

# Can we Pretrain a SotA Legal Language Model on a Budget From Scratch?

Joel Niklaus and Daniele Giofré

Thomson Reuters Labs, Zug, Switzerland

firstname.lastname@thomsonreuters.com

## Abstract

Even though many efficient transformers have been proposed, only few such models are available for specialized domains. Additionally, since the pretraining process is extremely costly in general – but even more so as the sequence length increases – it is often only in reach of large research labs. One way of making pretraining cheaper is the Replaced Token Detection (RTD) task, by providing more signal during training compared to MLM, since the loss can be computed over all tokens. In this work, we train Longformer models with the efficient RTD task on long-context legal data to showcase that pretraining efficient LMs is possible using less than 12 GPU days. We evaluate the trained models on challenging summarization tasks requiring the model to summarize complex long texts. We find that both the small and base models outperform their baselines on the in-domain BillSum and out-of-domain PubMed tasks in their respective parameter range. We publish our models as a resource for researchers and practitioners.

## 1 Introduction

Pretrained transformer models have achieved excellent performance across various Natural Language Processing (NLP) tasks such as Text Classification (TC), Named Entity Recognition (NER), Question Answering (QA) and summarization (Devlin et al., 2019; Yang et al., 2020; He et al., 2021; Zhang et al., 2020a).

Transfer learning is to a large extent responsible for this success (Howard and Ruder, 2018). Usually, transformer models are pretrained in a self-supervised way on large unlabeled corpora (Devlin et al., 2019; Radford et al., 2018). Pretraining is very resource intensive (especially for large models), thus making it costly and only available for large organizations (Sharir et al., 2020). The Masked Language Modeling (MLM) task has been

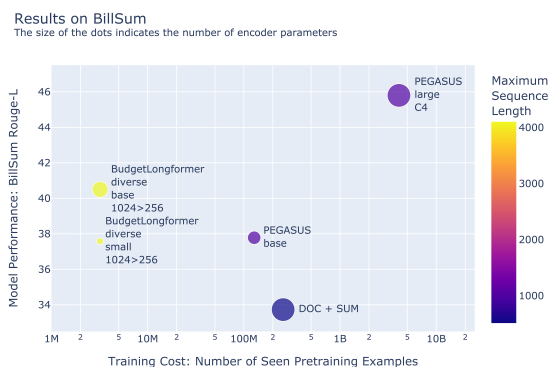


Figure 1: Results on BillSum (log-scaled x-axis)

very successful, with many models adopting the task in pretraining (Devlin et al., 2019; Liu et al., 2019; Beltagy et al., 2020; Zaheer et al., 2021). Since typically only 15% of the tokens are masked, the loss can be computed for those tokens only.

Clark et al. (2020) introduced the Replaced Token Detection (RTD) task, enabling loss computation on all tokens for efficient training. On the GLUE benchmark (Wang et al., 2018), ELECTRA matches RoBERTa (Liu et al., 2019) and XLNet (Yang et al., 2020) using 1/4 their compute. Although ELECTRA’s training strategy seems very promising, to the best of our knowledge, only few works have adopted the RTD task so far (He et al., 2021; Kanakarajan et al., 2021).

On another note, domain-specific pretraining has been shown to improve downstream performance in many domains such as law (Chalkidis et al., 2020; Xiao et al., 2021), biology (Lee et al., 2019), scientific articles (Beltagy et al., 2019), clinical documents (Li et al., 2022), or even code (Chen et al., 2021a). Despite the vast amount of legal text and the importance of training legal models for downstream tasks, there has yet to be domain-specific pretraining coupled with the RTD task for law.

Pretraining on legal documents is especially challenging, given that legal documents tend to span multiple pages (ranging from 10s to 100s of pages, which translates to tens of thousands tokens). This

is incompatible with current transformer architectures (Vaswani et al., 2017) as they often prohibit efficient processing of sequences longer than 512 tokens on current hardware due to the quadratic time and memory requirement of the attention mechanism. To solve this problem, a rich body of research investigates how transformers can be adapted to efficiently process longer input (Tay et al., 2020b; Child et al., 2019; Beltagy et al., 2020; Zaheer et al., 2021; Roy et al., 2021; Kitaev et al., 2020; Tay et al., 2021; Lee-Thorp et al., 2021).

Longformer (Beltagy et al., 2020) is one of these efficient transformer architectures for long sequences, leveraging windowed and global attention. So far, to the best of our knowledge, there does not yet exist a public Longformer model pretrained on English legal data<sup>1</sup>, although Xiao et al. (2021) have proven the effectiveness of the Longformer in dealing with long legal text in many Chinese-related tasks. This work aims to fill this gap.

To test the ability to grasp long-distance dependencies in the text, we mainly evaluated our models on the task of automatic (abstractive) summarization. It consists 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).

In particular, we used the BillSum dataset (Kornilova and Eidelman, 2019), as a domain-specific summarization task, and the PubMed dataset (Cohan et al., 2018), to evaluate the model’s ability outside the legal context (i.e., in the biomedical context). On BillSum, we achieve a new state-of-the-art (SOTA) (see Figure 1) in our parameter range. On Pubmed, we obtain comparable metrics even though the Language Model (LM) has only been pretrained on legal data and the tokenizer is also optimized for legal data (see Figure 2).

We emphasize that this performance was achieved with minimal pretraining due to the combination of the RTD task and the Longformer infrastructure making our LM very attractive from the perspective of building costs. For example, our model saw 3.2M examples during pretraining, while RoBERTa (Liu et al., 2019) or PEGASUS-large (Zhang et al., 2020a) saw 4.1B examples

(nearly 1300x more). For reference, RoBERTa was trained for 1024 GPU days (>42x more than our base model), while our small and base models only used 12 and 24 GPU days respectively (16GB NVIDIA V100 GPUs for all models).<sup>2</sup>

## Contributions

The contributions of this paper are three-fold:

- We train and release a new model pretrained on recently published curated English legal text (Henderson et al., 2022), capable of handling input spans longer than 512 tokens out of the box.
- Using Longformer and RTD, dubbed Budget-Longformer, we achieve a new SOTA on BillSum and PubMed compared to models of the same size. Our small model even outperforms a transformer base model (Vaswani et al., 2017) containing almost 4 times more encoder parameters (110M vs. 29M). On BillSum it performs on par with a PEGASUS base model (Zhang et al., 2020a) whose encoder is also almost 4 times larger and has been pretrained specifically for the abstractive summarization task in mind.
- We verified that pretraining with the RTD task is suitable for down-stream summarization tasks by evaluating our model on an out-of-domain benchmark (PubMed), obtaining comparable results with summarization-specific architectures.

## Main Research Questions

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

**RQ1:** *Is it possible to train a LM with domain (e.g. legal) expertise efficiently from scratch, reducing costs?*

**RQ2:** *How does our model compare with other models on the challenging legal domain-specific BillSum summarization benchmark?*

**RQ3:** *How well does our model compare with other models on the biomedical out-of-domain PubMed summarization benchmark?*

## 2 Related Work

### Domain-Specific Language Models

Previous work showed that domain-specific pretraining achieves promising results 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), clinical

<sup>1</sup>On the web there is a model based on Longformer in the legal domain, but it offers no model card (<https://huggingface.co/saibo/legal-longformer-base-4096>). Also, concurrent to our work, Mamakas et al. (2022) trained legal Longformer models, but they are private. Additionally, concurrently, Hua et al. (2022) trained Reformer (Kitaev et al., 2020) models with the RTD task on legal data.

<sup>2</sup>Although Zhang et al. (2020a) do not report the compute used, we expect it to be similar to RoBERTa.

documents (Li et al., 2022), or even code (Chen et al., 2021a).

Gururangan et al. (2020) show that continued pretraining on a RoBERTa checkpoint on biomedical data, scientific articles in computer science, and reviews, clearly improves downstream performance in the respective domain-specific datasets. The effect was less pronounced on news domain datasets, presumably because RoBERTa has seen many news articles during pretraining already.

### Long Document Processing

In the past few years, a vast amount of research has been devoted to addressing the problem of quadratic time and memory complexity associated with the dense attention mechanism (Vaswani et al., 2017), practically limiting the maximum sequence length severely (often to 512 tokens) (Tay et al., 2020b; Child et al., 2019; Beltagy et al., 2020; Zaheer et al., 2021; Roy et al., 2021; Kitaev et al., 2020; Tay et al., 2021; Lee-Thorp et al., 2021). These research works have given rise to a new class of transformers, referred to as sparse transformers or efficient transformers (Tay et al., 2020b). Reducing the cost associated with the computation of the dense attention matrix while maintaining the same performance is the core idea behind efficient transformers. This is often achieved by introducing sparsity in the attention matrix in a variety of ways that may be fixed pattern such as local (windowed) attention (Child et al., 2019; Beltagy et al., 2020), global attention (Zaheer et al., 2021) or learnable patterns such as routing attention (Roy et al., 2021) and LSH attention (Kitaev et al., 2020) or a random pattern (Zaheer et al., 2021; Tay et al., 2021). Recently, Lee-Thorp et al. (2021) proposed to use Fourier transforms instead of the attention layer. Tay et al. (2020b) provide a comprehensive list of efficient transformers and the detailed description of their attention mechanism. (Tay et al., 2020a) proposed a series of tasks designed for testing the capabilities of these different models suitable for longer inputs. However, this so-called “Long Range Arena” considers mostly artificial tasks, with the goal of evaluating the models independent of any pretraining.

### Efficient Pretraining

ELECTRA-style pretraining (Clark et al., 2020) has been shown to reduce training cost substantially, while matching the performance of SOTA LMs. ELECTRA leverages a smaller generator

model (discarded after pretraining), that changes some tokens. The larger discriminator model (used for down-stream tasks) must predict for each token if it was changed by the generator or not, similar to how Generative Adversarial Networks (GANs) are trained (Goodfellow et al., 2014). This enables the loss to be relevant for every token, leading to much faster and thus more efficient training.

## 3 Datasets

### 3.1 Pile of Law

Henderson et al. (2022) recently released a large-scale English corpus suitable for pretraining LMs. It contains 256 GB of diverse legal text in English from various jurisdictions and judicial bodies including for example bills, court decisions and contracts from the US, Canada, and Europe even though the focus clearly lies on US data. While there are 28 US datasets available (253.25 GB or 99%), there is only 1 Canadian dataset<sup>3</sup> (243 MB or 0.09%), 3 European datasets<sup>4</sup> (2.3 GB or 0.9%), and 2 international datasets<sup>5</sup> (212 MB or 0.08%). The non-US datasets only cover the categories “Legal Case Opinions and Filings”, “Laws” and “Conversations”, but do not cover categories “Legal Analyses”, “Contracts / Business Documents” and “Study Materials”, whereas the US data is much more diverse and covers all categories.

### 3.2 BillSum

Kornilova and Eidelman (2019) introduced a legislative summarization dataset covering 21K 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 5K to 20K characters ( $\sim$  1K to 4K tokens<sup>6</sup>) with their summaries being up to 5K characters ( $\sim$  1K tokens) long (see Appendix C for more details).

### 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

<sup>3</sup>Canadian Court Opinions (ON, BC)

<sup>4</sup>European Court of Human Rights Opinions, EUR-LEX and European Parliament Proceedings Parallel Corpus

<sup>5</sup>World Constitutions and U.N. General Debate Corpus

<sup>6</sup>Our experiments show that using our tokenizer, one token corresponds to 5.33 characters on average.

words. Thus, similar to the BillsSum 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 C for more details).

## 4 BudgetLongformer

In the legal domain, it is especially important that models can handle long input. So far, there does not exist an English legal model capable of handling more than 512 tokens. Since many tasks in legal NLP are formulated as TC problems, a hierarchical architecture has been used frequently to process long documents (Chalkidis et al., 2019; Niklaus et al., 2021, 2022, 2023). This simple hierarchical architecture, however, cannot be easily adapted to solve the more complex sequence-to-sequence tasks like token classification or summarization because it compresses the long input sequence into a single token. For this reason, in this work, we pre-train a more versatile Longformer model. To make pretraining more affordable, we trained the well-proven Longformer model (Beltagy et al., 2020) with the RTD task proposed by Clark et al. (2020).

### 4.1 Longformer

We opted for the Longformer method over other efficient transformer architectures because it seems to work robustly<sup>7</sup> and is heavily used in the literature (Xiao et al., 2021; Dai et al., 2022; Maroudas et al., 2022). Longformer (Beltagy et al., 2020) proposed three sparse attention mechanisms: Sliding Window Attention, Dilated Sliding Window Attention and Global + Sliding Window. We follow their recommendations and use the Global + Sliding Window attention mechanism because we pretrain an encoder-only model.

### 4.2 Replaced Token Detection

Inspired by GAN training (Goodfellow et al., 2014), the RTD task adapts this training framework to NLP. The drawback of training with MLM is that the loss can only be computed for the masked tokens (usually 15%). With RTD training, a smaller generator model (usually 1/3 the size of the discriminator) solves the MLM task. The discriminator receives the predictions of the generator and determines for each token, whether it is original or changed by the generator. This leads to the loss being computed for each token for the discriminator,

<sup>7</sup>164 models on huggingface hub as of January 3rd 2023

PileOfLaw Subset	Dataset Size	# Words	# Documents
caselaw			
<b>CL Opinions</b>	<b>59.29GB</b>	<b>7.65B</b>	<b>3.39M</b>
diverse			
<b>Total</b>	<b>73.04GB</b>	<b>8.91B</b>	<b>2.1M</b>
CL Opinions	8.74GB	1.13B	500K
CL Docket Entries	17.49GB	1.80B	500K
U.S. State Codes	6.77GB	829.62M	157
U.S. Code	0.27GB	30.54M	43
EUR-Lex	1.31GB	191.65M	106K
Edgar Contracts	7.26GB	0.97B	500K
Atticus Contracts	31.2GB	3.96B	488K

Table 1: The datasets used for pretraining our models. CL is short for Court Listener

thus transporting more information per forward-pass and leading to more efficient training.

## 5 Experimental Setup

In this section, we describe how we set up the experiments. For all experiments, we used the huggingface transformers library (Wolf et al., 2020) available under an Apache 2.0 license and AMP mixed precision training and evaluation to reduce costs and GPU memory.

### 5.1 Tokenizer

We trained a byte-level BPE tokenizer (Wang et al., 2019) akin to Beltagy et al. (2020) with a large 64K token vocabulary to encode complex legal language well. We trained the tokenizer using the huggingface tokenizers library<sup>8</sup> on the entire PileOfLaw training split ( $\sim 192\text{GB}$ ,  $\sim 22.5\text{B}$  tokens,  $\sim 7.5\text{M}$  documents), covering a wide array of English (mostly US) legal texts without preprocessing/cleaning due to the high-quality data.

### 5.2 Pretraining

Henderson et al. (2022) have experienced difficulties when the language model was trained on the entire PileOfLaw. We believe that the highly imbalanced dataset concerning text types (contracts, court decisions, legislation, etc.) could have been a reason for the training instability.<sup>9</sup> This led us to do a sanity check by training only on caselaw first and then to subselect only the most important and largest subsets of the PileOfLaw for training the diverse model, leading to stable pretraining (see Section 6). On the contrary, on the summarization tasks, the diverse model – which includes more

<sup>8</sup><https://github.com/huggingface/tokenizers>

<sup>9</sup>However, the large model size could also explain the training instability.



lexical and layout diversity of documents – turns out to perform better and train more robustly.

We trained the *caselaw* models on the training subset of “Court Listener Opinions” from the PileOfLaw (59.3 GB, 7.65B words, 3.39M documents). The *diverse* models were trained on caselaw (“Court Listener Opinions” & “Court Listener Docket Entries”), legislation (“US Code”, “State Codes” & “EURLEX”) and contracts (“Atticus Contracts” & “EDGAR Contracts”). To balance the training data, we limited the number of documents to 500K (this affects Court Listener Opinions, Court Listener Docket Entries and EDGAR Contracts (see Table 1 for more details). Our validation set consisted of 1000 randomly selected examples from the respective training set.<sup>10</sup> To maximally use the available data, we concatenated all the examples and cut them off in slices of the model’s maximum sequence length (4096) – in batches of 1000 examples with multiprocessing to speed up data preparation. We dropped the last slice, since it will not contain 4096 tokens.

We trained both a small (29M parameters) and a base (159M parameters) model for each configuration (caselaw and diverse data). To reach 100K steps it took 68 hours (a bit less than 3 days) for the small model and 135 hours (a bit more than 5 days) for the base model on 4 16GB NVIDIA V100 GPUs. The achieved training and evaluation losses are shown in Table 7 in Appendix A. Interestingly, we find that the diverse models achieve lower training and evaluation losses. Please find more training details in Appendix A. Due to budget constraints, we trained for a maximum of 200K steps. Surprisingly, lower pretraining loss from 200K-step models did not transfer to downstream tasks. We hypothesize that a larger batch size might lead to improvements when training longer.

### 5.3 Downstream Benchmarks

For downstream finetuning, we paired our pre-trained encoder model with a randomly initialized BART-base decoder model (Lewis et al., 2020).<sup>11</sup> For BillSum, we set the maximum input length to 1024 and the maximum target length to 256 to save compute. However, many summaries get cut off at 256 tokens. This is why we took our best model

<sup>10</sup>We used such a small validation set to save compute.

<sup>11</sup>Interestingly, the randomly initialized decoder yielded better results than when we used the weights from the pretrained huggingface checkpoint at <https://huggingface.co/facebook/bart-base>.

and trained it with maximum input length 4096 and maximum target length 1024 (see results in Table 5 and examples in Table 12). For PubMed, we set the maximum input length to 4096 and the maximum generation length to 512. Due to high training costs, we only trained our models with one random seed (42). Our models contain 29M (small) and 159M (base) parameters in the encoder and 96M parameters in the decoder, resulting in a total of 125M (small) and 255M (base) parameters.

### 5.4 Ablation Studies

We run two ablation studies on the BillSum dataset, testing the influence of the pretraining corpus and the number of pretraining steps. To reduce computational costs, we set the maximum input and generation lengths to 1024 and 128 respectively.

# Steps	Size	Rouge-1 ↑	Rouge-2 ↑	Rouge-L ↑
100K	small	51.62	30.84	40.22
200K	small	49.02	27.02	36.98
100K	base	56.10	36.50	45.17
200K	base	55.30	35.47	44.30

Table 2: Models pretrained on caselaw only.

Corpus	Size	Rouge-1 ↑	Rouge-2 ↑	Rouge-L ↑
caselaw	small	51.62	30.84	40.22
diverse	small	53.61	33.54	42.50
caselaw	base	56.10	36.50	45.17
diverse	base	54.87	35.63	44.21

Table 3: Models pretrained for 100K steps.

#### Pretraining Steps

Though train and evaluation losses decrease steadily with more pretraining steps (see Table 7), surprisingly, models trained longer underperform on the BillSum benchmark (see Table 2). We hypothesize the low pretraining batch size caused fast convergence to a local optimum, inhibiting further progress. Consequently, we use the 100K steps model checkpoints.

#### Pretraining Corpus

In total, we trained 4 models (small and base each on the caselaw and diverse corpora). In Table 3 we perform an ablation on the pretraining corpus. The results are inconclusive, with the diverse corpus outperforming for the small models and the caselaw corpus outperforming for the base models. The caselaw models were unstable during finetuning and even failed completely for some learning

rates. Together with the fact that the diverse models reached lower pretraining losses (see Table 7), we focus on the diverse models for our experiments.

We acknowledge the necessity of more ablations. Because of limited compute, we opted for the safest and cheapest choices instead of ablating them (e.g. windowed and global attention, RTD pretraining task). Additionally, we put a focus on providing our models as a resource for further research in this area and for practitioners in the field of legal NLP. We thus leave further ablations for future work (w.r.t. pretraining task, more general domain corpora, efficient transformer method, etc.).

## 6 Results

In this section, we present results for the BillSum and PubMed datasets, conducting error analysis on generated summaries. Table 4 compares models in detail. All further experiments utilize models trained on the diverse dataset.

### 6.1 BillSum

We achieve a new SOTA on BillSum in the small and base parameter range and outperform models with almost 12 times more encoder parameters and others having seen more than 1200 times more pretraining examples. The results on BillSum are presented in Figure 1 and Table 5.

We observe that even our small diverse model clearly exceeds the baseline of the original article (DOC + SUM), even though their model is based on BERT-large, containing almost 12 times more encoder parameters and pretrained for 10x more steps. Even more surprisingly, our small diverse model is on par with the PEGASUS-base model (Zhang et al., 2020a) (37.58 vs. 37.78 Rouge-L), pretrained using the Gap-Sentences task specifically designed for abstractive summarization. PEGASUS-base contains almost 4 times more encoder parameters and has seen 40 times more training examples during pretraining (128M vs. 3.2M; see Table 4). Most surprisingly, it even outperforms an LED large model<sup>12</sup> (37.58 vs. 34.23 Rouge-L) using a much longer input length (16384 vs. 1024), containing more than 8 times as many encoder parameters (257M vs. 29M) and having seen more than 1200 times more examples during pretraining.

By scaling up our model to the base size and increasing the maximum input and generation length

<sup>12</sup>[https://huggingface.co/Artifact-AI/led\\_large\\_16384\\_billsum\\_summarization](https://huggingface.co/Artifact-AI/led_large_16384_billsum_summarization)

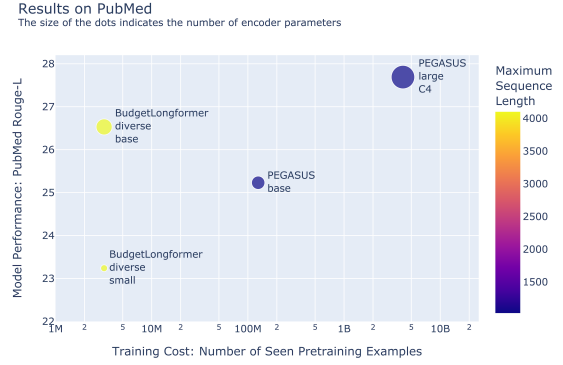


Figure 2: Results on PubMed (log-scaled x-axis)

to 4096 and 1024 tokens respectively, we even approach the performance of PEGASUS-large (43.23 vs. 45.8 Rouge-L). PEGASUS-large has seen three orders of magnitude more training examples during its pretraining in comparison to our model (4.1B vs. 3.2M) and contains almost twice as many encoder parameters (301M vs. 159M).

