.80
16	AD	0.60	0.75	0.67
16	REG	0.50	1.00	0.67
16	AUTH	1.00	0.80	0.89
16	SCOPE	0.67	0.67	0.67
32	NON	1.00	0.75	0.86
32	AD	1.00	0.75	0.86
32	REG	0.40	0.40	0.40
32	AUTH	0.45	0.71	0.56
32	SCOPE	0.71	0.62	0.67

The results of using SetFit+LegalBERT are shown in Table 3. Accuracy increased from 8-labels to 16-labels and decreased slightly from 16 to 32-labels. It appears that too many samples may have led to overfitting. Even though accuracy in assigning purpose type decreased from 16 to 32-labels, there was better discrimination between purpose and non-purpose.

To achieve better performance, we also tried to implement systematic hyperparameter optimization using Optuna [20] and Bayesian methods. We employed Optuna to automate tuning across a wide range of parameters, including learning rate (1e-5 to 5e-5), batch size (8, 16, 32), epochs (up to 5), and weight decay (0.01 to 0.1).

However, limitations in the SetFit framework restricted control over certain parameters, necessitating a more focused approach. Bayesian optimization was applied to explore stable hyperparameter regions, narrowing learning rates to 3e-5 to 5e-5, batch sizes to 16 and 32, and weight decay to a refined range informed by prior results. Despite these efforts, the variations showed limited impact on performance, and training time remained a significant challenge, indicating diminishing returns on further tuning.

GliClass is a newer classification model that is capable of using a set of labels to classify text without any training data. Like SetFit however, a minimal set of examples for learning significantly improves accuracy. We used the same labels as before, to compare GliClass performance. The results were similar to the BERT-based model, with 16 samples showing the best performance, as shown in Table 4.

Table 4
Confusion matrix between GliClass labels and categories assigned manually

	GliClass Predictions				
Manual Labels	NON	AD	REG	AUTH	SCOPE
NON	14	0	0	2	0
AD	1	7	3	2	3
REG	2	0	11	2	1
AUTH	1	0	3	10	2
SCOPE	2	0	5	2	7
Precision	0.70	1.00	0.50	0.56	0.54
Recall	0.88	0.44	0.69	0.62	0.44
F1-score	0.78	0.61	0.58	0.59	0.48
Support	16	16	16	16	16

Table 5
Confusion matrix between categories assigned to regulations by the GliClass and LegalBERT models

LegalBERT Classifications	GliClass Classifications				
	NON	AD	REG	AUTH	SCOPE
NON	958509	31664	19633	20890	1036
AD	5562	7906	4834	3779	739
REG	31226	8095	11694	3949	731
AUTH	43567	5204	3653	5337	756
SCOPE	830	304	371	171	12

It was clear from these results that we would do better optimizing a model that simply identified 'purpose' and 'non-purpose' sections. The confusion matrix in Table 5 shows that in our dataset of 1,170,452 state regulations containing text content, out of the 138,720 regulations where LegalBERT assigned a 'purpose' classification, GliClass assigned 'non-purpose' 81,185 times. Out of the 130,758 regulations where GliClass assigned a 'purpose' classification, LegalBERT assigned 'non-purpose' 73,223 times.

4.2. Binary Classifier

We updated the training data to add labels limited to 'Purpose' and 'Non-purpose'. The goal of the new classifier was to maximize the F1-score for the 'Non-purpose' label. We pretrained a new model using a LegalBERT base on a curated dataset of legal texts that only included binary labels. We conducted fine-tuning exclusively on a binary-labeled dataset, emphasizing balanced representation of purpose and non-purpose examples. We also began providing the classifier with metadata including the text of the regulations' headings. We split our hand-annotated data into separate datasets for pretraining LegalBERT, training SetFit, validation, and testing.

Hyperparameter optimization played a crucial role in the fine-tuning process, with batch sizes of 16 and 32, epochs ranging from 3 to 5, and learning rates in the range of 1e-5 to 5e-5 systematically tested to identify the best configuration. We adjusted class weights to address class imbalance, ensuring the model prioritized non-purpose sections without neglecting purpose sections. We then ran SetFit as normal with all five classes. We observed that all three models performed the best on a sample size of 16 labels, indicating the potential impacts of overfitting. Most of them were best at identifying the purpose-administrative label and non-purpose label.

