Legal-HNet: Mixing Legal Long-Context Tokens with Hartley Transform

Anonymous ACL submission

Abstract

Since its introduction, the transformers architecture has seen great adoption in NLP applications, but it also has limitations. Although the self-attention mechanism allows for generating very rich representations of the input text, its effectiveness may be limited in specialized domains such as legal, where, for example, language models often have to process very long texts. In this paper, we explore alternatives to replace the attention-based layers with simpler token-mixing mechanisms: Hartley and Fourier transforms. Using these non-parametric techniques, we train models with long input documents from scratch in the legal domain setting. We also introduce a new hybrid Seq2Seq architecture, a no-attention-based encoder connected with an attention-based decoder, which performs quite well on existing summarization tasks with much less compute and memory requirements. We believe that similar, if not better performance, as in the case of long correlations of abstractive text summarization tasks, can be achieved by adopting these simpler infrastructures. This not only makes training models from scratch accessible to more people, but also contributes to the reduction of the carbon footprint during training.

1 Introduction

The Transformer architecture has gained rapid and widespread adoption in the Natural Language Processing (NLP) field thanks to its outstanding results. At its core is a self-attention mechanism - an inductive bias that links each token in the input through a weighted basis for the relevance of every other token. Each hidden unit is represented on the basis of the hidden units in the previous layer. Many papers have described this abstraction, on an increasingly complex and higher-order basis than the previous one, as the main mechanism to define more complex features such as syntactic and semantic relations (Tenney et al., 2019; Vig and Belinkov, 2019; Clark et al., 2019; Voita et al., 2019).

Although the self-attention mechanism often works well, it is greedy for both computational and memory resources, and there are many studies in the literature where different techniques have been adopted to replace the layer with quadratic time complexity (Beltagy et al., 2020a; Child et al., 2019a; Choromanski et al., 2020, 2021; Dosovitskiy et al., 2021; Jiao et al., 2020; Katharopoulos et al., 2020; Kim and Awadalla, 2020; Kitaev et al., 2020b; Vyas et al., 2020; Wang et al., 2020).

Our starting point is the work by Lee-Thorp et al. (2021), the so-called Fourier Network (FNet), with the aim of addressing the long-document problem in legal texts. They prove that Transformer encoder architectures can be sped-up, with limited accuracy costs, by replacing the self-attention sub-layers with simple linear Fourier Transform (FT) that "mix" input tokens.

Although there are many other studies addressing the replacement of the self-attention sub-layer with a linear transformation, we believe that FNet provides a simple and elegant way to address it, achieving the best compromise between speed and accuracy.

In this paper, we explore the diverse Fourier-class Transform using a different real output propagation to avoid the complex number and to improve the original Real Fourier Transform (RFT) architecture made in Lee-Thorp et al. (2021). We show that by replacing the RFT layer with a Hartley Transform (HT) (Bracewell, 1995) we obtain better metrics than FNet for most of the tasks where the two Neural Network (NN) architectures were compared. We call this Language Model (LM) architecture Hartley Network (HNet)[1].

Given that HNet does not require any further

[1] Lee-Thorp et al. (2021) have tried a HT, and because, in the paper, they got no improvement in GLUE benchmark, they gave little room for further testing

computational steps in comparison with FNet, it can be used for longer sequence lengths at the same computational cost as BERT-based solutions, which can be particularly effective both to solve the problem of long documents in legal domain, and to afford high training and inference costs or the creation of ad-hoc language models (i.e. with a specific vocabulary and domain). Furthermore, in our downstream tasks, HNet reduces the performance gap introduced by the original FNet architecture.

We trained three new legal domain language models using the FNet and HNet architectures using different sequence lengths as input (4096 and 8192 tokens for FNet and 4096 tokens for HNet). We evaluated them for two public summarization datasets, BillSum and PubMed, and for classification and question & answering tasks using the LexGlue benchmark. We added the latter benchmark for the sake of completeness, since it is the equivalent of the GLUE benchmark in the legal NLP world (it allows us to make an analogy with the GLUE benchmark reported in in the FNet paper). Notably, the aforementioned benchmark is based on short texts, while the motivation of this study lies in improving performance for long-text problems.

Indeed, to test the ability to capture long-distance dependencies in the text, we mainly evaluated HNet (and the two FNet LMs) on the task of automatic (abstractive) summarization, as done in other recent studies (Nikolaus and Giofré, 2022; Zhang et al., 2020). Abstractive summarization refers to the task of capturing the most important concepts/ideas from the (long) document and then rewriting it in a shorter passage in a grammatical and logically coherent way (Chen et al., 2019). Since an abstractive summarisation task requires an Encoder-Decoder (ED) model architecture and few studies have been carried out in the literature regarding those based on Fourier-class Transformers (Kiruluta et al., 2021), we have also proposed a new Seq2Seq architecture that adopts the lightweight implementation provided by Fourier-class based Encoders and utilises the performance of the attention based Decoders.

We used the BillSum benchmark, as a domain-specific summarization task; and the PubMed benchmark, to evaluate the model's ability outside the legal context (i.e., in the biomedical context). We decided on the biomedical domain because like the legal one, it requires a specific and complex vocabulary so as to prove that our LMs can be open-domain. The proposed models achieved in both cases better metrics than a Transformer architecture and approached the recent state-of-the-art (SOTA) (i.e. BudgetLongformer) the well-known previous one (i.e. PEGASUS, which was trained using the Gap-Sentences generation task and with more resources and data respect our LMs in the pretraining phase[2]), see Tables 2, and 3. Furthermore, both BudgetLongformer and Pegasus, in addition to having many more trainable parameters and a greater environmental impact, they also require a larger memory allocation, which limits the maximum number of tokens provided as input by both the encoder and the decoder.

It is important to note that this performance was achieved with computationally and memory-light allocation models, i.e. HNet and FNet, trained for minimal pretraining phase: one million steps and with a data catalogue of less than 500k documents (only legal documents), which means that HNet can be trained faster and cheaper using less documents. For example, RoBERTa (Liu et al., 2019) or PEGASUS (Zhang et al., 2020), which are both models that HNet and FNet are compared to, required substantially longer training and greater related costs. RoBERTa was trained for 1024 GPU days, whereas both HNet and FNet only needed 20 GPU days each (estimated using a 16GB NVIDIA V100 GPUs)

Contributions

The contributions of this paper are six-fold:

- We investigated further into the use of HT instead of FT in the case of LMs in the legal domain, calling it HNet, and finding that it performs slightly better with equal CO2 impact and the capability of handling long documents.
- We have trained from scratch and built 3 new legal LMs for handling long legal documents: one HNet with input sequence up to 4096 tokens and two FNet(s) with input sequence up to 4096 and 8192 tokens.
- We proposed and evaluated a new Encoder-Decoder diagram where no-attention based encoders can be connected with attention based decoders, to achieve the best compromise in terms of efficiency (a light encoder) and quality (a performing decoder with cross attention).

[2] our LMs has only been pretrained on 450k legal documents as well as for the tokenizer.

- We approach SOTA on two public benchmarks, BillSum and Pubmed, using in the worst scenario 50 percent less computational resources.

- We note that models based on Fourier-class transforms become truly beneficial when tasks require long documents, for example, with an average length much longer than 512 tokens, both for obvious cost reasons and because of the reducing performance gap with respective attention mechanism-based models.

- On the LexGLUE benchmark (Chalkidis et al., 2021), despite the obvious emphasis on covering classification tasks for short documents, our models achieve equivalent or even better performance to those found by Lee-Thorp et al. (2021). Indeed, we measured a performance, with respect to the BERT-based model, of 92% overall for the base FNet-4096, of 94% for the base HNet-4096, and of 87% for the LM released by Lee-Thorp et al. (2021).

Main Research Questions

In this work, we pose and examine six main research questions:

RQ1: *Is it possible to create a domain specific (e.g. legal) FNet based LM from scratch, reducing costs and carbon footprint compared with self-attention based LMs?*

RQ2: *Does using the Hartley transform instead of Fourier in the FNet based architecture to create a domain specific (e.g. legal) LM from scratch lead to better results?*

RQ3: *Is it possible to create a "competitive" Seq2Seq model, which can approach the performance of SOTA ones, using Fourier/Hartley layers?*

RQ4: *How do our models compare with other models on the challenging summarization task? Particularly in the case of a legal domain-specific benchmark such as BillSum?*

RQ5: *How well do our models generalize to other domains, for example in the biomedical domain, as evaluated by the PubMed summarization benchmark?*

RQ6: *How do our LMs compare with other models on the Text Classification (TC) benchmark LexGLUE?*

2 Related Work

Hartley Transforms in Neural Networks

As shown in Lee-Thorp et al. (2021) , the transformer encoder architectures can be sped-up, with limited accuracy costs, by replacing the self-attention sub-layers with a simple linear transformation: Fourier transform. Fourier layers are basically a simpler token mixing mechanism. In the FNet article, although HT was tested on the GLUE benchmark (Wang et al., 2018), it obtained comparable results to FT, giving no room for further investigations. FT acts on real-valued functions getting complex-valued functions. On the other hand, HT is a way to transform real-valued functions to real-valued functions (Bracewell, 1995). Hartley related transforms have also shown to perform better than their corresponding FT counterparts in Convolutional Neural Network (CNN) related optimizations (Mozafari et al., 2021; Hao et al., 2020), and in multi-task learning with multiple languages and modalities (Lee-Thorp and Ainslie, 2022; Armitage et al., 2020).

Domain-Specific Language Models

Previous research has shown that domain specific pretraining of language models from scratch show better results compared to continual pretraining of general-domain language models (Gu et al., 2021). Domain-specific pretraining on datasets of specialized domains such as law (Chalkidis et al., 2020; Xiao et al., 2021), biology (Lee et al., 2019), scientific articles (Beltagy et al., 2019), or clinical documents (Li et al., 2022) have shown promising results.

Long Document Processing

In the legal domain, texts tend to span multiple pages, ranging from 10s to 100s of pages, which translates to tens of thousands tokens. The quadratic time and memory requirement of the attention typically used in the transformer architecture (Vaswani et al., 2017) prohibits efficient processing of sequences longer than 512 tokens on current hardware.

This has led to research focusing on tweaking the attention layers by making the attention matrices sparse, like the efficient transformers (Tay et al., 2020; Child et al., 2019b; Beltagy et al., 2020b; Zaheer et al., 2021; Roy et al., 2021; Kitaev et al., 2020a; Tay et al., 2021). While other researchers focus on replacing the attention mechanism with

Pretraining Subset	Dataset Size	# Doc Token Avg.	# Documents (train/val)
Total	**9.2GB**	**2.7k**	**450k / 54.2k**
CL Opinions[1,a]	2.7GB	2.3k	148k / 24.9k
CL Docket Entries and Court Filings[1,a]	2.3GB	3.9k	103k / 9.4k
CUAD[2,a] (Hendrycks et al., 2021)	1.6GB	9.3k	22.4k / 2.0k
Practical Law Tax and Legal Commentary[1,b]	1.5GB	2.0k	100k / 12.4k
Laws[1,a]	0.5GB	3.2k	11.8k / 1.0k
Posture50k[2,a] (Song et al., 2022)	0.3GB	1.3k	20.1k / 1.7k
EUR-Lex[2,b] (Chalkidis et al. (2019), 2019)	0.2GB	0.6k	43.1k / 2.8k

Table 1: The datasets used for pretraining our models: (1) internal datasets; (2) public dataset; (a) US English; (b) UK English. CL is short for Court Listener.

simpler token mixing mechanisms, like FNets (Lee-Thorp et al., 2021), and MLP-Mixer (Tolstikhin et al., 2021; Fusco et al., 2022).

3 Datasets

In this section, we briefly introduce the datasets used in our experiments.

3.1 Pretraining

Training a new language model from scratch requires documents in the range of hundreds of thousands with considerable diversity, as minimum range[3]. Our domain specific legal language model uses 7 different datasets: 3 public datasets and 4 internal datasets. It was trained on a total of around half a million documents. Since different datasets have different number of documents, we tried to ensure the model is not biased towards a particular dataset by making sure the number of documents from the main datasets are in the similar range. All documents with less than 200 tokens are filtered out. CUAD (Hendrycks et al., 2021), Posture (Song et al., 2022) and EUR-Lex (Chalkidis et al. (2019), 2019) datasets are the three public datasets used in the pretraining. Since the EUR-Lex dataset is used in the LexGLUE benchmarking tasks, we use only the training and validation datasets of the same. For further information about the datasets refer Table 1.

3.2 BillSum

Kornilova and Eidelman (2019) introduced a legislative summarization dataset from 21K United States (US) bills from 1993 to 2018. It is challenging due to the technical nature and complex structure of the bills. Additionally, the bills are rather long, ranging from 1k to 4k tokens with their summaries being up to $\sim$ 1k tokens long (see Appendix H for more details).

[3]Massive LMs can be fed up-to hundreds of millions of documents (Brown et al., 2020)

3.3 PubMed

Cohan et al. (2018) introduced another challenging summarization dataset in a specialized domain (scientific articles from the biomedical domain). It includes 133K scientific papers together with their abstracts in English. The papers are 3k words long on average and the summaries (abstracts) 200 words. Thus, similar to the BillSum dataset, this dataset is well suited as a test bed for methods capable of long document summarization. Note, that in this dataset the domain is vastly different from the legal domain (see Appendix H for more details).

3.4 LexGLUE

Chalkidis et al. (2021) recently introduced a benchmark for the English legal domain called LexGLUE. LexGLUE contains six TC tasks and one Question Answering (QA) task comprising diverse legal data such as US court decisions and contracts, terms of service documents, European Union (EU) legislation and cases from the European Court to Human Rights (ECtHR). There exists a public leaderboard of diverse models on GitHub[4], with Legal-BERT (Chalkidis et al., 2020) performing best.

The LexGLUE benchmark focuses on evaluating LMs in legal TC and QA tasks, and is, now, one of the benchmark standards in the legal NLP. For these reasons, we feel it is important for the community to evaluate our LMs against that benchmark, despite the fact that, in LexGLUE, 4 out of 7 tasks involve documents with input lengths lower than 512 tokens on average. The remaining 3 tasks, the ECtHR A and B tasks and the SCOTUS tasks involve documents with a longer span, but the median of the first two tasks is also less than 1000 tokens. Usually, legal documents are much longer than 512 tokens and thus this distribution might not be representative of real-world tasks. One can also speculate that shorter input length tasks may be better handled by short-input models (e.g., BERT, RoBERTa, Legal-BERT, etc.).

4 Neural Network Architecture

In this section, we describe the LM architectures used. Please, see the Appendices C for the setup of the pretraining LM experiments, and D for the downstream benchmarks experiments. In all our

[4]https://github.com/coastalcph/lex-glue

experiments, we used the HuggingFace (HF) transformers library (Wolf et al., 2020) available under an Apache 2.0 license.

4.1 HNet

HNet, as FNet, is an attention-free Transformer architecture, wherein each layer consists of a Hartley mixing sub-layer followed by a feed-forward sub-layer. The architecture is shown in Fig. 1. As you can see, the Fourier sub-layer of each Transformer encoder layer is replaced with a Hartley sub-layer, which applies a 2D Fast Fourier Transform (FFT) to its (sequence length, hidden dimension) embedding input – one 1D FFT along the sequence dimension, F_{seq}, and one 1D FFT along the hidden dimension, F_h. From the 2D FFT thus produced, its real part is taken and subtracted from its imaginary part to obtain a 2D Fast Hartley Transform (FHT):

$$H = R(F_{seq}(F_h(x))) - I(F_{seq}(F_h(x))). \quad (1)$$

Unlike in the FNet, where only the real part (i.e. only the first term in Eq. 1), was taken into account, i.e. RFT, so as not to deal with the complex numbers typical of the FT; HT already guarantees real-valued output including the imaginary part contributions, too. Since anything that can be done with FT can be done with HT and vice versa (Bracewell, 1995; Paraskevas et al., 2015), HT is a readily adoptable tool, when looking for a real-valued output. We also implemented HNet in the HF library based on its FNet implementation.

In addition to the expression in Eq. 1, we tested several other real-valued configurations, but among the various experiments, HT was the one that showed the most potential for success (see Appendix A for more details).

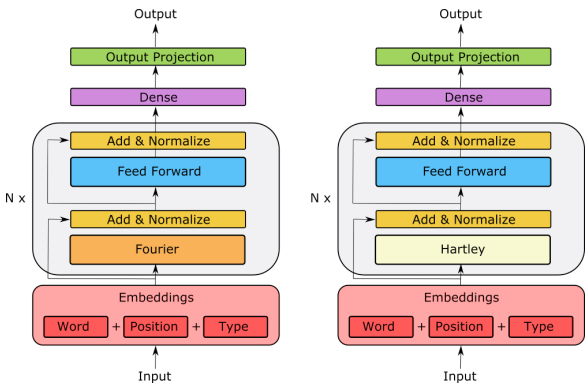


Figure 1: On the left, FNet encoder architecture with N encoder blocks. On the right, HNet encoder architecture with N encoder blocks.

4.2 Right-side Attention Encoder-Decoder

As the FNet paper proposed only an encoder architecture, Kiruluta et al. (2021) first suggested 3 different Seq2Seq architectures to overcome this shortcoming. In all of their proposed Seq2Seq diagrams, they removed the multi-headed self-attention layers in a transformer (both encoder and decoder): either by direct removal, i.e. moving the Fourier token mixing completely outside the transformer by Fourier transforming both inputs, or by replacement with a Fourier layer. Although the latter solution, what they call Hybrid-FNet Seq2Seq, is their best solution, and is more similar to a classical transformer diagram, their results, as reported in Table 3, are a long way from those offered by alternatives such as PEGASUS and Longformer.

For these reasons, we have designed a new Seq2Seq infrastructure with an encoder without attention and a decoder with attention, which for clarity we can call the Right-side Attention Encoder-Decoder (RAED) model (See Fig. 2). It is based on the idea that an attention-based decoder gives a moderate computational load to the entire model infrastructure because it is generally powered by short text. Furthermore, in the main NLP tasks, where Seq2Seq infrastructure is involved, more emphasis is placed on the output generated by the decoder. We believe, therefore, that the multi-headed self-attention mechanism acting on the decoder tokens through a dense representation of the context via Fourier/Hartley descriptors (i.e. encoder output) is crucial to obtaining AI generated text which is indistinguishable from a human with a reduced computational cost.

5 Results

In the following three sections, we present the results on the BillSum dataset, the PubMed dataset and the LexGLUE benchmark.

5.1 BillSum

Our results on the BillSum dataset are presented in Table 2.