To conclude, it appears that pretraining with the RTD on (high-quality) in-domain data can be an effective and computationally cheap alternative to a summarization-specific model trained on web text (i.e. PEGASUS). Whether the gain is due to in-domain pretraining or the RTD task is inconclusive, and we leave these experiments for future work.

### 6.2 PubMed

We achieve a new SOTA on PubMed in the small and base parameter range and almost reach the performance of a PEGASUS large model pretrained with a summarization-specific task. The results on PubMed are presented in Figure 2 and Table 6.

Similar to the results on BillSum, our small model clearly outperforms the transformer-base model (23.24 vs. 19.02 Rouge-L) and approaches the PEGASUS-base model (23.24 vs. 25.2 Rouge-L) despite not being specifically pretrained for summarization and having seen significantly fewer examples during pretraining (3.2M vs. 128M). Similar again, our base model outperforms PEGASUS-base (26.53 vs. 25.23 Rouge-L) and almost reaches the performance of PEGASUS-large (26.53 vs. 27.69 Rouge-L) while having seen 1280 times fewer examples during pretraining (3.2M vs. 4.1B).

Our model is pretrained on the narrower domain of legal text, rather than broader C4 data used by PEGASUS. Furthermore, our model and tokenizer had no exposure to medical data in pretraining. This, combined with the high quality of legal data used in pretraining, may explain our model’s good out-of-domain performance, similar to the findings

Model Name	Source	P. Steps	P. BS	# P. Examples	# Enc. Par	# Dec. Par	MaxSeqLen	Vocab
DOC + SUM	(Kornilova and Eidelman, 2019)	1000K	256	256M	340M	–	512	30K
Transformer base	(Zhang et al., 2020a)	–	–	–	159M	187M	1024	96K
PEGASUS base	(Zhang et al., 2020a)	500K	256	128M	159M	187M	1024	96K
PEGASUS large (C4)	(Zhang et al., 2020a)	500K	8192	4096M	301M	368M	1024	96K
LED large	(Beltagy et al., 2020)	500K	8192	4096M	257M	254M	16384	50K
LongT5 xl	(Guo et al., 2022)	1000K	2048	2048M	1224M	1626M	16384	32K
BudgetLongformer small	ours	100K	32	3.2M	29M	96M	4096	64K
BudgetLongformer base	ours	100K	32	3.2M	159M	96M	4096	64K

Table 4: Comparison of the evaluated models. For more information on the baselines, refer to the cited papers. (Abbreviations: P.: Pretraining, BS: Batch Size, Enc.: Encoder, Dec.: Decoder, Par: Parameters.)

Model	Size	MaxInLen	MaxGenLen	Rouge-1 $\uparrow$	Rouge-2 $\uparrow$	Rouge-L $\uparrow$
BudgetLongformer	small	1024	256	<b>49.85</b>	<b>29.63</b>	<b>37.58</b>
Transformer	base	512	256	44.05	21.30	30.98
PEGASUS	base	512	256	51.42	29.68	37.78
BudgetLongformer	base	1024	256	52.70	32.97	40.50
BudgetLongformer	base	4096	1024	<b>55.45</b>	<b>36.68</b>	<b>43.23</b>
DOC + SUM	large	512	512	40.80	23.83	33.73
PEGASUS (C4)	large	1024	256	<b>57.20</b>	<b>39.56</b>	<b>45.80</b>
LED	large	16384	1024	47.84	26.34	34.23

Table 5: Results on BillSum. Best results per model size are in bold.

of Taylor et al. (2022). Even though our pretraining data is out-of-domain PubMed – whereas C4, likely contains medical data – compared to PEGASUS, our models perform similarly on PubMed as on BillSum. This makes us believe the gains stem mainly from the RTD pretraining task.

Krishna et al. (2022) find that pretraining on the downstream corpus can achieve similar results as pretraining on a large upstream corpus, significantly cutting costs. Finetuning a small model on BillSum cost us approx. half-day of a 16GB V100 GPU. Pretraining the small model for 100K steps cost approx. 12 GPU days<sup>13</sup>. Pretraining and finetuning a smaller model with the RTD task on a task specific corpus might be a suitable alternative to finetuning a larger general model, yielding similar performance with shorter inference time and costs.

### 6.3 Error Analysis

We conducted an error analysis by manually inspecting 25 random summaries. Example summaries are shown in Appendix D.

**Coherence** The inspected summaries were well-structured and emulated the specific style of the reference summaries in the respective domains<sup>14</sup>.

<sup>13</sup>For the base model the numbers are approx. double

<sup>14</sup>In future work, we will corroborate these findings by performing human evaluations with domain experts.

**Consistency** We find the summaries mostly factually aligned with the source.<sup>15</sup> However, sometimes it copies formulas from the source text, but then mixes up numbers.<sup>16</sup>

**Fluency** Generally, we find the summaries to be fluent<sup>17</sup> and grammatically correct.

**Relevance** In general, the model summaries contain important content from the source document. However, we find repetitions to be a repeating issue in both BillSum and PubMed summarization. In the BillSum task, the model occasionally uses the same start of the sentence multiple times instead of providing a longer list<sup>18</sup>. It correctly imitates the lists often given in BillSum summaries, but then seems to struggle with continuing lists to more entries. Other times it manages well to formally continue the lists, but repeats list items. In the PubMed task, in one particular summary, a phrase

<sup>15</sup>e.g. “imaging guidance improved the accuracy of intra-articular injections of the knee ( 96.7% versus 81.0%,  $p < 0.001$  ) and shoulder ( 97.3% versus 65.4%,  $p < 0.001$  )”

<sup>16</sup>e.g. “[ a1c ( % ) = [ 0.021 mbg ( mg / dl ) + 4.3,  $r = 0.92$  ] + 4.3,  $r = 0.58$  ]”

<sup>17</sup>Repetitions are discussed in “Relevance”

<sup>18</sup>e.g. “**amends the agricultural marketing act of 1946 to** terminate the authority of the secretary of agriculture (usda) to: (1) livestock processing plant processing plant slaughter, and (2) slaughtering plant slaughter. **amends the agricultural marketing act of 1946 to:** (1) revise minimum reporting requirements; and (2) revise reporting requirements”

Model	Size	MaxInLen	MaxGenLen	Rouge-1 ↑	Rouge-2 ↑	Rouge-L ↑
BudgetLongformer	small	4096	512	<b>34.98</b>	<b>13.56</b>	<b>23.24</b>
Transformer	base	512	256	33.94	7.43	19.02
PEGASUS	base	512	256	39.98	15.15	25.23
BudgetLongformer	base	4096	512	<b>41.16</b>	<b>18.15</b>	<b>26.53</b>
PEGASUS (C4)	large	1024	256	<b>45.49</b>	<b>19.90</b>	<b>27.69</b>
LongT5	xl	16384	512	<b>50.23</b>	<b>24.76</b>	<b>46.67</b>

Table 6: Results on PubMed. Best results per model size are in bold.

gets repeated 10 times. Even in a high scoring example (Rouge1: 62.2, RougeL: 48.5, 464 tokens summary length), a sentence is repeated three times. Here, in contrast to BillSum, the repetitions are also occurring on a lower level.<sup>19</sup>

Generally, the problems are similar in the BillSum and the PubMed tasks; however, they are less pronounced in the in-domain BillSum dataset.

## 7 Conclusions and Future Work

### 7.1 Answers to Main Research Questions

**RQ1:** *Is it possible to train a LM with domain (e.g. legal) expertise efficiently from scratch, reducing costs?* Yes, this work demonstrates the feasibility of pretraining a domain-expertise LM from scratch with minimal compute, matching performance of methods exposed to three orders of magnitude more pretraining examples. Particularly when a high-performing large teacher model is unavailable, our method is advisable.

**RQ2:** *How does our model compare with other models on the challenging legal domain-specific BillSum summarization benchmark?* Our LMs compare favorably to baselines on the challenging domain-specific summarization benchmark BillSum, necessitating long input processing. Our small model outperforms the larger PEGASUS-base, and our base model almost reaches the performance of the larger PEGASUS-large. Both baselines have been pretrained with much more compute and data, and additionally with a pretraining task crafted specifically for summarization.

**RQ3:** *How well does our model compare with other models on the biomedical out-of-domain PubMed summarization benchmark?* Our results on the out-of-domain PubMed summarization benchmark show that our models compare favorably to baselines. Again, our small model

outperforms PEGASUS-base and our base model approaches PEGASUS large.

### 7.2 Conclusion

In this work, we show that we can successfully pretrain Longformer models with the RTD task on a Budget. Using very little pretraining, we can achieve SOTA performance on the challenging legal summarization task BillSum, outperforming PEGASUS, that has been pretrained specifically for summarization. Our model even outperforms PEGASUS on the out-of-domain PubMed dataset involving biomedical research articles. To sum up, we present a simple and extremely cheap way of pretraining a long-context LM in cases without the availability of a large teacher model.

### 7.3 Future Work

Future work could test our models on further legal downstream benchmarks such as LexGLUE (Chalkidis et al., 2021), ClassActionPrediction (Semo et al., 2022), CUAD (Hendrycks et al., 2021) or MultiLexSum (Shen et al., 2022). Additionally, one can test whether the out-of-domain results hold on other out-of-domain summarization datasets, such as BigPatent (Sharma et al., 2019) or ArXiv (Cohan et al., 2018). Future work could further scale up the models in terms of batch size, pre-training steps, parameter count and data size to test what further gains can be achieved. Additionally, to further save compute and enhance models, one could explore warm-starting ELECTRA pretraining from existing checkpoints.. The difficulty, of course, lies in getting a suitable generator and discriminator, trained with the same tokenizer. One possible setup might be Longformer-base as the generator and Longformer-large as the discriminator. Finally, one can investigate the use of other efficient transformers with the RTD task.

<sup>19</sup>e.g. “hemoglobin glycated hemoglobin ( hba1c )”



## Limitations

ELECTRA-style training has the disadvantage of the setup being slightly more complicated, requiring a generator and a discriminator. Additionally, the generator should be smaller than the discriminator to ensure stable training. This makes it difficult to warm start from available checkpoints, since two models of different sizes are required. Often, small models are not released, which makes it difficult to warm-start base models using the RTD task. We leave the direction of warm starting a large discriminator with a base generator to future work.

Except for EUR-LEX (1.31 GB or 1.8% of our diverse dataset), our models have only seen US data during the pretraining phase. So, while these models are expected to work well on US data or datasets with similar content such as heavily influenced by the US or mainly common-law based, legal data from Europe for example is expected to look very different (mainly civil-law based except for the UK) and often translated from the original European languages. Thus, our models are not expected to transfer well to such kind of data.

Because of insufficient compute, we were not able to scale up our models in terms of parameter size, batch size and number of pretraining steps. So while we can show that our approach scales well from the small to the base model, it is unknown if this continues to even larger model sizes. Although it is expected to produce better results, we do not know if using a higher batch size and more pretraining steps boosts performance significantly. Additionally, the lacking compute budget made evaluating on more and especially large datasets like BigPatent impossible. Therefore, we cannot give any conclusions at this point to whether our results are robust across a wide range of datasets.

So far, we did not evaluate our summarization models using newer reference-based metrics such as BERTScore (Zhang et al., 2020b) or BARTScore (Yuan et al., 2021), or reference-free metrics such as SUPERT (Gao et al., 2020) or Semantic Distribution Correlation (SDC) (Chen et al., 2021b). However, our baselines used ROUGE only, requiring us to rerun experiments for comparison using newer scores, straining our low compute budget.

So far, we did not have the resources to conduct a thorough human expert evaluation of the quality of our summarization outputs. Such an evaluation would be needed for production systems and for better comparison of models. However,

it also requires highly educated medical experts (for PubMed) or lawyers with specific expertise in US bills (for BillSum) respectively, and thus a prohibitively high amount of resources.

For comparing the efficiency of pretraining, number of FLOPs would probably be best. We compared the models’ efficiency based on the number of seen examples during pretraining, due to ready availability (most papers report batch size and number of steps, but few papers report FLOPs).

## 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 method makes significantly reduces the compute requirements and thus the carbon footprint.

As with any large LM there is the risk of it producing biased or unfair output. Researchers using the model should put into place respective safeguards to identify biased and/or toxic language.

## References

- 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. 2020. [Longformer: The Long-Document Transformer](#). *arXiv:2004.05150 [cs]*. ArXiv: 2004.05150.
- Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. 2019. [Neural Legal Judgment Prediction in English](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4317–4323, Florence, Italy. Association for Computational Linguistics.
- Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. [LEGAL-BERT: The Muppets straight out of Law School](#). *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.
- Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray,

- Nick Ryder, Mikhail Pavlov, Alethea Power, Lukasz Kaiser, Mohammad Bavarian, Clemens Winter, Philippe Tillet, Felipe Petroski Such, Dave Cummings, Matthias Plappert, Fotios Chantzis, Elizabeth Barnes, Ariel Herbert-Voss, William Hebgen Guss, Alex Nichol, Alex Paino, Nikolas Tezak, Jie Tang, Igor Babuschkin, Suchir Balaji, Shantanu Jain, William Saunders, Christopher Hesse, Andrew N. Carr, Jan Leike, Josh Achiam, Vedant Misra, Evan Morikawa, Alec Radford, Matthew Knight, Miles Brundage, Mira Murati, Katie Mayer, Peter Welinder, Bob McGrew, Dario Amodei, Sam McCandlish, Ilya Sutskever, and Wojciech Zaremba. 2021a. [Evaluating Large Language Models Trained on Code](#). *arXiv:2107.03374 [cs]*. ArXiv: 2107.03374.
- Wang Chen, Piji Li, and Irwin King. 2021b. [A Training-free and Reference-free Summarization Evaluation Metric via Centrality-weighted Relevance and Self-referenced Redundancy](#). 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 404–414, Online. Association for Computational Linguistics.
- Yangbin Chen, Yun Ma, Xudong Mao, and Qing Li. 2019. [Multi-Task Learning for Abstractive and Extractive Summarization](#). *Data Science and Engineering*, 4(1):14–23.
- Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. 2019. [Generating Long Sequences with Sparse Transformers](#). *arXiv:1904.10509 [cs, stat]*. ArXiv: 1904.10509.
- Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. [ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators](#). *arXiv:2003.10555 [cs]*. ArXiv: 2003.10555.
- 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.
- Xiang Dai, Ilias Chalkidis, Sune Darkner, and Desmond Elliott. 2022. [Revisiting Transformer-based Models for Long Document Classification](#). *arXiv:2204.06683 [cs]*. ArXiv: 2204.06683.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#). *arXiv:1810.04805 [cs]*. ArXiv: 1810.04805.
- Yang Gao, Wei Zhao, and Steffen Eger. 2020. [SUPERT: Towards New Frontiers in Unsupervised Evaluation Metrics for Multi-Document Summarization](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1347–1354, Online. Association for Computational Linguistics.
- Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. [Generative Adversarial Networks](#). ArXiv:1406.2661 [cs, stat].
- Mandy Guo, Joshua Ainslie, David Uthus, Santiago Ontanon, Jianmo Ni, Yun-Hsuan Sung, and Yinfei Yang. 2022. [LongT5: Efficient text-to-text transformer for long sequences](#). In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 724–736, Seattle, United States. Association for Computational Linguistics.
- Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. [Don’t Stop Pretraining: Adapt Language Models to Domains and Tasks](#). *arXiv:2004.10964 [cs]*. ArXiv: 2004.10964.
- Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2021. [DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing](#). *arXiv:2111.09543 [cs]*. ArXiv: 2111.09543.
- Peter Henderson, Mark S. Krass, Lucia Zheng, Neel Guha, Christopher D. Manning, Dan Jurafsky, and Daniel E. Ho. 2022. [Pile of Law: Learning Responsible Data Filtering from the Law and a 256GB Open-Source Legal Dataset](#). ArXiv:2207.00220 [cs].
- Dan Hendrycks, Collin Burns, Anya Chen, and Spencer Ball. 2021. [CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review](#). ArXiv:2103.06268 [cs].
- Jeremy Howard and Sebastian Ruder. 2018. [Universal Language Model Fine-tuning for Text Classification](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 328–339, Melbourne, Australia. Association for Computational Linguistics.
- Wenyue Hua, Yuchen Zhang, Zhe Chen, Josie Li, and Melanie Weber. 2022. [LegalRelectra: Mixed-domain Language Modeling for Long-range Legal Text Comprehension](#). ArXiv:2212.08204 [cs].
- Kamal raj Kanakarajan, Bhuvana Kundumani, and Malaikannan Sankarasubbu. 2021. [BioELECTRA: Pretrained Biomedical text Encoder using Discriminators](#). In *Proceedings of the 20th Workshop on Biomedical Language Processing*, pages 143–154, Online. Association for Computational Linguistics.
- Nikita Kitaev, Łukasz Kaiser, and Anselm Levskaya. 2020. [Reformer: The Efficient Transformer](#). *arXiv:2001.04451 [cs, stat]*. ArXiv: 2001.04451.

- 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.
- Kundan Krishna, Saurabh Garg, Jeffrey P. Bigham, and Zachary C. Lipton. 2022. [Downstream Datasets Make Surprisingly Good Pretraining Corpora](#). ArXiv:2209.14389 [cs].
- 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, 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.
- 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.
- Dimitris Mamakas, Petros Tsotsi, Ion Androutsopoulos, and Ilias Chalkidis. 2022. [Processing Long Legal Documents with Pre-trained Transformers: Modding LegalBERT and Longformer](#). ArXiv:2211.00974 [cs].
- Stelios Maroudas, Sotiris Legkas, Prodromos Malakasiotis, and Ilias Chalkidis. 2022. [Legal-Tech Open Diaries: Lesson learned on how to develop and deploy light-weight models in the era of humongous Language Models](#). ArXiv:2210.13086 [cs].
- Joel Niklaus, Ilias Chalkidis, and Matthias Stürmer. 2021. [Swiss-Judgment-Prediction: A Multilingual Legal Judgment Prediction Benchmark](#). In *Proceedings of the Natural Legal Language Processing Workshop 2021*, pages 19–35, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Joel Niklaus, Veton Matoshi, Pooja Rani, Andrea Galassi, Matthias Stürmer, and Ilias Chalkidis. 2023. [LEXTREME: A Multi-Lingual and Multi-Task Benchmark for the Legal Domain](#). ArXiv:2301.13126 [cs].
- Joel Niklaus, Matthias Stürmer, and Ilias Chalkidis. 2022. [An Empirical Study on Cross-X Transfer for Legal Judgment Prediction](#). In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 32–46, Online only. Association for Computational Linguistics.
- David Patterson, Joseph Gonzalez, Quoc Le, Chen Liang, Lluís-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.
- Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving Language Understanding by Generative Pre-Training. page 12.
- 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.
- Gil Semo, Dor Bernsohn, Ben Hagag, Gila Hayat, and Joel Niklaus. 2022. [ClassActionPrediction: A Challenging Benchmark for Legal Judgment Prediction of Class Action Cases in the US](#). In *Proceedings of the Natural Legal Language Processing Workshop 2022*, pages 31–46, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.
- Or Sharir, Barak Peleg, and Yoav Shoham. 2020. [The Cost of Training NLP Models: A Concise Overview](#). ArXiv:2004.08900 [cs].
- Eva Sharma, Chen Li, and Lu Wang. 2019. [BIG-PATENT: A Large-Scale Dataset for Abstractive and Coherent Summarization](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2204–2213, Florence, Italy. Association for Computational Linguistics.
- 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].
- 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, Samira Abnar, Yikang Shen, Dara Bahri, Philip Pham, Jinfeng Rao, Liu Yang, Sebastian Ruder, and Donald Metzler. 2020a. [Long Range Arena: A Benchmark for Efficient Transformers](#). *arXiv:2011.04006 [cs]*. ArXiv: 2011.04006.
- Yi Tay, Mostafa Dehghani, Dara Bahri, and Donald Metzler. 2020b. [Efficient Transformers: A Survey](#). *arXiv:2009.06732 [cs]*. ArXiv: 2009.06732.
- Ross Taylor, Marcin Kardas, Guillem Cucurull, Thomas Scialom, Anthony Hartshorn, Elvis Saravia, Andrew Poulton, Viktor Kerkez, and Robert Stojnic. 2022. [Galactica: A Large Language Model for Science](#). ArXiv:2211.09085 [cs, stat].
- 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.
- 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.
- Changhan Wang, Kyunghyun Cho, and Jiatao Gu. 2019. [Neural Machine Translation with Byte-Level Subwords](#). ArXiv:1909.03341 [cs].
- 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.
- Weizhe Yuan, Graham Neubig, and Pengfei Liu. 2021. [BARTScore: Evaluating Generated Text as Text Generation](#). *arXiv:2106.11520 [cs]*. ArXiv: 2106.11520.
- 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. 2020a. [PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization](#). *arXiv:1912.08777 [cs]*. ArXiv: 1912.08777.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020b. [BERTScore: Evaluating Text Generation with BERT](#). *arXiv:1904.09675 [cs]*. ArXiv: 1904.09675.