When we reviewed the results, Kentucky stood out from the other states because nearly all its regulations were classified as 'purpose' regulations, but we found that that classification was accurate. In Kentucky, the regulations assigned the 'Section' type in the Public.Resource.Org dataset corresponded to full regulations, and they contained multiple provisions that Kentucky labeled as 'Sections'. They

also followed a convention of beginning with a 'Necessity, Function, and Conformity' section that explicitly outlined the rationale behind the regulation, its intended purpose, and its alignment with statutory requirements. Each of these regulations was passed as a single document to the classifier, and the classifier usually detected its discussion of the regulation's purpose.

For instance, the Kentucky regulation with the heading 'Avian influenza' begins with the following purpose statement, including a citation to the statute that the regulation implements. "NECESSITY, FUNCTION, AND CONFORMITY: KRS 257.070 requires that importation of animals into Kentucky complies with administrative regulations promulgated by the board. This administrative regulation establishes requirements for entry into Kentucky to prevent the introduction and spread of avian influenza virus into Kentucky domestic poultry." (302 KAR 20:250)

Kentucky posed a significant challenge for earlier models, as they struggled to identify these sections accurately due to their reliance on more generalized patterns and keywords. However, the binary classifier, with its focused architecture and domain-specific pretraining, successfully adapted to this context.

5. Conclusion and Future Work

The classification of regulatory texts into Purpose and Non-Purpose categories is a critical step toward improving the accessibility and analysis of regulations across jurisdictions. This project addressed the inherent challenges posed by the complexity, inconsistency, and variability of regulatory language across states. By combining models such as LegalBERT, SetFit, and GliClass with targeted improvements in data preparation and training methodologies, we made significant strides in achieving accurate and generalizable classifications. Pretraining LegalBERT on a corpus tailored to regulatory texts enhanced its ability to capture domain-specific nuances, enabling it to outperform out-of-the-box versions. The introduction of metadata as contextual cues reduced false positive classifications and improved the overall reliability of the models. The binary classifier's success in Kentucky, where most regulations discuss a regulatory purpose in the Necessity, Function, and Conformity section, demonstrated the LegalBERT classifier's ability to adapt to state-specific regulatory frameworks. This validation underscores the model's potential for application across diverse jurisdictions with varying regulatory structures.

Future research work may focus on expanding training datasets, optimizing model architectures, and extracting the purpose text from 'purpose' regulations. This work has broader implications beyond regulatory classification. With domain-specific pretraining, the machine learning techniques used in this research can streamline legal analysis, support policymaking, and enhance public understanding of regulations.

Acknowledgments

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] D. H. Berman, C. D. Hafner, Representing teleological structure in case-based legal reasoning: the missing link, in: Proceedings of the 4th international conference on Artificial intelligence and

law, ICAIL '93, Association for Computing Machinery, New York, NY, USA, 1993, pp. 50–59. URL: https://dl.acm.org/doi/10.1145/158976.158982. doi:10.1145/158976.158982.

[2] T. Bench-Capon, G. Sartor, A model of legal reasoning with cases incorporating theories and values, Artificial Intelligence 150 (2003) 97–143. URL: https://linkinghub.elsevier.com/retrieve/pii/S0004370203001085. doi:10.1016/S0004-3702(03)00108-5.

[3] M. Grabmair, Modeling purposive legal argumentation and case outcome prediction using argument schemes in the value judgment formalism, 2016. URL: https://d-scholarship.pitt.edu/27608/.

[4] C. Benzmüller, D. Fuenmayor, B. Lomfeld, Modelling Value-Oriented Legal Reasoning in LogiKEy, Logics 2 (2024) 31–78. URL: https://www.mdpi.com/2813-0405/2/1/3. doi:10.3390/logics2010003, number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[5] J. P. Wallner, A. Wyner, T. Zurek, Value-Based Reasoning in ASPIC+, in: C. Reed, M. Thimm, T. Rienstra (Eds.), Frontiers in Artificial Intelligence and Applications, IOS Press, 2024. URL: https://ebooks.iospress.nl/doi/10.3233/FAIA240332. doi:10.3233/FAIA240332.

[6] A. Chorley, T. Bench-Capon, An empirical investigation of reasoning with legal cases through theory construction and application, Artif Intell Law 13 (2005) 323–371. URL: https://doi.org/10.1007/s10506-006-9016-y. doi:10.1007/s10506-006-9016-y.