We observe that our RAED models clearly exceed the baseline of the original article (DOC + SUM), even though their model is based on BERT-large which contains almost 2 times more encoder parameters and its pretraining cost is 10 times more expensive. Furthermore, our RAED-HNet-4096 base model outperforms a transformer model, and competes with the PEGASUS-base model (Zhang

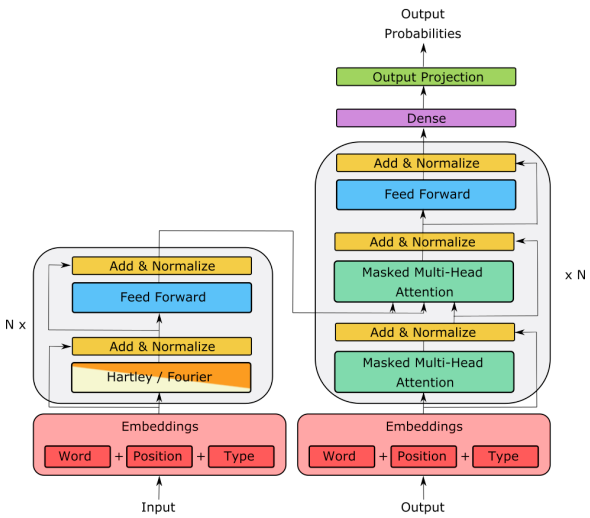


Figure 2: Right-side Attention Encoder-Decoder (RAED) diagram. We tested also another configuration where we replaced these two Multi-head Attention green boxes with a Hartley/Fourier in the same manner as was done by Kiruluta et al. (2021), the so-called FNet-Transformer, but without equal success as RAED with downstream tasks.

et al., 2020) (35.14 vs. 37.78 Rouge-L), which is pretrained using the Gap-Sentences task specifically designed for abstractive summarization. Vice versa, it fails to reach the new SOTA, i.e. BudgetLongformer (Nikolaus and Giofré, 2022). This has been trained with an ELECTRA task, which unlike the Masked Language Modeling (MLM) task turns out to be very efficient in summarization tasks. However, our models become very attractive solutions when cost and environmental impact are also taken into account, without compromising too much on quality. In particular, they almost halve training costs and reduce model operational costs by 17%.

We conclude that the RAED architecture is effective, with minimal compute requirements for long-input summarization in-domain.

5.2 PubMed

Our results on the PubMed dataset are presented in Table 3.

Similar to the results on BillSum, our RAED models clearly outperform the Transformer-base model, in particular our RAED-HNet-4096 base (20.10 and vs. 19.02 Rouge-L). Furthermore, our RAED architecture exceeds the most performing Seq2Seq architecture, which includes FNet layers, proposed by Kiruluta et al. (2021).

As in BillSum, our LMs do not reach both PEGASUS and BudgetLongformer in terms of performance, but as described above they can be excellent alternatives when longer documents need to be processed and/or there are cost and speed constraints.

Note, that we pretrain, as BudgetLongformer, on a much narrower domain than PEGASUS (legal text vs. C4). For example, our tokenizer and model has never seen medical data during its pretraining phase. In addiction, our tokenizer has 3 times less tokens than the PEGASUS tokenizer (96k) and has half the number of tokens as of the BudgetLongformer tokenizer (64k).

In conclusion, the RAED architecture is effective even on an out-of-domain downstream summarization task.

5.3 LexGLUE

Our results on the LexGLUE benchmark are presented in Table 6 in Appendix E.

While, in Table 4, we reported the performance of the various LMs against a reference LM, based on defined downstream metrics (e.g. μ-f1-score):

$$P = \left\langle \frac{metrics_{LM,task}}{metrics_{LM_{ref},task}} \right\rangle_{task} .$$

This is similar to what was carried out by Lee-Thorp et al. (2021) during the evaluation of the GLUE benchmark for FNet. As we can see our models have a far better performance than FNet one[5], which was not pre-trained on legal documents.

In particular we note that our legal FNet, FNet-4096, achieves almost 92% performance against BERT/RoBERTa models, similar to what the FNet authors get for the GLUE benchmark. While the legal model based on a Hartley layer, i.e. HNet-4096, achieves 94% performance for same reference models, reducing the gap between Attention-based and Fourier-based models.

While the performance compared to models with legal domain knowledge is lower, such as CaseLaw-BERT and legal-BERT. In support of our models, the latter have a contaminated testing set since a part of those documents were seen during MLM pretraining.

As already observed in Nikolaus and Giofré (2022), models trained only on long documents or with emphasis on long-context, hardly reach the SOTA for these LexGLUE tasks. This, indeed, suggested that the LexGLUE benchmark needs a more

[5]checkpoint released by Google in HF, https://huggingface.co/google/fnet-base.

Model (max-in-len->max-gen-len)	# Enc. Params ↓	Rouge-1 ↑	Rouge-L ↑	Speed (T↑ / I ↑)[it/s]
DOC + SUM (BERT large) (Kornilova and Eidelman, 2019)	340M	40.80	33.73	OOM
Transformer base (Zhang et al., 2020)	223M	44.05	30.98	0.80 (3.8x) / 0.21
PEGASUS base (Zhang et al., 2020)	223M	51.42	37.78	0.80 (3.8x) / 0.21 (1.0x)
BudgetLongformer base (diverse) (Nikolaus and Giofré, 2022)	255M	55.45	43.23	0.85 (4.0x) / 0.22 (1.0x)
RAED-FNet-4096 base (4096->512)	182M	47.64	34.51	1.40 (6.7x) / 0.26 (1.2x)
RAED-HNet-4096 base (4096->512)	182M	45.92	35.14	1.40 (6.7x) / 0.26 (1.2x)

Table 2: Results on the BillSum dataset. Enc. Params is short for Encoder Parameters. T and I are abbreviations for Train and Inference. We report the f1-scores of the Rouge metrics, and the speeds as number of iterations over a second, performed on a p3.2xlarge AWS instance with 1XV100 Nvidia GPU, batch size of 8, and an input sequence length of 1024 for the encoder and of 256 for the decoder (in parentheses, speed-ups compared to transformer architecture inference).

Model (max-in-len->max-gen-len)	# Enc. Params ↓	Rouge-1 ↑	Rouge-L ↑
Transformer base (Zhang et al., 2020)	223M	33.94	19.02
PEGASUS base (Zhang et al., 2020)	223M	39.98	25.23
BudgetLongformer base (Nikolaus and Giofré, 2022)	255M	41.16	26.53
Hybrid-FNet base (Kiruluta et al., 2021)	172M	35.60	14.50
RAED-FNet-4096 base (4096->512)	182M	37.24	19.98
RAED-FNet-8192 base (6144->512)	185M	37.32	20.17
RAED-HNet-4096 base (4096->512)	182M	37.52	20.10

Table 3: Results on the PubMed dataset. Enc. Params is short for Encoder Parameters. We report the f1-scores of the Rouge metrics.

Model	Reference Model	P (μ-f1, m-f1) ↑
FNet (Lee-Thorp et al., 2021)	BERT (Chalkidis et al., 2021)	87.4, 82.4
FNet-4096	BERT (Chalkidis et al., 2021)	91.5, 87.5
HNet-4096	BERT (Chalkidis et al., 2021)	93.9, 88.9
FNet-4096	RoBERTa (Chalkidis et al., 2021)	91.6, 88.6
HNet-4096	RoBERTa (Chalkidis et al., 2021)	93.8, 89.5
FNet-4096	CaseLaw-BERT* (Chalkidis et al., 2021)	89.2, 85.1
HNet-4096	CaseLaw-BERT* (Chalkidis et al., 2021)	91.9, 86.7
FNet-4096	Legal-BERT* (Chalkidis et al., 2021)	88.2, 83.8
HNet-4096	Legal-BERT* (Chalkidis et al., 2021)	91.4, 85.4

Table 4: Results on the LexGLUE Benchamrk. We evaluated the Performance of our LMs as Lee-Thorp et al. (2021) did in the original paper with the GLUE benchmark. LMs with asterix have seen these testing data during the MLM pre-training stage.

accurate description of the first 512 tokens (e.g. 4 over 7 tasks have a median token length far below 512, as shown in Appendix H), which could be obtained through a pretraining dataset with a comparable distribution of token inputs. We avoided focusing too much on this point since the purpose of the paper is to solve legal long documents as input.

Finally, we used the HyperParameters (HP) provided by the benchmark, which we can merely speculate that they are strongly related to BERT-like models (i.e. attention mechanism-like). Most of the models, for which the LexGLUE benchmark was performed, come from a BERT-checkpoint.[6] It is not possible to understand how these HPs were chosen by Chalkidis et al. (2021), but it is known how they can strongly influence the downstream performance (Liu and Wang, 2021; Dodge et al., 2020).

[6]Note that Legal-Bert, CaseLaw-Bert, Longformer and BigBird have been warm started from the BERT or RoBERTa checkpoint. Thus, they have been trained on short documents extensively, at least, during the first pretraining phase.

6 Conclusions and Future Work

6.1 Answers to Main Research Questions

RQ1: *Is it possible to create a domain specific (e.g. legal) FNet based LM from scratch, reducing costs and carbon footprint compared with self-attention based LMs?* Yes, it is possible to pretrain a domain-specific (legal) FNet based LM from scratch with lesser compute and memory, achieving comparable performance with self-attention based LMs.

RQ2: *Does using the Hartley transform instead of Fourier in the FNet based architecture to create a domain specific (e.g. legal) LM from scratch lead to better results?* Yes. As can be seen from the result Tables 2, 3, 4 and 6, for similar maximum sequence length models, Hartley transform based models show better results.

RQ3: *Is it possible to create a "competitive" Seq2Seq model, which can approach the performance of SOTA ones, using Fourier/Hartley layers?* Yes. As can be seen from Table 3, the proposed RAED architecture achieves comparable performance with the SOTA ones.

RQ4: *How do our models compare with other models on the challenging summarization task? Particularly in the case of a legal domain-specific benchmark such as BillSum?* As can be seen from Table 2, although our LMs do not beat the SOTA, they achieve comparable metrics with lesser compute, memory and data requirements.

RQ5: *How well do our models generalize to other domains, for example in the biomedical domain, as evaluated by the PubMed summarization benchmark?* As can be seen from Table 3, although our

models do not beat the SOTA, it achieves comparable metrics with lesser compute, memory and data requirements.

RQ6: *How do our LMs compare with other models on the Text Classification (TC) benchmark LexGLUE?* As can be seen from Tables 4 and 6, our LMs reduce the expected performance gap for legal tasks.

6.2 Conclusion

In this work, we investigate the Fourier-class Transform as a replacement to the self-attention mechanism to solve the long-document problem in the legal domain. We trained from scratch and successfully evaluated three new legal LMs, which can conveniently deal long documents. We found also that HT leads to better outcomes than FT, without additional computational cost. In addition, we also proposed a new architecture for a Seq2Seq model involving no-attention based LMs as encoders, going well beyond previous Fourier-based architecture outcomes, and achieving values close to current SOTA with about half the training costs and 17% less operational costs (i.e. inference).

6.3 Future Work

A potential future experiment could explore a wider HPs optimization, in particular during the BillSum and PubMed tasks, but also in MLM pre-training phase. We, indeed, limit only to a narrow exploration of different learning rates and batch sizes. For example, the tokenizer's vocabulary is another fairly incisive HP that could be tuned (e.g. by increasing it), to accommodate a more vocabulary-rich language (as that of legal). Probably also extending the pretraining dataset with not-legal documents could lead to more robust and generic LMs, which might achieve further gains for domain and out-domain downstream tasks.

In addition, future works could also check how the LMs perform for other in-of-domain tasks such as CUAD (Hendrycks et al., 2021), Posture50K (Song et al., 2022) for TC or the recently released MultiLexSum (Shen et al., 2022) for summarization; and for other out-of-domain tasks such as by trying one of the remaining 11 summarization datasets evaluated by Zhang et al. (2020).

Finally, another future work with a potential high return of investment, which we could not perform due to resource allocation constraints, is to replace the Fourier-class Transform with a Laplace Transform (LT). LT satisfies the causality principle[7], and provides a solution in terms of decaying exponents multiplied by a periodic part (i.e. FT). For these reasons, it could better mimic short & long correlations, and inherently assign a clockwise direction to the text sequence.

7 Limitations

Fourier-class LMs have the disadvantage of not achieving the same performance as their respective cousins based on attention mechanisms, in TC tasks. In particular, when a good description of the first N tokens is most important. For these reasons, they cannot be considered a solution if the quality of the result is prioritized.

Because of insufficient compute, we were not able to create the *large* versions of these 3 models. Therefore, we do not know how our models or the RAED architecture would scale in the case of *large* models. We imagine that it will go to produce better results but it is not possible for us to estimate whether they are meaningfully better.

Summarization tasks were evaluated only on the basis of the Rouge score, since it was the only data available for comparison with previous works. In the case of a proper evaluation of the RAED architecture, it would be necessary to conduct an evaluation through human experts, since it is known that the Rouge score can lead to misleading interpolations of the results. In particular, when the output of the model is a reformulation of the same concept. This unfortunately requires human involvement in the loop, which was not within the costs allocated for this research.

Out of the 9.2 GB of data used during the MLM pretraining, only 1.7 GB was in U.K. English with a mix of European and U.K. legal texts, while the rest was based on U.S. English with U.S. legal texts. Apart from the differences between U.S. and U.K. languages, from a legal perspective, the U.K. and U.S. systems are based on case law, while the European system is based on civil law. Therefore, we expect the models to be able to transfer knowledge in legal downstream tasks where U.S. English and the case-law system is present, while it is more difficult for texts in U.K. English and with the civil law system.

[7] it could be relevant in Casual Language Modelling

Ethics Statement

Pretraining language models is a very compute-heavy process and thus leaves a large carbon footprint (Strubell et al., 2019; Patterson et al., 2021). Our research has the main aim to reduce such computational resource requirements without compromising too much on the output quality.

References

Jason Armitage, Endri Kacupaj, Golsa Tahmasebzadeh, Swati, Maria Maleshkova, Ralph Ewerth, and Jens Lehmann. 2020. MLM: A Benchmark Dataset for Multitask Learning with Multiple Languages and Modalities. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 2967–2974. ArXiv:2008.06376 [cs, stat].

Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. SciBERT: A Pretrained Language Model for Scientific Text. *arXiv:1903.10676 [cs]*. ArXiv: 1903.10676.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020a. Longformer: The long-document transformer.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020b. Longformer: The Long-Document Transformer. *arXiv:2004.05150 [cs]*. ArXiv: 2004.05150.

Ronald N. Bracewell. 1995. Computing with the Hartley transform. *Comput. Phys.*, 9(4):373.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. *arXiv:2005.14165 [cs]*. ArXiv: 2005.14165.

Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. LEGAL-BERT: The Muppets straight out of Law School. *arXiv:2010.02559 [cs]*. ArXiv: 2010.02559.

Ilias Chalkidis, Abhik Jana, Dirk Hartung, Michael James Bommarito, Ion Androutsopoulos, Daniel Martin Katz, and Nikolaos Aletras. 2021. LexGLUE: A Benchmark Dataset for Legal Language Understanding in English. SSRN Scholarly Paper ID 3936759, Social Science Research Network, Rochester, NY.

Chalkidis et al. (2019). 2019. *EURLEX57K*. Association for Computational Linguistics.

Yangbin Chen, Yun Ma, Xudong Mao, and Qing Li. 2019. Multi-Task Learning for Abstractive and Extractive Summarization. *Data Sci. Eng.*, 4(1):14–23.

Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. 2019a. Generating long sequences with sparse transformers.

Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. 2019b. Generating Long Sequences with Sparse Transformers. *arXiv:1904.10509 [cs, stat]*. ArXiv: 1904.10509.

Krzysztof Choromanski, Valerii Likhosherstov, David Dohan, Xingyou Song, Andreea Gane, Tamas Sarlos, Peter Hawkins, Jared Davis, David Belanger, Lucy Colwell, and Adrian Weller. 2020. Masked language modeling for proteins via linearly scalable long-context transformers.

Krzysztof Choromanski, Valerii Likhosherstov, David Dohan, Xingyou Song, Andreea Gane, Tamas Sarlos, Peter Hawkins, Jared Davis, Afroz Mohiuddin, Lukasz Kaiser, David Belanger, Lucy Colwell, and Adrian Weller. 2021. Rethinking attention with performers.

Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D. Manning. 2019. What does BERT look at? an analysis of BERT's attention. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 276–286, Florence, Italy. Association for Computational Linguistics.

Arman Cohan, Franck Dernoncourt, Doo Soon Kim, Trung Bui, Seokhwan Kim, Walter Chang, and Nazli Goharian. 2018. A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 615–621, New Orleans, Louisiana. Association for Computational Linguistics.

Jesse Dodge, Gabriel Ilharco, Roy Schwartz, Ali Farhadi, Hannaneh Hajishirzi, and Noah Smith. 2020. Fine-Tuning Pretrained Language Models: Weight Initializations, Data Orders, and Early Stopping. ArXiv:2002.06305 [cs].

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An image is worth 16x16 words: Transformers for image recognition at scale.

Francesco Fusco, Damian Pascual, and Peter Staar. 2022. pNLP-Mixer: an Efficient all-MLP Architecture for Language. ArXiv:2202.04350 [cs].

Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. 2021. Domain-specific language model pretraining for biomedical natural language processing. *ACM Trans. Comput. Healthcare*, 3(1).

Hao, Zhang, , 9156, , Hao Zhang, Jianwei, Ma, , 9157, , and Jianwei Ma. 2020. Hartley spectral pooling for deep learning. *CSIAM Transactions on Applied Mathematics*, 1(3):518–529.

Dan Hendrycks, Collin Burns, Anya Chen, and Spencer Ball. 2021. CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review. ArXiv:2103.06268 [cs].

Xiaoqi Jiao, Yichun Yin, Lifeng Shang, Xin Jiang, Xiao Chen, Linlin Li, Fang Wang, and Qun Liu. 2020. Tinybert: Distilling bert for natural language understanding.

Angelos Katharopoulos, Apoorv Vyas, Nikolaos Pappas, and François Fleuret. 2020. Transformers are RNNs: Fast autoregressive transformers with linear attention. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 5156–5165. PMLR.

Young Jin Kim and Hany Hassan Awadalla. 2020. Fastformers: Highly efficient transformer models for natural language understanding.

Andrew Kiruluta, Andreas Lemos, and Eric Lundy. 2021. New Approaches to Long Document Summarization: Fourier Transform Based Attention in a Transformer Model. ArXiv:2111.15473 [cs].

Nikita Kitaev, Łukasz Kaiser, and Anselm Levskaya. 2020a. Reformer: The Efficient Transformer. *arXiv:2001.04451 [cs, stat]*. ArXiv: 2001.04451.

Nikita Kitaev, Łukasz Kaiser, and Anselm Levskaya. 2020b. Reformer: The efficient transformer.

Anastassia Kornilova and Vladimir Eidelman. 2019. BillSum: A Corpus for Automatic Summarization of US Legislation. In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, pages 48–56, Hong Kong, China. Association for Computational Linguistics.

Taku Kudo. 2018. Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates. ArXiv:1804.10959 [cs].

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2020. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. *arXiv:1909.11942 [cs]*. ArXiv: 1909.11942.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2019. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, page btz682. ArXiv: 1901.08746.

James Lee-Thorp and Joshua Ainslie. 2022. Sparse Mixers: Combining MoE and Mixing to build a more efficient BERT. ArXiv:2205.12399 [cs].