## A Hyperparameters and Training Details

Model	Data	# Steps	Train Loss	Eval Loss
small	caselaw	50K	14.61	15.78
small	caselaw	100K	13.93	15.07
small	caselaw	150K	13.63	14.77
small	caselaw	200K	13.38	14.49
small	diverse	50K	13.75	12.70
small	diverse	100K	12.78	11.66
small	diverse	150K	12.28	11.29
small	diverse	200K	12.05	11.03
base	caselaw	50K	12.40	13.76
base	caselaw	100K	11.67	12.99
base	caselaw	150K	11.31	12.58
base	caselaw	200K	11.02	12.27
base	diverse	50K	10.70	10.01
base	diverse	100K	9.86	9.22
base	diverse	150K	9.42	8.79
base	diverse	200K	9.20	8.56

Table 7: Training and Evaluation losses for the different trained models. Note that these losses are the addition of the loss of the generator and the loss of the discriminator. Since the loss of the discriminator is much smaller, it is scaled by a factor of 50 to stabilize training.

In this section, we present additional details regarding training and the chosen hyperparameters.

### A.1 Pretraining

We pretrained our models with batch size 128 and learning rate 5e-4 and 3e-4 for the small and base models respectively. We used a Longformer attention window of 256. As described in by [Clark et al. \(2020\)](#), we used 10000 warm up steps and a 4 and 3 times smaller generator than the discriminator in the small and base version respectively. In contrast to [Clark et al. \(2020\)](#), we reduced the generator’s depth (number of hidden layers) instead of its width (embedding size, hidden size and intermediate size). We used a MLM probability of 25% for



the generators. The pretraining losses are shown in Table 7.

## A.2 Downstream Benchmarks

We finetuned on the summarization datasets using early stopping on the validation set with patience of 3 epochs. We used a batch size of 32 and learning rate of  $7e-5$  after tuning in  $\{5e-4, 9e-5, 7e-5, 5e-5, 3e-5, 1e-5\}$ . We used the bart-base default config for num\_beams (4) and no\_repeat\_ngram\_size (3).

Overall, we found the diverse models to be more robust in finetuning with less failed runs and typically higher performance.

## A.3 Compute Costs

For running the pretraining, we used an AWS p3.8xlarge instance with 4 16GB NVIDIA V100 GPUs. Training the four models to 200K steps each, took approx. 36 days or 144 GPU days in total (almost 6 days and almost 12 days for the small and base models respectively). Previous debug runs additionally consumed approx. 12 GPU days. For running the finetuning experiments, we used an AWS p3.16xlarge instance with 8 16GB NVIDIA V100 GPUs. Running the BillSum, and PubMed experiments including debugging and hyperparameter tuning took approximately 25 and 7 GPU days in total respectively. Putting it all together, we trained our models for 176 16GB NVIDIA V100 GPU days.

## B Library Versions

We used the following versions to the libraries in a pip requirements.txt format:

```
datasets==2.4.0
huggingface-hub==0.9.0
nltk==3.7
pandas==1.3.5
rouge-score==0.1.2
scikit-learn==1.0.2
scipy==1.7.3
tokenizers==0.12.1
torch==1.12.1
tqdm==4.64.0
transformers==4.21.1
```

## C Data Details

We used our own tokenizer to calculate the number of tokens. 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.

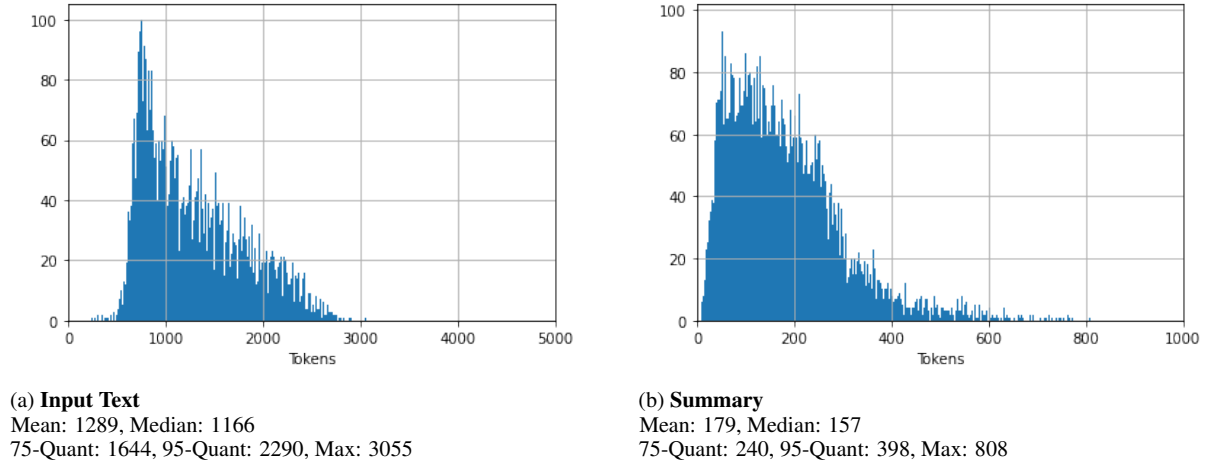


Figure 3: Histograms for the BillSum training set (18949 samples).

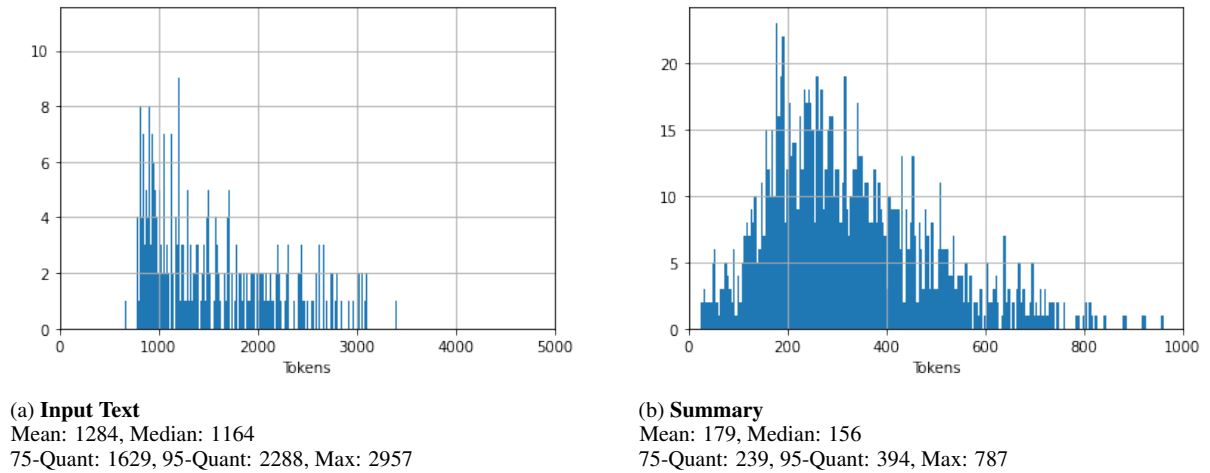
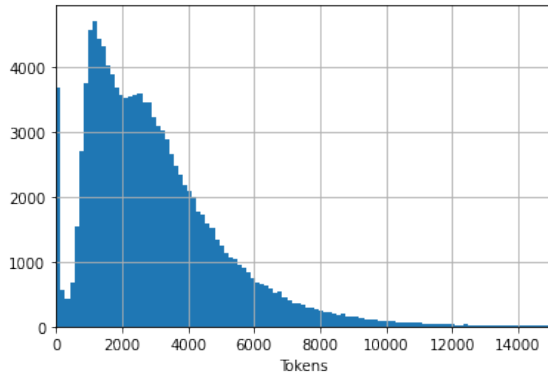


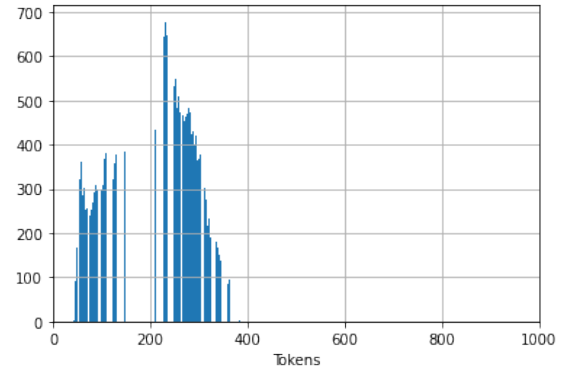
Figure 4: Histograms for the BillSum test set (3269 samples).

## D Examples

Example summaries are displayed in Tables [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [13](#), [15](#), and [16](#). 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.

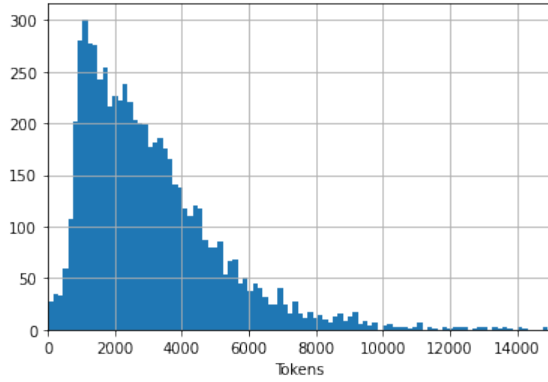


(a) **Input Text**  
Mean: 3044, Median: 2572  
75-Quant: 3996, 95-Quant: 7057, Max: 109759

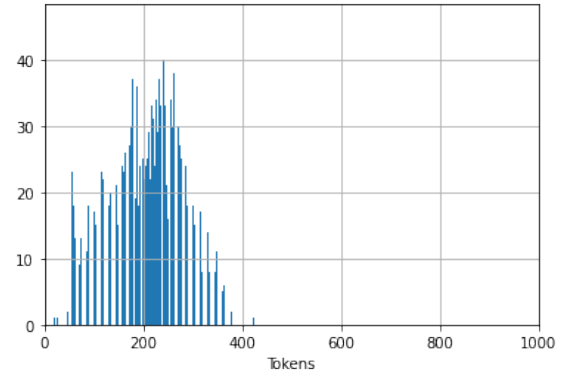


(b) **Summary**  
Mean: 202, Median: 208  
75-Quant: 262, 95-Quant: 326, Max: 391

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

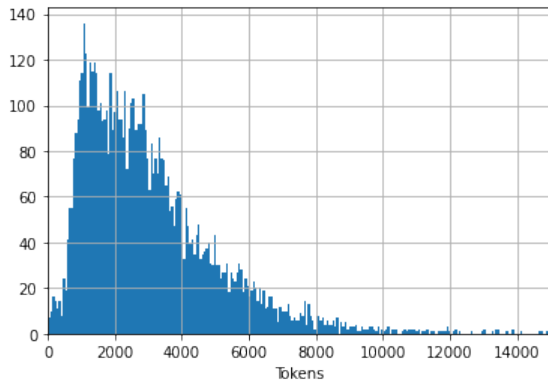


(a) **Input Text**  
Mean: 3112, Median: 2609  
75-Quant: 4011, 95-Quant: 6968, Max: 119269

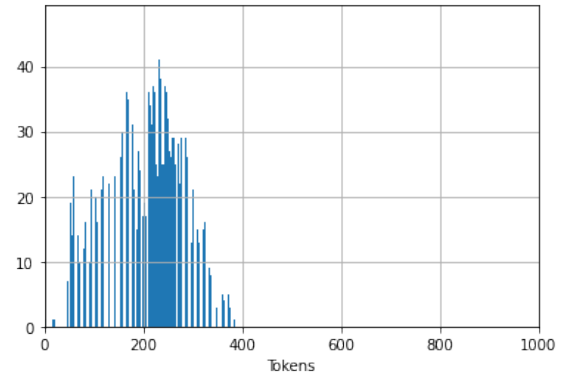


(b) **Summary**  
Mean: 203, Median: 209  
75-Quant: 263, 95-Quant: 330, Max: 518

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



(a) **Input Text**  
Mean: 3093, Median: 2596  
75-Quant: 3964, 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).



Bottom 5% example (Sorted by rougeL)	
Document	<p>SECTION 1. SHORT TITLE. This Act may be cited as the "Child Citizenship Act of 2000".</p> <p>TITLE I—CITIZENSHIP FOR CERTAIN CHILDREN BORN OUTSIDE THE UNITED STATES</p> <p>SEC. 101. AUTOMATIC ACQUISITION OF CITIZENSHIP FOR CERTAIN CHILDREN BORN OUTSIDE THE UNITED STATES.</p> <p>(a) In General.—Section 320 of the Immigration and Nationality Act (8 U.S.C. 1431) is amended to read as follows:</p> <p>"children born outside the united states and residing permanently in the united states; conditions under which citizenship automatically acquired</p> <p>"Sec. 320. (a) A child born outside of the United States automatically becomes a citizen of the United States when all of the following conditions have been fulfilled: "(1) At least one parent of the child is a citizen of the United States, whether by birth or naturalization. "(2) The child is under the age of eighteen years. "(3) The child is residing in the United States in the legal and physical custody of the citizen parent pursuant to a lawful admission for permanent residence. "(b) Subsection (a) shall apply to a child adopted by a United States citizen parent if the child satisfies the requirements applicable to adopted children under section 101(b)(1).". (b) Clerical Amendment.—The table of sections of such Act is amended by striking the item relating to section 320 and inserting the following: "Sec. 320. Children born outside the United States and residing permanently in the United States; conditions under which citizenship automatically acquired."</p> <p>SEC. 102. ACQUISITION OF CERTIFICATE OF CITIZENSHIP FOR CERTAIN CHILDREN BORN OUTSIDE THE UNITED STATES.</p> <p>(a) In General.—Section 322 of the Immigration and Nationality Act (8 U.S.C. 1433) is amended to read as follows:</p> <p>"children born and residing outside the united states; conditions for acquiring certificate of citizenship</p> <p>"Sec. 322. (a) A parent who is a citizen of the United States may apply for naturalization on behalf of a child born outside of the United States who has not acquired citizenship automatically under section 320. The Attorney General shall issue a certificate of citizenship to such parent upon proof, to the satisfaction of the Attorney General, that the following conditions have been fulfilled: "(1) At least one parent ...</p>
Gold	Provides for issuance of a certificate of naturalization for a child born outside of the United States when the following conditions are met: (1) at least one parent is a U.S. citizen who has been present in the United States for not less than five years, at least two of which were after having attained the age of 14, or who has a citizen parent meeting such requirements; (2) the child is under 18 years old; and (3) the child is residing outside the United States in the legal and physical custody of the citizen parent, is temporarily and lawfully present in the United States, and is maintaining such lawful status. Applies such provision to an adopted child meeting certain definitional requirements who is adopted by a U.S. citizen parent. Title II: Protections for Certain Aliens Voting Based on Reasonable Belief of Citizenship - Amends the Immigration and Nationality Act respecting unlawful voting or false U.S. citizenship claims by permanent resident aliens under 16 years old having natural or adoptive U.S. citizen parents, to provide exceptions from certain provisions regarding deportability, moral character, inadmissibility or related criminal penalties.
Model	table of contents: title i: citizenship for certain title ii: immigration and naturalization provisions title i: citizenship for certain children born outside the united states - amends the immigration and nationality act (ina) to revise naturalization requirements with respect to child born outside the united states. (sec. 102) amends the immigration and nationality act to revise requirements with respect to: (1) naturalization as a citizen of the united states; (2) naturalization of a u.s. citizen; (3) citizenship; (4) citizenship; (5) naturalization service; and (6)
Metrics	Rouge1: 35.9, Rouge2: 15.54, RougeL: 22.56, RougeLsum: 20.51, Summary length (tokens): 129
Bottom 25% example (Sorted by rougeL)	
Document	<p>SECTION 1. SHORT TITLE. This Act may be cited as the "Effective Terrorists Prosecution Act of 2006".</p> <p>SEC. 2. DEFINITION OF UNLAWFUL ENEMY COMBATANT.</p> <p>Paragraph (1) of section 948a of title 10, United States Code (as enacted by the Military Commissions Act of 2006 (Public Law 109-366)), is amended to read as follows: "(1) Unlawful enemy combatant.—The term 'unlawful enemy combatant' means an individual who directly participates in hostilities as part of an armed conflict against the United States who is not a lawful enemy combatant. The term is used solely to designate individuals triable by military commission under this chapter."</p> <p>SEC. 3. DETERMINATION OF UNLAWFUL ENEMY COMBATANT STATUS BY COMBATANT STATUS REVIEW TRIBUNAL NOT DISPOSITIVE FOR PURPOSES OF JURISDICTION OF MILITARY COMMISSIONS.</p> <p>Section 948d of title 10, United States Code (as enacted by the Military Commissions Act of 2006 (Public Law 109-366)), is amended— (1) by striking subsection (c); and (2) by redesignating subsection (d) as subsection (c).</p> <p>SEC. 4. EXCLUSION FROM TRIAL BY MILITARY COMMISSION OF STATEMENTS OBTAINED BY COERCION.</p> <p>Section 948f of title 10, United States Code (as enacted by the Military Commissions Act of 2006 (Public Law 109-366)), is amended by striking subsections (c) and (d) and inserting the following new subsection (c): "(c) Exclusion of Statements Obtained by Coercion.—A statement obtained by use of coercion shall not be admissible in a military commission under this chapter, except against a person accused of coercion as evidence that the statement was made."</p> <p>SEC. 5. DISCRETION OF MILITARY JUDGE TO EXCLUDE HEARSAY EVIDENCE DETERMINED TO BE UNRELIABLE OR LACKING IN PROBATIVE VALUE.</p> <p>Section 949a(b)(2)(E)(ii) of title 10, United States Code (as enacted by the Military Commissions Act of 2006 (Public Law 109-366)), is amended by striking "if the party opposing the admission of the evidence demonstrates that the evidence is unreliable or lacking in probative value" and inserting "if the military judge determines, upon motion by counsel, that the evidence is unreliable or lacking in probative value".</p> <p>SEC. 6. DISCRETION OF MILITARY JUDGE TO TAKE CERTAIN ACTIONS IN EVENT THAT A SUBSTITUTE FOR CLASSIFIED EXCULPATORY EVIDENCE IS INS...</p>
Gold	Effective Terrorists Prosecution Act of 2006 - Amends federal armed forces provisions enacted by the Military Commissions Act of 2006 to, among other things: (1) exclude from military commission (commission) trials statements obtained by coercion; (2) allow a commission military judge to exclude hearsay evidence determined to be unreliable or lacking in probative value; (3) provide for review of commission decisions by the U.S. Court of Appeals for the Armed Forces rather than the Court of Military Commission Review; (4) revise generally provisions concerning the implementation of treaty obligations with respect to the U.S. prosecution of enemy combatants; (5) restore habeas corpus rights for individuals detained by the United States; and (6) provide for expedited judicial review of provisions of the Military Commissions Act of 2006.
Model	effective terrorists prosecution act of 2006 - amends federal armed forces law to revise the definition of " unlawful enemy combatant" to include an individual who directly participated in hostilities as part of an armed conflict against the united states who is not a lawful enemy combatant. amends the military pay reform and reform act of 2006 to provide that a complaint obtained by mail or the military court of appeals for the u.s. military court of appeals for the armed forces who is not a lawful enemy combatant. directs the u.s. military department of defense to review the record in each case, except against a person aggrieved by prosecution or
Metrics	Rouge1: 51.64, Rouge2: 24.64, RougeL: 33.8, RougeLsum: 37.56, Summary length (tokens): 129
Top 75% example (Sorted by rougeL)	
Document	<p>SECTION 1. SHORT TITLE. This Act may be cited as the "Census Address List Improvement Act of 1994".</p> <p>SEC. 2. ADDRESS INFORMATION REVIEWED BY LOCAL GOVERNMENTS. (a) In General.—Chapter 1 of title 13, United States Code, is amended by adding after section 15 the following new section: "Sec. 16. Address information reviewed by States and local governments "(a) The Secretary, to assist efforts to ensure the accuracy of censuses and surveys under this title, shall— "(1) publish standards defining the content and structure of address information which States and local units of general purpose government may submit to the Secretary to be used in developing a national address list; "(2)(A) develop and publish a timetable for the Bureau to receive, review, and respond to submissions of information under paragraph (1) before the decennial census date; and "(B) provide for a response by the Bureau with respect to such submissions in which the Bureau specifies its determinations regarding such information and the reasons for such determinations; and "(3) be subject to the review process developed under section 3 of the Census Address List Improvement Act of 1994 relating to responses pursuant to paragraph (2). "(b)(1) The Secretary— "(A) shall provide officials who are designated as census liaisons by a local unit of general purpose government with access to census address information for the purpose of verifying the accuracy of the address information of the Bureau for census and survey purposes; and "(B) together with such access, should provide an explanation of duties and obligations under this title. "(2) Access under paragraph (1) shall be limited to address information concerning addresses within the local unit of general purpose government represented by the census liaison or an adjacent local unit of general purpose government. "(3) The Bureau should respond to each recommendation made by a census liaison concerning the accuracy of address information, including the determination (and reasons therefor) of the Bureau regarding each such recommendation. "(4) For the purposes of paragraph (1), in a case in which a local unit of general purpose government is within another local unit of general purpose government and is not independent of the enclosing unit, the census liaison shall be ...</p>
Gold	Census Address List Improvement Act of 1994 - Directs the Secretary of Commerce to: (1) publish standards defining the content and structure of address information which States and local governments may submit to the Secretary to be used in developing a national address list; (2) develop and publish a timetable for the Bureau to receive, review, and respond to the submitted information before the decennial census date; (3) provide for a response by the Bureau that specifies its determinations regarding such information and the reasons for such determinations; and (4) be subject to the review process developed under this Act relating to such responses. Directs the Secretary to provide officials who are designated as census liaisons by local governments with access to census address information for the purpose of verifying the accuracy of the Bureau's address information for census and survey purposes and together with such access, provide an explanation of duties and obligations under this Act. Limits such access to the addresses within the local government represented by the census liaison or an adjacent local government. Requires the Bureau to respond to each recommendation made by a census liaison concerning the accuracy of address information, including the determination (and reasons therefor) of the Bureau regarding each such recommendation. Prohibits a census liaison from using information made available under this Act for purposes other than the purposes specified in this Act. Makes provisions that require, with exceptions, that such information be treated as confidential applicable to local government census liaisons. Imposes a fine and up to five years' imprisonment on whoever being or having been a census liaison wrongfully discloses such information. Requires: (1) the Administrator of the Office of Information and Regulatory Affairs, acting through the Chief Statistician, to develop an appeals process for those States and local governments which desire to appeal determinations of the Bureau; and (2) the Postal Service to provide to the Secretary for use by the Bureau such address, address- related, and point of postal delivery information, including postal delivery codes, determined by the Secretary to be appropriate for any census or survey being conducted by the Bureau.
Model	census address list improvement act of 1994 - directs the secretary of transportation to: (1) publish standards amending the content and structure of address information which states and local governments may submit to the secretary to the secretary to the secretary for such a national address list; (2) provide for a response by a census liaison with respect to such reports; and (3) provide for a response by a census liaison with respect to such reports. requires the secretary to: (1) provide officials who are designated as census liaison with access to census information; and (2) report annually to the congressional committees on the accuracy of
Metrics	Rouge1: 62.26, Rouge2: 41.9, RougeL: 50.0, RougeLsum: 52.83, Summary length (tokens): 129
Top 95% example (Sorted by rougeL)	
Document	<p>SECTION 1. SHORT TITLE. This Act may be cited as the "National Geologic Mapping Reauthorization Act of 1996".</p> <p>SEC. 2. FINDINGS.</p> <p>Congress finds that— (1) in enacting the National Geologic Mapping Act of 1992 (43 U.S.C. 31a et seq.), Congress found, among other things, that— (A) during the 2 decades preceding enactment of that Act, the production of geologic maps had been drastically curtailed; (B) geologic maps are the primary data base for virtually all applied and basic earth-science investigations; (C) Federal agencies, State and local governments, private industry, and the general public depend on the information provided by geologic maps to determine the extent of potential environmental damage before embarking on projects that could lead to preventable, costly environmental problems or litigation; (D) the lack of proper geologic maps has led to the poor design of such structures as dams and waste- disposal facilities; (E) geologic maps have proven indispensable in the search for needed fossil fuel and mineral resources; and (F) a comprehensive nationwide program of geologic mapping is required in order to systematically build the Nation's geologic-map data base at a pace that responds to increasing demand; (2) the geologic mapping program called for by that Act has not been fully implemented; and (3) it is time for this important program to be fully implemented.</p> <p>SEC. 3. REAUTHORIZATION AND AMENDMENT.</p> <p>(a) Definitions.—Section 3 of the National Geologic Mapping Act of 1992 (43 U.S.C. 31b) is amended— (1) by striking "As used in this Act:" and inserting "In this Act:"; (2) by redesignating paragraphs (2), (3), (4), and (5) as paragraphs (3), (4), (5), and (6), respectively; (3) by inserting after paragraph (1) the following: "(2) Association.—The term 'Association' means the Association of American State Geologists."; and (4) in each paragraph that does not have a heading, by ...</p>
Gold	National Geologic Mapping Reauthorization Act of 1996 - Amends the National Geologic Mapping Act of 1992 to establish a national cooperative geologic mapping program between the U.S. Geological Survey and State geological surveys. Establishes a geologic mapping advisory committee to advise the Director of the U.S. Geological Survey on planning and implementation of the geologic mapping program. Authorizes appropriations.
Model	national geologic mapping reauthorization act of 1996 - amends the national geologic mapping act of 1992 to establish a national cooperative geologic mapping program within the united states geological survey (usgs) to be administered and administered through the association. establishes a national cooperative geologic mapping program between the united states geological survey and the association. authorizes appropriations.
Metrics	Rouge1: 74.14, Rouge2: 56.14, RougeL: 65.52, RougeLsum: 67.24, Summary length (tokens): 69

