

Effectively Leveraging BERT for Legal Document Classification

Nut Limsopatham
Microsoft Corporation
Redmond, WA

`nut.limsopatham@microsoft.com`

Abstract

Bidirectional Encoder Representations from Transformers (BERT) has achieved state-of-the-art performances on several text classification tasks, such as GLUE and sentiment analysis. Recent work in the legal domain started to use BERT on tasks, such as legal judgement prediction and violation prediction. A common practise in using BERT is to fine-tune a pre-trained model on a target task and truncate the input texts to the size of the BERT input (e.g. at most 512 tokens). However, due to the unique characteristics of legal documents, it is not clear how to effectively adapt BERT in the legal domain. In this work, we investigate how to deal with long documents, and how is the importance of pre-training on documents from the same domain as the target task. We conduct experiments on the two recent datasets: ECHR Violation Dataset and the Overruling Task Dataset, which are multi-label and binary classification tasks, respectively. Importantly, on average the number of tokens in a document from the ECHR Violation Dataset is more than 1,600. While the documents in the Overruling Task Dataset are shorter (the maximum number of tokens is 204). We thoroughly compare several techniques for adapting BERT on long documents and compare different models pre-trained on the legal and other domains. Our experimental results show that we need to explicitly adapt BERT to handle long documents, as the truncation leads to less effective performance. We also found that pre-training on the documents that are similar to the target task would result in more effective performance on several scenario.

1 Introduction

Recent advances in deep learning contribute to effective performances of several natural language processing (NLP) tasks on legal text documents, such as, violation prediction (Chalkidis et al., 2020), overruling prediction (Zheng et al., 2021), legal judgement prediction (Chalkidis et al., 2019),

legal information extraction (Chalkidis et al., 2018), and court opinion generation (Ye et al., 2018).

Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) has gained attentions from the NLP community due to its effectiveness on several NLP tasks (Chalkidis et al., 2019, 2020; Zheng et al., 2021). Importantly, the effectiveness of BERT is mainly due to the transfer learning ability that leverages semantic and syntactic knowledge from pre-training on a large non-labeled corpus (Devlin et al., 2019; Chalkidis et al., 2020; Zheng et al., 2021). However, Chalkidis et al. (2019) reported that BERT could not effectively handle long documents in the European Court of Human Rights (ECHR) dataset. In addition, pre-training BERT is costly. We need access to a special type of machines to pre-train BERT on a large corpora (Devlin et al., 2019; Liu et al., 2019; Zheng et al., 2021).

In this work, we investigate how to effectively adapt BERT to handle long documents, and how importance of pre-training on in-domain documents. Specifically, we will focus on two legal document prediction tasks, including ECHR Violation Dataset (Chalkidis et al., 2021) and Overruling Task Dataset (Zheng et al., 2021). The ECHR Violation Dataset provides a multi-label classification task. Given a list of facts described in free-texts, the task is to identify which articles of the European Convention were violated. The Overruling Task is a binary classification task to predict whether a legal statement will be later overruled by the same or higher ranking court (Zheng et al., 2021). We will discuss more about the two tasks in Section 4.2.

The main contributions of this paper are three-fold:

1. We investigate how to effectively adapt BERT to deal with long documents (i.e. documents containing more than 512 tokens).
2. We analyse the impacts of pre-training on different types of documents, especially in-

domain documents, on the performance of a fine-tuned BERT model.

3. We thoroughly evaluate the approaches to adapt BERT on long documents and pre-trained models to identify best practises for using BERT in legal document classification tasks.

The remainder of the paper is organised as follows. Section 2 further discusses related work and positions our work in the literature. Section 3 describes the two research questions we aim to answer in this paper and how we will find the answers. Sections 4 and 5 discuss our experimental setup and results. Section 6 provides more insight from the experimental results and answers the two research questions. Finally, we provide concluding remarks in Section 7.