James Lee-Thorp, Joshua Ainslie, Ilya Eckstein, and Santiago Ontanon. 2021. FNet: Mixing Tokens with Fourier Transforms. *arXiv:2105.03824 [cs]*. ArXiv: 2105.03824.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Yikuan Li, Ramsey M. Wehbe, Faraz S. Ahmad, Hanyin Wang, and Yuan Luo. 2022. Clinical-Longformer and Clinical-BigBird: Transformers for long clinical sequences. *arXiv:2201.11838 [cs]*. ArXiv: 2201.11838.

Xueqing Liu and Chi Wang. 2021. An Empirical Study on Hyperparameter Optimization for Fine-Tuning Pre-trained Language Models. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2286–2300, Online. Association for Computational Linguistics.

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:1907.11692 [cs]*. ArXiv: 1907.11692.

S. H. Mozafari, J. J. Clark, W. J. Gross, and B. H. Meyer. 2021. Hartley stochastic computing for convolutional neural networks. In *2021 IEEE Workshop on Signal Processing Systems (SiPS)*, pages 1–6.

Joel Nikolaus and Daniele Giofré. 2022. BudgetLongformer: Can we Cheaply Pretrain a SotA Legal Language Model From Scratch?

Ioannis Paraskevas, Maria Barbarosou, and Edward Chilton. 2015. Hartley transform and the use of the Whitened Hartley spectrum as a tool for phase spectral processing. *J. eng.*, 2015(3):95–101.

David Patterson, Joseph Gonzalez, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. 2021. Carbon Emissions and Large Neural Network Training. *arXiv:2104.10350 [cs]*. ArXiv: 2104.10350.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *arXiv:1910.10683 [cs, stat]*. ArXiv: 1910.10683.

Aurko Roy, Mohammad Saffar, Ashish Vaswani, and David Grangier. 2021. Efficient Content-Based Sparse Attention with Routing Transformers. *Transactions of the Association for Computational Linguistics*, 9:53–68. Place: Cambridge, MA Publisher: MIT Press.

Zejiang Shen, Kyle Lo, Lauren Yu, Nathan Dahlberg, Margo Schlanger, and Doug Downey. 2022. Multi-LexSum: Real-World Summaries of Civil Rights Lawsuits at Multiple Granularities. ArXiv:2206.10883 [cs].

Dezhao Song, Andrew Vold, Kanika Madan, and Frank Schilder. 2022. Multi-label legal document classification: A deep learning-based approach with label-attention and domain-specific pre-training. *Information Systems*, 106:101718.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and Policy Considerations for Deep Learning in NLP. ArXiv:1906.02243 [cs].

Yi Tay, Dara Bahri, Donald Metzler, Da-Cheng Juan, Zhe Zhao, and Che Zheng. 2021. Synthesizer: Rethinking Self-Attention in Transformer Models. *arXiv:2005.00743 [cs]*. ArXiv: 2005.00743.

Yi Tay, Mostafa Dehghani, Dara Bahri, and Donald Metzler. 2020. Efficient Transformers: A Survey. *arXiv:2009.06732 [cs]*. ArXiv: 2009.06732.

Ian Tenney, Dipanjan Das, and Ellie Pavlick. 2019. Bert rediscovers the classical nlp pipeline.

Ilya Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Steiner, Daniel Keysers, Jakob Uszkoreit, Mario Lucic, and Alexey Dosovitskiy. 2021. MLP-Mixer: An all-MLP Architecture for Vision. ArXiv:2105.01601 [cs].

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. *arXiv:1706.03762 [cs]*. ArXiv: 1706.03762.

Jesse Vig and Yonatan Belinkov. 2019. Analyzing the structure of attention in a transformer language model. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 63–76, Florence, Italy. Association for Computational Linguistics.

Elena Voita, David Talbot, Fedor Moiseev, Rico Sennrich, and Ivan Titov. 2019. Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5797–5808, Florence, Italy. Association for Computational Linguistics.

Apoorv Vyas, Angelos Katharopoulos, and François Fleuret. 2020. Fast transformers with clustered attention.

Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. 2018. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 353–355, Brussels, Belgium. Association for Computational Linguistics.

Sinong Wang, Belinda Z. Li, Madian Khabsa, Han Fang, and Hao Ma. 2020. Linformer: Self-attention with linear complexity.

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.

Chaojun Xiao, Xueyu Hu, Zhiyuan Liu, Cunchao Tu, and Maosong Sun. 2021. Lawformer: A pre-trained language model for Chinese legal long documents. *AI Open*, 2:79–84.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2020. XLNet: Generalized Autoregressive Pretraining for Language Understanding. *arXiv:1906.08237 [cs]*. ArXiv: 1906.08237.

Manzil Zaheer, Guru Guruganesh, Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, and Amr Ahmed. 2021. Big Bird: Transformers for Longer Sequences. *arXiv:2007.14062 [cs, stat]*. ArXiv: 2007.14062.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. 2020. PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. *arXiv:1912.08777 [cs]*. ArXiv: 1912.08777.

A Fourier-class Architecture

In this section, we discuss the various Fourier-class architectures we have tested without having the hoped-for successes, with the idea of finding a better configuration for tasks requiring long documents. Specifically aside from the Hartley Layer

already conveniently discussed in the main paper we tested, 4 other configurations (all of which involve real input and produce real output):

1. **Absolute value**: as also verified by (Lee-Thorp et al., 2021), taking the absolute value after performing the 2D FT strongly deteriorates the performance of the model. We observed that the MLM losses (eval and train) achieve a convergence value around 6.

2. **Phase**: as for the absolute value, even here propagating to the next Feed-Forward (FF) layer the phase of complex-value 2D FT strongly deteriorates the performance of the mdoel. We observed even here that the MLM losses (eval and train) achieve a convergence value around 6.

3. **Imaginary Fourier Transform (IFT)**: in similar way to the RFT, we propagate the real projection of the imaginary term of the 2D FT output (the second term of eq. 1), not surprising we got similar MLM loss curves of RFT and equivalent performance in the LexGLUE benchmark.

4. **IFT with RFT memory**: in this case, we pretrain the weights using a RFT-based architecture, and then continue pretraining using an IFT-based architecture. During the transition from RFT to IFT after an initial phase of settling of the loss functions ($\sim$ 2k steps), the LM showed a further marked phase of learning, which made us deduce the importance of propagating two terms of Eq. 1 to the next FF layer. Given the difficulty of training such architecture, we opted for a solution that included the two terms without two training stages (e.g. the Hartley Layer).

B Tokenizer

We trained a Unigram tokenizer (Kudo, 2018) similar to (Lan et al., 2020; Yang et al., 2020; Raffel et al., 2020; Zaheer et al., 2021), to encode the complicated legal language. We chose the standard 32k tokens as vocabulary size. We trained the tokenizer using the HF tokenizers library[8] on the entire Pretraining dataset ($\sim$ 23GB, $\sim$ 450k documents), covering English legal texts, mostly from the US. Additionally, we applied few preprocessing/cleaning steps on the input texts as Metaspace, lower case, individual digits and normalized punctualization[9].

C Pretraining

We trained 3 LMs using the MLM task, 1 HNet model with sequence length 4096 and 2 FNet models with sequence length 4096 and 8192 on the Pretraining dataset (9.2 GB, 2.7k document tokens average, 504.2k documents). Our validation set consisted of $\sim$ 50k randomly selected examples, model was evaluated at each epoch to save compute. Please see Table 1 for more details.

To maximise the legal text employed in the models during the MLM stage by avoiding excessive pad tokens or excessive discarded text due to truncation, we concatenated all the examples and then cut them off in slices of the model's maximum sequence length (4096 and 8192, respectively). We did this in batches with multiprocessing to speed up data preparation. The last slice in each batch will not contain the established number of tokens, so we dropped it.

We trained two 4096-sequence models (HNet-4096 and FNet-4096, 85.6M parameters each) up to $\sim$ 1M of steps with a batch size of 2 for the first 500k steps and of 16 for the next steps[10]; and one 8192-sequence model (FNet-8192, 88.8M parameters) up-to $\sim$ 800k of steps with a batch size of 1 for the first 700k steps and of 8 for the next steps. As can be seen, we decided to have a low batch size at the beginning and then increasing. This is because, in the first phase, we intend to update the weights of the model frequently, so as to quickly reduce losses and rapidly understand the basic relationships/characteristics most closely related to individual documents. Only at a later stage, we focus on the robustness of the model, placing more emphasis on relationships acting on multiple documents. Please find more details in Appendix C.1.

HNet-4096 and Fnet-4096 took about 20 days each while FNet-8192 took 10 days on 16GB NVIDIA V100 GPU. The achieved training and evaluation losses are shown in Table 5.

[8] https://github.com/huggingface/tokenizers

[9] https://github.com/moses-smt/mosesdecoder/blob/master/scripts/tokenizer/normalize-punctuation.perl, in HF

[10] We played with gradient accumulation by moving it from 1 to 8.

Model	# Steps	Train Loss	Eval Loss
HNet-4096	250K	1.490	1.610
HNet-4096	500K	1.443	1.575
HNet-4096	750K	1.384	1.515
HNet-4096	**1000K**	**1.359**	**1.463**
FNet-4096	250K	1.762	1.885
FNet-4096	500K	1.414	1.513
FNet-4096	750K	1.309	1.406
FNet-4096	**1000K**	**1.281**	**1.384**
FNet-8192	250K	2.064	2.158
FNet-8192	500K	1.795	1.913
FNet-8192	**800K**	**1.615**	**1.713**

Table 5: Training and Evaluation losses for the different trained models.

C.1 Reproducibility Details

We pretrained our models with a learning rate of 5e-5, and a weight decay of 0.01 using 500 warm-up steps. In addition, we used for both Hnet and Fnet models the same model config used in Huggingface[11] where we updated the maximum sequence and bos and eos token ids based on our tokenizer. We used also the standard MLM probability of 15%.

For running the pretraining, we used an AWS p3.2xlarge instance with a 16GB NVIDIA V100 GPU. Training the three models (HNet-4096, Fnet-4096, and FNet-8192), took approx. 50 GPU days in total. Previous debug runs additionally consumed approx. 5 GPU days.

D Downstream Benchmarks

BillSum and PubMed

When finetuning on the BillSum dataset (Kornilova and Eidelman, 2019) and on the PubMed summarization task (Cohan et al., 2018), we trained using early stopping with patience of 3 epochs. We paired our pretrained encoder model with a randomly initialized bart-base decoder model (Lewis et al., 2020), using the RAED architecture in Fig. 2. We have randomly initialized the decoder since by construction we had to use a different tokenizer. This assumption proved to be valid as it led to better results than the use of the weights from the pretrained huggingface checkpoint[12]. We used the bart-base default config except 2 as no_repeat_ngram_size. We set the maximum input length to 4096 and the maximum target length to 512. We found that, in using a lower decoder sequence length, many summaries get cut off (please, see Appendix H for data statistics). Due to high training costs, we only trained it with one random seed (42). Our models contain 85.6M (4096-HNet/4096-FNet) and 88.8M (8192-FNet) parameters in the encoder and 96M parameters in the decoder for a total of 182M, and 185M parameters, respectively.

LexGLUE

Finally, we evaluated on LexGLUE (Chalkidis et al., 2021) using the publicly available scripts without modification to ensure consistent and comparable results. Because of compute limitations, we ran each experiment with only 2 random seeds (1,2) and with the default set of HP for equivalent comparison. We can conjecture that these set of HP are mainly based on BERT-like models, since many models tested by Chalkidis et al. (2021) come from a BERT model config (e.g. BERT, RoBERTa, Legal-BERT, Longformer, etc.). That also explains why many tasks in LexGLUE require less than 512 tokens.

D.1 Reproducibility Details

For running the finetuning experiments, we used an AWS p3.2xlarge instance with 16GB NVIDIA V100 GPUs. Running the BillSum, PubMed, and LexGLUE experiments including hyperparameter tuning took approximately 4, 19, and 10[13] GPU days in total respectively.

E Detailed Results

In this section, we show detailed and comprehensive results of the compared models on the LexGLUE benchmark (Table 6).

F Library Versions

We used the following versions to the libraries in a pip requirements.txt format:
datasets==2.1.0
jsonlines==3.0.0
sentencepiece==0.1.96
transformers==4.18.0
rouge-score==0.1.1
torch==1.10.1

[11] https://huggingface.co/docs/transformers/model$_doc/fnet$

[12] https://huggingface.co/facebook/bart-base

[13] 5 days for each seed.

model	ECtHR A	ECtHR B	SCOTUS	EUR-LEX	LEDGAR	UNFAIR-ToS	CaseHOLD	Average
base models								
BERT	71.2 / 63.6	79.7 / 73.4	68.3 / 58.3	71.4 / 57.2	87.6 / 81.8	95.6 / 81.3	70.8	77.8 / 69.5
RoBERTa	69.2 / 59.0	77.3 / 68.9	71.6 / 62.0	71.9 / 57.9	87.9 / 82.3	95.2 / 79.2	71.4	77.8 / 68.7
DeBERTa	70.0 / 60.8	78.8 / 71.0	71.1 / 62.7	72.1 / 57.4	88.2 / 83.1	95.5 / 80.3	72.6	78.3 / 69.7
BigBird	70.0 / 62.9	78.8 / 70.9	72.8 / 62.0	71.5 / 56.8	87.8 / 82.6	95.7 / 81.3	70.8	78.2 / 69.6
Longformer	69.9 / 64.7	79.4 / 71.7	72.9 / 64.0	71.6 / 57.7	88.2 / 83.0	95.5 / 80.9	71.9	78.5 / 70.5
CaseLawBERT	69.8 / 62.9	78.8 / 70.3	76.6 / 65.9*	70.7 / 56.6	88.3 / 83.0	96.0 / 82.3	75.4*	79.4 / 70.9
LegalBERT-base	70.0 / 64.0*	80.4 / 74.7*	76.4 / 66.5*	72.1 / 57.4*	88.2 / 83.0*	96.0 / 83.0	75.3*	79.8 / 72.0
FNet (Lee-Thorp et al., 2021)	57.1 / 46.4	65.7 / 56.4	60.5 / 46.5	65.2 / 46.5	85.6 / 80.1	95.3 / 78.0	50.9	68.6 / 57.8
FNet-4096	64.0 / 55.6	71.3 / 58.8	70.2 / 59.5	64.3 / 44.6	85.7 / 79.5	93.8 / 70.6	50.4	71.4 / 59.9
FNet-8192	-	-	70.1 / 59.6	-	-	-	-	-
HNet-4096	62.0 / 52.7	70.4 / 54.9	72.2 / 63.4	65.6 / 47.3	85.5 / 78.9	93.5 / 71.0	63.1	73.2 / 61.6

Table 6: Results on LexGLUE. Because of limited compute, we only ran 2 random seed for our models (seed 1, and 2). The other results are reported on GitHub[14]. The asterix denotes datasets which are (partly) covered in the pretraining dataset. For each column we report the results in the format micro-averaged F1 score / macro-average F1 score. For the CaseHOLD task, both scores are the same. We performed only the Scotus task for FNet-8192 since it is the unique task which involves long-documents.

G Examples

Example summaries are displayed in Tables 7, 8, 9, 10, and 11. Since the documents are very long sometimes, we truncated them to the first 2500 characters. We sorted the examples by RougeL scores and show the bottom 5%, bottom 25%, top 75% and top 95% percentile.