Table 8: Examples of the BillSum dataset using the model billsum-1024-128 small diverse

Bottom 5% example (Sorted by rougeL)	
Document	SECTION 1. NATIONAL GUARD SUPPORT FOR BORDER CONTROL ACTIVITIES. (a) Operation Jump Start.– (1) In general.–Not fewer than 6,000 National Guard personnel shall continue to be deployed along the international border between the United States and Mexico under Operation Jump Start until the date on which the Federal Government has achieved operational control of such border (as defined in section 2(b) of the Secure Fence Act of 2006 (Public Law 109- 367)). (2) Exemption.–National Guard personnel deployed pursuant to paragraph (1) shall not be included in the calculation to determine compliance with– (A) limits on end strength; or (B) limits on the number of National Guard personnel that may be placed on active duty for operational support. (b) Support Authorized.–Chapter 1 of title 32, United States Code, is amended by inserting after section 112 the following: “Sec. 112a. Border control activities “(a) Funding Assistance.–The Secretary of Defense may provide funds to the Governor of a State who submits to the Secretary a State border control activities plan satisfying the requirements of subsection (c). Such funds shall be used for the following: “(1) The pay, allowances, clothing, subsistence, gratuities, travel, and related expenses, as authorized by State law, of personnel of the National Guard of that State used, while not in Federal service, for the purpose of border control activities. “(2) The operation and maintenance of the equipment and facilities of the National Guard of that State used for the purpose of border control activities. “(3) The procurement of services and equipment, and the leasing of equipment, for the National Guard of that State used for the purpose of border control activities. However, the use of such funds for the procurement of equipment may not exceed \$5,000 per item, unless approval for procurement of equipment in excess of that amount is granted in advance by the Secretary of Defense. “(b) Use of Personnel Performing Full-Time National Guard Duty.– (1) Under regulations prescribed by the Secretary of Defense, personnel of the National Guard of a State may, in accordance with the State...
Gold	Provides for continued National Guard support for border control activities along the U.S.- Mexico border under Operation Jump Start until operational control of such border has been achieved. Authorizes the Secretary of Defense to provide funding to a state that submits to the Secretary of State a state border control activities plan that meets certain requirements. Limits the number of National Guard that may be so deployed.
Model	(this measure has not been amended since it was passed by the senate on december 21, 2011. the summary of that version is repeated here.) states that not fewer than six,000, national guard personnel shall continue to be deployed along the international border between the united states and mexico under operation jump start up to six,000 shall not be included in the annual calculation to determine compliance with: (1) limits on end strength or limits on the number of national guard personal that may be placed on active duty for operational support; and (2) the pay, allowances, and benefits of the national guard of a state shall
Metrics	Rouge1: 46.33, Rouge2: 14.86, RougeL: 22.6, RougeLsum: 32.77, Summary length (tokens): 129
Bottom 25% example (Sorted by rougeL)	
Document	SECTION 1. SHORT TITLE. This Act may be cited as the “International Commerce Enhancement Act”. SEC. 2. REFERENCES. Except as otherwise specifically provided in this Act, whenever in this Act an amendment or repeal is expressed as the amendment or repeal of a section or other provision, the reference shall be considered to be made to a section or other provision of the Arms Export Control Act (22 U.S.C. 2751 et seq.). SEC. 3. FOREIGN AND NATIONAL SECURITY POLICY OBJECTIVES AND RESTRAINTS. (a) Value of Defense Articles and Services.–Section 3(d) (22 U.S.C. 2753(d)) is amended in paragraphs (1) and (3)(A)– (1) by striking “\$14,000,000” each place it appears and inserting “\$25,000,000”; and (2) by striking “\$50,000,000” each place it appears and inserting “\$85,000,000”. (b) Transfers With Respect to NATO and Major Non-NATO Countries.– Section 3(d) (22 U.S.C. 2753(d)) is amended– (1) in paragraph (2)– (A) in subparagraph (A), by striking “Except as provided in subparagraph (B), unless” and inserting “Unless”; and (ii) in subparagraph (B) to read as follows: “(B) Subparagraph (A) shall not apply in the case of a proposed transfer to the North Atlantic Treaty Organization, or any member country of such Organization, Japan, Australia, or New Zealand.”; and (iii) in subparagraph (C), by striking “or (B)”; and (2) in paragraph (3)– (A) in the second sentence of subparagraph (A), by striking “shall be submitted” and all that follows through “unless the President” and inserting “shall be submitted at least 30 calendar days before such consent is given in the case of a transfer to a country other than a country which is a member of the North Atlantic Treaty Organization, Japan, Australia, or New Zealand, unless the President”; (B) in the third sentence of subparagraph (A), by striking “(thus waiving the requirements of clause (i) or (ii), as the case may be, and of subparagraph (B))”; and (C) in subparagraph (B)– (i) by striking ...
Gold	International Commerce Enhancement Act - Amends the Arms Export Control Act to increase threshold values of major defense equipment or defense articles or related training or other defense services whose transfer or lease to foreign countries would require a presidential certification to Congress. Exempts from specified congressional oversight requirements any such transfers to North Atlantic Treaty Organization (NATO) countries, Japan, Australia, or New Zealand.
Model	international commerce enhancement act - amends the arms export control act to: (1) increase the amount of defense articles and services from \$10 million to \$10 million to \$10 million the value of defense articles and services (currently, \$10 million); (2) prohibit the transfer of defense articles or services to the north atlantic treaty organization (nato); and (3) prohibit the transfer of defense articles or services from the north atlantic treaty organization (nato) defense articles or services. repeals the requirement that the transfer of defense articles or services from defense articles or services from the north atlantic treaty (nato)
Metrics	Rouge1: 44.72, Rouge2: 21.38, RougeL: 34.78, RougeLsum: 38.51, Summary length (tokens): 129
Top 75% example (Sorted by rougeL)	
Document	SECTION 1. SHORT TITLE. This Act may be cited as the “Family Education Reimbursement Act of 2005”. SEC. 2. FAMILY EDUCATION REIMBURSEMENT ACCOUNTS. (a) Establishment.–The Secretary of Education, in consultation with the Secretary of Health and Human Services, shall– (1) establish a Family Education Reimbursement Account Program under which, at the direction of the parent of each displaced student who signs up under subsection (d), the Secretary provides reimbursement to enable the student or preschool-age child to attend the school or preschool program of his or her parent’s choice during the 2005-2006 school year; (2) of the amount available to carry out this section for fiscal year 2006, use not more than one third of one percent of such amount for administrative expenses, including outreach, support services, and dissemination of information; and (3) contract with a nongovernmental entity to administer and operate the program. (b) Reimbursement.– (1) In general.–In carrying out this section, the Secretary– (A) shall allow the parent of the participating displaced student to select the school or preschool program to be attended by the student during the 2005- 2006 school year; (B) at the direction of the parent, shall provide reimbursement to that school or preschool program on a quarterly basis; and (C) in the case of a public school, may provide such reimbursement to the appropriate local fiscal agent for the school. (2) Amount.–In providing reimbursement under paragraph (1), the Secretary shall– (A) determine the amount of reimbursement to a school or preschool program based on the number of weeks during which the participating displaced student attended the school or preschool program during the preceding quarter; (B) subject to subparagraph (C), provide the same amount of reimbursement to each school and preschool program for each week of attendance by one participating displaced student; (C) not provide reimbursement...
Gold	Family Education Reimbursement Act of 2005 - Directs the Secretary of Education to establish a Family Education Reimbursement Account Program for families of students displaced by Hurricane Katrina or Hurricane Rita. Provides for reimbursing parents for costs of such students or preschool-age children attending schools or preschool programs, chosen by the parents, during the 20052006 school year. Requires the Secretary to make a contract with a nongovernmental entity to administer and operate the program.
Model	family education reimbursement act of 2005 - directs the secretary of education to establish a family education reimbursement account program under which, at the direction of the parent of each displaced student who signs up under the 2005- 2006 school year, the secretary of education shall: (1) provide reimbursement to the student or preschool child to attend the school or breakfast program of his or her parent’s choice during the 2005- 2006- 2006 school year; (2) contract with a nongovernmental entity to administer and operate the program; and (3) contract with a nongovernmental entity to administer the program. requires the secretary to develop and implement a web-
Metrics	Rouge1: 63.74, Rouge2: 44.44, RougeL: 52.75, RougeLsum: 61.54, Summary length (tokens): 129
Top 95% example (Sorted by rougeL)	
Document	SECTION 1. SHORT TITLE. This Act may be cited as the “Pharmaceutical Products Price Equity Act”. SEC. 2. FINDINGS. The Congress finds the following: (1) Modern life enhancing and life sustaining drugs are the products of decades of scientific research and practice. (2) The refining and production of products by American pharmaceutical enterprises must be appropriately compensated. (3) The pricing of American pharmaceutical products varies greatly from country to country with the corporate owners of such products voluntarily assuming no losses by always selling their products above the cost of production. (4) Considering the vital life and death nature of pharmaceutical products, it is appropriate that the United States Government, in harmony with other nations belonging to the World Trade Organization, should engage in the reasonable regulation of the prices of pharmaceutical products. SEC. 3. REASONABLE LIMITS ON PROFITS OF PHARMACEUTICAL PRODUCTS. (a) Presidential Authority.–The President shall issue such orders and regulations, and establish such procedures and reporting requirements, as the President determines to be appropriate to ensure that no pharmaceutical product (as defined by the President in such regulation or order) is sold to any consumer in the United States at a price that is more than 6 percent above the average retail price at which such pharmaceutical product is sold in the 5 most industrialized, free-market countries, other than the United States, as determined by the President. (b) Delegation.–The President may delegate the performance of any function under this section to any officer of any Federal department or agency who has been appointed by the President, by and with the consent of the Senate. (c) Confidentiality of Information.–All information reported to or otherwise obtained by any person exercising any authority under this section which contains or relates to a trade secret or other matter referred to in section 1905 of title 18, United States Code, shall be considered confidential for purposes of that section, except that such information may be disclosed to any other officer or employee of the United States involved in carrying out this section solely for the purpose of carrying out, and enforcing compliance with, this section. (d) Subpoena Powe...
Gold	Pharmaceutical Products Price Equity Act - Directs the President to issue orders and regulations, and establish procedures and reporting requirements, to ensure that no pharmaceutical product is sold to any consumer in the United States at a price that is more than six percent above the average retail price at which such pharmaceutical product is sold in the five most industrialized, free-market countries, other than the United States. Sets forth enforcement and civil liability provisions.
Model	pharmaceutical products price equity act - directs the president to issue orders and regulations to ensure that no pharmaceutical product is sold to any consumer in the united states at a price more than six percent above the average retail price at which such product is sold in the five most industrialized, free-market countries, other than the united states. authorizes the president to delegate the performance of any such function to any federal department or agency who has been appointed by the president, by and with the consent of the senate, by and with the consent of the senate.
Metrics	Rouge1: 70.11, Rouge2: 63.95, RougeL: 68.97, RougeLsum: 68.97, Summary length (tokens): 113

Table 9: Examples of the BillSum dataset using the model billsum-1024-128 base diverse