2 Related Work

Legal documents, such as EU & UK legislation, European Court of Human Rights (ECHR) cases, Case Holdings On Legal Decisions (CaseHOLD) are normally written in a descriptive language in a non-structured text format and have unique characteristics that are different from those of other domains. In order to advance Legal NLP research, several tasks and datasets have been developed, including violation prediction on the ECHR dataset (Chalkidis et al., 2020), court overruling (Zheng et al., 2021), legal docket classification (Nallapati and Manning, 2008) and court view generation (Ye et al., 2018). In this work, we focus on text classification, which is a main research area of legal NLP.

Bidirectional Encoder Representations from Transformers (BERT) is a language representation model that is optimized during pre-training by self-training using a masked language model prediction and a next sentence prediction as a joint training objective (Devlin et al., 2019). As shown in Figure 1, BERT model architecture is built upon a multi-layer bidirectional Transformer encoder of Vaswani et al. (2017), where the number of input tokens is limited to 512. Pre-training BERT enables effective transfer learning from a large dataset before fine-tuning the model on a specific task (Devlin et al., 2019; Vaswani et al., 2017). Importantly, Devlin et al. (2019) used this transfer learning method to achieve the state-of-the-art performance on several NLP datasets, such as GLUE (Wang et al., 2018), SQuAD (Rajpurkar et al., 2016), Concept Nor-

malisation (Limsopatham and Collier, 2015, 2016) and Novel Named Entity Recognition (Derczynski et al., 2017). In particular, when fine-tuning BERT, we normally add a classification layer (either Soft-Max or Sigmoid) on the C (or CLS) representation in BERT output layer, in order to compute the prediction probabilities as in Figure 1. In the legal domain, Zheng et al. (2021) found that pre-training BERT on legal documents before fine-tuning on particular tasks lead to a better performance than pre-training BERT on general documents. However, Chalkidis et al. (2019) found that BERT did not perform well on the violation prediction task due to the length of the documents that are mostly longer than 512 tokens. They dealt with the long legal documents by using a hierarchical BERT technique (Chalkidis et al., 2019). Difference from the previous work, we investigate the effectiveness of variances of pre-trained BERT-based models and compare several methods to handle the long legal documents in legal text classification.

Several attempts (Beltagy et al., 2020; Zaheer et al., 2020; Pappagari et al., 2019) have been made to enable BERT-like models to work on documents with more than 512 tokens. For example, Beltagy et al. (2020) and Zaheer et al. (2020) used several different attention-mechanism techniques, such as global attentions and sliding window attentions to enable learning on a longer number of tokens. Pappagari et al. (2019) investigated different approaches to apply BERT on sequence chunks of texts in a document before aggregating the features using techniques, such as max pooling and mean pooling. In this work, we adapt these techniques to learn how to effectively use BERT on long legal documents.

3 Research Questions

This section discusses research questions we aim to answer in this paper.

- RQ1 For legal text classification, does pre-training on the in-domain documents lead to a more effective performance than pre-training on general documents?

To answer the first research question, we compare the performance of variances of BERT-based models that are pre-trained on general documents or different types of legal documents. Examples of the models are BERT (Devlin et al., 2019) and LEGAL-BERT (Chalkidis et al., 2020). The complete list of models will be described in Section 4.3.

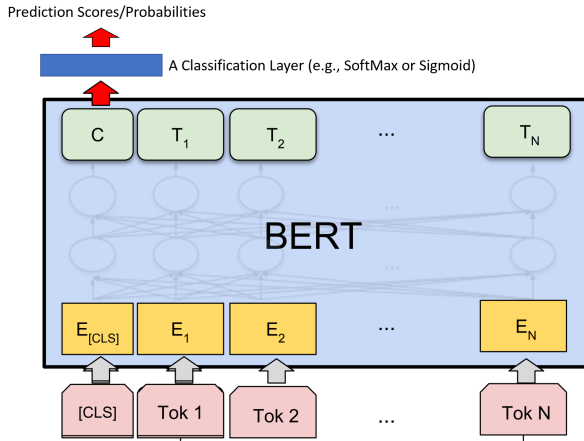


Figure 1: An example of fine-tuning BERT model on a classification task.