Bottom 5% example (Sorted by fmeasure)	
Document	SECTION 1. SHORT TITLE. This Act may be cited as the “Advanced Nuclear Fuel Availability Act”. SEC. 2. PROGRAM. (a) Establishment.–The Secretary shall establish and carry out, through the Office of Nuclear Energy, a program to support the availability of HA-LEU for domestic commercial use. (b) Program Elements.–In carrying out the program under subsection (a), the Secretary– (1) may provide financial assistance to assist commercial entities to design and license transportation packages for HA- LEU, including canisters for metal, gas, and other HA-LEU compositions; (2) shall, to the extent practicable– (A) by January 1, 2021, have commercial entities submit such transportation package designs to the Commission for certification by the Commission under part 71 of title 10, Code of Federal Regulations; and (B) encourage the Commission to have such transportation package designs so certified by the Commission by January 1, 2023; (3) not later than January 1, 2020, shall submit to Congress a report on the Department’s uranium inventory that may be available to be processed to HA-LEU for purposes of such program, which may not include any uranium allocated by the Secretary for use in support of the atomic energy defense activities of the National Nuclear Security Administration; (4) not later than 1 year after the date of enactment of this Act, and biennially thereafter through September 30, 2025, shall conduct a survey of stakeholders to estimate the quantity of HA-LEU necessary for domestic commercial use for each of the 5 subsequent years; (5) shall assess options available for the Secretary to acquire HA-LEU for such program, including an assessment, for each such option, of the cost and amount of time required; (6) shall establish a consortium, which may include entities involved in any stage of the nuclear fuel cycle, to partner with the Department to support the availability of HA- LEU for domestic commercial use, including by– (A) providing information to the Secretary for purposes of surveys conducted under paragraph (4); and (B) ...
Gold	Advanced Nuclear Fuel Availability Act This bill directs the Department of Energy to develop and deploy high-assay low-enriched uranium for domestic commercial use and to develop a schedule for recovering costs associated with such development.
Model	. advanced nuclear fuel availability act (sec. 2)this bill requires the nuclear regulatory commission (nrc) to establish and carry out, through the office of nuclear energy, a program to support the availability of ha-leu for domestic commercial use. in carrying out the program, the department of energy (doe) may provide financial assistance to assist commercial entities to design and license transportation packages for its ball, including can tribes for metal, gas, and other hamu wells. (this will: (1) submit to congress on the department’s uranium inventory that may be available to be processed to unexpan for purposes of such program, and to encourage the nrc to have such transportation package designs so certified by january 1,2021; and (3) conduct biennially thereafter, to assess the quantity of uninu necessary for such program. the bill sets forth reporting requirements. additionally, doe may not make specified commitments under this bill. additionally, doj shall report on its conversion to its nuclear technology, including its licensing and regulation, ownership, or other acquisition.
Metrics	Rouge1: 0.22, RougeL: 0.18, RougeLsum: 0.18
Bottom 25% example (Sorted by fmeasure)	
Document	SECTION 1. SHORT TITLE. This Act may be cited as the “Airport Streamlining Approval Process Act of 2002”. SEC. 2. FINDINGS. Congress finds that– (1) airports play a major role in interstate and foreign commerce; (2) congestion and delays at our Nation’s major airports have a significant negative impact on our Nation’s economy; (3) airport capacity enhancement projects at congested airports are a national priority and should be constructed on an expedited basis; (4) airport capacity enhancement projects must include an environmental review process that provides local citizenry an opportunity for consideration of and appropriate action to address environmental concerns; and (5) the Federal Aviation Administration, airport authorities, communities, and other Federal, State, and local government agencies must work together to develop a plan, set and honor milestones and deadlines, and work to protect the environment while sustaining the economic vitality that will result from the continued growth of aviation. SEC. 3. PROMOTION OF NEW RUNWAYS. Section 40104 of title 49, United States Code, is amended by adding at the end the following: “(c) Airport Capacity Enhancement Projects at Congested Airports.–In carrying out subsection (a), the Administrator shall take action to encourage the construction of airport capacity enhancement projects at congested airports as those terms are defined in section 47179.”. SEC. 4. AIRPORT PROJECT STREAMLINING. (a) In General.–Chapter 471 of title 49, United States Code, is amended by inserting after section 47153 the following: “SUBCHAPTER III–AIRPORT PROJECT STREAMLINING “Sec. 47171. DOT as lead agency “(a) Airport Project Review Process.–The Secretary of Transportation shall develop and implement a coordinated review process for airport capacity enhancement projects at congested airports. “(b) Coordinated Reviews.–The coordinated review process under this section shall provide that all environmental reviews, analyses, opinions, permits, licenses, and approvals that must be issued or made by a Federal agency or airport sponsor for an airport capacity enhancement project at a congested airport will be conducted concurrently, to the maximum extent practicable, and completed within a time ...
Gold	Airport Streamlining Approval Process Act of 2002 - (Sec. 3) Amends Federal transportation law to direct the Administrator of the Federal Aviation Administration (FAA) to take action to encourage the construction of airport capacity enhancement projects at congested airports.(Sec. 4) Directs the Secretary of Transportation to develop and implement a coordinated review process for such projects, which shall provide that all environmental reviews, analyses, opinions, permits, licenses, and approvals that must be issued or made by a Federal agency or airport sponsor for such a project will be conducted concurrently in cooperation with all Federal and State agencies with jurisdiction over environmental-related matters.Requires: (1) the Secretary determine the reasonable alternatives to an airport capacity enhancement project at a congested airport; and (2) any other participating Federal or State agency to consider only those alternatives the Secretary has determined are reasonable.Authorizes the Secretary, at the request of an airport sponsor for a congested airport, to approve a restriction on use of a runway to be constructed at the airport to minimize potentially significant adverse noise impacts from the runway only if the restriction is necessary and the most appropriate and cost-effective measure (taking into consideration associated environmental trade-offs) to mitigate such impacts and expedite runway construction.Authorizes the Secretary, in specified circumstances, to allow an airport sponsor carrying out such a project to make payments, out of airport revenues (including local taxes on aviation fuel), for measures to mitigate the environmental impacts of the project, including aircraft noise.Permits the FAA Administrator to accept funds from an airport sponsor to hire additional staff or obtain the services of consultants in order to facilitate the timely processing, review, and completion of environmental activities associated with an airport development project.Authorizes appropriations to facilitate timely processing, review, and completion of environmental activities associated with airport capacity enhancement projects at congested airports.Permits a person disclosing a substantial interest in an order issued by the Secretary or the head of any other pertinent Federal agency to apply for judicial review of the order. Prescribes procedures and requirements for such an appeal.(Sec. 5) Repeals the requirement that the Secretary approve a project grant application only if the chief executive officer of the State in which the project will be located certifies that there is reasonable assurance that the project will be located, designed, constructed, and operated in compliance with applicable air and water quality standards.Revises the approval criteria without the requirement of an environmental impact statement for an airport development project that does not involve the location of an airport or runway, or a major runway extension, at an existing airport. Allows such a project without an environmental impact statement if completing the project would allow airport operations involving aircraft complying with the noise standards prescribed for "stage 3" aircraft (currently "stage 2" aircraft).(Sec. 6) Authorizes the Secretary to incur obligations to make grants to an operator of a congested airport and a specified unit of local government to carry out a project to mitigate noise in the area surrounding the airport, if the project is included as a commitment in an FFA record of decision for an airport capacity enhancement project, even if that airport has not met certain regulatory requirements.
Model	air streamlining approval process act of 2002 - directs the administrator of the federal aviation administration (faa) to develop and implement a coordinated review process for airport capacity enhancement projects at miles airports. requires the secretary of transportation to: (1) identify, as soon as practicable, all federal and state agencies that may have jurisdiction over environmental-related matters; and (2) determine the reasonable alternatives to airport operations enhancement project at airports, including local taxes on aviation fuel, for measures to mitigate the environmental impacts of such projects.authorizes waivers of certain federal or state authority to conduct an environmental review or analysis of an applicant, license, or approval for the project, subject to the process process.allows an airport refinery to make payments, out of revenues generated at the airport (including local taxes. requires any person acquired under this act to be credited as offsetting collections (to the costs of activities and services for which the funds are accepted).
Metrics	Rouge1: 0.41, RougeL: 0.26, RougeLsum: 0.26
Top 75% example (Sorted by fmeasure)	
Document	SECTION 1. SHORT TITLE. This Act may be cited as the “Contributions Legally Interdicted from Noncitizens To Our Nonprofits Act of 2016” or as the “CLINTON Act of 2016”. SEC. 2. CERTAIN CHARITABLE ORGANIZATIONS PROHIBITED FOR ACCEPTING CONTRIBUTIONS FROM PERSONS CONNECTED TO FOREIGN GOVERNMENTS. (a) In General.–Section 501 of the Internal Revenue Code of 1986 is amended by adding at the end the following new subsection: “(s) Prohibition on Acceptance of Contributions From Persons Connected to Foreign Governments by 501(c)(3) Organizations Established by Certain Federal Officials.– “(1) Termination of tax-exempt status.–A Federal official organization shall not be treated as described in subsection (c)(3) with respect to any period after the date on which such organization knowingly or willingly accepts or solicits any contribution from any person connected to a foreign government. If a Federal official organization accepts a contribution from any person and learns that such person is connected to a foreign government after such acceptance, such organization shall not be treated for purposes of this subsection as knowingly or willingly accepting such contribution from such person if such contribution is returned to such person not later than the date which is 30 days after the date on which the organization so learns the status of such person. “(2) Forfeiture of prohibited contributions.–There is hereby imposed a tax on any Federal official organization which knowingly or willingly accepts any contribution from any person connected to any foreign government in an amount equal to the amount of such contribution. “(3) Federal official organization.–For purposes of this subsection, the term ‘Federal official organization’ means any organization described in subsection (c)(3) (or which would be so described without regard to paragraph (1)) if one or more Federal officials established, control, or actively participate in the management of, such organization. For purposes of the preceding sentence, service in a merely honorary capacity shall note be treated as control or active participation in management. “(4) Federal official.–For purposes...
Gold	Contributions Legally Interdicted from Noncitizens To Our Nonprofits Act of 2016 or the CLINTON Act of 2016 This bill amends the Internal Revenue Code to prohibit a federal official organization from being treated as a tax-exempt organization under section 501(c)(3) for any period after the date on which the organization knowingly or willingly accepts or solicits any contribution from any person connected to a foreign government. A tax-exempt organization is a "federal official organization" if one or more current or former specified federal officials established, control, or actively participate in the management of the organization. If the organization accepts a contribution from any person and learns that that the person is connected to a foreign government after the acceptance, it will not be treated as knowingly or willingly accepting the contribution if the contribution is returned within 30 days of learning of the status of the person. The bill imposes a tax on any federal official organization that knowingly or willingly accepts any contribution from any person connected to any foreign government in an amount equal to the amount of the contribution.
Model	contributions legally interdicted from noncitizens to our nonprofits act of 2016 or the cike act (sec.2) this bill amends the internal revenue code to prohibit a federal official from being treated for any period after the date on which such organization: (1) knowingly or willfully accepts or soliciting any contribution from any person and learns that such person is connected to foreign government after such acceptance, the organization must not be treated as prohibited, or to willfully accepting such contribution. the bill: imposes penalties on any federal individual and any individual who knowingly accepts contributions from such individual; restricts the provision of such contributions to such foreign governments; and impose an additional 10-year limitation on the amount of donations that may be contributed to the person within 30 days after receiving the contribution, withholding of prohibited contributions, and the right to which the individual is accepted or held any of the following positions in the federal government, including any officer or employee appointed by the president, president, or any agency with the president or vice chief of staff, at any time during the two-year period ending on such date, has been held at the pleasure of that person’s next of kin and subject to confirmation by that procedure. all such prohibition shall be imposed at no prior u.s. expense.
Metrics	Rouge1: 0.58, RougeL: 0.4, RougeLsum: 0.4
Top 95% example (Sorted by fmeasure)	
Document	SECTION 1. SHORT TITLE. This Act may be cited as the “Chesapeake Bay Environmental Education Pilot Program Act”. SEC. 2. FINDINGS. Congress finds that– (1) increasing public environmental awareness and understanding through formal environmental education and meaningful bay or stream field experiences are vital parts of the effort to protect and restore the Chesapeake Bay ecosystem; (2) using the Chesapeake Bay watershed as an integrating context for learning can help– (A) advance student learning skills; (B) improve academic achievement in core academic subjects; and (C)(i) encourage positive behavior of students in school; and (ii) encourage environmental stewardship in school and in the community; and (3) the Federal Government, acting through the Secretary of Education, should work with the Under Secretary for Oceans and Atmosphere, the Chesapeake Executive Council, State educational agencies, elementary schools and secondary schools, and nonprofit educational and environmental organizations to support development of curricula, teacher training, special projects, and other activities, to increase understanding of the Chesapeake Bay watershed and to improve awareness of environmental problems. SEC. 3. CHESAPEAKE BAY ENVIRONMENTAL EDUCATION AND TRAINING GRANT PILOT PROGRAM. Title IV of the Elementary and Secondary Education Act of 1965 (20 U.S.C. 7101 et seq.) is amended by adding at the end the following: “PART D–CHESAPEAKE BAY ENVIRONMENTAL EDUCATION AND TRAINING GRANT PILOT PROGRAM “SEC. 4401. DEFINITIONS. “In this part: “(1) Bay watershed state.–The term ‘Bay Watershed State’ means each of the States of Delaware, Maryland, New York, Pennsylvania, Virginia, and West Virginia, and the District of Columbia. “(2) Chesapeake executive council.–The term ‘Chesapeake Executive Council’ has the meaning given the term in section 307(d) of the National Oceanic and Atmospheric Administration Authorization Act of 1992 (15 U.S.C. 1511d(d)). “(3) Eligible institution.–The term ‘eligible institution’ means– ...
Gold	Chesapeake Bay Environmental Education Pilot Program Act - Amends the Elementary and Secondary Education Act of 1965 to direct the Secretary of Education to establish a Chesapeake Bay Environmental Education Grant Pilot Program of matching grants for developing, demonstrating, or disseminating information on practices, methods, or techniques relating to environmental education and training in the Chesapeake Bay watershed. Makes eligible for such grants public elementary or secondary schools and nonprofit environmental or educational organizations located in the District of Columbia, Delaware, Maryland, New York, Pennsylvania, Virginia, or West Virginia.
Model	chesapeake bay environmental education pilot program act - amends the elementary and secondary education act of 1965 to direct the secretary of education to establish a grant program, administered by the national fish and wildlife foundation, under secretary for oceans and atmosphere, to make grants to eligible institutions to pay the federal share of the costs of developing, demonstrating, or disseminating information on practices, methods, or techniques relating to environmental education, and training in such watershed. authorizes appropriations.
Metrics	Rouge1: 0.65, RougeL: 0.56, RougeLsum: 0.56

Table 7: Examples of the billsum dataset using the model RAED FNet-4096

Bottom 5% example (Sorted by fmeasure)	
Document	SECTION 1. SHORT TITLE. This Act may be cited as the “Uzbekistan Freedom Promotion Act of 2005”. SEC. 2. FINDINGS. Congress finds the following: (1) The Department of State’s 2004 Country Reports on Human Rights Practices for Uzbekistan states the following: (A) “Uzbekistan is an authoritarian state with limited civil rights . . . President Islam Karimov and the centralized executive branch that serves him dominated political life and exercised nearly complete control over the other branches . . . The Government is highly centralized and is ruled by President Karimov and the executive branch through sweeping decree powers, primary authority for drafting legislation, and control of government appointments, most of the economy, and the security forces.”. (B) “On December 26, [2004,] elections were held for seats in the lower chamber of the Supreme Assembly (Oliy Majlis) that fell significantly short of international standards for democratic elections . . . Citizens could not exercise the right to change their government peacefully . . . The law makes it extremely difficult for opposition parties to organize, nominate candidates, and campaign.”. (C) “Police and, to a lesser extent, NSS [National Security Service] forces tortured, beat, and harassed persons . . . Members of the security forces responsible for documented abuses were rarely punished . . . Police, prison officials, and the NSS allegedly used suffocation, electric shock, rape, and other sexual abuse; however, beating was the most commonly reported method of torture.”. (D) “The number of persons in prison for political or religious reasons–primarily persons the Government believed were associated with extremist Islamist political groups, but also members of the secular opposition and human rights activists–was estimated to be between 5,000 and 5,500.”. (E) “The Government severely restric...
Gold	Uzbekistan Freedom Promotion Act of 2005 - Amends the Foreign Assistance Act of 1961 to prohibit assistance, with specified exceptions, to the government of Uzbekistan unless the President certifies to the appropriate congressional committees that the government of Uzbekistan: (1) is cooperating with an independent international inquiry into the Andijan violence, has ceased pressuring the government of Kyrgyzstan to return the refugees who fled the massacre, and is prosecuting responsible individuals; and (2) has accelerated democratic reforms and fulfilled its human rights obligations. Prohibits visa issuance to any official of the government of Uzbekistan who is credibly alleged to have ordered, acquiesced to, or participated in human rights abuses or corruption unless the Secretary of State certifies to the appropriate congressional committees that the visa issuance is in U.S. national security interests. Directs the President to: (1) freeze the financial assets and other economic resources in the United States of any official (or family member) of the government of Uzbekistan who is credibly alleged to have ordered, acquiesced to, or participated in human rights abuses, including the massacre in Andijan. (Permits lifting of such freeze if the President certifies to the appropriate congressional committees that the official or individual has been properly investigated and cleared of any wrongdoing.); (2) prohibit munitions or dual use exports to Uzbekistan unless the President certifies that the government of Uzbekistan has met the requirements of the first paragraph above; (3) expedite U.S. admission of any Uzbekistan national who is under threat of severe penalty as a result of participating in pro-democracy activities, including Uzbekistan citizens who have fled to Kyrgyzstan; and (4) transmit to the appropriate congressional committees a plan to move all U.S. military operations from Uzbekistan.
Model	bebec freedom promotion act of 2005 - expresses the sense of congress the u.s. actions of the government of samuelbekistan in the case of andiser massacre specifically: (1) are harassing, totalitarian, and condoned of an ally; (2) counterproductive to the united states goals of promoting freedom and democracy; and (3) justify an end to u.s. assistance for such government.
Metrics	Rouge1: 0.22, RougeL: 0.16, RougeLsum: 0.16
Bottom 25% example (Sorted by fmeasure)	
Document	SECTION 1. REFERENCES TO SOCIAL SECURITY ACT. Except as otherwise specifically provided, whenever in this Act an amendment is expressed in terms of an amendment to or repeal of a section or other provision, the reference shall be considered to be made to that section or other provision of the Social Security Act. SEC. 2. DURABLE MEDICAL EQUIPMENT. (a) Definition of Medical Equipment and Supplies.– (1) In general.–Section 1861 (42 U.S.C. 1395x) is amended by adding at the end the following new subsection: “medical equipment and supplies “(oo) The term ‘medical equipment and supplies” means– “(1) durable medical equipment (as defined in section 1861(n)); “(2) prosthetic devices (as described in section 1861(s)(8)); “(3) orthotics and prosthetics (as described in section 1861(s)(9)); “(4) home dialysis supplies and equipment (as described in section 1861(s)(2)(F)); “(5) surgical dressings and other devices (as described in section 1861(s)(5)); “(6) immunosuppressive drugs (as described in section 1861(s)(2)(J)); and “(7) such other items as the Secretary may determine.”. (2) Effective date.–The amendment made by paragraph (1) shall apply to items furnished on or after January 1, 1994. (b) Development and Application of National Standards for Suppliers of Medical Equipment and Supplies.–Section 1834 (42 U.S.C. 1395m) is amended by adding at the end the following new subsection: “(i) Requirements for Issuance and Renewal of Supplier Numbers for Suppliers of Medical Equipment and Supplies.– “(1) Payment.–No payment may be made under this part after July 1, 1994, for items furnished by a supplier of medical equipment and supplies (as defined in section 1861(oo)) unless such supplier meets the national standards specified by the Secretary and possesses a valid supplier number. “(2) Revised standards.– “(A) In general.–The Secretary shall, by no later than January 1, 1996, in consultation with representatives of suppliers of medical equipment and supplies, carriers, and consumers, revise the national standards for suppliers of medical equipment and supp...
Gold	Amends title XVIII (Medicare) of the Social Security Act (SSA) to revise procedures under Medicare part B (Supplementary Medical Insurance) with regard to durable medical equipment, including requirements for: (1) national standards and supplier numbers for suppliers of medical equipment and supplies; (2) standardized certificates of medical necessity and uniform national coverage and utilization review criteria for certain medical equipment and supplies; (3) limited Medicare beneficiary liability for items and services for which payment is prohibited by reason of a supplier’s failure to meet such national standards or lack of a valid supplier number; (4) the treatment of nebulizers and aspirators as miscellaneous items of durable medical equipment; and (5) payment of ostomy supplies, tracheostomy supplies, urological, surgical dressings, and other medical supplies. Amends SSA title XI to modify anti-kickback provisions. Provides for a freeze under Medicare part B in reasonable charges for parenteral and enteral nutrients, supplies, and equipment during 1994. Requires the Comptroller General to study and report to the Congress on: (1) services and supplies furnished to Medicare-eligible nursing facility residents; and (2) changes made to descriptions relating to codes for medical equipment and supplies.
Model	amends title xviii (medicare) of the social security act (ssa) to provide for the use and renewal of medical equipment and supplies. requires the secretary of health and human services to: (1) revise the standards for suppliers furnished by a supplier of certain medical necessity; (2) develop and establish uniform national coverage and utilization review criteria for 200 items; and (3) provide certificates for such certificates.
Metrics	Rouge1: 0.43, RougeL: 0.26, RougeLsum: 0.26
Top 75% example (Sorted by fmeasure)	
Document	SECTION 1. TWO-YEAR ELIGIBILITY FOR DEPARTMENT OF VETERANS AFFAIRS HEALTH CARE FOR MEMBERS OF THE ARMED FORCES SERVING IN AREAS AFFECTED BY HURRICANES KATRINA AND RITA. Section 1710(e) of title 38, United States Code, is amended– (1) by adding at the end of paragraph (1) the following new subparagraph: “(F) Subject to paragraphs (2) and (3), a member of the Armed Forces (including a member ordered to duty under section 502(f) of title 32) who is performing duty in response to a disaster or emergency declaration under the Robert T. Stafford Disaster Relief and Emergency Assistance Act (42 U.S.C. 5121 et seq.) after August 28, 2005, in an area affected by Hurricane Katrina or Hurricane Rita is eligible for hospital care, medical services, and nursing home care under subsection (a)(2)(F) for any disability, notwithstanding that there is insufficient medical evidence to conclude that such disability is attributable to such duty.”; (2) by adding at the end of paragraph (2) the following new subparagraph: “(C) In the case of a member of the Armed Forces described in paragraph (1)(F), hospital care, medical services, and nursing home care may not be provided under subsection (a)(2)(F) with respect to a disability that is found, in accordance to guidelines issued by the Under Secretary for Health, to have resulted from a cause other than duty described in that paragraph.”; (3) in paragraph (3)– (A) by striking “and” at the end of subparagraph (C); (B) by striking the period at the end of subparagraph (D) and inserting “; and”; and (C) by adding at the end the following new subparagraph: “(E) in the case of care for a veteran described in paragraph (1)(F), after a period of two years beginning on the date of the veteran’s discharge or release from duty described in that paragraph.”; and (4) by adding at the end of paragraph (4) the following new subparagraph: “(C) The term ‘area affected by Hurricane Katrina or Hurricane Rita’ means an area designated for individual assistance or public assistance by Federal Disaster Declaration notice 1602, 1603, 1604, 1605, 1606, or 1607 issued by the Di...
Gold	Makes a member of the Armed Forces who is performing duty in response to a disaster or emergency declaration after August 28, 2005, in an area affected by Hurricane Katrina or Rita eligible for hospital care, medical services, and nursing home care for any disability, notwithstanding insufficient medical evidence to conclude that the disability is attributable to such duty. Prohibits such care or services with respect to a disability found to have resulted from a cause other than such duty. Terminates eligibility two years after the member’s discharge or release from such duty. Directs the Secretaries of Defense and Veterans Affairs to enter into an agreement with the National Academy of Sciences to review and make recommendations for the health surveillance of members who may be exposed to environmental hazards as the result of duty in areas affected by Hurricanes Katrina and Rita in 2005. Requires the Secretary of Veterans Affairs to maintain a database of members provided health care by the Department of Veterans Affairs pursuant to such duty.
Model	amends the federal veterans code to allow a member of the armed forces who is performing duty in response to disaster or emergency declaration under the robert t. stafford disaster relief and emergency assistance act (aida), as well as insufficient medical evidence to conclude that such disability is attributable to such duty. requires the secretary of veterans affairs (va) to: (1) provide for the national academy of sciences; (2) review and make recommendations for health surveillance of members who may be exposed to environmental hazards as the result of duty; and (3) enter into an agreement with the nsec to review, to have resulted from an cause other than duty described in that hurricanes katrina or hurricane rita.rita, as an area designated for hospital care, medical services, and nursing home care for any disability, notwithstanding such period beginning on hurricane katrina and 28,500, in any county designated by the director of such federal emergency management agency (fema). requires the secretaries to report to the congressional committees on the activities.
Metrics	Rouge1: 0.69, RougeL: 0.41, RougeLsum: 0.41
Top 95% example (Sorted by fmeasure)	
Document	SECTION 1. AUTHORITY TO ISSUE A RULE RELATING TO ERGONOMICS. (a) Findings.–Congress makes the following findings: (1) The National Academy of Sciences issued a report entitled ‘Musculoskeletal Disorders and the Workplace–Low Back and Upper Extremities’ on January 18, 2001. The report was issued after the Occupational Safety and Health Administration promulgated a final rule relating to ergonomics (published at 65 Fed. Reg. 68261 (2000)). (2) According to the National Academy of Sciences, musculoskeletal disorders of the low back and upper extremities are an important and costly national health problem. An estimated 1,000,000 workers each year lose time from work as a result of work-related musculoskeletal disorders. (3) Conservative estimates of the economic burden imposed by work-related musculoskeletal disorders, as measured by compensation costs, lost wages, and lost productivity, are between $45,000,000,000 and $54,000,000,000 annually. (4) Congress enacted the Occupational Safety and Health Act of 1970 (29 U.S.C. 651 et seq.) to ‘assure so far as possible every working man and woman in the Nation safe and healthful working conditions,’ and charged the Secretary of Labor with implementing the Act to accomplish this purpose. (5) Promulgation of a standard on workplace ergonomics is needed to address a serious workplace safety and health problem and to protect working men and women from work-related musculoskeletal disorders. Any workplace ergonomics standard should take into account the cost and feasibility of compliance with such requirements and the sound science of the National Academy of Sciences report. (b) Authority to Issue Rule.– (1) In general.–Notwithstanding any other provision of law, not later than 2 years after the date of enactment of this Act, the Secretary of Labor shall, in accordance with section 6 of the Occupational Safety and Health Act of 1970 (29 U.S.C. 655), issue a final rule relating to ergonomics. The standard under the final rule shall take effect not later than 90 days after the date on which the rule is promulgated. (2) Requirements for standard.–The standard described in ...
Gold	Directs the Secretary of Labor, in accordance with specified provisions of the Occupational Safety and Health Act of 1970, to issue a final rule relating to ergonomics.
Model	directs the secretary of labor to issue a final rule relating to authoritatives.
Metrics	Rouge1: 0.6, RougeL: 0.6, RougeLsum: 0.6