Bottom 5% example (Sorted by rougeL)	
Document	<p>–(1) For purposes of subsection (a)(2) and this subsection, the term ‘joint resolution’ means only a joint resolution introduced by a qualifying Member specified in paragraph (2) after the date on which the report of the President under subsection (a)(1) is received by the Congress: –(A) the matter after the resolving clause of which is as follows: ‘That the Congress hereby concurs in the certification of the President relating to deployment of a National Missile Defense system as submitted to Congress pursuant to section 4(b) of the National Missile Defense Act of 1999;’; ‘(B) which does not have a preamble; and ‘(C) the title of which is as follows: ‘Joint resolution relating to deployment of a National Missile Defense system.’; –(2) For purposes of this subsection, a qualifying Member described in this paragraph is– ‘(A) in the case of the House of Representatives, the majority leader or minority leader of the House of Representatives or a Member of the House of Representatives designated by the majority leader or minority leader; and ‘(B) in the case of the Senate, the majority leader or minority leader of the Senate or a Member of the Senate designated by the majority leader or minority leader.’; ‘(3) The provisions of paragraphs (3) through (8) of section 4(c) of the National Missile Defense Deployment Criteria Act of 2001 shall apply to a joint resolution under this subsection in the same manner as to a joint resolution under such section.”.</p> <p>SEC. 4. LIMITATION ON OBLIGATION OF FUNDS FOR PROCUREMENT FOR NATIONAL MISSILE DEFENSE SYSTEM.</p> <p>(a) Limitation.—No funds appropriated to the Department of Defense for procurement may be obligated for the National Missile Defense system unless— (1) the President submits to Congress a report concerning testing of the National Missile Defense system against countermeasures that includes a certification described in subsection (b); and (2) a joint resolution concurring in the President’s certification in such report is enacted as provided for in this section. (b) Presidential Certification.—A certification described in this subsection is a certification by the President that— (1) an adequate testing program for the National Missile _</p>
Gold	<p>National Missile Defense Deployment Criteria Act of 2001 - Amends the National Missile Defense Act of 1999 to allow deployment of a national missile defense system (system) only if: (1) the system is technologically feasible; (2) system cost in relation to other Department of Defense (DOD) priorities will not lead to an overall reduction in national security by reducing resources available for other defense priorities; (3) the system will not diminish overall U.S. national security; (4) the system will not threaten to disrupt relations with U.S. nuclear allies, U.S. European allies, Russia, the People’s Republic of China, and other nations; and (5) the threat of a long-range ballistic missile attack from a nation of concern is clearly demonstrated.Prohibits the President from directing DOD to deploy a system unless and until: (1) the President certifies to Congress that the above deployment conditions have been met; and (2) a joint resolution is enacted concurring in the President’s certification.Prohibits DOD procurement funds from being obligated for a system unless: (1) the President certifies to Congress that adequate system tests have been undertaken to meet identified threats against countermeasures; and (2) a joint resolution is enacted concurring in the President’s certification.Requires the Secretary of Defense to direct the Ballistic Missile Defense Organization to: (1) include specified system countermeasures in system ground and flight testing conducted before the system becomes operational; and (2) determine the extent to which the exoatmospheric kill vehicle and the system can reliably discriminate between warheads and such countermeasures.</p>
Model	prohibits funds appropriated to the department of defense (dod) for procurement from being obligated for the national missile defense system unless the president certifies to congress that: (1) an adequate testing program for the system is in place to meet the threats identified in the report; and (2) an adequate ground and flight testing of the system has been conducted against the system that are likely to be used against the system and that other countries have or are likely to acquire.
Metrics	Rouge1: 40.69, Rouge2: 16.67, RougeL: 20.0, RougeLsum: 20.0, Summary length (tokens): 94
Bottom 25% example (Sorted by rougeL)	
Document	<p>TITLE I- FEDERAL AIRPORTS SECURITY ENHANCEMENT ACT</p> <p>SEC. 101. SHORT TITLE.</p> <p>This title may be cited as the “Federal Airports Security Enhancement Act”.</p> <p>SEC. 102. ESTABLISHMENT OF AIRPORT SECURITY COMMITTEES.</p> <p>The Act of July 5, 1994 (49 U.S.C. 44935), is amended— (1) by striking section 44901 subparagraph (b) and inserting the following:</p> <p>“SEC. 103. EMPLOYMENT STANDARDS AND TRAINING:”</p> <p>(2) by striking section 44935 subparagraph (b) and inserting the following: “(a) Review and Recommendations.—The Administrator of the Federal Aviation Administration shall establish Security Committees at each airport location to be composed of representatives of the air carriers, airport operators, other interested parties and at least one representative from the Federal Protective Service, the Federal Bureau of Investigation, the Federal Bureau of Administration and one member from each local jurisdiction that the airport may be located in or that may have jurisdictional authority for the airport facility. Each Airport Security Committee shall meet at least quarterly and shall make recommendations for minimum security countermeasures to the Administrator. The Federal Protective Service shall have primary responsibility for conducting an ongoing basis security surveys and formulating recommendations to the Security Committee. The Administrator shall prescribe appropriate changes in existing procedures to improve that performance.”</p> <p>SEC. 103. SCREENING PASSENGERS AND PROPERTY.</p> <p>The Act of July 5, 1994 (49 U.S.C. 44935), is amended by striking section 44901, subparagraph (a), and inserting the following: “(a) General Requirements.—The Administrator of the Federal Aviation Administration shall prescribe regulations requiring screening of all passengers and property that will be carried in a cabin of an aircraft in air transportation or intrastate air transportation. The screening must take place before boarding and be carried out by a weapon detecting facility or procedure used or operated by an employee or agent of the Federal Protective Service. The Administrator— “(1) shall require that sufficient Federal Police Officers are posted at airport facilities to provide patrol duties during all hours of operations as well as supervise screening personnel; “(2) shall maintain sufficient numbers of Special Agents to provid..</p>
Gold	<p>Federal Airports Security Enhancement Act - Amends Federal aviation law to direct the Administrator of the Federal Aviation Administration (FAA) to establish at each airport a Security Committee which shall make recommendations for minimum security counter-measures. Requires the Administrator, on the basis of such recommendations, to prescribe appropriate changes to improve the performance of existing airport security procedures.Requires the screening of passengers and property that will be carried in a cabin of an aircraft to be carried out by Federal Protective Service employees or agents. (Currently, screening is carried out by employees or agents of an air carrier, interstate air carrier, or foreign air carrier).Authorizes the Administrator of the General Services Administration (GSA) to appoint police officers and special agents (currently, special policemen and nonuniformed special policemen) for the policing of all Federal buildings (including buildings under the control of the GSA). Sets forth certain additional powers of such officers and agents, including the authority to carry firearms and to police areas adjacent to Federal property.Establishes the Federal Protective Service as a separate operating service of the GSA. Calls for at least 1,000 full-time equivalent Service police officers to be assigned to areas outside of airport operations. Requires the Commissioner of the Service to prescribe minimum employment and training standards to be applied in the contracting of security personnel for the policing of buildings and areas controlled by the United States and GSA. Authorizes GSA to recover airport security costs from the FAA.</p>
Model	table of contents: title i: federal airports security enhancement act title ii: miscellaneous provisions general federal airports security enhancement act - title i: federal airports security enhancement - amends the federal aviation act of 1992 to direct the administrator of the federal aviation administration (faa) to prescribe regulations requiring screening of all passengers and property that will be carried in a port of aircraft in air transportation or intrastate air transportation. (sec. 102) directs the administrator to prescribe regulations requiring screening of all passengers and property that will be carried out by the federal protective service, the federal bureau of investigation (fbi), the federal bureau of investigation (fbi), and one member from each local jurisdiction that the aircraft may be located in or that may have jurisdictional authority for the airport of an aircraft in air transportation or intrastate air transportation. (sec. 103) directs the administrator to prescribe regulations requiring screening of all passengers and property that will be carried out by a weapon detection facility or procedure used or operated by an employee or agent of the federal protective service. (sec. 103) authorizes the administrator to enter into agreements with state and local law enforcement authorities to obtain authority for, jointly with state and local law enforcement authorities. (
Metrics	Rouge1: 52.44, Rouge2: 22.84, RougeL: 29.7, RougeLsum: 47.8, Summary length (tokens): 256
Top 75% example (Sorted by rougeL)	
Document	<p>SECTION 1. SHORT TITLE.</p> <p>This Act may be cited as the “Patent and Trademark Office Authorization Act of 2002”.</p> <p>SEC. 2. AUTHORIZATION OF AMOUNTS AVAILABLE TO THE PATENT AND TRADEMARK OFFICE.</p> <p>(a) In General.—There are authorized to be appropriated to the United States Patent and Trademark Office for salaries and necessary expenses for each of the fiscal years 2003 through 2008 an amount equal to the fees estimated by the Secretary of Commerce to be collected in each such fiscal year, respectively, under— (1) title 35, United States Code; and (2) the Act entitled “An Act to provide for the registration and protection of trademarks used in commerce, to carry out the provisions of certain international conventions, and for other purposes”, approved July 5, 1946 (15 U.S.C. 1051 et seq.) (commonly referred to as the Trademark Act of 1946). (b) Estimates.—Not later than February 15, of each fiscal year, the Undersecretary of Commerce for Intellectual Property and the Director of the Patent and Trademark Office (in this Act referred to as the Director) shall submit an estimate of all fees referred to under subsection (a) to be collected in the next fiscal year to the chairman and ranking member of— (1) the Committees on Appropriations and Judiciary of the Senate; and (2) the Committees on Appropriations and Judiciary of the House of Representatives.</p> <p>SEC. 3. ELECTRONIC FILING AND PROCESSING OF PATENT AND TRADEMARK APPLICATIONS.</p> <p>(a) Electronic Filing and Processing.—Not later than December 1, 2004, the Director shall complete the development of an electronic system for the filing and processing of patent and trademark applications, that— (1) is user friendly; and (2) includes the necessary infrastructure to— (a) allow examiners and applicants to send all communications electronically; and (b) allow the Office to process, maintain, and search electronically the contents and history of each application. (b) Authorization of Appropriations.—Of amounts authorized under section 2, there are authorized to be appropriated to carry out subsection (a) of this section not more than \$50,000,000 for each of fiscal years 2003 and 2004. Amounts made available under this subsection shall..</p>
Gold	<p>Patent and Trademark Office Authorization Act of 2002 - Authorizes appropriations to the U.S. Patent and Trademark Office for salaries and expenses for FY 2003 through 2008 in an amount equal to all patent and trademark fees estimated by the Secretary of Commerce for each fiscal year.(Sec. 2) Requires the Under Secretary of Commerce for Intellectual Property and the Director of the Office (Director), by February 15 of each fiscal year, to report an estimate of all fees to be collected in the next fiscal year to the chairman and ranking member of specified congressional committees.(Sec. 3) Requires the Director, by December 1, 2004, to complete the development of an electronic system for the filing and processing of patent and trademark applications that: (1) is user friendly; and (2) includes the necessary infrastructure to allow examiners and applicants to send all communications electronically, and the Office to process, maintain, and search electronically the contents and history of each application. Authorizes appropriations for FY 2003 and 2004 for development of such system.(Sec. 4) Requires the Secretary, in each of the five calendar years following the enactment of this Act, to report to specified congressional committees on the progress made in implementing the 21st Century Strategic Plan issued on June 3, 2002, and on any amendments made to it.(Sec. 5) Amends Federal patent law to provide that previous citation by or to, or consideration by the Office of, a patent or printed publication does not preclude the existence of a substantial new question of patentability in patent reexamination proceedings.(Sec. 6) Revises requirements for appeals in inter partes reexamination proceedings to allow a third-party requester to appeal to the U.S. Court of Appeals for the Federal Circuit, or be a party to any appeal taken by the patent owner, with respect to any final decision favorable to the patentability of any original or proposed amended or new claim of the patent.Allows a third-party requester to appeal a decision of the Board of Patent Appeals and Interferences.Provides that a third-party requester in an inter partes reexamination proceeding dissatisfied with the final decision in an appeal to the Board may appeal the decision only to the U.S. Court of Appeals for the Federal Circuit.</p>
Model	patent and trademark office authorization act of 2002 - authorizes appropriations to the u.s. patent and trademark office for fy 2003 through 2008. requires the director of the patent and trademark office to: (1) complete the development of an electronic system for the filing and processing of patent and trademark applications; and (2) submit an annual report to the congressional committees on progress made in implementing the 21st century strategic plan issued under the federal patent and trademark programs.
Metrics	Rouge1: 48.99, Rouge2: 39.86, RougeL: 44.3, RougeLsum: 48.32, Summary length (tokens): 94
Top 95% example (Sorted by rougeL)	
Document	<p>SECTION 1. SHORT TITLE.</p> <p>This Act may be cited as the “Guidance, Understanding, and Information for Dual Eligibles (GUIDE) Act”.</p> <p>SEC. 2. FINDINGS; PURPOSE.</p> <p>(a) Findings.—The Congress finds the following: (1) Nearly 8,800,000 Americans were eligible for benefits under the Medicare program and for medical assistance under Medicaid (dual eligible beneficiaries) in fiscal year 2005. Of these “dual eligible beneficiaries”, almost 40 percent have cognitive impairments, including Alzheimer’s disease, dementia, serious mental illnesses, and intellectual disabilities. Until December 31, 2005, dual eligible beneficiaries received outpatient prescription drug benefits through medical assistance under Medicaid. On January 1, 2006, drug coverage for dual eligibles switched from Medicaid to Medicare. (2) In 2008, 53 percent of dual eligible beneficiaries had medication access problems and of those, 27 percent experienced significant adverse clinical events. (3) Individuals with medication access issues experience significantly more adverse clinical events. Among dual eligible beneficiaries with mental illness who had medication access problems, 27 percent experienced significant adverse clinical events, which included emergency room visits and hospitalizations. (4) In total, over 1,000,000 dual eligible beneficiaries and low-income subsidy beneficiaries were automatically auto- enrolled to new benchmark prescription drug plans under part D of the Medicare program between 2006 and 2007. (5) Community providers are at the front line of helping the most vulnerable dual eligible beneficiaries obtain prescription drug coverage under the Medicare program and navigate complex enrollment and low-income subsidy eligibility requirements under such program. (b) Purpose.—It is the purpose of this bill to help low-income persons with cognitive impairments to enroll in and navigate the prescription drug benefit under the Medicare program by providing front line community providers who serve the population daily with financial assistance to conduct vigorous education and outreach and direct case management.</p> <p>SEC. 3. MEDICARE PRESCRIPTION DRUG OUTREACH DEMONSTRATION PROGRAM FOR DUAL E..</p>
Gold	<p>Guidance, Understanding, and Information for Dual Eligibles (GUIDE) Act - Directs the Secretary of Health and Human Services to establish a three-year demonstration program under which the Secretary awards grants and contracts to appropriate, qualified community programs and clinics for individuals with intellectual or developmental disabilities, or certain programs under the Public Health Services Act, to employ qualified social workers and case managers to provide one-on-one counseling about benefits under part D (Voluntary Prescription Drug Benefit Program) of title XVIII (Medicare) of the Social Security Act (SSA) to a full-benefit dual eligible individual (eligible for benefits under both Medicare and SSA title XIX (Medicaid) who has one or more mental disabilities.</p>
Model	guidance, understanding, and information for dual eligible beneficiaries with intellectual or developmental disabilities act - directs the secretary of health and human services (hhs) to establish a three-year demonstration program under which the secretary awards grants and contracts to qualified community programs and clinics for individuals with intellectual or developmental disabilities or such programs to provide medicare prescription drug assistance to individuals with intellectual or developmental disabilities or such programs.
Metrics	Rouge1: 60.87, Rouge2: 47.25, RougeL: 58.7, RougeLsum: 58.7, Summary length (tokens): 80

Table 10: Examples of the BillSum dataset using the model billsum-1024-256 small diverse

Bottom 5% example (Sorted by rougeL)	
Document	SECTION 1. SHORT TITLE. This Act may be cited as the "Health Coverage Tax Credit Extension Act of 2015". SEC. 2. EXTENSION AND MODIFICATION OF HEALTH COVERAGE TAX CREDIT. (a) Extension.—Subparagraph (B) of section 35(b)(1) of the Internal Revenue Code of 1986 is amended by striking "before January 1, 2014" and inserting "before January 1, 2020". (b) Coordination With Credit for Coverage Under a Qualified Health Plan.—Subsection (g) of section 35 of the Internal Revenue Code of 1986 is amended—(1) by redesignating paragraph (11) as paragraph (13), and (2) by inserting after paragraph (10) the following new paragraphs: "(11) Election.—“(A) In general.—A taxpayer may elect to have this section apply for any eligible coverage month. “(B) Timing and applicability of election.—Except as the Secretary may provide,—“(i) an election to have this section apply for any eligible coverage month in a taxable year shall be made not later than the due date (including extensions) for the return of tax for the taxable year, and “(ii) any election for this section to apply for an eligible coverage month shall apply for all subsequent eligible coverage months in the taxable year and, once made, shall be irrevocable with respect to such months. “(12) Coordination with premium tax credit.—“(A) In general.—An eligible coverage month to which the election under paragraph (11) applies shall not be treated as a coverage month (as defined in section 36B(c)(2)) for purposes of section 36B with respect to the taxpayer. “(B) Coordination with advance payments of premium tax credit.—In the case of a taxpayer who makes the election under paragraph (11) with respect to any eligible coverage month in a taxable year or on behalf of whom any advance payment is made under section 7527 with respect to any month in such taxable year— ...
Gold	Health Coverage Tax Credit Extension Act of 2015 This bill extends the tax credit for health insurance costs of a taxpayer and qualifying family members through 2019. The tax credit for health insurance costs is a refundable tax credit equal to 72.5% of the cost of qualified health coverage paid by an eligible individual [defined as an individual who is receiving a trade adjustment allowance, is eligible for the alternative trade adjustment assistance program, or is over age 55 and receives pension benefits from the Pension Benefit Guaranty Corporation (PBGC)]. The bill requires a taxpayer to make an election to have the tax credit apply for any eligible coverage month during a taxable year. An eligible coverage month is a month in which an eligible individual is benefited by qualified health insurance, does not have other specified coverage, and is not imprisoned. The bill also directs the Departments of the Treasury, Health and Human Services, and Labor and the PBGC to conduct a public outreach, including on the Internet, to inform individuals eligible for the tax credit for health insurance costs on the extension of such credit and the availability of the election to claim such credit retroactively for coverage months beginning after December 31, 2013.
Model	health coverage tax credit extension act of 2015 this bill amends the internal revenue code, with respect to health care coverage, to: (1) extend through 2020 the tax credit for advance payments to individuals, (2) allow advance payments of advance payments of advance payments of advance payments, and (3) extend through 2018 the tax credit for advance payments of advance payments to individuals.
Metrics	Rouge1: 26.37, Rouge2: 11.07, RougeL: 21.25, RougeLsum: 25.64, Summary length (tokens): 82
Bottom 25% example (Sorted by rougeL)	
Document	SECTION 1. EXTENSION. (a) In General.—Chapter 5 of subtitle B of the Agricultural Marketing Act of 1946 (7 U.S.C. 1636 et seq.) is amended by adding at the end the following new section: "SEC. 260. TERMINATION OF AUTHORITY. “The authority provided by this subtitle terminates on September 30, 2010.”. (b) Conforming Amendment and Extension.—Section 942 of the Livestock Mandatory Reporting Act of 1999 (7 U.S.C. 1635 note; Public Law 106-78) is amended by striking "terminate on September 30, 2005" and inserting "other than section 911 of subtitle A and the amendments made by that section) terminate on September 30, 2010". SEC. 2. DEFINITIONS. (a) Base Market Hog.—Section 231(4) of the Agricultural Marketing Act of 1946 (7 U.S.C. 1635i(4)) is amended to read as follows: “(4) Base market hog.—The term ‘base market hog’ means a barrow or gilt for which no discounts are subtracted from and no premiums are added to the base price.”. (b) Bows.—Section 231(5) of such Act (7 U.S.C. 1635i(5)) is amended to read as follows: “(5) Boar.—The term ‘boar’ means a sexually-intact male swine.”. (c) Packer of Sows and Boars.—Section 231(12) of such Act (7 U.S.C. 1635i(12)) is amended by— (1) striking subparagraph (B) and inserting the following new subparagraph: “(B) for any calendar year, the term includes only—“(i) a swine processing plant that slaughtered an average of at least 100,000 swine per year during the immediately preceding five calendar years; and “(ii) a person that slaughtered an average of at least 200,000 sows, bows, or any combination thereof, per year during the immediately preceding five calendar years; and”, and (2) in subparagraph (C)—(A) by inserting “or person” after “swine processing plant”; (B) by inserting “or person” after “plant capacity of the processing plant”; and (C) by inserting “or person” after “determining whether the processing plant”. SEC. 3. REPORTING; BARROWS AND GILTS. Section 232(c) of the Agricultural Marketing Act of 1946 (7 U.S.C. 1635(c)) is amended to read as follows: “(c) Daily Reporting; Barrows and Gilts.—“(1) Prior day report.—“(A) In general.—The corporate officers or officially d...
Gold	Amends the Agricultural Marketing Act of 1946 to extend the provisions of the Livestock Mandatory Price Reporting Act of 1999 through September 30, 2010. Amends swine price reporting provisions.
Model	amends the agricultural marketing act of 1946 to terminate the authority of the secretary of agriculture (usda) to: (1) livestock processing plant processing plant slaughter, and (2) slaughtering plant slaughter. amends the agricultural marketing act of 1946 to: (1) revise minimum reporting requirements; and (2) revise reporting requirements. amends the agricultural marketing act of 1946 and the agricultural marketing act of 1946 to: (1) revise reporting requirements; and (2) revise reporting requirements.
Metrics	Rouge1: 33.66, Rouge2: 18.18, RougeL: 31.68, RougeLsum: 29.7, Summary length (tokens): 105
Top 75% example (Sorted by rougeL)	
Document	SECTION 1. SHORT TITLE. This Act may be cited as the "Maritime Administration Authorization Act for Fiscal Year 2001". SEC. 2. AUTHORIZATION OF APPROPRIATIONS FOR FISCAL YEAR 2001. Funds are hereby authorized to be appropriated, as Appropriations Acts may provide, for the use of the Department of Transportation for the Maritime Administration as follows: (1) For expenses necessary for operations and training activities, not to exceed \$80,240,000 for the fiscal year ending September 30, 2001. (2) For the costs, as defined in section 502 of the Federal Credit Reform Act of 1990, of guaranteed loans authorized by title XI of the Merchant Marine Act, 1936 (46 U.S.C. App. 1271 et seq.), \$500,000,000, to be available until expended. In addition, for administrative expenses related to loan guarantee commitments under title XI of that Act, \$4,179,000. SEC. 3. AMENDMENTS TO TITLE IX OF THE MERCHANT MARINE ACT, 1936. (a) Title IX of the Merchant Marine Act, 1936 (46 U.S.C. App. 101 et seq.) is amended by adding at the end thereof the following: "SEC. 910. DOCUMENTATION OF CERTAIN DRY CARGO VESSELS. “(a) In General.—The restrictions of section 901(b)(1) of this Act concerning a vessel built in a foreign country shall not apply to a newly constructed drybulk or breakbulk vessel over 7,500 deadweight tons that has been delivered from a foreign shipyard or contracted for construction in a foreign shipyard before the earlier of—“(1) the date that is 1 year after the date of enactment of the Maritime Administration Authorization Act for Fiscal Year 2001; or “(2) the effective date of the OECD Shipbuilding Trade Agreement Act. “(b) Compliance With Certain U.S.-Build Requirements.—A vessel timely contracted for or delivered pursuant to this section and documented under the laws of the United States shall be deemed to have been United-States built for purposes of sections 901(b) and 901(b) of this Act if—“(1) following delivery by a foreign shipyard, the vessel has any additional shipyard work necessary to receive its initial Coast Guard certificate of inspection performed in a United States shipyard; “(2) the vessel is not documented in another country before being documented under the laws of the United States; “(3)...
Gold	(Sec. 3) Amends the Merchant Marine Act, 1936 to declare that certain restrictions concerning a vessel built in a foreign country shall not apply to a newly constructed drybulk or breakbulk vessel over 7,500 deadweight tons that has been delivered from a foreign shipyard or contracted for construction in a foreign shipyard before the earlier of two specified dates. Deems U.S.-built any vessel timely contracted for or delivered and documented under U.S. law, if certain conditions are met. (Sec. 4) Directs the Secretary of State, in coordination with the Secretary of Transportation, to initiate discussions in all appropriate international forums to establish an international standard for the scrapping of vessels in a safe and environmentally sound manner. Directs the Secretary of Transportation to develop, and report to specified congressional committees on, a program for the scrapping of obsolete National Defense Reserve Fleet Vessels. Amends the National Maritime Heritage Act of 1994 to extend, through September 30, 2006, the authority of the Secretary to dispose of certain vessels in the National Defense Reserve Fleet. Requires that such vessels be disposed of in the most cost effective manner to the United States, taking into account the need for disposal, the environment, and safety concerns. Amends Federal law to authorize the expenditure of funds from the National Defense Sealift Fund for costs related to the scrapping of National Defense Reserve Fleet vessels. Names vessels in the National Defense Reserve Fleet that may be scrapped in the United States or a foreign country. (Sec. 5) Requires the Maritime Administration (in its annual report to Congress and its estimated annual budget) to state separately the amount, source, intended use, and nature of any funds (other than funds appropriated to the Administration or to the Secretary for use by the Administration) administered, or subject to oversight, by the Administration. (Sec. 6) Amends Federal maritime law to authorize the Secretary of Transportation to make a grant to a National Maritime Enhancement Institute for maritime and maritime intermodal research as if the Institute were a university transportation center. (Sec. 7) Directs the Secretary to study maritime research and technology development, and report the results, including any recommendations, to Congress. Authorizes appropriations. (Sec. 8) Authorizes the Secretary to convey all right, title, and U.S. interest in the U.S.S. GLACIER (formerly of the National Defense Reserve Fleet) to the Glacier Society, Inc., Bridgeport, Connecticut.
Model	maritime administration authorization act for fiscal year 2001 - authorizes appropriations for the department of transportation (dot) for fy 2001 for: (1) operations and training activities; (2) training activities; and (3) administrative expenses.amends the merchant marine act, 1936 to make appropriations for fy 2001 through 2001 for the maritime administration.amends the merchant marine act, 1936 to apply certain restrictions concerning a vessel located in a foreign country to a newly constructed dry or breakable vessel over seven,500 feet that has been delivered from a foreign shipyard or contracted for construction in a foreign shipyard before the earlier of: (1) one year after enactment of this act, or (2) the effective date of the international maritime administration act. directs the secretary of state in coordination with the secretary of transportation to initiate discussions in all appropriate international forums in order to establish an international standard for the scrapping of vessels in a safe and environmentally sound manner. directs the secretary of state to initiate discussions in all appropriate international forums to establish an international standard for the scrapping of vessels in a safe and environmentally sound manner.
Metrics	Rouge1: 61.19, Rouge2: 41.5, RougeL: 47.76, RougeLsum: 57.21, Summary length (tokens): 222
Top 95% example (Sorted by rougeL)	
Document	SECTION 1. SMALL BUSINESS EXPENSING PROVISIONS MADE PERMANENT. (a) Increase in Small Business Expensing Made Permanent.— (1) In general.—Subsection (b) of section 179 of the Internal Revenue Code of 1986 (relating to limitations) is amended—(A) by striking “\$25,000 (\$125,000 in the case of taxable years beginning after 2006 and before 2011)” in paragraph (1) and inserting “\$500,000”, and (B) by striking “\$200,000 (\$500,000 in the case of taxable years beginning after 2006 and before 2011)” in paragraph (2) and inserting “\$1,000,000”. (2) Conforming amendment.—Section 179(b) of such Code is amended by striking paragraph (7). (b) Expensing for Computer Software Made Permanent.—Clause (ii) of section 179(d)(1)(A) of such Code is amended by striking “and which is placed in service in a taxable year beginning after 2002 and before 2011”. (c) Inflation Adjustment.— (1) So much of subparagraph (A) of section 179(b)(5) of such Code as precedes clause “(A) thereof is amended to read as follows: “(A) In general.—In the case of any taxable year beginning in a calendar year after 2009, the \$500,000 and \$1,000,000 amounts in paragraphs (1) and (2) shall each be increased by an amount equal to—”. (2) Section 179(b)(5)(A)(ii) of such Code is amended by striking “2006” and inserting “2008”. (d) Effective Date.—The amendments made by this section shall apply to taxable years ending after the date of the enactment of this Act. SEC. 2. DEDUCTION FOR PURCHASE OF DOMESTICALLY MANUFACTURED AUTOMOBILES. (a) In General.—Part VII of subchapter B of chapter 1 of the Internal Revenue Code of 1986 (relating to additional itemized deductions for individuals) is amended by redesignating section 224 as section 225 and by inserting after section 223 the following new section: "SEC. 224. DEDUCTION FOR PURCHASE OF DOMESTICALLY MANUFACTURED AUTOMOBILES. “(a) Allowance of Deduction.—In the case of an individual, there shall be allowed as a deduction an amount equal to the cost of any qualified automobile placed in service by the taxpayer during the taxable year. “(b) Limitation Per Vehicle.—The amount of the ded...
Gold	Amends the Internal Revenue Code to: (1) increase and make permanent the expensing allowance for depreciable business assets; and (2) allow a tax deduction, up to \$10,000, for the purchase of a motor vehicle manufactured in the United States. Terminates such tax deduction after 2010.
Model	amends the internal revenue code to make permanent: (1) the increased expensing allowance for depreciable business assets; and (2) the tax deduction for the purchase of manufactured manufactured automobiles.
Metrics	Rouge1: 72.0, Rouge2: 46.58, RougeL: 64.0, RougeLsum: 64.0, Summary length (tokens): 40