We fine-tune the models on the violation prediction and court overruling prediction tasks. We provide detailed information about the tasks in Section 4.2.

RQ2 How to adapt BERT-based models to effectively deal with long documents in legal text classification?

For RQ2, we discuss the performances of several BERT variances (including truncating long documents from the front or from the back), as well as hierarchical BERT models (Pappagari et al., 2019) that learn to combine output vectors of BERT using different strategies, such as, max pooling (Krizhevsky et al., 2012), and mean pooling (Krizhevsky et al., 2012) before applying a classification layer.

4 Experimental Setup

In Section 3, we have discussed the two main research questions to be investigated in this paper. In this section, we discuss the hyper-parameters of our models in Section 4.1. Then, we provide the details of the two legal text classification datasets (Section 4.2) and the variances of the BERT models (Section 4.3) used in the experiments.

4.1 Hyper-parameters

We use the transformers library¹ to develop and train BERT models in our experiments. For all experiments, we fine-tune the models using AdamW optimizer (Loshchilov and Hutter, 2017), learning rate of 5e-5 and a linear learning-rate scheduler. We

¹<https://huggingface.co/transformers/>

used a batch size of 16 and fine-tune the models on individual tasks for 5 epochs².

4.2 Datasets

4.2.1 ECHR Violation (Multi-Label) Dataset

The dataset contains 11k cases from the European Convention of Human Rights public database (Chalkidis et al., 2021). Each case contains a list of paragraphs representing facts in the case. The task is to predict which of the human right articles of the Convention are violated (if any) in a given case. The number of target labels are 40 ECHR articles (Chalkidis et al., 2021).

Table 1 provides statistical information of the ECHR Violation (Multi-Label) dataset. In particular, the dataset is separated into 3 folds: training, development and testing with the number of data points (cases) of 9,000, 1,000 and 1,000, respectively. On average, the number of tokens within a case is between 1,619 - 1,926, which are more than 512 tokens supported by BERT.

This is a multi-label classification task where we follow Chalkidis et al. (2021) and evaluate the classification performance in terms of micro-F1 score.

4.2.2 Overruling Task Dataset

This dataset is composed of 2,400 data-points, which are legal statements that are either overruled or not overruled by the same or the higher ranked court (Sulea et al., 2017; Zheng et al., 2021).

We show the statistics of the Overruling Task Dataset in Table 2. The average and the maximum number of tokens within a statement (i.e. case) is 21.94 and 204, respectively. Therefore, the BERT model should directly support this dataset without any alteration.

Follow Zheng et al. (2021), the task is modeled as a binary classification, where we conduct a 10 folds cross-validation on the dataset. Finally, we report the average of the F1-score across the 10 folds with a standard deviation value (Zheng et al., 2021).

4.3 Model Variances

Next, we discuss the variances of adapting *Model, which is a pre-trained BERT-based model from Table 3, to deal with long documents in the experiments. The used methods are as follows:

²Our preliminary results showed that 5 epochs resulted in most effective performances for most of the used models.

Fold	# Cases	Max # Words	Min # Words	Avg. # Words	Max # Labels	Min # Labels	Avg. # Labels
Training	9,000	35,426	69	1619.24	10	0	1.8
Development	1,000	14,493	84	1,784.03	7	0	1.7
Testing	1,000	15,919	101	1,925.73	6	1	1.7

Table 1: Statistics: ECHR Violation (Multi-Label).

# Cases	2,400
Max # Words	204
Min # Words	1
Avg. # Words	21.94
Ratio of Negative:Positive Labels	1:1.03

Table 2: Statistics: Overruling Task Dataset.