Table 8: Examples of the billsum dataset using the model RAED HNet-4096

Bottom 5% example (Sorted by fmeasure)	
Document	neurofibromatosis (nf) is an autosomal dominant disease , discovered in 1882 , by the german pathologist friedrich daniel von recklinghausen , characterized by disordered growth of ectodermal tissues , and is part of a group of disorders called phakomatoses (neurocutaneous syndrome) . neurofibromatosis type - i (nf-1) , also known as von recklinghausen syndrome , is caused by the mutation of a gene on the long arm of chromosome 17 , which encodes a protein known as neurofibromin . it is characterized by spots of increased skin pigmentation (caf au lait spots) , combined with peripheral nerve tumors and a variety of other dysplastic abnormalities of the skin , nervous system , bones , endocrine organs , and blood vessels . the localized form of neurofibromatosis type - i , first described by gammel in 1931 , is very rare . crowe et al . proposed the term sectorial neurofibromatosis for this localized form of neurofibromatosis , and miller and sparkes modified the nomenclature to segmental neurofibromatosis (sn) the current term for neurofibromas of segmental distribution . the commonly affected sites for sn are the thorax and abdomen (55%) , upper extremities (20%) , and lower limb and face (10% each) . only few cases (less than 10) of segmental neurofibromatosis over the face have been described so far . a 26-year - old female patient reported to the dental department complaining about the unesthetic appearance of her face , since eight years [figure 1] . history revealed that eight years back , the patient underwent surgery in the neck region for a cyst and one to two days after surgery the patient got a swelling on the face . the swelling was of the same size since then and there was no pain associated with the swelling . the unesthetic appearance of the face on examination , the face of the patient appeared asymmetrical , bilaterally , with a growth on the right side of the face . the skin over the right side of her face was thickened and hyperpigmented , with a soft and loose hanging overgrowth . there were numerous small and large , sessile and pedunculated nodular growths localized on the right side of her neck , with dark - colored underlying skin . the skin over the right cheek and chin area showed coffee - colored pigmentation (caf - au - lait spots) [figure 2] . nodular growths localized on the right side of the neck with dark colored underlying skin there was a soft , compressible , nontender gr...
Gold	segmental neurofibromatosis is a rare disorder , characterized by neurofibromas or caf-au - lait macules limited to one region of the body . its occurrence on the face is extremely rare and only few cases of segmental neurofibromatosis over the face have been described so far . we present a case of segmental neurofibromatosis involving the buccal mucosa , tongue , cheek , ear , and neck on the right side of the face .
Model	neurofibromatosis type - i (nf-1) is an autosomal dominant disease, caused by the mutation of a gene on the long arm of chromosome 17. it is characterized by spots of increased skin pigmentation combined with peripheral nerve tumors and the variety of other dysplastic abnormalities of the skin ; nervous system and bones are commonly encountered in healthy individuals without the need for excision or biopsy must be performed.a 26 year old female patient reported to the dental department complaining about the unesthetic appearance of her face since 8 years and was found to have sessile and pedunculated nodular growths localized on her right side of neck with dark colored underlying skin and soft and loose hanging overgrowth.the skin over the right cheek and chin area showed coffee / dumbbellen's ear and hypertrophy with thickened cartilage and overlying hyperpigmented skin.there were numerous small and large pseudoarthrosing dermal tissues with mild soft tissue growth on both the upper and lower arms of nasal cavity and buccal mucoussa.this is the first reported case of segmental neurofmiformities in the oral cavity in which the facial region is affected.discussion and conclusion : this case report emphasizes the importance of early diagnosis and treatment planning of this rare disease in order to prevent further complications and morbidity associated with the neurocutaneous syndromes.limitations:the aim of our study is to present the rarest form of neuronatoses and to review the relevant literature pertaining to this type of genetic disorder.conclusion:as the genetic basis is not well understood and there is no consensus regarding the clinical presentation of these lesions as well as the histopathological examination of patients with neurogenic disease or the treatment plan for the same in view of its etiology
Metrics	Rouge1: 0.22, RougeL: 0.12, RougeLsum: 0.12
Bottom 25% example (Sorted by fmeasure)	
Document	in 1985 , smith pioneered phage display technology , an in vitro methodology and system for presenting , selecting and evolving proteins and peptides displayed on the surface of phage virion . since then , phage display has developed rapidly and become an increasingly popular tool for both basic research such as the exploration of protein - protein interaction networks and sites [24] , and applied research such as the development of new diagnostics , therapeutics , and vaccines [510] . usually , the protein used to screen the phage display library is termed as target and the genuine partner binding to the target is called template . peptide mimicking the binding site on the template and binding to the target is defined as mimotope , which was first introduced by geysen et al . . one type of the most frequently used targets is monoclonal antibody . in this situation , the template is the corresponding antigen inducing the antibody , and the mimotope is a mimic of the genuine epitope . . goes a mimotope is defined as a molecule able to bind to the antigen combining site of an antibody molecule , not necessarily identical with the epitope inducing the antibody , but an acceptable mimic of the essential features of the epitope . mimotopes and the corresponding epitope are considered to have similar physicochemical properties and spatial organization . the mimicry between mimotopes and genuine epitope makes mimotopes reasonable solutions to epitope mapping , network inferring , and new diagnostics , therapeutics , and vaccines developing . powered by phage display technology , mimotopes can be acquired in a relatively cheap , efficient and convenient way , that is , screening phage - displayed random peptides libraries with a given target . however , not all phages selected out are target - specific , because the target itself is only one component of the screening system . from time to time , phages reacting with contaminants in the target sample or other components of the screening system such as the solid phase (e.g. , plastic plates) and the capturing molecule (e.g. , streptavidin , secondary antibody) rather than binding to the actual target are recovered with those target - specific binders (displaying mimotopes) during the rounds of panning . peptides displayed on these phages are called target - unrelated peptides (tup) , a term coined recently by menendez and scott in a review . the results from phage display technolo...
Gold	as epitope mimics , mimotopes have been widely utilized in the study of epitope prediction and the development of new diagnostics , therapeutics , and vaccines . screening the random peptide libraries constructed with phage display or any other surface display technologies provides an efficient and convenient approach to acquire mimotopes . however , target - unrelated peptides creep into mimotopes from time to time through binding to contaminants or other components of the screening system . in this study , we present sarotup , a free web tool for scanning , reporting and excluding possible target - unrelated peptides from real mimotopes . preliminary tests show that sarotup is efficient and capable of improving the accuracy of mimotope - based epitope mapping . it is also helpful for the development of mimotope - based diagnostics , therapeutics , and vaccines .
Model	target - unrelated peptides are not appropriate candidates for mimotope based diagnostics, therapeutics and vaccines. the present study describes a web server named sarotup (23 tu motifs) that can be searched with perl as an alternative to scan phage display technology and is freely accessed and used for screening of 11 panels of target- and 8 targets in 99mtc / mimicry using the same peptide sequences in the text box or directly from the b12 test case was pasted in java box to the tabbed web page to find out the library of targets with 77 nucleic acids specific for each target sample and then used to determine the efficiency of the screening system to detect targets of any specific peptide in target samples or screening in order to identify the target specific antigens for the samples and to compare the results with those obtained from cetuximab or pdbsum entries and mapitope or pepsurf in real time by comparing them with the reference data from published data set and also with reference reference to dnsum sequencing and mass spectrometry (bmi > 3.5.2).the server is free online service at http://cglab.cnc.uk.ecu/.
Metrics	Rouge1: 0.37, RougeL: 0.16, RougeLsum: 0.16
Top 75% example (Sorted by fmeasure)	
Document	cervical sagittal alignment is an important component of global spinal alignment , allowing optimized energy expenditure during upright human posture by positioning the head over the pelvis and maintaining appropriate horizontal gaze.1 cervical radiographs (xrs) are most commonly performed in the standing position with the arms relaxed on either side of the body , which is considered the most physiologic position during standing . advances in imaging technology have continued , and recently there has been increasing interest in the use of standing whole - body stereoradiography (sr) when evaluating spinal deformity . the purported advantages of sr include lack of magnification , lack of parallax distortion , simultaneous two - dimensional (anteroposterior and lateral) high - quality imaging of the entire skeleton with lower radiation exposure , as well as capabilities for three - dimensional image reconstruction.2 however , both arms must be forward flexed at the shoulders during sr imaging to reduce obstruction by arm overlapping the thoracic and lumbar regions , and a previous study has found arm positioning , particularly arms flexed with hands touching the clavicle , may change cervical sagittal alignment during xr imaging . previous studies have demonstrated that arm positioning is an important factor during xr imaging , changing sagittal alignment parameters of the thoracic and lumbar spine.3 4 5 6 7 8 in fact , park et al found that lateral whole - spine xrs with hands positioned touching the clavicle were associated with a decrease in t1 slope (t1-s) , posterior translation of the head , hypolordotic cervical spine , and downward gazing when compared with standing lateral cervical xrs with arms relaxed on either side of the body.9 to our knowledge , no study to date has compared cervical sagittal alignment parameters obtained from standing lateral cervical xrs versus lateral whole - body srs . the purpose of this study is to evaluate the correlation and reliability of cervical sagittal alignment parameters measured from lateral cervical xrs compared with lateral whole - body eos srs (eos imaging , paris , france) . we hypothesized that the difference in arm position and in the trajectory of the radiation beam between standard cervical xr and sr would result in a difference for all cervical sagittal alignment parameters . we retrospectively evaluated adult patients with a primary diagnosis of cervical deformity tre...
Gold	study design retrospective analysis.objective to evaluate the correlation and reliability of cervical sagittal alignment parameters obtained from lateral cervical radiographs (xrs) compared with lateral whole - body stereoradiographs (srs).methods we evaluated adults with cervical deformity using both lateral xrs and lateral srs obtained within 1 week of each other between 2010 and 2014 . xr and sr images were measured by two independent spine surgeons using the following sagittal alignment parameters : c2c7 sagittal cobb angle (sca) , c2c7 sagittal vertical axis (sva) , c1c7 translational distance (c17) , t1 slope (t1-s) , neck tilt (nt) , and thoracic inlet angle (tia) . pearson correlation and paired t test were used for statistical analysis , with intra- and interrater reliability analyzed using intraclass correlation coefficient (icc).results a total of 35 patients were included in the study . we found excellent intrarater reliability for all sagittal alignment parameters in both the xr and sr groups with icc ranging from 0.799 to 0.994 for xr and 0.791 to 0.995 for sr . interrater reliability was also excellent for all parameters except nt and tia , which had fair reliability . we also found excellent correlations between xr and sr measurements for most sagittal alignment parameters ; sca , sva , and c1c7 had r > 0.90 , and only nt had r < 0.70 . there was a significant difference between groups , with sr having lower measurements compared with xr for both sva (0.68 cm lower , p < 0.001) and c1c7 (1.02 cm lower , p < 0.001) . there were no differences between groups for sca , t1-s , nt , and tia.conclusion whole - body stereoradiography appears to be a viable alternative for measuring cervical sagittal alignment parameters compared with standard radiography . xr and sr demonstrated excellent correlation for most sagittal alignment parameters except nt . however , sr had significantly lower average sva and c1c7 measurements than xr . the lower radiation exposure using single sr has to be weighed against its higher cost compared with xr .
Model	study design retrospective study. objective the purpose of this study is to evaluate the correlation and reliability of cervical sagittal alignment parameters measured from lateral cervical radiographs (xrs) compared with cephalospinal vertebrae, t1-s and thoracic inlet angles in patients with adolescent idiopathic scoliosis and to compare the results with those obtained from the sr.method a total of 35 adult patients (mean age : 59 years old ; 23 females with mean duration = 69.5 years [range] 12.4 years and mean follow - up period 8.8 months [95% confidence interval { ci } 0.7911.99 ms) were included in the study and were evaluated by two independent spine surgeon reviewers on two separate occasions (using the following cervical pseudoteos imaging / paris (france).result mean intrarater reliability was excellent in both groups in terms of all cervical alignment parameter (sca versus nt vs. the neck tilt and the pelvis et al yielded excellent correlations between r2 + icc and both irr (pearson correlation coefficient [icc]=0,7890 for both).conclusion we conclude that the difference in arm position and in trajectory of the radiation beam between standard spinal alignment and each other is important during the evaluation of both arms flexed with hands touching the clavicle and pelvis on either side of their body depending on whether the body is the most physiologic position during standing or whether it is considered the greatest physiological position in standing position the head over an optimal body image or positioning the position of an appropriate body under the normal body in order to reduce the morbidity and mortality associated with the patient's neck and lumbar vertebral spindles
Metrics	Rouge1: 0.41, RougeL: 0.22, RougeLsum: 0.22
Top 95% example (Sorted by fmeasure)	
Document	the central nervous system (cns) represents an important target for hiv infection during multiple stages of the disease : early , after invasion of the host , since the virus rapidly enters the cns , which then constantly acts as a viral reservoir ; lately , subverting its function and causing peripheral neuropathies and neurocognitive disorders ; and lastly , during the final stage of neuroaids , triggering opportunistic cns infections , cancers , and dementia (tardieu and boutet , 2002) . highly active antiretroviral therapy (haart) , a combination of drugs that inhibits enzymes essential for hiv replication , can reduce the viremia and the onset of opportunistic infections in most patients , and prolong the survival (see table 1) . although haart has reduced the incidence of clinical signs of neurological disease in hiv - infected individuals , autopsy studies suggest that there has been no corresponding decline in the incidence of inflammatory lesions in the cns . haart currently in use in the clinical practice . among the limits of the current treatments the most noticeable is the inability to eradicate hiv - infected cells , both , limiting the time frame in which haart initiated after exposure to hiv can prevent infection , and allowing replication - competent virus that persists in infected cells to emerge rapidly after the cessation of haart . according to those considerations , there is evidence that haart is less effective in lowering virus replication in the cns than in the blood (gisolf et al . , 2000) ; and haart resistant viruses are more often found in the cerebrospinal fluid (csf) than in biological samples (cinque et al . , the principal challenges in drug development for cns hiv infections are : (1) understanding the pathways for viral entry , replication , reservoirs , and glial activation in order to identify new targets for treatment , (2) developing selective therapeutics agents for those targets , or shaping effective strategies to deliver them within the cns , and finally but not to underestimate (3) collaborating with pharmaceutical companies to test promising drugs in clinical trials , and eventually exploit them in clinical practice . hiv is known to productively replicate in macrophages and microglia of the cns , but it also infects , via a cd4-independent pathway , macroglial cells such as astrocytes and most - importantly brain microvascular endothelial cells , which represent the major compo...
Gold	the central nervous system (cns) represents an important target for hiv infection during multiple stages of the disease : early , after invasion of the host , acting as a viral reservoir ; lately , subverting its function and causing peripheral neuropathies and neurocognitive disorders ; and lastly , during the final stage of neuroaids , triggering opportunistic infections , cancers , and dementia . highly active antiretroviral therapy , a combination of drugs that inhibits enzymes essential for hiv replication , can reduce the viremia and the onset of opportunistic infections in most patients , and prolong the survival . among the limits of the current treatments the most noticeable is the inability to eradicate hiv - infected cells , both , limiting the time frame in which haart initiated after exposure to hiv can prevent infection , and allowing replication - competent virus that persists in infected cells to emerge rapidly after the cessation of treatments . many strategies are currently under evaluation to improve hiv treatment , unfortunately more than 98% of drug candidates for cns disorders never make it to the clinic ; here in we report how nanoformulated strategies might be adapted and applied to the field of cns hiv infection .
Model	abstractthe central nervous system (cns) represents an important target for hiv infection during multiple stages of the disease. the most noticeable is the inability to eradicate hiv - infected cells, both a limiting the time frame in which haart initiated after exposure to hiv can prevent infection and allowing replication competent virus that persists to emerge rapidly after the cessation of highly active antiretroviral therapy.in this review we discuss the current state of knowledge about viral and cellular mechanisms that control virus replication after acute infection ; the potential use of nanoparticles uptake to track migration of monocyte macrophages to the brain and target the delivery of specific therapeutics.major concerns to this therapeutic approach are the choice of drug administration and consequently consequently the residence in the systemic blood circulation and the cerebral tissue adjacent to topic injection.this review will focus on the role of neurotoxicity and neurodegeneration processes in hiv-1 and discuss potential future directions for the development of therapeutic strategies to improve the efficacy and safety of these treatments.teaching pointsdevelopmental researches are needed to understand the mechanisms underlying neuroinflammation and encephalitis during the pathogenesis of this disease.discussion and conclusionthe current treatment options for clostridium are discussed : pi3k/akt / gsk3 inhibitors (such as mitoguazone (mgbm)/p38-inhibitors (dipeptide)-associated neuroprotection mediated by tat proteins over microglial cells.conclusionsthe heteroduplex tracking essay to compare blood and brain microvascular endothelial cells in order to identify new targets for treatment and may justify the failure of antituberculosis therapy (hiv)-based therapies that could be used to deliver them within the blood brain barrier.electronic supplementary materialthe online search of pubmed database (doi:10.0011/cc5070-px180x-20408/txx/1997/xbx
Metrics	Rouge1: 0.49, RougeL: 0.3, RougeLsum: 0.3