Table 11: Examples of the BillSum dataset using the model billsum-1024-256 base diverse



Bottom 5% example (Sorted by rougeL)	
Document	<p>SECTION 1. SHORT TITLE. This Act may be cited as the "Public Health Equity Act".</p> <p>SEC. 2. FINDINGS. Congress finds that-- (1) all communities and individuals are entitled to protection from occupational and other exposure to substances that are hazardous to the public health; (2) hazardous substances have had a disproportionate impact on the public health of poor and ethnic minority communities and individuals, resulting in exclusion from participation in, denial of benefits under, and discrimination under, programs and activities receiving Federal financial assistance; and (3) each Federal agency has an obligation to ensure that all Federally assisted programs and activities that affect human health do not directly or through contractual arrangements use criteria, methods, or practices that cause discrimination on the ground of race, color, or national origin.</p> <p>SEC. 3. PUBLIC HEALTH EQUITY. The Public Health Service Act (42 U.S.C. 201 et seq.) is amended by adding at the end thereof the following new title:</p> <p>"TITLE XXVII--PUBLIC HEALTH EQUITY</p> <p>"SEC. 2701. DEFINITIONS. "As used in this title: "(1) Activity; program--The term 'program or activity' means any operation of-- "(A)(i) a department, agency, special purpose district, or other instrumentality of a State or of a local government; or "(ii) the entity of such State or local government that distributes such assistance and each such department or agency (and each other State or local government entity) to which the assistance is extended, in the case of assistance to a State or local government; "(B)(i) a college, university, or other postsecondary institution, or a public system of higher education; or "(ii) a local educational agency (as defined in section 198(a)(10) of the Elementary and Secondary Education Act of 1965), system of vocational education, or other school system; "(C)(i) an entire corporation, partnership, or other private organization, or an entire sole ...</p>
Gold	Public Health Equity Act - Amends the Public Health Service Act to add a new title, title XXVII: Public Health Equity. Prohibits, under such title, acts with discriminatory effect as well as intentionally discriminatory acts that result in disproportionate exposure to a covered substance on the basis of race, color, or national origin. Defines a covered substance to include certain hazardous materials, substances, contaminants, or chemicals listed, identified, or defined in specified laws.
Model	public health equity act - amends the public health service act to require the president to ensure that no person shall be excluded from participation in, be denied the benefits of, or being subject to discrimination under, any program or activity on the ground of race, color, or national origin. requires the president to ensure that no person shall be excluded from participation in, be denied the benefits of, or being subject to discrimination under, any program or activity on the ground of race, color, or national origin. requires the secretaries of labor, health and human services, the administrator of the health and human services, the administrator, and any other head of a federal agency with responsibility for providing federal financial assistance to a program or activity to issue regulations implementing such nondiscrimination requirements. requires such regulations to: (1) declare that no person shall be excluded from participation in, be denied the benefits of, or be subject to discrimination under, any program or activity on the ground of race, color, or national origin; and (2) address actions of programs or activities that result in disproportionate exposure to a covered substance on the basis of race, color, or national origin.
Metrics	Rouge1: 28.89, Rouge2: 20.15, RougeL: 22.96, RougeLsum: 26.67, Summary length (tokens): 239
Bottom 25% example (Sorted by rougeL)	
Document	<p>SECTION 1. SHORT TITLE; REFERENCES TO TITLE 38, UNITED STATES CODE. (a) Short Title.--This Act may be cited as the "Veterans Programs Improvement Act of 2003". (b) References.--Except as otherwise expressly provided, wherever in this Act an amendment is expressed in terms of an amendment to a section or other provision, the reference shall be considered to be made to a section or other provision of title 38, United States Code.</p> <p>SEC. 2. INCREASE IN RATES OF DISABILITY COMPENSATION AND DEPENDENCY AND INDEMNITY COMPENSATION.</p> <p>(a) Rate Adjustment.--The Secretary of Veterans Affairs shall, effective on December 1, 2003, increase the dollar amounts in effect for the payment of disability compensation and dependency and indemnity compensation by the Secretary, as specified in subsection (b). (b) Amounts To Be Increased.--The dollar amounts to be increased pursuant to subsection (a) are the following: (1) Compensation.--Each of the dollar amounts in effect under section 1114. (2) Additional compensation for dependents.--Each of the dollar amounts in effect under section 1115(1). (3) Clothing allowance.--The dollar amount in effect under section 1162. (4) New dic rates.--Each of the dollar amounts in effect under paragraphs (1) and (2) of section 1311(a). (5) Old dic rates.--Each of the dollar amounts in effect under section 1311(a)(3). (6) Additional dic for surviving spouses with minor children.--The dollar amount in effect under section 1311(b). (7) Additional dic for disability.--Each of the dollar amounts in effect under subsections (c) and (d) of section 1311. (8) DIC for dependent children.--Each of the dollar amounts in effect under sections 1313(a) and 1314. (c) Determination of Increase.-- (1) The increase under subsection (a) shall be made in the dollar amounts specified in subsection (b) as in effect on November 30, 2003. (2) Except as provided in paragraph (3), each such amount shall be increased by the same percentage as the percentage by which benefit amounts payable under title II of the Social Security Act (42 U.S.C. 401 et seq.) are increased effective December 1, 2003, as a result of a determination under section 215(i) of such Act (42 U.S.C. ...</p>
Gold	Veterans Programs Improvement Act of 2003 - Directs the Secretary of Veterans Affairs to increase, as of December 1, 2003, the rates of veterans' disability compensation, additional compensation for dependents, the clothing allowance for certain disabled adult children, and dependency and indemnity compensation for surviving spouses and children. Makes the effective date for the award of death pension the same as that for the award of death compensation or dependency and indemnity compensation. Excludes lump-sum insurance proceeds from income for purposes of eligibility for veterans' pensions. Prohibits the payment of veterans' disability compensation for an alcohol- or drug-abuse related disability even if the alcohol or drug abuse is secondary to a service-connected disability. Provides alternative beneficiaries for National Service Life Insurance and United States Government Life Insurance proceeds when the first beneficiary does not make a claim. Provides burial benefit eligibility for a veteran's surviving spouse who remarries following the veteran's death. Makes permanent the authority for the State cemetery grants program. Repeals the Department of Veterans Affairs Education Loan program. Includes self-employment training under the Montgomery GI Bill.
Model	veterans programs improvement act of 2003 - directs the secretary of veterans affairs, effective on december 1, 2003, to increase the rates of disability and dependency and indemnity compensation (dic) through the department of veterans affairs (va), to: (1) increase the rates of disability compensation and dependency and indemnity compensation; (2) provide for additional compensation for dependents; (3) provide for additional compensation for dependents; (4) exclude lump-sum sales of any life insurance policy or policies on a veteran for purposes of pension benefits; (5) exclude lump-sum sales of any life insurance policy or policies on a veteran for purposes of pension benefits; (6) exclude lump-sum life insurance proceeds from the determinations of annual income for pension purposes; (7) provide for alternative beneficiaries for certain veterans' life insurance policies or policies on a veteran's service-connected disability; and (8) authorize the secretary to approve a program of self-employment on-employment in the department of veterans affairs education loan program.amends the veterans' advisory committee on education to: (1) repeal the requirement that a claimant and the claimant's representative is necessary to complete an application is not received by the secretary within one year from the date of such notification; (2) make permanent the same authority for state cemetery grants program; and (3) authorize the secretary to approve a program of self-employment on-employment in the department of america known as the department of veterans affairs.
Metrics	Rouge1: 60.71, Rouge2: 29.79, RougeL: 33.88, RougeLsum: 50.82, Summary length (tokens): 297
Top 75% example (Sorted by rougeL)	
Document	<p>SECTION 1. SHORT TITLE. This Act may be cited as the "Cameron Gulbransen Kids and Cars Safety Act of 2003".</p> <p>SEC. 2. EVALUATION OF DEVICES AND TECHNOLOGY TO REDUCE CHILD INJURY AND DEATH FROM PARKED OR UNATTENDED MOTOR VEHICLES.</p> <p>(a) In General.--The Secretary of Transportation shall evaluate-- (1) devices and technologies intended to reduce the incidence of child injury and child death occurring outside of parked motor vehicles in nontraffic, noncrash events, including backing-over incidents, that are caused by such vehicles, and determining which of those methods is the most effective; and (2) currently available technology to prevent injury and death of children left unattended inside of parked motor vehicles, including injury or death due to hyperthermia, power windows, or power sunroofs. (b) Report.--The Secretary of Transportation shall submit a report on the findings and determinations of the evaluation under this section to the Congress by not later than one year after the date of the enactment of this Act. (c) Completion of Rulemaking Regarding Power Windows.--The Secretary of Transportation shall by not later than 6 months after the submission of the report under subsection (b) complete any rulemaking begun before the date of the enacting of this Act regarding power windows and power window switches.</p> <p>SEC. 3. DATABASE FOR TRACKING THE NUMBER AND TYPES OF INJURIES AND DEATHS IN NONTRAFFIC, NONCRASH EVENTS.</p> <p>(a) Establishment.--The Secretary of Transportation shall establish a database of (or modify an existing database to include), and collect data regarding, the numbers and types of injuries and deaths in nontraffic, noncrash events involving motor vehicles. (b) Included Information.--The Secretary of Transportation shall collect and include in such database the following information: (1) The types, makes, models, and model years of motor vehicles involved in nontraffic, noncrash events. (2) Whether there was an operator of each motor vehicle in such events. (3) The age of each operator of such motor vehicles. (4) The age of each individual who suffered injury or death in such events. (5) Whether each motor vehicle had technology installed to detect individuals and objects behind it. (6) ...</p>
Gold	Cameron Gulbransen Kids and Cars Safety Act of 2003 - Directs the Secretary of Transportation to: (1) evaluate devices and technologies to reduce child injuries and deaths occurring outside of parked motor vehicles in non-traffic, non-crash events or inside of parked vehicles when children are left unattended; (2) establish a database of, and collect data on, the number and types of injuries and deaths in such events; (3) evaluate technologies for detecting and preventing collisions with individuals and objects behind motor vehicles; (4) prescribe safety standards to require devices for detecting individuals and objects behind motor vehicles; and (5) prescribe safety standards for power windows and power sunroofs, including requirements for child-safe switches and auto reverse technology.
Model	tamarisk kids and cars safety act of 2003 - directs the secretary of transportation (dot) to evaluate: (1) devices and technologies intended to reduce the incidence of child injury and death occurring inside distant motor vehicles in nontraffic, noncrash events, and determine which are the most effective; and (2) currently available technology to prevent injury and death of children left behind the motor vehicles. directs the secretary to: (1) establish a database of, and collect data regarding, the number and types of injuries and deaths in nontraffic, noncrash events involving motor vehicles; and (2) prescribe motor vehicle safety standards.
Metrics	Rouge1: 63.59, Rouge2: 37.21, RougeL: 50.69, RougeLsum: 49.77, Summary length (tokens): 132
Top 95% example (Sorted by rougeL)	
Document	<p>SECTION 1. FINDINGS.</p> <p>The Congress finds the following: (1) As a Member of Congress from the Tenth Congressional District of Texas, as Majority Leader of the U.S. Senate, Vice- President and President of the United States, Lyndon Baines Johnson's accomplishments in the fields of civil rights, education, and economic opportunity rank among the greatest achievements of the past half century. (2) As President, Lyndon Johnson proposed, championed, led to passage, and signed into law on August 6, 1965, the Voting Rights Act of 1965, which swept away barriers impeding millions of Americans from meaningful participation in American political life. (3) On July 30, 1965, President Johnson signed into law the Social Security Amendments Act of 1965, popularly known as Medicare, which has transformed the delivery of health care in the United States and which, along with Social Security, reduced the rate of poverty among the elderly from 28.5 percent in 1966 to 9.1 percent in 2012. (4) On July 2, 1964, President Johnson secured passage and signed into law the most sweeping civil rights legislation since Reconstruction, the Civil Rights Act of 1964, which prohibits discrimination in employment, education, and public accommodations based on race, color, religion, or national origin. (5) On November 8, 1965, President Johnson signed into law the Higher Education Act, which provided need-based financial aid to students in the form of scholarships, work-study grants, and loans, and thus made higher education more accessible to populations of persons who were previously unable to attend college because of economic circumstances. (6) On October 3, 1965, President Johnson signed into law the Immigration and Naturalization Act of 1965, which transformed the Nation's immigration system by abolishing the racially based quota system that had defined American immigration policy for four decades and replaced it with a policy whose central purpose was family reunification, with a preference for immigrants with specific skill sets. (7) According to Robert A. Caro, the preeminent biographer of Lyndon Baines Johnson, with the s...</p>
Gold	This bill directs the Speaker of the House and the President pro tempore of the Senate to arrange for the posthumous award of a Congressional Gold Medal to Lyndon Baines Johnson in recognition of his contributions to the nation, including passage of the Voting Rights Act of 1965, the Social Security Amendments Act (Medicare) of 1965, the Civil Rights Act of 1964, the Higher Education Act of 1965, and the Immigration and Naturalization Act of 1965. Requires such medal to be given to the Lyndon Baines Johnson Library and Museum following its award, where it will be available for display and research.
Model	this bill directs the speaker of the house of representatives and the president pro tempore of the senate to arrange for the posthumous award, on behalf of congress, of a gold medal to lyrics to lyrics in recognition of his contributions to the nation, including recognition of his contributions to the nation, including recognition of the landmark voting rights act of 1965, the civil rights act of 1964, the higher education act of 1965, and the immigration and naturalization act of 1965.
Metrics	Rouge1: 72.83, Rouge2: 62.64, RougeL: 68.48, RougeLsum: 68.48, Summary length (tokens): 97

Table 12: Examples of the BillSum dataset using the model billsum-4096-1024 base diverse