- **RR-*Model*** – Remove tokens in the rear of the input texts if the length is more than 512 and fine-tune the model on each classification task (similar to vanilla BERT (Devlin et al., 2019)).
- **RF-*Model*** – Remove tokens in the front of the input texts if the length is more than 512 and fine-tune the model on each classification task.
- **MeanPool-*Model*** – Apply the model on every chunk of n tokens before using a mean function to average the features from the same dimensions of the output vector representations of the chunks. Then, use a classification layer for each classification task. In this work, we set $n = 200$.
- **MaxPool-*Model*** – Apply the model on every chunk of n tokens before using a max function to select features from each dimension, based on the highest scores among the same dimensions of the output vector representations of the chunks, as a final vector representation. Then, use a classification layer for each classification task.

In addition, we include other two baselines that use different attention techniques, in order to cope with document longer than 512 tokens:

- **BigBird** – Fine-tuning the BigBird from Zaheer et al. (2020), which was pre-trained using English language corpora, such as BookCorpus and English portion of the CommonCrawl News, on each classification task. BigBird is a variance of BERT that uses several attention techniques, such as, random attention,

window attention and global attention, so that it can deal with documents longer than 512 tokens.

- **LongFormer** – Fine-tuning the LongFormer from Beltagy et al. (2020), which was pre-trained using BookCorpus and English Wikipedia, on each classification task. LongFormer is a variance of BERT that uses several attention techniques, such as, sliding window attention, dilated sliding window, and global attention, so that it can handle documents longer than 512 tokens.

5 Experimental Results

In this section, we compare the variances of BERT-based models on the ECHR Violation Dataset (Section 5.1) and Overruling Task Dataset (Section 5.2), respectively.

5.1 ECHR Violation Dataset

Table 4 reports the performances in terms of Micro-F1 score on different approaches to deal with long legal documents.

First, when comparing the performance of different BERT pre-trained models, we found that **-ECHR-Legal-BERT* outperformed the other pre-trained models across all of the methods used for adapting BERT to deal with long documents. This finding supported that pre-training BERT on the documents that are more similar to the task would lead to a better performance. Please note that **-ECHR-Legal-BERT* was pre-training using documents from the ECHR Violation Dataset, as mentioned in Table 3. Moreover, we observed that **-RoBERTa* performs comparably to **-Harvard-Law-BERT*. This provides an insight that if the in-domain documents (or documents similar to the task) are limited, pre-training the model on a large corpus could also lead to an effective performance.

Next, as shown in Table 4, when comparing the performance of *RR-** and *RF-** in Table 4, we found that the micro F-1 scores of *RR-** (e.g. 0.6466 for BERT) is worse than those of the corresponding *RF-** (e.g. 0.6803 for BERT). This shows

BERT	The BERT (bert-base-uncased) from Devlin et al. (2019) , which were pre-trained using BookCorpus and English Wikipedia.
ECHR-Legal-BERT	The BERT (bert-base-uncased) from Chalkidis et al. (2020) , which were pre-trained using legal documents including the ECHR dataset.
Harvard-Law-BERT	The BERT (bert-base-uncased) from Zheng et al. (2021) , which were pre-trained using the entire Harvard Law case corpus.
RoBERTa	The RoBERTa (roberta-base) from Liu et al. (2019) , which were pre-trained using English language corpora, such as BookCorpus and English portion of the CommonCrawl News. RoBERTa is a variance of BERT which trains only to optimize the dynamic masking language model.

Table 3: Pre-trained BERT-based Models used in the experiment.