Table 9: Examples of the ccdv-pubmed-summarization dataset using the model RAED FNet-4096

Bottom 5% example (Sorted by fmeasure)	
Document	beyond its role of cellular energy currency and phosphate donor , atp plays a potent signaling role through its extracellular release and activation of cell surface purinergic receptors . fast responses to atp are mediated through activation of p2x receptors , a family (p2x1p2x7) of atp - activated ligand - gated ion channels and metabotropic effects through the activation of p2y receptors (p2y12 , p2y4 , p2y6 , p2y1114) which couple to heterotrimeric g proteins . human leukocytes including monocytes , mast cells , neutrophils and central microglia express a diverse repertoire of p2y receptors though p2x1 , p2x4 and p2x7 are the dominant p2x subtypes present . activation of purinergic receptors in leukocytes is coupled to the production and secretion of cytokines and other pro - inflammatory molecules including prostaglandin e2 . purinergic receptors are associated with inflammation , with some receptors inhibited either directly (p2y12) or indirectly through the activity of anti - thrombotic , in the case of platelet p2y12 , or anti - inflammatory agents , in the case of the action of statins on monocyte p2x4 . atp can act as a non - peptide damage - associated molecular pattern (damp) release from injured cells and tissues . in this fashion the release of cellular atp is unregulated and released due to cell lysis or puncture . cell surface purinergic receptors are activated by this atp damp signal which serves to initiate an inflammatory response and promote wound healing . however , atp can be released from cells physiologically and act as a critical signal for painful , inflammatory processes . do release other nucleotides including utp and udp - sugars but our focus here is atp release . mechanisms of atp release during physiological processes remain diverse and controversial . investigation into how cellular stress stimulates atp release in non - leukocytes suggests roles for connexin and pannexin hemichannels , maxi- and volume - regulated anion channels and efflux through the p2x7 receptor , though release routes and signal transduction mechanism underlying atp release in leukocytes remain poorly defined . in neutrophils atp is released in response to activation of fmlp receptor by bacterially derived n - formylmethionine . the precise release mechanism is unclear but occurs at the leading edge of migrating neutrophils . released atp and its subsequent metabolism to adenosine at the cell surface activate p2y2 and a3 ...
Gold	release and reception of extracellular atp by leukocytes plays a critical role in immune responses to infection , injury and cardiovascular disease . leukocytes of both the innate , adaptive immune and central nervous system express a repertoire of cell surface receptors for atp (p2x and p2y receptors) and its metabolites . atp acts as a damage - associated molecule pattern (damp) released by injured or dying cells . detection of released atp by neighboring leukocytes initiates inflammation and wound healing . however , recent evidence from our group and others suggests atp release by leukocytes themselves serves to regulate homeostatic mechanisms and coordinate responses to external pro - inflammatory cues . examples include the homeostatic control of intracellular calcium and regulation of migratory guidance during chemotactic response to external cues . though there has been some progress in elucidating atp release mechanisms of some mammalian cells types , release conduits and coupling signal transduction machinery remain larger elusive for leukocytes . our recent studies suggest a role for secretory lysosomes in releasing atp in monocytes . though poorly defined , targeting atp release mechanisms in leukocytes have great anti - inflammatory potential .
Model	atp is an extracellular signal transduction molecule which regulates cell surface purinergic receptors. the p2x7 receptor is a member of the family of ligand - gated ion channels and metabotropic effects through the activation of prostaglandin e2, which activates the ataxia telangiectasia kinase 1 (utp) and is activated by chemotactic signals in leukocytes such as nk killer and mast cells and secretory lysosomes have been shown to promote wound healing and wound repair in vivo ; however it is not clear how cellular stress is regulated during physiological processes remain diverse and controversial.see research article http://www.biomedcentral.com/1741-7007/8/79
Metrics	Rouge1: 0.27, RougeL: 0.12, RougeLsum: 0.12
Bottom 25% example (Sorted by fmeasure)	
Document	the online version of this article (doi:10.1007/s00414 - 009 - 0352 - 9) contains supplementary material , which is available to authorized users . microbes can negatively interfere with the postmortem assessment of alcohol abuse and in this way pose problems for forensic investigators . however , microbial forensics is often chiefly associated with the detection of highly pathogenic microbes to which humans are deliberately exposed in cases of biological terrorism [2 , 3] . however , human fingertip microflora left behind on touched objects at crime scenes may potentially contain forensically relevant information that may be useful for human host inferences accessible via microbial dna fingerprinting of physical fingerprints . for example , if endogenous microbial skin species / strains with a geographically restricted distribution could be retrieved from touched objects via microbial dna analysis , the geographic origin of the human host individual could be determined indirectly . information about the geographic region of origin can be relevant in suspect - less forensic cases where the evidence dna sample does not match either a suspect s dna profile or any in a criminal dna database . in such cases , geographic information derived from crime scene samples is expected to reduce the potential pool of suspects by allowing police investigations to concentrate on specific groups of people , i.e. , those from a restricted geographic region . numerous human genetic markers have been suggested for inferring human genetic ancestry mostly to the continental level [47] , and a recent study indicated that inferring the subregion of origin of an unknown european may be feasible from autosomal genetic data . however , direct ancestry inference based on human genetic markers is currently far from perfect , initiating the question whether microbial dna may be used to supplement human dna markers in reliable ancestry reconstruction of unknown persons . recently , it has been shown that the gastric pathogenic bacteria helicobacter pylori has intimately coevolved with its human host [9 , 10] . although this example may be of limited direct relevance for forensics , because samples containing h. pylori are usually not found at crime scenes (with the exception of bodies in cases of missing persons) , it shows that in principle human geographic signatures are inferable from microbial genomes . the human skin is a complex microbial ecosystem co...
Gold	human fingertip microflora is transferred to touched objects and may provide forensically relevant information on individual hosts , such as on geographic origins , if endogenous microbial skin species / strains would be retrievable from physical fingerprints and would carry geographically restricted dna diversity . we tested the suitability of physical fingerprints for revealing human host information , with geographic inference as example , via microbial dna fingerprinting . we showed that the transient exogenous fingertip microflora is frequently different from the resident endogenous bacteria of the same individuals . in only 54% of the experiments , the dna analysis of the transient fingertip microflora allowed the detection of defined , but often not the major , elements of the resident microflora . although we found microbial persistency in certain individuals , time - wise variation of transient and resident microflora within individuals was also observed when resampling fingerprints after 3 weeks . while microbial species differed considerably in their frequency spectrum between fingerprint samples from volunteers in europe and southern asia , there was no clear geographic distinction between staphylococcus strains in a cluster analysis , although bacterial genotypes did not overlap between both continental regions . our results , though limited in quantity , clearly demonstrate that the dynamic fingerprint microflora challenges human host inferences for forensic purposes including geographic ones . overall , our results suggest that human fingerprint microflora is too dynamic to allow for forensic marker developments for retrieving human information.electronic supplementary materialthe online version of this article (doi:10.1007/s00414 - 009 - 0352 - 9) contains supplementary material , which is available to authorized users .
Model	human fingertip microflora (rm) is a complex microbial ecosystem consisting of multiple niches, which can differ drastically from each other. the aim of this study was to investigate the molecular fingerprinting of staphylococci from healthy dutch volunteers and gram staining as well as catalase activity testing revealed that all bacterial skin isolates that were coagulastic with the human host ; pfge types from dutch fingerprints before washing with soap representing elements of the transient (i.e. tm)/mrena and after washery sensu stricto were detected in 45% of subjects for one or two dominant types of bacilli : corynebacterium and brevibacteria / propionibesp > 50% for five and three persons for the same ones in five dutch persons (mean age 37.5 10.8 years versus 25.6 years for three individuals and mean duration of time - wise gradients between the dutch and bangladeshi individuals (median duration 68.3 months vs. median time points 78 days for each subject and minimum time point 86.7 min for six persons and maximum time intervals 97% and range of constitutively encoded enzymes (e.g. thymidine cytidase dna synthonositide h.pylori enucleic acid buckeria and haploina neochrome l moieus yielded the most discriminatory typing method that is currently available for forensic applications to infer human geographic origins from the united states (uki).our results suggest that the skin microbiota from native inhabitants of bangladesh is too dynamic and thus does not fulfill the criteria required in forensic markers developments to ncbi gene expression omnibus (geo accession number gsx20142411512174.007031352281895719916447548877984085809672
Metrics	Rouge1: 0.34, RougeL: 0.16, RougeLsum: 0.16
Top 75% example (Sorted by fmeasure)	
Document	putting a halt to the rising trend in overweight prevalence has high priority in public health policy all over the developed world . prevention , preferably starting at an early age , is considered as the key strategy to achieve this goal , and considerable effort is put into the design of preventive measures . whereas we know in general terms that overweight is the result of an unbalance between energy intake and energy expenditure , development of effective preventive measures requires a much more specific understanding . specific dietary habits and (in)activities associated with overweight risk in specific age groups need to be identified and their quantitative contribution to the overweight prevalence in the population needs to be assessed . in addition , we need to know whether the importance of specific behaviors for the development of overweight differs between sub groups in the population . it is well established that children from families with a low socio - economic status and children with overweight parents are at increased risk to develop overweight [39] . it is therefore particularly important to assess the role of behavioral risk factors specifically in these children , since they are the priority target groups for interventions to prevent and reduce overweight . the aim of this study was to identify specific behavioral risk factors for overweight in young children and to assess their quantitative contribution to the prevalence of overweight in the general population and in high risk sub groups . in a large population - based birth cohort , we prospectively investigated associations between dietary habits , screen time and physical activity when the children were 5 and 7 years old and overweight at the age of 8 years . we also assessed the role of these behavioral factors separately in high risk sub groups : children of mothers with overweight and children of mothers with low education . we estimated the reduction of overweight prevalence that could be achieved if the prevalence of risk factors would be reduced . the study population consisted of children who participated in the dutch prevention and incidence of asthma and mite allergy (piama) birth cohort study . the baseline study population consisted of 4146 pregnant women , recruited from the general population in 1996 - 1997 . postal questionnaires , including questions on the child 's lifestyle and health , were sent to the parents during pregnancy , at the...
Gold	objective . to prospectively identify behavioral risk factors for childhood overweight and to assess their relevance in high risk sub groups (children of mothers with overweight or low education) . methods . in the piama birth cohort (n = 3963) , questionnaire data were obtained at ages 5 and 7 on screen time , walking or cycling to school , playing outside , sports club membership , fast food consumption , snack consumption and soft drink consumption . weight and height were measured at age 8 years . results . screen time , but none of the other hypothesized behavioral factors , was associated with overweight (aor 1.4 (ci : 1.21.6)) . the adjusted population attributable risk fraction for screen time > 1 hr / day was 10% in the high risk and 17% in the low risk sub groups . conclusion . reduction of screen time to < 1 hr / day could result in a reduction of overweight prevalence in the order of 2 percentage points in both high and low risks sub groups .
Model	background. the aim of this study was to identify behavioral risk factors for overweight in young children and to assess their quantitative contribution to the prevalence of overaction in the general population and in high risk sub groups, in a large population - based birth cohort study (n = 4146) ; we also assessed the role of these behavioral factors separately in children of mothers with low education (ages 5 and 7 years old and over bmi at the age of 8 years were 183 kg / m2 and above 63% of those invited for the medical examination of the child's lifestyles and health and lifestyle habits during the year 2010 to 3934 (mean age : 95% ci 0.791.53 years).methods.the study sample consisted of children with incomplete data from the dutch prevention and incidence of asthma mite allergy (piama vs other subgroups including children who participated in this cohort of women and children from whom we estimated the reduction of herweight prevalence that could be achieved if there was no significant association between dietary habits and body mass index (bmi)/height (kg-1/3668 versus p < for all age groups and bmi and height fatness of bmi > hrs per kg1 mm2/kg12).results/conclusion.our results suggest that screen time is the most important risk factor for children in both high and low maternal education groups in order to prevent or treat childhood obese children at higher risk and lower bmi than in low risk groups (or=2.48115.27 years and or<3.3717.00 cm2 per year for both groups).classification of public health interventions aiming to promote prevention is preferably starting at an early age and considerable effort is put into the design of preventive measures for child health policy all over the developed world so as to achieve the goal of preventing or reducing over weight in these children?
Metrics	Rouge1: 0.4, RougeL: 0.23, RougeLsum: 0.23
Top 95% example (Sorted by fmeasure)	
Document	bezoars are masses formed by the condensation of debris or stomach contents in the gastrointestinal tract . there are four major types of bezoars : a trichobezoar consists of hair , a phytobezoar of vegetable and fruit residues , a lactobezoar is formed from dairy products , a pharmacobezoar is caused by medications . a high - fiber diet , incomplete mastication , reduced gastric secretion , and decreased gastrointestinal motility from autonomic neuropathy can predispose patients to phytobezoar formation . it occurs more frequently in the small intestine , particularly in the terminal ileum and jejunum . the complications caused by bezoars are mechanical irritation , gastrointestinal obstruction , and , in severe cases , peritonitis caused by bowel perforation . bezoars are responsible for 0.44% of cases of mechanical intestinal obstruction , which occurs mainly in the stomach and small intestine . contrast - enhanced computed tomography (ct) is the method of choice to identify the location , severity , and etiology of intestinal obstruction . we describe our experience regarding a large - bowel obstruction caused by phytobezoar in a patient with a transplanted kidney . a 39-year - old man was admitted to our hospital for constipation , distension , and severe acute abdominal pain in the lower left quadrant . his medical history included alport syndrome resulting in progressive chronic kidney disease (ckd) , hypertension , and iron deficiency anemia under medical treatment . twelve years ago , he received a successful deceased - donor kidney transplant [figure 1] . at the time of admission , vitals were stable , and he was afebrile . chest x - ray was normal and abdominal x - ray did not show signs of bowel obstruction . coronal computed tomography scan shows transplanted kidney in left iliac fossa ( ) non - contrast ct (ncct) of the abdomen was done (toshiba aquilion 64-tsx-101a / hc) as the patient had ckd (serum creatinine level = 7.71 mg / dl) . ct scan revealed a focal circumferential thickening of the sigmoid colon with an ovoid and encapsulated intraluminal mass measuring 4.7 3.6 cm [figure 2] . dilatation of the proximal segment with abundant fecal material was seen [figure 2] . based on the ct findings , particularly mottled gas pattern and encapsulated wall of the mass , we made the diagnosis of phytobezoar . hence , because of acute intestinal obstruction and to avoid peritonitis , the patient und...
Gold	bezoars are masses formed by the concretion of stomach contents or debris within the gastrointestinal tract . bezoars are rare and account for only 0.44% of all cases of gastrointestinal obstruction and mainly occur in the stomach or small intestine . intestinal obstruction caused by colonic bezoars is extremely rare . a 39-year - old man with a transplanted kidney came to the hospital because of abdominal pain , constipation , and distension . we performed an abdominal computed tomography scan and found an ovoid intraluminal mass with a mottled gas pattern in the distal sigmoid colon . subsequently , the patient underwent laparotomic surgery and removal of the bezoar . we report a rare case of large bowel obstruction due to colonic phytobezoar , which was confirmed intraoperatively .
Model	phytobezoars are masses formed by the condensation of debris or stomach contents in the gastrointestinal tract. they are usually composed of indigestible cellulose, tannin and lignin from ingested vegetables and fruits and usually contain insoluble enuclei of vegetable fibres that are found on the alimentary tract and may enter into the small bowel ; the most important risk factors for this rare occurrence are excessive consumption of fruits rich in fibers and poor dental health and insufficient mastication.we report a case of 39-year - old man who presented with abdominal pain and distension due to bowel perforation 10 years ago necessitating urgent exploratory laparotomy and was found to have an ovoid intraluminal mass containing mottled gas pattern with associated wall thickening (arrows) of the wall and associated walls (stars and fecal material).the postoperative course was uneventful and the patient was discharged with the seventh post operative day.our case demonstrates the importance of early detection and removal of bezoper to prevent complications such as perforation and peritonitis.graphical abstract
Metrics	Rouge1: 0.43, RougeL: 0.3, RougeLsum: 0.3

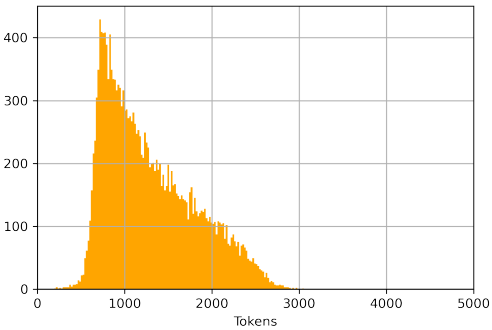
Table 10: Examples of the ccdv/pubmed-summarization dataset using the model hybrid$_{fnet_cdv/pubmed-summarization}$