[7] E. de Maat, R. Winkels, Automated Classification of Norms in Sources of Law, in: E. Francesconi, S. Montemagni, W. Peters, D. Tiscornia (Eds.), Semantic Processing of Legal Texts: Where the Language of Law Meets the Law of Language, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 170–191. URL: https://doi.org/10.1007/978-3-642-12837-0_10. doi:10.1007/978-3-642-12837-0_10.

[8] M. Gray, J. Savelka, W. Oliver, K. Ashley, Using LLMs to Discover Legal Factors, in: Legal Knowledge and Information Systems, IOS Press, 2024, pp. 60–71. URL: https://ebooks.iospress.nl/doi/10.3233/FAIA241234. doi:10.3233/FAIA241234.

[9] J. Drápal, H. Westermann, J. Šavelka, Using Large Language Models to Support Thematic Analysis in Empirical Legal Studies, in: Legal Knowledge and Information Systems, IOS Press, 2023, pp. 197–206. URL: https://ebooks.iospress.nl/doi/10.3233/FAIA230965. doi:10.3233/FAIA230965.

[10] R. Sarkar, A. K. Ojha, J. Megaro, J. Mariano, V. Herard, J. P. McCrae, Few-shot and Zero-shot Approaches to Legal Text Classification: A Case Study in the Financial Sector, in: N. Aletras, I. Androutsopoulos, L. Barrett, C. Goanta, D. Preotiuc-Pietro (Eds.), Proceedings of the Natural Legal Language Processing Workshop 2021, Association for Computational Linguistics, Punta Cana, Dominican Republic, 2021, pp. 102–106. URL: https://aclanthology.org/2021.nllp-1.10/. doi:10.18653/v1/2021.nllp-1.10.

[11] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: J. Burstein, C. Doran, T. Solorio (Eds.), 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), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: https://aclanthology.org/N19-1423/. doi:10.18653/v1/N19-1423.

[12] J. Šavelka, K. D. Ashley, The unreasonable effectiveness of large language models in zero-shot semantic annotation of legal texts, Front. Artif. Intell. 6 (2023) 1279794. URL: https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2023.1279794/full. doi:10.3389/frai.2023.1279794, publisher: Frontiers.

[13] A. Izzidien, Using the interest theory of rights and Hohfeldian taxonomy to address a gap in machine learning methods for legal document analysis, Humanit Soc Sci Commun 10 (2023) 1–15. URL: https://www.nature.com/articles/s41599-023-01693-z. doi:10.1057/s41599-023-01693-z, publisher: Palgrave.

[14] H. Kataria, A. Gupta, NLP-Titan at SemEval-2023 Task 6: Identification of Rhetorical Roles Using Sequential Sentence Classification, in: A. K. Ojha, A. S. Doğruöz, G. Da San Martino, H. Tayyar Madabushi, R. Kumar, E. Sartori (Eds.), Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 1365–1370. URL: https://aclanthology.org/2023.semeval-1.189/. doi:10.18653/

       `v1/2023.semeval-1.189`.

[15] Public.Resource.Org, vLex, Fastcase, State Regulations Available In Bulk : By the People, For the People : Free Download, Borrow, and Streaming, 2024. URL: https://archive.org/details/state.regulations.bulk.

[16] YData, ydata-profiling, 2024. URL: https://docs.profiling.ydata.ai/latest/.

[17] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, I. Androutsopoulos, LEGAL-BERT: The Muppets straight out of Law School, in: T. Cohn, Y. He, Y. Liu (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2020, Association for Computational Linguistics, Online, 2020, pp. 2898–2904. URL: https://aclanthology.org/2020.findings-emnlp.261/. doi:`10.18653/v1/2020.findings-emnlp.261`.

[18] L. Tunstall, N. Reimers, U. E. S. Jo, L. Bates, D. Korat, M. Wasserblat, O. Pereg, Efficient few-shot learning without prompts, 2022. URL: https://arxiv.org/abs/2209.11055. doi:`10.48550/ARXIV.2209.11055`.

[19] GLiClass: Generalist and Lightweight Model for Sequence Classification, 2024. URL: https://huggingface.co/knowledgator/gliclass-small-v1.0.

[20] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 2623–2631. URL: https://doi.org/10.1145/3292500.3330701. doi:`10.1145/3292500.3330701`.