Approach	Micro F-1
RR-BERT	0.6466
RR-ECHR-Legal-BERT	0.6699
RR-Harvard-Law-BERT	0.6590
RR-RoBERTa	0.6656
RF-BERT	0.6803
RF-ECHR-Legal-BERT	0.7090
RF-Harvard-Law-BERT	0.6896
RF-RoBERTa	0.6925
MeanPool-BERT	0.7075
MeanPool-ECHR-Legal-BERT	0.7196
MeanPool-Harvard-Law-BERT	0.7009
MeanPool-RoBERTa	0.6949
MaxPool-BERT	0.7110
MaxPool-ECHR-Legal-BERT	0.7213
MaxPool-Harvard-Law-BERT	0.7010
MaxPool-RoBERTa	0.7000
BigBird	0.7308
LongFormer	0.7238

Table 4: Comparing the performances on ECHR Violation Dataset.

that for this ECHR Violation Dataset, the back sections of the cases are more important than the front sections. Importantly, removing texts at the back of the input as suggested by [Devlin et al. \(2019\)](#) could lead to a poor performance. In addition, the best approach, *MaxPool-ECHR-Legal-BERT*, achieved 0.7213 Micro F-1 score, which was significantly better than any of the *RR*-* and *RF*-*, supported that truncation worsened the classification performance.

Finally, we observed that *BigBird* and *LongFormer* (Micro-F1 score 0.7308 and 0.7238, respectively) outperformed the other baselines that adapted BERT to deal with longer documents. This supported that *BigBird* and *LongFormer* that were explicitly designed to with long documents using different variances of attention techniques could lead to a better performance than aggregating results from applying BERT on multiple chunks of text.

5.2 Overruling Task Dataset

In this section, we discuss the performances on the Overruling Task Dataset. As discussed in Section 4.2.2, the lengths of the documents in this dataset are shorter than 512 tokens. Therefore, we can directly use BERT without any changes.

Table 5 reported the performance in terms of F-1 score averaged across 10 folds cross-validation, along with the standard deviation (STD). From Table 5, we observed that *Harvard-Law-BERT* and *ECHR-Legal-BERT* achieved the best and the 2nd best performances (0.9756 and 0.9725, respectively). This supported the impacts of pre-training on the in-domain documents. Meanwhile, *RoBERTa* achieved the 3 rank (0.9683 F-1 score) demonstrated that if no in-domain documents available, pre-training on a large corpus could also be effective. These results are inline with the findings in Section 5.1.

On the other hand, *BigBird* and *LongFormer* (0.9570 and 0.9569, respectively) performed a marginally worse than the other approaches. This could be due to the fact that *BigBird* and *LongFormer* are explicitly modelled to deal with long documents. Specifically, for shorter documents al-

Approach	Mean F1 \pm STD
BERT	0.9656 \pm 0.010
ECHR-Legal-BERT	0.9725 \pm 0.005
Harvard-Law-BERT	0.9756 \pm 0.010
RoBERTa	0.9683 \pm 0.010
BigBird	0.9570 \pm 0.010
LongFormer	0.9569 \pm 0.009

Table 5: Comparing the performances, in terms of F1-score, of different BERT pre-trainings on the Overruling Task Dataset.

lowing multi-head attentions to freely attend to any tokens would lead to a more effective performance than restricting them to be on particular sliding windows or specific areas (e.g. global attentions or randomized attentions).

6 Discussions

In this section, we provide further discussions on the experimental results from Section 5, in order to answer the research questions posed in Section 3.

RQ1 For legal text classification, does pre-training on the in-domain documents lead to a more effective performance than pre-training on general documents?

Yes, based on the experiments on both datasets, the model pre-trained on documents in the legal domain (*ECHR-Legal-BERT* and *Harvard-Law-BERT*) achieved the highest performance as shown in Tables 4 and 5, respectively. In addition, as discussed in Sections 5.1 and 5.2, *RoBERTa* achieved competitive performances on both datasets supported that pre-training on a large dataset could be a good option, if in-domain data cannot be obtained.

RQ2 How to adapt BERT-based models to effectively deal with long documents in legal text classification?