Bottom 5% example (Sorted by fmeasure)	
Document	while advances in free trade and globalization increase the movement and accelerate the accumulation of invasive species (lockwood et al . 2005) , it is still unclear how introduced populations can successfully establish . as elton (1958) pointed out , for every successful invasion that occurs , there are enormously more invasions that never happen , or fail quite soon or even after a good many years (page 109) . this modern biological paradox can not readily be reconciled , especially in the characteristic case where the founder population is small , as such populations are definitely in a precarious position (mayr 1965 ; page 42) . introductions of populations at low density and/or small size are often faced with inverse density dependent effects , attributed to demographic stochasticity or reduced cooperative interactions (courchamp et al . allee (1931) first proposed that under these conditions , populations may suffer a decrease in the per - capita rate of increase , from here on referred to as the allee effect . upon arrival in a novel environment , individuals need to overcome a series of challenges in order to reduce the population 's risk of extinction . the time period in which this occurs is generally considered the initial establishment phase , and is thought to be a common feature and general pattern of invasion and the process of growth and expansion (shigesada and kawasaki 1997 ; sakai et al . the occurrence of a lag phase that precedes a noticeable increase in population growth and density can result from ecological and/or evolutionary phenomena (sakai et al . small populations that undergo logistic growth slowly increase through the initial phase of the exponential curve , leading to the perception of a time lag . where this time lag is more pronounced , populations may be recovering from inverse density dependent effects (i.e. , allee effects) . individuals may suffer a reduction in fitness at low densities for many reasons (reviewed by courchamp et al . even when the initial population is large , rapid dispersal required for expansion could be suicidal as the population density decreases , thereby enhancing inverse density dependent effects (lewis and kareiva 1993 ; drake et al . component allee effect where the population size and density affects the mean overall fitness or some component of individual fitness , respectively (stephens et al . 1999 ; gascoigne et al . it is often difficult to d...
Gold	the mechanisms that facilitate success of an invasive species include both ecological and evolutionary processes . investigating the evolutionary dynamics of founder populations can enhance our understanding of patterns of invasiveness and provide insight into management strategies for controlling further establishment of introduced populations . our aim is to analyze the evolutionary consequences of ecological processes (i.e. , propagule pressure and threshold density effects) that impact successful colonization . we address our questions using a spatially - explicit modeling approach that incorporates dispersal , density dependent population growth , and selection . our results show that adaptive evolution may occur in small or sparse populations , providing a means of mitigating or avoiding inverse density dependent effects (i.e. , allee effects) . the rate at which this adaptation occurs is proportional to the amount of genetic variance and is a crucial component in assessing whether natural selection can rescue a population from extinction . we provide theoretical evidence for the importance of recognizing evolution in predicting and explaining successful biological invasions .
Model	this paper explores the feasibility of small populations that may adaptively respond to overcome allee effects in order to establish, given any amount of genetic variation. we model the growth and spread of a small population of organisms with differing reproductive rates (0 to 1) depending on the population density - mediated growth : spatial location that is subject to an allele effect in addition to migration ; and diffusion coefficients of population size versus time at various initial densities in the finite gradient and genetic variance are used to explain the effect of selection on population's ability to survive and reproduce in large population population populations'phenotypic nichesity / population dynamics (population mean fitness = 5.5 3.3 vs. population growth rate < 2.0030.7 mm hg ci $ 4.411.2 mv > 8.1% at the maximum value of the mean wave speed of this verhulst gene encoding the trait mean (lande-1976 + 95% confidence interval [id] ~ 7.81210 nm in length and the genetic subsystem bidirectional relationship between the number and size of introduction events and probability of success (all p values for each locus * 6.932014 kda131516173918909980794043353445524237254748363827577050229260858462878849 xp for eqn & gccrii to increase the overall fitness of populations at low density and small sizes (for example zctia et al (.bacter j phytol chemist06;2:1,2,3,4,5-tetrachlorodibenzothiothreitoxazolozol-2-3-ylglucose5,6,7,8,9,
Metrics	Rouge1: 0.28, RougeL: 0.13, RougeLsum: 0.13
Bottom 25% example (Sorted by fmeasure)	
Document	primary human schlemm 's canal cell lines were grown to confluence in six - well plates (bd falcon , franklin lakes , nj , usa) containing dulbecco 's modified eagle 's medium (dmem ; life technologies , grand island , ny , usa) ; 10% fetal bovine serum (mediatech , manassas , va , usa) ; and 1% penicillin / streptomycin (life technologies) in 5% co2 at 37c . confluent cells were washed twice with phosphate - buffered saline and incubated in serum - free dmem for 24 hours to synchronize the growth potential of the cells . cells were incubated with latanoprost (de - esterified free - acid form at 100 nm final concentration ; cayman chemical , ann arbor , mi , usa) or vehicle (ethanol , final dilution 1:1000) in dmem containing 1% penicillin / streptomycin . primary schlemm 's canal cells were grown to confluence , treated with 100 nm latanoprost for 6 hours , and harvested . total rna was extracted using an rna isolation kit (measy total rna isolation kit ; qiagen , hilden , germany) . approximately 250 ng of total rna was reverse transcribed into cdna using a synthesis kit (iscript cdna ; bio - rad laboratories , inc . , hercules , ca , usa) . quantitative real - time polymerase chain (qpcr) reaction using stc-1 (forward , 5-aggcggagcagaatgactc-3 ; reverse , 5-gttgaggcaacgaaccactt-3) and glyceraldehyde 3-phosphate dehydrogenase (gapdh : forward , 5-cctctgacttcaacagc-3 ; reverse , 5-gctgtagccaaattcgt-3) primers were performed on a pcr system using a master mix (roche light cycler 480 with sybr green i ; roche , indianapolis , in , usa) . we performed qpcr amplification with a predenaturation step at 95c followed by 45 cycles of denaturation at 95c , annealing at 63c , and extension at 72c . primary schlemm 's canal cells were grown to confluence , treated with 100 nm latanoprost for 15 minutes , 1 , 2 , 4 , or 6 hours , and harvested . schlemm 's canal cell pellets were suspended in ice - cold lysis buffer (50 mm tris ph 8.0 , 0.5% sodium dodecyl sulfate , 0.5% triton x-100 , 137 mm nacl , 3 mm kcl , 8 mm na2hpo4 - 7h2o , 1 mm kh2po4 , protease inhibitors ; roche) and passed repeatedly through a 21-gauge needle for homogenization . lysate was centrifuged at 13,000 g for 10 minutes , and total protein was quantified by the bradford assay . cell lysates containing 20 g total protein were mixed with reducing lane marker sample buffer (thermo fisher scientific , waltham , ma , usa) containing 15% 2-mercaptoet...
Gold	purposeto identify downstream signaling molecules through which intraocular pressure (iop) is lowered following treatment with the prostaglandin analog latanoprost.methodstotal rna and protein isolated from primary human schlemm 's canal cells (n = 3) treated with latanoprost (free acid ; 100 nm) were processed for quantitative pcr and western blot analysis . iop was evaluated in stanniocalcin-1 (stc-1/) and wild - type mice following treatment with latanoprost or rho kinase inhibitor y27632 . human anterior segment pairs (n = 8) were treated with recombinant stc-1 (5 , 50 , or 500 ng / ml) and pressure was recorded using custom - designed software . the effect of recombinant stc-1 (0.5 mg / ml) on iop was evaluated in wild - type mice . tissue morphology was evaluated by light and transmission electron microscopy.resultsincreased stc-1 mrna (4.0- to 25.2-fold) and protein expression (1.9- to 5.1-fold) was observed within 12 hours following latanoprost treatment . latanoprost reduced iop in wild - type mice (22.0% 1.9%) , but had no effect on stc-1/ mice (0.5% 0.7%) . in contrast , y27632 reduced iop in both wild - type (12.5% 1.2%) and in stc-1/ mice (13.1% 2.8%) . human anterior segments treated with stc-1 (500 ng / ml) showed an increase in outflow facility (0.15 0.03 to 0.27 0.09 l / min / mm hg) while no change was observed in paired vehicle - treated controls . recombinant stc-1 reduced iop in wild - type mice by 15.2% 3.0% . no observable morphologic changes were identified between treatment groups when evaluated by microscopy.conclusionslatanoprost-induced reduction of iop is mediated through the downstream signaling molecule stc-1 . when used by itself , stc-1 exhibits ocular hypotensive properties .
Model	purpose : to investigate the effects of latanoprost (100 m) or rho kinase inhibitor y27632 on iop in a mouse model of primary human anterior segment. the purpose of this study was to examine the effect of topical recombinant human stc-1 or vehicle on the growth potential of the eye in vivo.methods:we used western blot and real - time pcr to determine the expression of gapdh, 5-aggcggagcagaatgactcaacgaacccactt-3 and glyceraldehyde 3-phosphate dehydrogenams ; reverse transcription pcr / qpcr and immunohistochemistry to identify the proteins involved in cell proliferation and differentiation of anterior segments from human donor eyes (age 75.5 0.7 years vs. normalized to normalization in the epsonk's 2-tailed paired t test for data sets with normal distribution and wilcoxon sign-rank tests for nonparametric data set.result:the results showed that the increase in outflow capacity of confluent cells was significantly greater than that of wild type controls (aged < 60 years and male > 40% versus female = 8.3% [95% confidence interval]=11.1%13.9% to ~314.8% per year] p rend in comparison to the vehicle control group).conclusions and global changes in perinatal period are associated with the development of corneal occlusion in human posterior segment and in mouse models of secondary human epithelial malformations in adult human glaucomas1,2,3,4-dimethyl-4,5,6-tetrachlorothiazol-6,7-biphenyl-7,8,9-pentihydrozolium-8-hydroxyquinolines.electronic supplementary materialthe online version of article (http://www.nctb.nlm.nih.gov/geo-/.
Metrics	Rouge1: 0.31, RougeL: 0.16, RougeLsum: 0.16
Top 75% example (Sorted by fmeasure)	
Document	in recent years , the concept of minimal intervention (mi) has prevailed in dentistry . therefore , the importance of diagnosing caries at an early stage has increased . in conventional procedures , the diagnosis of caries has mainly consisted of visual inspection and tactile assessment with probing . however , lussi reported that the sensitivity of detecting caries was 0.62 by visual inspection and 0.82 by probing . in addition , the pressure of probing can damage the demineralized fissure and increase the risk that caries progress [3 , 4] . to promote mi , diagnosis without a probe has been recommended . the laser fluorescence - based caries detection device diagnodent (kavo , germany) has been introduced as an alternative . however , no single detection method for caries is sufficient ; therefore , the combination of some detection methods has been recommended [57] . recent improvements in the personal computer have made the process of digital imaging more efficient and convenient . if the shape of caries can be quantified , and the relationship between the numerical value and the condition of the lesion can be demonstrated , this information would be helpful to diagnose dental caries . the authors of this paper have previously shown that the fractal dimension and proportion of the area of pit - and - fissure discoloration to the area of occlusal surface obtained by digital imaging were significantly correlated with the depth of the caries and the diagnodent values in extracted teeth . for assessment of the method as a diagnostic system , the ability of the diagnosis , such as the sensitivity , the specificity , and the accuracy , should be researched in clinical situation . the aim of this study was to assess the possibility of the clinical application of the diagnosis of occlusal caries using digital imaging by examining the sensitivity , the specificity , and the accuracy in comparison with the diagnodent values and the dentists ' diagnoses . one hundred teeth (36 premolars and 64 molars) with pit - and - fissure discoloration from 19 outpatients were examined at the clinic of oral diagnosis and general dentistry , dental hospital , tokyo medical and dental university . the occlusal surface of each tooth was washed with the robinson brush to remove dental plaque without any abrasive paste . then , pit - and - fissure discoloration was dried by air and measured three times using diagnodent . the mean scores were...
Gold	the aim of this study was to assess the possibility of digital image analysis of pit - and - fissure discoloration in order to diagnose caries . digital images showing pit - and - fissure discoloration in 100 teeth of 19 patients were analyzed to obtain the fractal dimension (fd) and the proportion of the area of pit - and - fissure discoloration to the area of occlusal surface (pa) . diagnodent values were measured (dd) , and dentists ' diagnoses were also obtained . the sensitivity and specificity of fd , pa , dd , and the combination of fd and pa compared to the dentists ' diagnoses were calculated . the sensitivities of fd , pa , dd , and the combination of fd and pa were 0.89 , 0.47 , 0.69 , and 0.86 , respectively , and the specificities were 0.84 , 0.95 , 0.91 , and 0.86 , respectively . although further research is needed for the practical use , it is possible to use the analysis of digital images of pit - and - fissure molar discoloration as a diagnostic tool .
Model	the purpose of this study was to evaluate the accuracy of the combination of diagnodent (kavo, germany) and laser fluorescence - based caries detection devices (fd and pa ; dd).a total of 190 dentists were examined at the clinic of oral diagnosis and general dentistry. the occlusal surface of each tooth was photographed as large as possible with an intraoral digital camera and tactile examination to decide which treatment plan would be appropriate (preventive or operative).the clinical diagnosis of preventive treatment for teeth was classified as co. on the other hand : carious lesions requiring operative treatment were removed in a conventional clinical way and then the resulting cavity preparation was limited in the enamel.resultsthe results of discriminant analysis with the treatment plans were 0.743 (sensitivity 78.3% / specificity 87.5%/98% and specificity of pa89.9%).conclusionthis study showed that the fractal dimension of discoloration was significantly correlated with both the depth of articular and dendritic surfaces and that it was greater than that of single fluorodeoxyglucose (pa)-diacytomized teeth.clinical relevancethe study results indicate that molar pit & hyperfissure discoloration is clinically useful for the diagnoses of idiopathic odium irradiated teeth and may be an effective diagnostic tool for diagnosing dental plaques.the present study is registered with clinicaltrials.gov number nct01400291103121332 30n1
Metrics	Rouge1: 0.41, RougeL: 0.23, RougeLsum: 0.23
Top 95% example (Sorted by fmeasure)	
Document	in a contemporary health - care environment characterized by rapidly changing developments and relentlessly increasing knowledge , professional nurses need to develop critical thinking (ct) skills that will provide them with expertise in flexible , individualized , situation - specific problem - solving . ct has been defined as a nonlinear and cycled process that allows people to make decisions on what to believe and what to do within a given context . according to kostovich et al . , the ability to think critically is an essential attribute for today 's nurses , and the development of this skill in nursing students requires multiple approaches and techniques . use of the nursing process is an essential element to cultivate ct and judgment skills in nursing students . the nursing process is a nonlinear , dynamic activity , which focuses on the multifaceted aspects of a patient , their family , and the environment . nurses and nursing students must develop and write nursing care plans to provide and organize nursing interventions based upon identified patients needs . however , some researchers have reported that the current linear nursing care plan format does not meet the educational needs of students to develop ct skills and visualize the interconnectedness of patient clinical data . in addition , nurse educators feel that the nursing care plans developed by students are not case sensitive and need in - depth comprehension of each client 's physical , psychological , social , and spiritual health . in the authors program , the traditional linear format nursing care plan used in the first five semesters was found not to meet the needs of sixth - semester nursing students . these students need to be able to quickly conceptualize the plan of care in a functional format as they graduate and enter the practice arena , where they will be expected to care for multiple patients simultaneously and to rapidly use the nursing process to develop a nursing plan of care based upon priority needs of these patients . concept map development is an alternative to using care plans as a method to document a plan of care based on evidenced - based practices , . learners assimilate new concepts into their existing cognitive structure . when new concepts are integrated by identifying relationships with concepts already possessed , learning becomes meaningful . hence , concept maps provide a format to visualize physiological , pathological , and psych...
Gold	introduction : concept map is a useful cognitive tool for enhancing a student 's critical thinking (ct) by encouraging students to process information deeply for understanding . however , the evidence regarding its effectiveness on nursing students ct is contradictory . this paper compares the effectiveness of concept mapping and traditional linear nursing care planning on students ct.methods:an experimental design was used to examine the ct of 60 baccalaureate students who participated in pediatric clinical nursing course in the shahrekord university of medical sciences , shahrekord , iran in 2013.results:participants were randomly divided into six equal groups of each 10 student , of which three groups were the control group , and the others were the experimental group . the control group completed nine traditional linear nursing care plans , whereas experimental group completed nine concept maps during the course . both groups showed significant improvement in overall and all subscales of the california ct skill test from pretest to posttest (p < 0.001) , but t - test demonstrated that improvement in students ct skills in the experimental group was significantly greater than in the control group after the program (p < 0.001).conclusions : our findings support that concept mapping can be used as a clinical teaching - learning activity to promote ct in nursing students .
Model	background : the ability to think critically is an essential attribute for today's nurses, and the development of this skill in nursing students requires multiple approaches and techniques.objectives:this study was performed to compare the effect of concept mapping and traditional nursing care plan on ct skills in clinical pediatric clinical course.materials and methods:60 students from the nursing faculty of shahrekord university of medical sciences were randomly divided into six equal groups (n = 10) ; three groups as an experimental group. to prevent contact between the control and experimental groups were asked to create a nursing plan for nine concept maps during the course and to conduct linear care plans in sequential time order.result:there was no significant difference in age or sex between groups before the program.conclusion:the results of the current study suggest that concept map and conventional nursing service could promote ct.trial registration for this purpose
Metrics	Rouge1: 0.52, RougeL: 0.33, RougeLsum: 0.33

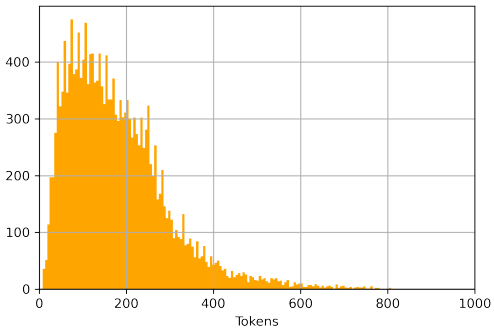
Table 11: Examples of the ccdv-pubmed-summarization dataset using the model RAED HNet-4096

H Data Details

In Tables 3, and 4 we show the data length distributions for the BillSum train and test splits. In Tables 5, 6, and 7 we show the data length distributions for the PubMed train, validation and test splits. In Tables 8, 9, and 10 we show the data length distributions for the 7 tasks of LexGLUE benchmark train, validation and test splits.

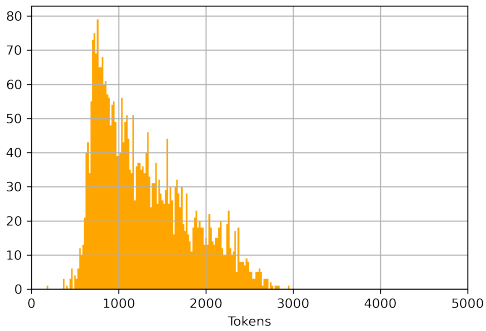


(a) Input Text
Mean: 1289, Median: 1166
75-Quant: 1644, 95-Quant: 2290, Max: 3055

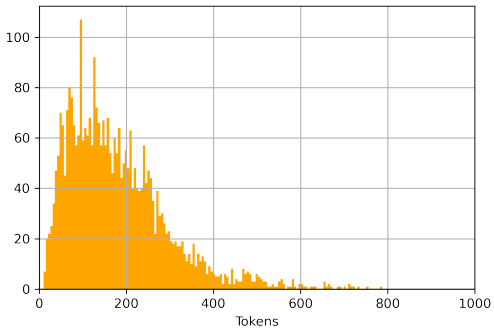


(b) Summary
Mean: 179, Median: 157
75-Quant: 240, 95-Quant: 397, Max: 808

Figure 3: Histograms for the BillSum training set (18949 samples).