Bottom 5% example (Sorted by rougeL)	
Document	in the last decade the amount of data regarding micrnas ( mirs ) and their target genes described in the literature has expanded tremendously . the volume of information on this new group of regulators ( i.e . , mirs ) has complicated attempts to integrate this data within existing metabolic and signalling networks . as regulators of gene expression in addition , a single mir can potentially regulate multiple different genes at the same time , leading to complex functional outcomes . however , from another perspective , the identification of groups of genes targeted by the same mir and the clustering of these genes within individual signalling pathways represents a means to understand the cross talk between multiple signalling networks and their role in a common biological process . the focus of this review is to summarize the validated groups of mirs functionally linked to the cross talk between tgf- , notch , and wnt signalling during the common biological process of epithelial - to - mesenchymal transition ( emt ) . in particular , this review will address whether the documented cross talk between these three important emt - associated pathways could be further reinforced by the identification of a signature of mirs , already depicted in the literature but not yet sharpened or clearly defined in this role . in the past years , many studies have elegantly described the role of tgf- , notch , and wnt pathways in promoting emt and emt - associated disorders including fibrosis and metastatic dissemination in cancer [ 16 ] . here we identify published and validated interactions between mirs and genes involved in tgf- , notch , and wnt signalling . this led to the discovery of a signature of 30 mirs each regulating all three of these signalling pathways . we then searched for additional validated genes targeted by these 30 mirs and then further clustered these into the tgf- , notch , and wnt signalling pathways . interestingly , in our attempt to identify mirs that were common to all three of these signalling pathways , we found that the 30-mir signature strongly reinforced existing evidence supporting cross talk between these three pathways during emt . in this review we used tarbase v6.0 , the largest currently available manually curated mir target gene database , which includes targets derived from specific and high throughput experiments . using tarbase v6.0 we searched the collection of manually curated , experimentally validated mir - gene intera...
Gold	epithelial - to - mesenchymal transition ( emt ) is a reversible process by which cancer cells can switch from a sessile epithelial phenotype to an invasive mesenchymal state . emt enables tumor cells to become invasive , intravasate , survive in the circulation , extravasate , and colonize distant sites . paracrine heterotypic stroma - derived signals as well as paracrine homotypic or autocrine signals can mediate oncogenic emt and contribute to the acquisition of stem / progenitor cell properties , expansion of cancer stem cells , development of therapy resistance , and often lethal metastatic disease . emt is regulated by a variety of stimuli that trigger specific intracellular signalling pathways . altered micrnas ( mir ) expression and perturbed signalling pathways have been associated with epithelial plasticity , including oncogenic emt . in this review we analyse and describe the interaction between experimentally validated mirs and their target genes in tgf- , notch , and wnt signalling pathways . interestingly , in this process , we identified a signature of 30 experimentally validated mirs and a cluster of validated target genes that seem to mediate the cross talk between tgf- , notch , and wnt signalling networks during emt and reinforce their connection to the regulation of epithelial plasticity in health and disease .
Model	micrnas ( mirs ) are a class of small noncoding rnas that regulate multiple gene expression at the same time , leading to complex functional outcomes . recent studies have shown that mirs regulate multiple different pathways at the same time , leading to complex functional outcomes . however , from another perspective , the identification of genes targeted by the same mir and the clustering of these genes within individual signalling pathways represents a means to understand the cross talk between multiple signalling pathways and their role in a common biological process . in this review , we summarize the published studies of mir- , notch , and wnt signalling in
Metrics	Rouge1: 20.22, Rouge2: 0.0, RougeL: 11.24, RougeLsum: 14.61, Summary length (tokens): 129
Bottom 25% example (Sorted by rougeL)	
Document	mhc class ii molecules are heterodimeric cell surface glycoproteins that bind exogenously derived antigenic peptides and present them to cd4 t cells 12 . class ii and chains are translocated into the endoplasmic reticulum ( er ) , where they form nonamers with invariant ( ii ) chain 3 . ii chain prevents the binding of immunogenic peptides due to the presence of a 14amino acid domain ( clip ) that occupies the peptide - binding groove of / dimers 3 . after ii degradation in the endocytic pathway , the mhc - encoded molecules hla - dm ( or h2-m in the mouse ) and hla - do ( h2-o ) facilitate the removal of clip from / dimers , allowing peptide binding 456 . ii chain has been implicated in functions such as er export , endosome targeting , and even b cell maturation 37 . two alternatively spliced ii isoforms exist ( p31 and p41 ) , distinguished by a 64-residue domain in the luminal portion of p41 8 . the isoforms are expressed differently in various apcs and regulate the presentation of certain antigen epitopes in b cells 9 . this difference may reflect protease inhibition by the amino acid insertion in p41 , as it has been shown to inhibit the lysosomal cysteine protease cathepsin 1 both in vitro and in vivo 910 . therefore , ii chain may contribute to the modulation of the proteolysis in the endocytic pathway and thus modulate antigen processing indirectly 1112 . we demonstrate here that ii chain deletion leads to the lysosomal degradation of h2-mb in apcs , suggesting that ii chain is required to prevent the proteolysis of h2-m and perhaps of other proteins . this feature may help explain how ii chain expression affects t cell selection and b cell maturation independently from its effect on mhc class ii traffic 131415 . c57b6/6 ( control ) and ii , ii p31 1416 , class ii , and class ii / ii mice ( the gift of p . murrack , university of colorado health sciences center , denver , co ) were kept in a pathogen - free environment for 78 wk before killing . splenocytes were obtained as described 7 . bone marrow derived dendritic cells ( dcs ) were cultured as described 17 . after purification , immature dcs were characterized by immunofluorescence and processed in parallel with the lps - treated dcs . the primers used here to detect i - a , h2-m , and h2-m are identical to the primers described previously 19 . 3 10 late dcs were pulse labeled with 7.5 mci / ml of [ s]methionine translabel ( scn ) and chased as described 17 . ...
Gold	the association of invariant ( ii ) chain with major histocompatibility complex ( mhc ) class ii dimers is required for proper antigen presentation to t cells by antigen - presenting cells . mice lacking ii chain have severe abnormalities in class ii transport , t cell selection , and b cell maturation . we demonstrate here that h2-m , which is required for efficient class ii antigenic peptide loading , is unexpectedly downregulated in splenocytes and mature dendritic cells ( dcs ) from i/i mice . downregulation reflects an increased rate of degradation in i/i cells . degradation apparently occurs within lysosomes , as it is prevented by cysteine protease inhibitors such as e64 , but not by the proteasome inhibitor lactacystin . thus , ii chain may act as a lysosomal protease inhibitor in b cells and dcs , with its deletion contributing indirectly to the loss of h2-m .
Model	mhc class ii molecules are heterodimeric cell surface glycoproteins that bind exogenously derived antigenic peptides and present them to cd4 t cells 12 . type ii chains are translocated into the endoplasmic reticulum ( er ) , where they form nonamers with invariant ( ii ) chain 3 . type ii chains are translocated into the endoplasmic reticulum ( er ) , where they form nonamers with invariant ( ii ) chain 3 . type ii chains are translocated into the endoplasmic reticulum ( er ) , where
Metrics	Rouge1: 27.67, Rouge2: 8.92, RougeL: 17.61, RougeLsum: 25.16, Summary length (tokens): 129
Top 75% example (Sorted by rougeL)	
Document	stroke , after myocardial infarction ( mi ) , is the second leading reason for mortality in iran as with many countries worldwide . the epidemiology of stroke has already been investigated in the american , european , african , and asian countries , no comprehensive study has yet investigated the epidemiology of stroke , particularly in mi patients . in iran , one of the largest countries in southwest asia , stroke and mi share many risk factors , most prevalent of which are smoking , dyslipidemia , type 2 diabetes , and hypertension . the risk factors for stroke and mi , especially smoking , hypertension , and dyslipidemia are highly prevalent in iran , as well , according to projections urbanism , increased life expectancy , reduction in childbirth , aging and elderly population , epidemiological changes , socioeconomic status , geographical conditions , and lifestyles such as poor diet , stress , and low mobility are the main causes of the burden of noncommunicable diseases , particularly stroke . because the determinants of stroke in different communities are various , we require knowledge about the risk factors and determinants of mortality in a community for effective planning and selection of appropriate strategies for the prevention and management of stroke and heart attack as the most important causes of death . since no comprehensive study has yet been investigated the status and mortality determinants of stroke in mi patients in iran , this study is conducted to determine and compare the determinants of mortality due to stroke in mi patients . in this retrospective cohort study , the data obtained from the mi registry of iran ' s cardiovascular diseases surveillance system were analyzed . around 20,750 hospitalized patients with mi with a new presentation ( hospitalized in 540 hospitals ) between april , 2012 and march , 2013 were enrolled . the study was approved by the management center of noncommunicable diseases and the department of cardiovascular diseases prevention of iran ' s ministry of health and medical education ( approval no . the research followed the principles of the declaration of helsinki ) ; the researchers did not conduct any interventions on the patients , and an institutional review board approved this research . the study protocol was approved by an independent scientific review committee ( at shahid beheshti university medical of sciences ) ( no . inclusion criteria were determined according to the world...
Gold	background : the data and determinants of mortality due to stroke in myocardial infarction ( mi ) patients are unknown , this study was conducted to evaluate the differences in risk factors for hospital mortality among mi patients with and without stroke history.materials and methods: this study was a retrospective , cohort study ; 20,750 new patients with mi from april , 2012 to march , 2013 were followed up and their data were analyzed according to having or not having the stroke history . stroke and mi were defined based on the world health organization ' s definition . the data were analyzed by logistic regression in stata software.results:of the 20,750 studied patients , 4,293 had stroke history . the prevalence of stroke in the studied population was derived 20.96% [ confidence interval [ ci ] 95% : 20.1321.24 ] . of the patients , 2,537 ( 59.1% ) had st - elevation mi ( stemi ) . mortality ratio in patients with and without stroke was obtained 18.8% and 10.3% , respectively . the prevalence of risk factors in mi patients with and without a stroke is various . the adjusted odds ratio of mortality in patients with stroke history was derived 7.02 ( 95% ci : 5.429 ) for chest pain resistant to treatment , 2.39 ( 95% ci : 1.972.9 ) for stemi , 3.02 ( 95% ci : 2.53.64 ) for lack of thrombolytic therapy , 2.2 ( 95% ci : 1.662.91 ) for heart failure , and 2.17 ( 95% ci : 1.62.9 ) for ventricular tachycardia.conclusion:with regards to the factors associated with mortality in this study , it is particularly necessary to control the mortality in mi patients with stroke history . more emphasis should be placed on the mi patients with the previous stroke over those without in the interventions developed for prevention and treatment , and for the prevention of avoidable mortalities .
Model	background : stroke and myocardial infarction ( mi ) share many risk factors , most prevalent of which are smoking , dyslipidemia , type 2 diabetes , and hypertension . the aim of this study was to determine and compare the determinants of stroke in mi patients.methods:in this retrospective cohort study , the data obtained from the mi registry of iran's cardiovascular diseases surveillance system were analyzed . patients with mi were defined by the date at mi diagnosis , hospital stay , and follow - up till discharge or death ( outcome ) . odds ratio ( or ) of mortality for clinical and demographic risk factors were calculated by logistic regression.results
Metrics	Rouge1: 55.88, Rouge2: 20.79, RougeL: 31.37, RougeLsum: 51.96, Summary length (tokens): 129
Top 95% example (Sorted by rougeL)	
Document	stroke commonly causes loss of motor function due to weakening of upper / lower extremity muscles1 . according to ryerson2 , use the affected upper extremity decreases because of the patient ' s dependency on the unaffected upper extremity for normal functions , which results in problems such as learned disuse , asymmetric postural patterns , contractures , and aggravated functional restrictions involving the affected upper extremity . therefore , to improve functions of the affected upper extremity in stroke patients , measures that maximize opportunities to use the affectedupper extremity are necessary . bilateral activities have been discussed as measures to improve the body symmetry and to reduce abnormal muscle tone3 , thereby promoting voluntary movement of the affected upper extremity4 . thus far , bilateral upper extremity coordination movements have been applied in the form of bilateral single exercises utilizing tasks such as figure imitation5 , robot arm upper extremity mechanisms6 and functional stretching7 , and bilateral complex exercises combined with the principle of motor learning , such as rhythmic acoustic8 , unaffected extremity weight addition9 , and active neuromuscular electrical stimulation in stroke patients10 . most previous studies have reported the positive effects of these exercises on motor function recovery in stroke patients . whitall et al.8 reported that when chronic hemiplegia patients underwent bilateral training to push and pull upper extremity apparatuses , including acoustic signals , their upper extremity functions were improved . in a study conducted by sumner et al.7 with 12 chronic stroke patients , the movement time of the patients affected side upper extremity decreased and upper extremity functions showed more improvement when they performed tasks to hold objects by stretching both hands simultaneously then when they performed tasks to hold objects by stretching only one hand . however , lee su - jin11 advised that the tasks in those studies involved mostly gross motor exercises of the upper extremity and simple repetitive training tasks , and that they were not sufficient to improve patients grasping ability to hold objects with various sizes , shapes , and weights using the shoulder , elbow , wrist , and fingers . although studies that applied diverse functional training tasks including activities of daily living to symmetric bilateral upper ...
Gold	[ purpose ] the purpose of this study was to investigate the effects of task - oriented bilateral arm training and repetitive bilateral arm training on upper limb function and activities of daily living in stroke patients . [ subjects ] forty patients with hemiplegia resulting from stroke were divided into a task oriented bilateral arm training group ( n=20 ) and a repetitive bilateral arm training group ( n=20 ) . [ methods ] the task - oriented group underwent bilateral arm training with 5 functional tasks , and the repetitive group underwent bilateral arm training with rhythmic auditory cueing for 30 minutes / day , 5 times / week , for 12 weeks . [ results ] the upper limb function and the ability to perform activities of daily living improved significantly in both groups . although there were significant differences between the groups , the task - oriented group showed greater improvement in upper limb function and activities of daily living . [ conclusion ] we recommend bilateral arm training as well as adding functional task training as a clinical intervention to improve upper limb function activities of daily living in patients with hemiplegia .
Model	[ purpose ] the purpose of this study was to compare the effects of symmetric bilateral upper extremity exercises using diverse functional training tasks with the effects of repetitive bilateral upper extremity exercises including acoustic signals on the recovery of upper extremity function and daily living activities in stroke patients. [ subjects and methods ] forty patients with hemiplegia resulting from stroke were randomly and equally assigned to a task - oriented bilateral arm training group ( bg : 11 males , 9 females ) and a repetitive bilateral arm training group ( rbg : 13 males , 7 females ) . both groups performed their exercises for
Metrics	Rouge1: 62.03, Rouge2: 34.59, RougeL: 50.27, RougeLsum: 60.96, Summary length (tokens): 129

Table 13: Examples of the PubMed dataset using the model pubmed-1024-128 small diverse

Bottom 5% example (Sorted by rougeL)	
Document	<p>cdnas containing the sequence of human wt - flt and human mutant flt498499insic were introduced into the pet-28a(+ ) expression vector ( novagen , emd chemicals inc . ) . the cdnas were cloned between the bamhi and xhoi sites , downstream from and in - frame with the sequence encoding an n - terminal his6 tag , to eliminate the his6 tag ( included in the expression vector ) , the sequence of the vector was modified by introducing the recognition sequence for cleavage by factor xa before the coding sequence of the ferritin genes . per amplification of the ferritin cdnas was performed using the upstream primer f1 5-tgg atc cat gaa agg tgg tat gag ctc cca gat t-3 and the downstream primer r1 5-tta tgc ctc gag ccc tat tac ttg gca agg-3. flt contains the factor xa sequence ( underlined ) . pet-28a(+ ) carrying wt - flt and mt - flt cdnas was transformed into h21 ( dc3 ) escherichia coli ( invitrogen ) . transformed cells were grown in luria broth medium ( b ) containing 30 g / ml kanamycin ( invitrogen ) at 37 c up to an absorbance of 0.91.0 at 600 nm . bacteria were induced to overexpress recombinant proteins by adding 1 mm isopropyl thio-d - galactopyranoside ( icn biotechnologies ) for 12 h at 25 c . purification of recombinant wt- and mt - flt homopolymers cells were harvested by centrifugation and frozen at -80 c . the cell pellets were suspended in 50 mm sodium phosphate , 500 mm nacl ( ph 7.4 ) , 1 mg / ml lysozyme , and a protease inhibitor mixture ( complete , roche applied science ) for 30 min . bacteria were disrupted by sonication , and the insoluble material was removed by centrifugation at 21,000 g for 30 min . the soluble fraction was purified by nickel iminodiacetic acid affinity chromatography using an akta purifier system ( ge healthcare ) . purified protein was eluted with 250 mm imidazole in 50 mm sodium phosphate ( ph 7.4 ) , 0.5 m nacl , recombinant proteins were diluted with 50 mm tris and 10% glycerol ( v / v ) down to an absorbance of 0.5 at 280 nm , and ferritins were cleaved from the his tag by digestion with factor xa protease ( ge healthcare ) ( 5 units / mg of protein ) . after being dialyzed against 50 mm tris , ph 8.0 , for 18 h , proteins were further purified by anion exchange chromatography ( mono q ) using a linear nacl elution gradient in 50 mm tris ( ph 8 ) . peak fractions were 95% pure based on s...</p>
Gold	<p>nucleotide insertions in the ferritin light chain ( fl ) polypeptide gene cause hereditary ferritinopathy , a neurodegenerative disease characterized by abnormal accumulation of ferritin and iron in the central nervous system . here we describe for the first time the protein structure and iron storage function of the flt mutant p.phc167serfsx26 ( mt - flt ) , which has a c terminus altered in sequence and extended in length . mt - flt polypeptides assembled spontaneously into soluble , spherical 24-mers that were ultrastructurally indistinguishable from those of the wild type . far - uv cd showed a decrease in helical content , and 8-anilino-1-naphthalenesulfonate fluorescence revealed the appearance of hydrophobic binding sites . near - uv cd and proteolysis studies suggested little or no structural alteration outside of the c - terminal region . in contrast to wild type , mt - flt homopolymers precipitated at much lower iron loading , had a diminished capacity to incorporate sites , and were less thermostable . however , precipitation was significantly reversed by addition of iron chelators both in vitro and in vivo . our results reveal substantial protein conformational changes localized at the 4-fold pore of mt - flt homopolymers and imply that the c terminus of the mt - flt polypeptide plays an important role in ferritin solubility , stability , and iron management . we propose that the protrusion of some portion of the c terminus above the spherical shell allows it to cross - link with other mutant polypeptides through iron bridging , leading to enhanced mutant precipitation by iron . our data suggest that hereditary ferritinopathy pathogenesis is likely to result from a combination of reduction in iron storage function and enhanced toxicity associated with iron - induced ferritin aggregates .</p>
Model	<p>apoferrin is a member of the nuclear receptor superfamily family of membrane proteins. apoferrin is a member of the family of apoferrin family. apoferrin is a member of the family of apoferrin family. apoferrin is a member of the family of apoferrin family. apoferrin is a member of the family of apoferrin family. apoferrin is a member of the family of</p>
Metrics	<p>Rouge1: 13.33, Rouge2: 1.35, RougeL: 12.0, RougeLsum: 12.0, Summary length (tokens): 129</p>
Bottom 25% example (Sorted by rougeL)	
Document	<p>intracellular signaling pathways transmit signals of various extracellular stimuli to their cytosolic and nuclear targets in order to induce biological responses , such as proliferation , differentiation , cell death and migration . when needed , the signals are transmitted from the cytoplasm to the nucleus via translocation of one or more components of each of the signaling pathways involved . thus , after stimulation , a large number of signaling proteins are rapidly translocated to the nucleus to induce and regulate many nuclear processes . however , despite the importance of stimulated nuclear signaling , the mechanisms by which these components reach the nucleus upon stimulation have been elucidated only for a few signaling pathways . classic nuclear shuttling is mediated by an importin - complex that binds to cargoes containing a nuclear localization signal ( nls ) , consisting of mono- or bi - partite clusters of basic amino acids [ 1 - 3 ] . this importin - complex often acts as a housekeeping mechanism that shuttles most nuclear proteins immediately to the nucleus after their translation , the relocalization of cargoes is followed by the dissociation of the proteins from the importins upon binding to rimpap , which exports the importins back to the cytoplasm , while the cargo remains in the nucleus . however , only a limited number of signaling proteins , such as raf and erk5 [ extracellular signal - regulated kinase 5 ] [ 8 - 10 ] , use this machinery for their stimulated nuclear shuttle . aside from this canonical mechanism , importin- or similar karyopherins , termed -like importins , operate by binding to non - canonical nlss to mediate translocation without the assistance of importin- . the mechanism by which these components operate is much less understood and is currently under investigation . recently , several -like importins have been implicated in the stimulated nuclear translocation of signaling proteins . here , we propose that this group of importins might be responsible for the swift nuclear shuttling of many proteins following various stimuli . the rapid and robust activation of erk1/2 allows the phosphorylation and modulation of the activity of more than 300 proteins , which are localized either in the cytoplasm or the nucleus [ 13 - 15 ] . these substrates are important for the induction and regulation of cellular processes , including proliferation , differentiation , and migration amongst others [ 16 - ...</p>
Gold	<p>classic nuclear shuttling is mediated by an importin-heterodimer that binds to cargoes containing a nuclear localization signal , and shuttles most nuclear proteins immediately after their translation . aside from this canonical mechanism , kariopheryn-s or -like importins operate by binding to non - canonical nuclear localization signals to mediate translocation without the assistance of importin- . the mechanism by which these components operate is much less understood and is currently under investigation . recently , several -like importins have been implicated in the stimulated nuclear translocation of signaling proteins . here , we propose that this group of importins might be responsible for the swift nuclear shuttling of many proteins following various stimuli .</p>
Model	<p>the nuclear translocation signal ( nl ) is a key regulator of many cellular processes, including proliferation, differentiation, cell death, migration, and migration. recent studies have shown that the nuclear translocation signal ( nl ) is required for the induction and regulation of many nuclear processes, including proliferation, differentiation, migration, and migration. here , we discuss the role of nl in the regulation of the nuclear translocation signal ( nl ) , as well as the mechanisms that regulate it.</p>
Metrics	<p>Rouge1: 33.33, Rouge2: 5.0, RougeL: 18.52, RougeLsum: 32.1, Summary length (tokens): 100</p>
Top 75% example (Sorted by rougeL)	
Document	<p>helicobacter pylori is a stomach bacterium that colonizes ~50% of people globally.1 h. pylori is the primary risk factor for gastric cancer the third highest cause of global cancer morbidity.2 h. pylori infection rates are highly dependent on socioeconomic status , ~80% of those living in low socioeconomic areas of latin america , asia , and eastern europe are infected , compared with &lt; 20% of asymptomatic caucasians in the usa.3 h. pylori infection is treatable with different regimens of antibiotics.4 and eradication of h. pylori is a recognized way to lower incidence of gastric cancer.5 however , recurrence of infection is variable.6,7 and the emergence of antibiotic resistance compromises treatment efficacy . thus , determining the best course of treatment is important to improve treatment efficacy and to reduce recurrence of h. pylori infection . unfortunately , there is no broad consensus about an optimal antibiotic therapy for the treatment of h. pylori . for example , meta - analyses of european and asian clinical data compared the standard triple therapy ( amoxicillin , clarithromycin , and a proton - pump inhibitor for 14 days ) with 5- or 10-day quadruple therapy regimens ( adding metronidazole or tinidazole to the triple therapy ) and found that quadruple therapies are both significantly more effective and cheaper than the triple therapy.8,10 however , we previously published a study comparing eradication therapies in seven sites of six latin american countries that showed that the 14-day triple therapy was superior to the 5-day concomitant quadruple therapy , and no different than the 10-day sequential quadruple therapy.11,12 these inconsistencies reflect localized differences in antibiotic use practices , such as the use of clarithromycin for upper respiratory infections.13 the differences in efficacy of antibiotic therapy are supported by primary antibiotic resistance data . for example , h. pylori resistance to amoxicillin varied widely between africa ( 65.6% ) , europe ( 0.5% ) , asia ( 11.6% ) , and the americas ( 2.2% ) 12 even in the same region , patterns of resistance differ : within central and latin america , reported average metronidazole resistance varies from 30% in argentina to 83% in colombia , and tetracycline resistance varies from 2% in brazil to 33% in colombia.14 as such , characterizing local resistance patterns is important for selecting therapies with the highest likelihood of success . our research focus...</p>
Gold	<p>objectivegastric carcinoma is the most common cancer and cause of cancer mortality in peru . helicobacter pylori , a bacterium that colonizes the human stomach , is a group 1 carcinogen due to its causal relationship to gastric carcinoma . while eradication of h. pylori can help prevent gastric cancer , characterizing regional antibiotic resistance patterns is necessary to determine targeted treatment for each region . thus , we examined primary antibiotic resistance in clinical isolates of h. pylori in lima , peru , materials and methods . pylori strains were isolated from gastric biopsies of patients with histologically proven h. pylori infection . primary antibiotic resistance was examined using c - test strips . isolates were examined for the presence of the caga pathogenicity island and the vaca m1/m2 alleles via polymerase chain reaction.resultseventy-six isolates were recovered from gastric biopsies . clinical isolates showed evidence of antibiotic resistance to 1 ( 27.6% , n=21/76 ) , 2 ( 28.9% , n=22/76 ) , or 3 antibiotics ( 40.8% ) . of 76 isolates , eight ( 10.5% ) were resistant to amoxicillin and clarithromycin , which are part of the standard triple therapy for h. pylori infection . no trends were seen between the presence of caga , vaca m1 , or vaca m2 and antibiotic resistance.conclusionthe rate of antibiotic resistance among h. pylori isolates in lima , peru , is higher than expected and presents cause for concern . to develop more targeted eradication therapies for h. pylori in peru , more research is needed to better characterize antibiotic resistance among a larger number of clinical isolates prospectively .</p>
Model	<p>helicobacter pylori infection is the third most common cause of global cancer morbidity. eradication of helicobacter pylori is a recognized way to lower incidence of gastric cancer. however, there is no broad consensus about an optimal antibiotic therapy for the treatment of helicobacter pylori. we assessed primary antibiotic resistance to helicobacter pylori among 76 isolates from a cohort of patients recruited in lima, peru. primary antibiotic resistance to metronidazole, amoxicillin, tetracycline, clarithromycin, levofloxacin, and rifamp</p>
Metrics	<p>Rouge1: 50.0, Rouge2: 17.28, RougeL: 32.93, RougeLsum: 47.56, Summary length (tokens): 129</p>
Top 95% example (Sorted by rougeL)	
Document	<p>in general , brain injury can occur due to sudden and severe head strike to a hard object , which can be mild , moderate or severe ( 1 ) . the main causes of head injury include traffic accidents , falling from heights , physical violence , accidents at work , inside home accidents and during exercise incidents . however , the most important cause of head trauma in iranian population is traffic accident ( 2 ) . among the warning signs of head trauma are nausea , vomiting , dizziness , headache , blurred vision and loss of balance , difficulty in sleeping , memory problems , tinnitus and fatigue ( 3 ) . nausea and vomiting are the most common complications after minor head trauma that in addition to severe harassment of patients increases the risk of aspiration and intracranial pressure rising . ondansetron is a serotonin 5-ht3 receptor antagonist , which connects to the peripheral and central receptors of serotonin ( 1 ) . this drug is mostly used in nausea and vomiting after chemotherapy and surgery ( 2 ) . it does not have any effect on dopamine receptors thus ; it does not have extra pyramidal effect ( 3 ) . this drug has a half - life of 2 - 7 hours and is metabolized in the liver where it changes into glucuronide and sulfate which is inactive . its most common side effects include headache , fatigue , diarrhea , constipation , dizziness and anxiety . the recommended dose for the treatment of nausea and vomiting is 4 - 8 milligrams ( 4 , 5 ) . metoclopramide as an old antiemetic is mostly used in high doses . before chemotherapy and for nausea and vomiting caused by various reasons ( 6 - 8 ) . this drug blocks the dopamine receptors on the peripheral and central dopamine receptors and increases the movement of the upper gastrointestinal tract without increasing secretion ( 9 , 10 ) . its intravenous absorption takes about 3 minutes and the peak of its effect is about 15 minute . this drug is metabolized in the liver and its half - life is approximately 4 - 5 hours ( 11 ) . its most common side effects include dystonia &lt; 10% , fatigue , drowsiness , and flushing . based on the above - mentioned reasons , the present study was aimed to compare the antiemetic effects of metoclopramide and ondansetron in the treatment of post head trauma nausea and vomiting . study design and setting the study was a controlled , randomized , double blind clinical trial , which was conducted in the first 6 months of 2014 in al - zahra a...</p>
Gold	<p>introduction : nausea and vomiting are the most common complications after minor head trauma that increases the risk of intracranial pressure rising . therefore , the present study was aimed to compare the antiemetic effects of metoclopramide and ondansetron in the treatment of post - traumatic nausea and vomiting . methods : the study was a controlled , randomized , double blind clinical trial , which was conducted in the first 6 months of 2014 in emergency department al - zahra and kashani hospitals in isfahan , iran . the patients with minor head trauma associated with nausea and vomiting were randomly divided into 2 groups : treatment with metoclopramide ( 10mg/2ml , slow injection ) and treatment with ondansetron ( 4mg/2ml , slow injection ) . the comparison between the 2 groups was done regarding antiemetic efficacy and side effects using spss 21 statistical software . results:120 patients with minor head trauma were distributed and studied into two groups of 60 patients ( mean age 35.614 ± 1 years ; 50.0% male ) . administration of both ondansetron and metoclopramide significantly reduced the severity of nausea ( p=0.001 ) . changes in the severity of nausea in both groups before and after the treatment revealed that nausea had been decreased significantly in both groups ( p &lt; 0.001 ) . the incidence of fatigue ( p=0.44 ) , headache ( p=0.58 ) and dystonia ( p=0.06 ) had no significant difference in the two groups but the incidence of drowsiness and anxiety in the metoclopramide group was significantly higher ( p &lt; 0.001 ) . conclusion : the present study indicated that the treatment effectiveness of ondansetron and metoclopramide are similar . however , incidence of drowsiness and anxiety in the metoclopramide was considerably higher . since these complications can have adverse effects on the treatment of patients with brain injury , it is suggested that it may be better to use ondansetron in these patients .</p>
Model	<p>background : head injury can occur due to sudden and severe head strike to a hard object, which can be mild, moderate or severe. metoclopramide as an old antiemetic is mostly used in high doses, before chemotherapy and for nausea and vomiting caused by various reasons. the present study was aimed to compare the antiemetic effects of metoclopramide and ondansetron in the treatment of post head trauma nausea and vomiting. methods : this randomized, double blind clinical trial was conducted in the first 6 months of 2014 in isfahan, iran. the patients were randomly divided into</p>
Metrics	<p>Rouge1: 59.69, Rouge2: 46.56, RougeL: 51.31, RougeLsum: 58.64, Summary length (tokens): 129</p>