From the performances of *RR-* and *RF-*, in Section 5.1, we found that truncating long documents (on either ends) lessened the classification performance due to the lost of data. From the experimental result reported in Table 4, *BigBird* and *LongFormer* (even though not pre-trained on in-domain documents) outperformed other approaches that adapted BERT to deal with long documents. This highlighted the importance of explicitly handling

long documents during designing the model architecture. Next, both *MaxPool-** and *MeanPool-** achieved F-1 performances that are markedly better than the other approaches. Therefore, it is the most desirable to use *BigBird* or *LongFormer* that were explicitly designed to deal with long legal documents. An alternative method but less effective is to apply BERT on chunks of n tokens before using appropriate function (e.g. max or mean) to aggregate the vector representation across all the chunks before applying a classification layer, as described in Section 4.3.

7 Conclusions

We have discussed the challenges of using BERT for text classification in the legal domains, and posed two research questions regarding the pre-train documents and how to cope with long documents. To answer the research questions, we conducted the experiments on the ECHR Violation Dataset and the Overruling Task Dataset. Our experimental result showed that the models pre-trained on the domain similar to the task enhanced the performance. In addition, the experiments on ECHR Violation Dataset supported that truncating or discarding parts of a document resulted in a poor performance. Importantly, *BigBird* and *LongFormer*, which explicitly handled long documents using different attention techniques, achieved the best performance on long legal document classification. Alternatively, applying BERT on chunks of texts before aggregating the vector representation across all of the chunks using an appropriate function (e.g. max or mean) could achieve a reasonable result.

References

- Iz Beltagy, Matthew E Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*.
- 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.
- Ilias Chalkidis, Ion Androutsopoulos, and Achilleas Michos. 2018. Obligation and prohibition extraction using hierarchical rnns. *arXiv preprint arXiv:1805.03871*.
- Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos.

2020. Legal-bert: “preparing the muppets for court”. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 2898–2904.
- Ilias Chalkidis, Manos Fergadiotis, Dimitrios Tsarapatsanis, Nikolaos Aletras, Ion Androutsopoulos, and Prodromos Malakasiotis. 2021. Paragraph-level rationale extraction through regularization: A case study on european court of human rights cases. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Mexico City, Mexico. Association for Computational Linguistics.
- Leon Derczynski, Eric Nichols, Marieke van Erp, and Nut Limsopatham. 2017. Results of the wnut2017 shared task on novel and emerging entity recognition. In *Proceedings of the 3rd Workshop on Noisy User-generated Text*, pages 140–147.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. pages 4171–4186.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105.
- Nut Limsopatham and Nigel Collier. 2015. Adapting phrase-based machine translation to normalise medical terms in social media messages. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1675–1680.
- Nut Limsopatham and Nigel Collier. 2016. Normalising medical concepts in social media texts by learning semantic representation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1014–1023.
- 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*.
- Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- Ramesh Nallapati and Christopher D Manning. 2008. Legal docket classification: Where machine learning stumbles. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 438–446.
- Raghavendra Pappagari, Piotr Zelasko, Jesús Villalba, Yishay Carmiel, and Najim Dehak. 2019. Hierarchical transformers for long document classification. In *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pages 838–844. IEEE.
- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392.
- Octavia-Maria Sulea, Marcos Zampieri, Shervin Malmasi, Mihaela Vela, Liviu P Dinu, and Josef Van Genabith. 2017. Exploring the use of text classification in the legal domain. *arXiv preprint arXiv:1710.09306*.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.
- Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R Bowman. 2018. Glue: A multi-task benchmark and analysis platform for natural language understanding. In *International Conference on Learning Representations*.
- Hai Ye, Xin Jiang, Zhunchen Luo, and Wenhan Chao. 2018. Interpretable charge predictions for criminal cases: Learning to generate court views from fact descriptions. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1854–1864.
- Manzil Zaheer, Guru Guruganesh, Kumar Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, et al. 2020. Big bird: Transformers for longer sequences. In *NeurIPS*.
- Lucia Zheng, Neel Guha, Brandon R Anderson, Peter Henderson, and Daniel E Ho. 2021. When does pre-training help? assessing self-supervised learning for law and the casehold dataset of 53,000+ legal holdings. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, pages 159–168.