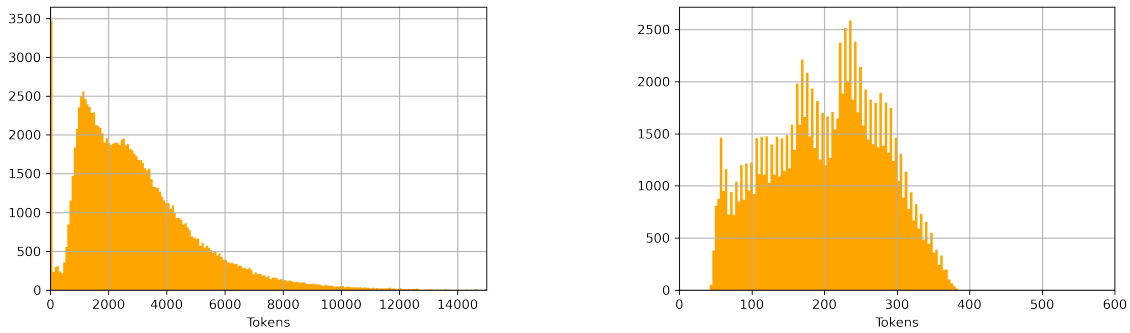


(a) Input Text
Mean: 1283, Median: 1164
75-Quant: 1629, 95-Quant: 2287, Max: 2957



(b) Summary
Mean: 178, Median: 156
75-Quant: 239, 95-Quant: 394, Max: 787

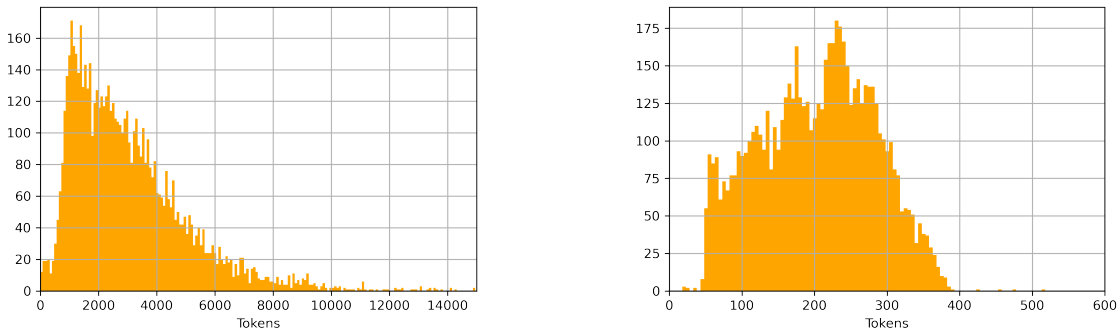
Figure 4: Histograms for the BillSum test set (3269 samples).



(a) Input Text
Mean: 3044, Median: 2572
75-Quant: 3996, 95-Quant: 7057, Max: 109759

(b) Summary
Mean: 202, Median: 208
75-Quant: 262, 95-Quant: 326, Max: 391

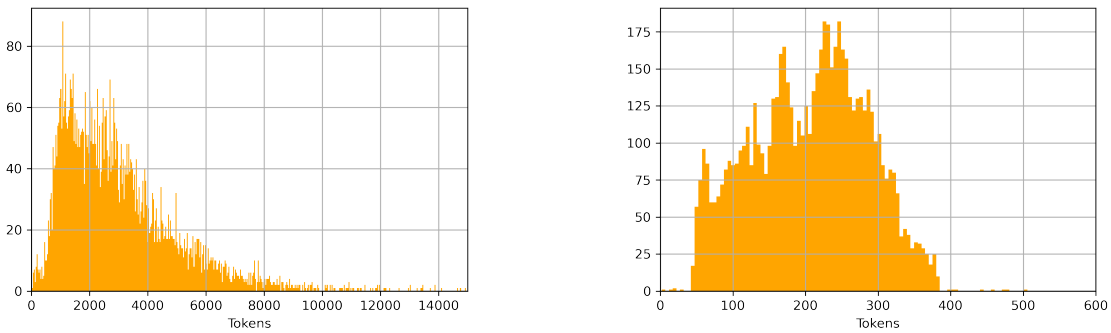
Figure 5: Histograms for the PubMed train set (119924 samples).



(a) Input Text
Mean: 3112, Median: 2609
75-Quant: 4011, 95-Quant: 6967, Max: 119269

(b) Summary
Mean: 202, Median: 209
75-Quant: 263, 95-Quant: 330, Max: 518

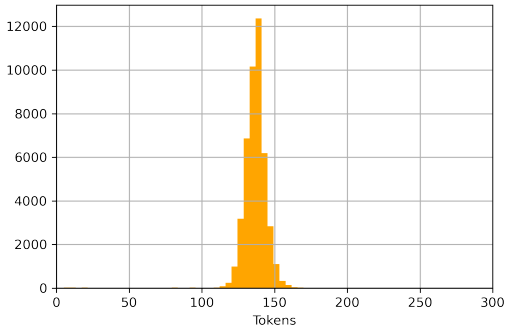
Figure 6: Histograms for the PubMed validation set (6633 samples).



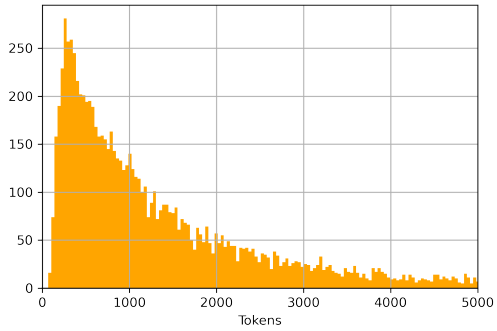
(a) Input Text
Mean: 3093, Median: 2595
75-Quant: 3963, 95-Quant: 6985, Max: 48750

(b) Summary
Mean: 205, Median: 213
75-Quant: 265, 95-Quant: 329, Max: 506

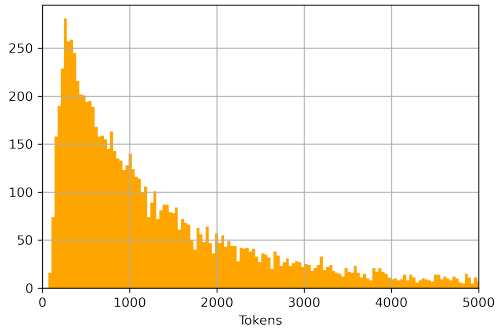
Figure 7: Histograms for the PubMed test set (6658 samples).



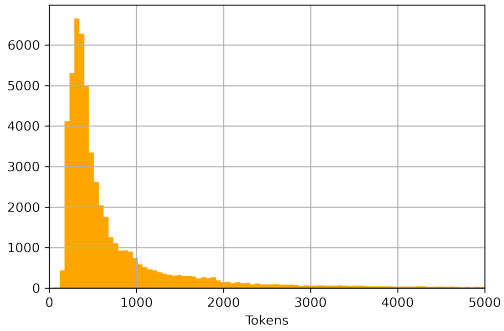
(a) CaseHold
Mean: 136, Median: 137
75-Quant: 141, 95-Quant: 148, Max: 207



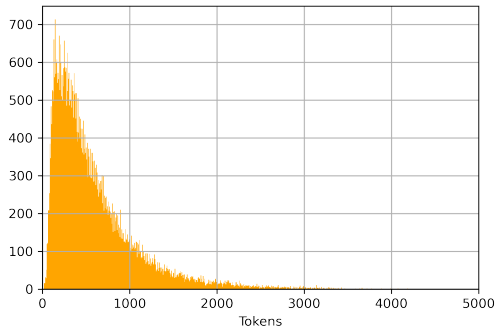
(b) ECTHR_A
Mean: 1619, Median: 984
75-Quant: 2002, 95-Quant: 5062, Max: 35416



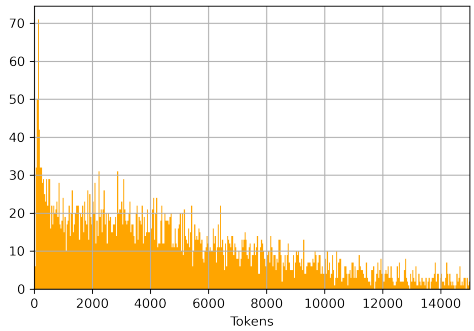
(c) ECTHR_B
Mean: 1619, Median: 984
75-Quant: 2002, 95-Quant: 5062, Max: 35416



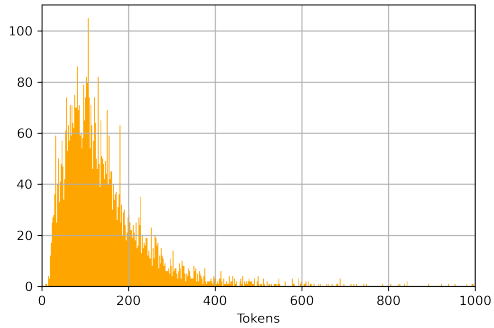
(d) EurLex
Mean: 1133, Median: 453
75-Quant: 871, 95-Quant: 4147, Max: 140103



(e) LEDGAR
Mean: 3044, Median: 2572
75-Quant: 3996, 95-Quant: 7057, Max: 109759

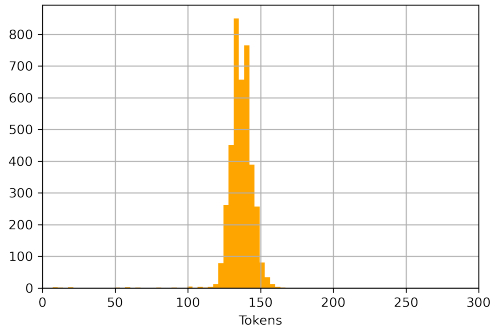


(f) SCOTUS
Mean: 5853, Median: 4364
75-Quant: 7933, 95-Quant: 16168, Max: 88566

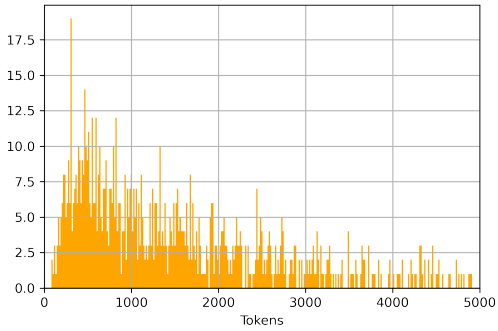


(g) Unfair_TOS
Mean: 145, Median: 120
75-Quant: 180, 95-Quant: 335, Max: 2093

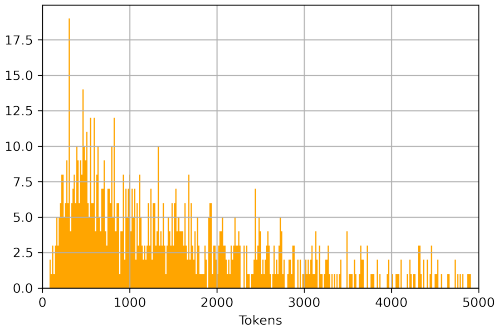
Figure 8: Histograms for the LexGLUE train set.



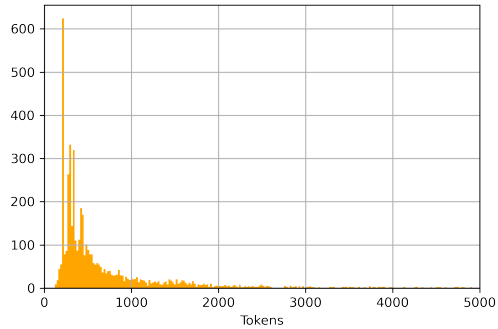
(a) CaseHold
Mean: 136, Median: 137
75-Quant: 142, 95-Quant: 148, Max: 185



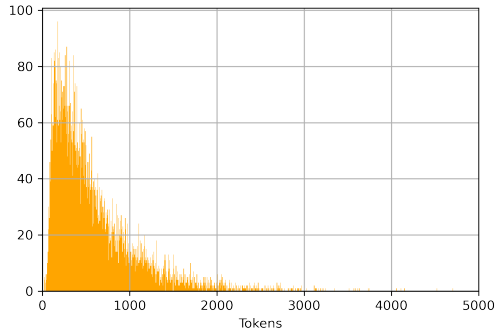
(b) ECTHR_A
Mean: 1619, Median: 984
75-Quant: 2002, 95-Quant: 5062, Max: 35416



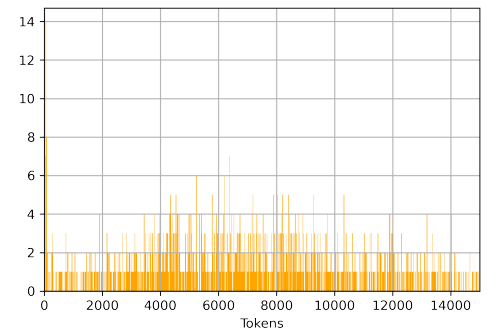
(c) ECTHR_B
Mean: 1784, Median: 1255
75-Quant: 2320, 95-Quant: 5119, Max: 14493



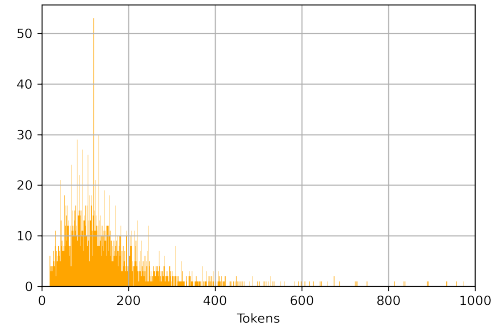
(d) EurLex
Mean: 1312, Median: 432
75-Quant: 910, 95-Quant: 5664, Max: 51944



(e) LEDGAR
Mean: 587, Median: 444
75-Quant: 775, 95-Quant: 1549, Max: 4706

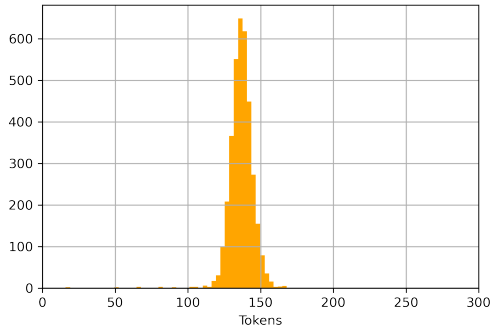


(f) SCOTUS
Mean: 8625, Median: 7591
75-Quant: 11558, 95-Quant: 19719, Max: 38098

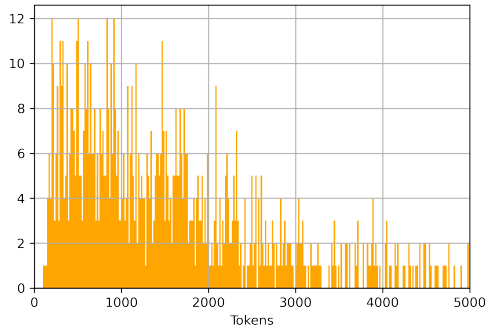


(g) Unfair_TOS
Mean: 154, Median: 125
75-Quant: 192, 95-Quant: 356, Max: 1104

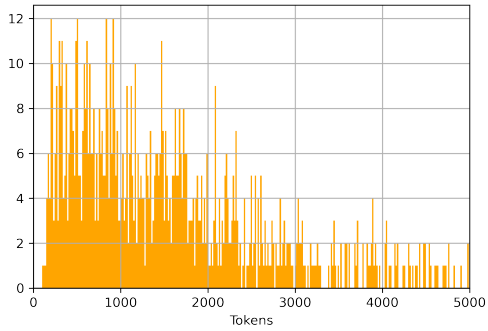
Figure 9: Histograms for the LexGLUE validation set.



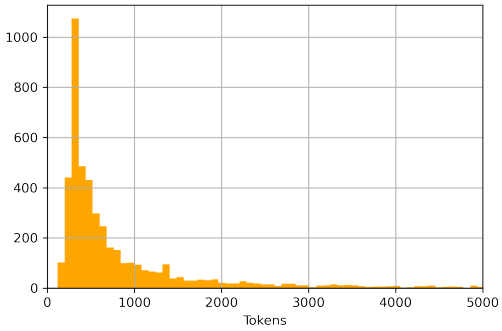
(a) CaseHold
Mean: 136, Median: 137
75-Quant: 141, 95-Quant: 148, Max: 168



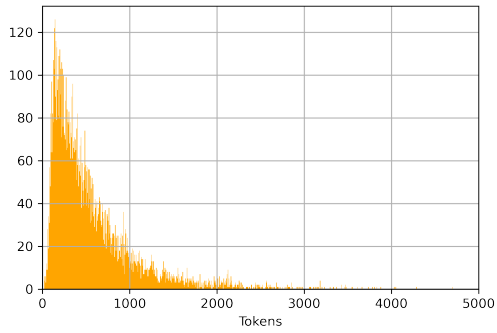
(b) ECTHR_A
Mean: 1925, Median: 1412
75-Quant: 2401, 95-Quant: 5700, Max: 15919



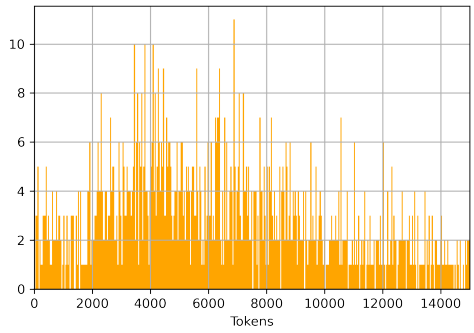
(c) ECTHR_B
Mean: 1925, Median: 1412
75-Quant: 2401, 95-Quant: 5700, Max: 15919



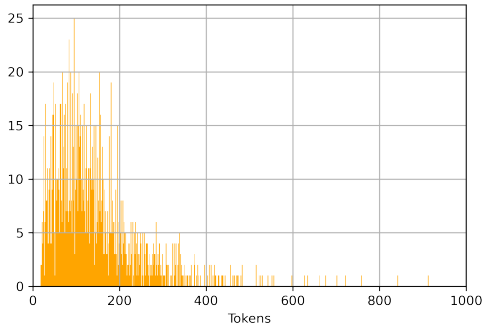
(d) EurLex
Mean: 1871, Median: 511
75-Quant: 1154, 95-Quant: 8177, Max: 200749



(e) LEDGAR
Mean: 562, Median: 407
75-Quant: 723, 95-Quant: 1563, Max: 5543



(f) SCOTUS
Mean: 8687, Median: 6883
75-Quant: 11296, 95-Quant: 21075, Max: 89379



(g) Unfair_TOS
Mean: 146, Median: 120
75-Quant: 182, 95-Quant: 338, Max: 1368

Figure 10: Histograms for the LexGLUE test set.