Table 14: Examples of the PubMed dataset using the model pubmed-1024-128 base diverse

## Bottom 5% example (Sorted by rougeL)

**Document** this study is an extension of a report on patient type 1 diabetes at children's hospital of new orleans (14) and was approved by the institutional review board at louisiana state university health sciences center, new orleans, louisiana. glucose data were downloaded from patient meters at each clinic visit, meter model and sampling protocols varied by patient preference and insurance provider. an average of three glucose measurements per day were recorded in a study using a similar self-monitoring protocol (7). a1c was measured by national glycohemoglobin standardization program (ngsp)-approved immunoassays (15) at the children's hospital (184 patients) or by commercial laboratories that presumably also used ngsp-approved methods (18 patients). patients were divided into 7 -moderate-, and 7 -high- subjects by a population regression equation  $\ln(a1c) = -0.022 \text{ mhg} [mg/dl] + 4.3, r = 0.57$  was derived using mean mhg and mean a1c from 202 patients collected at 1612 clinic visits as described in the same paper (7). the same equation was used to divide the same patients into 7 -low-, moderate-, and 7 -high- groups based on mean mhg using the regression equation. hgi values were calculated by subtracting predicted a1c from observed a1c measured at the same clinic visit. patients were divided into low-, moderate-, and high- hgi groups based on mean hgi tertile (33% rank low hgi,  $<0.41$ ;  $n = 67$ ; moderate hgi,  $0.41 \text{ to } 0.26$ ;  $n = 68$ ; high hgi,  $>0.26$ ;  $n = 67$ ). eag was calculated by inserting observed a1c into the aged linear regression equation  $\ln(a1c [mg/dl]) = 128.74 \text{ ln} [a1c [mg/dl]] + 146.7, r = 0.92$  (1). a rank low blood glucose index (mbgi) that quantifies the difference between mhg and eag was calculated by subtracting observed mhg from eag. descriptive statistics and linear regression analyses were generated using graphpad prism v. 4.05 (graphpad software, san diego, ca). glucose data from 184 patients from 1612 clinic visits and an average of three glucose measurements per day were recorded in a study using a similar self-monitoring protocol (7). a1c was measured by national glycohemoglobin standardization program (ngsp)-approved immunoassays (15) at the children's ho-

**Gold** objective a1c - derived average glucose is recommended reporting a1c in estimated average glucose (eAG) equivalents. we compared eAG with self - monitored mean blood glucose (mbg) to determine whether eAG is systematically biased due to biological variation in the relationship between mbg and a1c.research design and methodology and a1c are recorded from charts of 202 pediatric type 1 diabetic patients at 1,612 clinic visits. patients were divided into groups with low, moderate, or high a1c bias based on a hemoglobin glycation index (hgi)result:mean sd values for mbg versus eAG were as follows: total population, 194/34 vs. 196/36 mg/d; low - hgi group, 186/31 vs. 163/20 mg/d; moderate - hgi group, 195/28 vs. 193/19 mg/d; and high - hgi group, 230/31 vs. 230/31 mg/d.conclusion:se underestimated mbg in low hgi patients and overestimated mbg in high hgi patients. disagreement between eAG and mbg downloaded from patient glucose meters will cause confusion if eAG is implemented for clinical use.

Model

objectiveto investigate the relationship between hemoglobin glycated hemoglobin ( hba1c ), hemoglobin glycated hemoglobin ( hba1c ), and hemoglobin glycated hemoglobin ( hba1c ) in children with type 1 diabetes.research design and methodsa1c was measured by national glycohemoglobin standardization program ( ngsp)-approved immunoassays at the children's hospital of new orleans. hba1c was measured by national glycohemoglobin standardization program ( ngsp)-approved immunoassays at the children's hospital of new orleans. hba1c was measured by national glycohemoglobin standardization program ( ngsp)-approved immunoassays at the children's hospital of new orleans. a population regression equation [ a1c ( % ) = [ 0.021 mbg ( mg / dl ) + 4.3, r = 0.92 ] + 4.3, r = 0.58 ] was derived using mean hba1c ( % ). a population regression equation [ a1c ( % ) = [ 0.021 mbg ( mg / dl ) + 4.3, r = 0.92 ] + 4.3, r = 0.92 ] was

Metrics	Rouge1: 19.4, Rouge2: 6.06, RougeL: 10.03, RougeLsum: 13.38, Summary length (tokens): 256
---------	---

## Bottom 25% example (Sorted by rougeL)

**Document** the principal aim of this laboratory is the synthesis of conjugated unsaturated ketones as candidate antineoplastic agents. These compounds interact with thiols but in general, they have little or no affinity for amino and hydroxyl groups which are found in nucleic acids (1-3). Hence, thiol alkylators may not have the genotoxic properties associated with a number of contemporary anticancer drugs (4), however after an initial chemical screening, certain neoplasms are more vulnerable to a subsequent cytotoxic effect than various non-malignant cells (5, 6), hence by mounting the 1,5-diaryl-3-oxo-1,4-pentadienyl pharmacophore  $[Ar-C(=O)-C=C-C(=O)-Ar']$  on heterocyclic and cycloaliphatic scaffolds, two sequential interactions with cellular thiols can take place which may be more detrimental to tumours than normal tissues, such considerations led to the development of 3,5-bis(heteroaryl)-4-piperidines **1a-d** which demonstrated potent cytotoxic properties with the  $IC_{50}$  values in the low micromolar range against cell lines: C-lymphocytes and C-lymphocytes (7). The piperidine moiety was chosen because of its ability to form hydrogen bonds with the amino group of the primary binding site and a side chain on the piperidine nitrogen may align at an auxiliary binding site which could enhance cytotoxic properties. In order to evaluate this hypothesis, a novel series of  $\alpha$ -aryl-3,5-bis(heteroaryl)-4-piperidine derivatives **2-7** were synthesized (8). In these compounds, the side chain contains atoms and groups capable of forming hydrogen and van der Waals bonds and also an ionic bond in series **3-6** which could bind at an auxiliary binding site when the potencies of the compounds in series **3-6** were compared with that of the analogs **1a-d** which have the same aryl substituent, the  $IC_{50}$  values of the amides **3-6** were lower in 48% of the comparisons made while equipotency was noted in 35% of the cases (8). The proposed primary sites of cytotoxicity are the primary binding site and the auxiliary binding site. The physicochemical properties of the 3,5-bis(heteroaryl)-4-piperidine derivatives **1-7** that influence cytotoxic potencies, and second, to derive predictive  $2D$ -QSAR models that will be helpful to design new analogs with improved bioactivity. *...*

1-[4-(2-alkylaminoethoxy) phenylcarbamyl]-5,5-bis(arylidene)-4-piperidones are a novel class of potent cytotoxic agents. These compounds demonstrate low micromolar to submicromolar IC<sub>50</sub> values against human molar 468 and cell L-symphocytes and murine leukemia L1210 cells. In this study, a comparative *in vitro* investigation was performed on a series of 3,5-bis(arylidene)-4-piperidones, using different chemometric tools to develop the best predictive models for the development of analogs with improved cytotoxicity. all the qsr models were validated in internal validation tests, the qsr models obtained by a  $q^2$  ps method were considered the best as compared to mlr method, the best qsr model obtained by a  $q^2$  ps analysis on L1210,  $q^2$  and mol468 demonstrated good predictivity with  $\Delta$ pred values ranging from 0.94–0.80, molecular density, topological ( $\chi$  and  $\chi^2$ ) and geometrical indices of the molecules were found to be the most important factors for determining cytotoxic properties.

[illegible]

Metrics	Rouge1: 21.38, Rouge2: 5.06, RougeL: 16.35, RougeLsum: 16.35, Summary length (tokens): 253
---------	--

## Top 75% example (Sorted by rougeL)

**Document** educational evaluation (ee) is a formal process performed to evaluate the quality of effectiveness and/or value of a program, process, goal or curriculum.12 it deals with data collection and assessment of the progress of academic programs,34 by considering some principles related to educational measurements and data collection, it can result in a better understanding of such progress.57 during the past thirty years, theorists have presented numerous methods of evaluation, worthen and sanders2 mentioned that more than 50 different evaluation approaches have been developed in recent decades. among these, methods based on internal criteria are known as the ones that can interpret the scientific, educational, and therapeutic authenticity of different educational groups.48 this is greatly welcomed by the academic community and is widely spread to all universities in the world, that is because this method provided a scientific, appropriate, precise, timely, and valid basis regarding the interpreting of the results of the evaluation.36 the present study was inspired by the research of the council secretariat of ministry of health, treatment and medical education of iran in 1995/8 c.e. which has its most effective, value, and results when it can provide needed information to individuals which are directly related, as well as those who may be benefited from its results.346 educating dental professionals consists of theoretical and practical (clinical, paraclinical, and laboratory) courses, differing in duration, and educational curriculum among different countries. it might vary from 4 years (e.g., in india, turkey, and russia) to 6 years or more (e.g., in iran, consisting of 2 years of only basic medical sciences and 4 + years of dentistry courses), due to numerous practical educational units in dentistry education and with regard to expensive but very critical protocols for the dental resources in the world.346 therefore, the evaluation of the dental curriculum, on the other hand, the quality of dental services play an important role in public health, considering these issues, dentistry education needs to be cost-effective in terms of optimizing its quality, to date, the pace of advancements in dentistry...

**Gold** background: educational evaluation is a praxis which deals with data collection and assessment of academic activities progress. in this research, educational evaluation of dentistry school of tehran university of medical sciences, which trains students in undergraduate and residency courses, was studied.methods this descriptive study was done with a model of educational evaluation in ten steps and 13 fields including purposes and mission objectives, management and organization, academic board members, students, human resources and support, educational, research, health and treatment spaces, educational, diagnostic, research and laboratory tools, educational, research, health and treatment programs and courses, process of teaching and learning, evaluation and assessment, alumni, and patients satisfaction. data were collected using observation, interviews, questionnaires, and checklists.results results of the study were mainly qualitative and in some cases quantitative, based on defined optimal situation, the total mean of qualitative results of educational evaluation of dentistry school in all 13 fields was 55.98% which is relatively diverse. in the case of quantitative ones, results of some fields such as treatment quality of patients and educational facilities of students (61.32% and 60.16% respectively) conclusion: according to the results, educational goals and missions, educational and research facilities and spaces which were identified as the weakest areas need to be considered and paid more serious attention

[illegible]

Metrics	Rouge1: 39.32, Rouge2: 12.68, RougeL: 27.67, RougeLsum: 34.47, Summary length (tokens): 256
---------	---

## Top 95% example (Sorted by rougeL)

**Document** the study population included patients over 18 who had an initial cabg or combined cabg and open chest aortic valve replacement (avr), from april 1, 1998 to october 31, 2011. in 2010, the date of first cardiac surgery was the index date, and eligible patients were followed for 1 year with respect to major outcomes, and 5 years for mortality. preoperative data were included for 1 year prior to surgery, and outcomes for 1 year postoperatively. patients for whom sex, age, height, weight were missing, and patients living outside of ontario or of unknown residence were excluded. cardiac care network of ontario (ccn) data were used to identify baseline characteristics such as cardiac ejection fraction, number of grafts bypassed, prior myocardial infarction (mi), emergency or elective surgery, and other co-morbidities. ccn data and the following datasets were combined from iccs using deterministic linkage by unique iccs key number identifiers: ontario health insurance plan, canadian institute of health information (cihi) discharge abstracts (dab), and the iccs coronary artery bypass graft registry (cabg registry). patients were subsequently selected from the cihi discharge abstract database, a data for which other cardiac procedures had been performed during the same admission were excluded (eg, percutaneous coronary intervention or other valve procedures). bmi was calculated as weight (kg)/height (m), and patients were divided into groups: underweight (bmi <20 kg/m), normal weight (bmi 20.0 to 24.9 kg/m), overweight (bmi 25.0 to 29.9 kg/m), obese (bmi 30.0 to 34.9 kg/m), and morbidly obese (bmi ≥34.9 kg/m), closely based on world health organization (who) and health canada guidelines.[12] all of the following comorbidities were assessed for presence within 1 year prior to index date: diabetes, smoking history (current or ever smoked), peripheral vascular disease (pvd), pulmonary disease (copd), chronic kidney disease (ckd), chronic liver disease (clfd, from chli), hypertension, elective or emergent surgery, creatinine, within 30 days prior to surgery, the following cardiac characteristics were captured: ejection fraction, prior mi, left main coronary disease, and previous cabg at date of surgery.

**Gold** backgroundthe obesity paradox reflects an observed relationship between obesity and decreased morbidity and mortality, suggesting improved health outcomes for obese individuals. studies examining the relationship between high body mass index (BMI) and adverse outcomes have reported conflicting results methods and results study population (n=78 762) was comprised of adult patients who had undergone first-time coronary artery bypass (CABG) or combined CABG and aortic valve replacement (AVR) surgery from april 1, 1998 to october 31, 2011 in ontario (data from the institute for clinical evaluation and comparative research) and 5-year mortality was defined by the following categories (underweight < 20, normal weight 20 to 24.9, overweight 25 to 29.9, obese 30 to 34.9, morbidly obese > 34.9) were compared using bivariate analyses and cox multivariate regression analysis to investigate multiple confounders on the relationship between BMI and adverse outcomes. a reverse j-shaped curve was found between BMI and mortality with their respective hazard ratios, independent of confounding variables, 30-day, 1-year, and 5-year survival rates were highest for the obese group of patients [99.1% 95% confidence interval [ci ], 98.9 to 99.2], 97.6% [95% ci, 97.3 to 97.8], and 90.0% [95% ci, 89.5 to 90.5], respectively, and perioperative complications tended lower. underweight and morbidly obese patients had higher mortality and incidence of adverse outcomes.conclusionsoverweight and obese patients had lower mortality and adverse perioperative outcomes after cardiac surgery compared with normal weight, underweight, and morbidly obese patients. the obesity paradox was confirmed for overweight and moderately obese patients. this may impact health resource planning, shifting the focus to morbidly obese and underweight patients prior to, during, and after cardiac surgery.

**Model** background the aim of this study was to investigate the relationship between body mass index (bmi) and mortality in patients undergoing coronary artery bypass grafting (cabg) or open chest aortic valve replacement (avr). methods and results we identified 1 year patients undergoing primary cabg or combined cabg and open chest aortic valve replacement (avr) from april 1, 1998 to october 31, 2011. patients were divided into groups : underweight (bmi < 20 kg / m<sup>2</sup>), normal weight (bmi 25.0 to 29.9 kg / m<sup>2</sup>), overweight (bmi 25.0 to 29.9 kg / m<sup>2</sup>), obese (bmi 30.0 to 29.9 kg / m<sup>2</sup>), and morbidly obese (bmi > 34.9 kg / m<sup>2</sup>). patients were followed for 1 year with respect to major outcomes, and 5 years for mortality. cox proportional hazards regression analysis was used to investigate multiple confounders on the relationship between bmi and mortality, providing hazard ratios and 95% confidence intervals (ci). patients with bmi < 20 kg / m<sup>2</sup> were more likely to die during the 5 years of follow - up ( hazard ratio [ hr ]

Metrics	Rouge1: 56.82, Rouge2: 26.29, RougeL: 40.91, RougeLsum: 50.0, Summary length (tokens): 256
---------	--

Table 15: Examples of the PubMed dataset using the model pubmed-4096-512 small diverse



Bottom 5% example (Sorted by rougeL)	
Document	<p>in august , 4 months before presentation , a 35-year - old white woman of scots and english descent developed reddish urine for several days followed by eruption of vesicles and blisters on the dorsal surfaces of her hands and fingers , the sides of her nose , her upper anterior chest , she worked as a landscaping contractor and noticed that lesions occurred on areas exposed to sunlight , but application of sunscreen neither diminished the rate at which new lesions appeared , nor promoted healing of older lesions . her skin was fragile in areas of the lesions and the lesions healed slowly , often with scarring . she also developed dark brown pigmentation and the growth of fine black hair over her cheeks . she consumed three glasses of wine each week and had smoked electronic cigarettes for approximately 6 months , having changed from tobacco cigarettes . she had donated three units of blood for transfusion , but none in several years . she had no menses in the 12 months before presentation due to the effects of a contraceptive vaginal ring ( nuvaring ; etonogestrel / ethinyl estradiol ) . a dermatologist performed a punch biopsy of two skin lesions on her left forefinger and referred her for hematology evaluation and treatment . physical examination confirmed the presence of new vesicles and bullae with erythematous bases , some as large as 1 cm in diameter , and older lesions in various stages of erosion , resolution , and scarring in the anatomical distribution described above . lesions were most prominent on the dorsal surfaces of the hands and fingers ( fig . milia were scattered over areas affected with bullous lesions and were especially prominent on skin overlying finger joints . ( a ) before therapy , there were vesicles , bullae with erythematous bases , older lesions in various stages of erosion , resolution , scarring , and punctate milia , all typical of pct . ( b ) after iron depletion with phlebotomy , there were no active skin lesions , although hyperpigmentation , scarring , and milia ( especially over dorsal aspects of second and third interphalangeal joints ) persisted . punch biopsy specimens of skin were deposited in immunofluorescence transport medium , flash frozen , and cut for manual immunofluorescence staining . the sections were probed with fluorescein - labeled anti - human antibodies specific for igg , iga , igm , c3 , c5b-9 , and fibrinogen . complete blood count ( including ...</p>
Gold	<p>porphyria cutanea tarda ( pct ) is characterized by decreased uroporphyrinogen decarboxylase activity in hepatocytes , uroporphyrin i and heptacarboxyl porphyrin iii accumulation , photosensitivity dermatitis , and increased storage iron . in women , estrogen therapy , including oral contraceptives , postmenopausal hormone replacement , and tamoxifen for breast cancer treatment , is a risk factor for pct . we report the case of a woman who presented with pct , hfe c282y homozygosity , and hepatic iron overload and was using a contraceptive vaginal ring containing ethinyl estradiol , an estrogen . we discuss this case in the context of characteristics of other persons with pct , including common hfe mutations , iron overload , and estrogen exposure .</p>
Model	<p>a 35-year - old white woman of scots and english descent developed reddish urine for several days followed by eruption of vesicles and blisters on the dorsal surfaces of her hands and fingers, the sides of her nose, and her upper anterior chest, knees, and legs. skin biopsy specimens revealed linear glassy iga and igg deposition along the epidermal basement membrane zone and superficial dermal blood vessels, forming a doughnut pattern. the patient was treated with a levonorgestrel - releasing intrauterine contraceptive system ( skylla ). eight months after presentation, she underwent implantation of a levonorgestrel - releasing intrauterine contraceptive system ( skylla ). eight months after presentation, she underwent implantation of a levonorgestrel - releasing intrauterine contraceptive system ( skylla ).</p>
Metrics	<p>Rouge1: 20.56, Rouge2: 0.94, RougeL: 11.22, RougeLsum: 17.76, Summary length (tokens): 176</p>
Bottom 25% example (Sorted by rougeL)	
Document	<p>this was a multicenter , community - based , retrospective observational study of patients with pns , ranging from 8 to 20 mm in diameter , presenting to 18 geographically representative outpatient pulmonary clinics across the united states . the study was approved at 15 sites by a central institutional review board and at three sites by local institutional review board approval . four hundred forty sites were identified based on investigator databases and claims data from a large insurance carrier whose coverage population was representative of the overall us population . of these , 77 sites expressed interest in participating , and 48 sites went on to sign confidentiality agreements . of these , 17 did not request additional information , leaving 31 sites undergoing qualification review . eighteen outpatient pulmonary clinics were chosen to participate based on the following criteria : ( 1 ) management of patients with pns , ( 2 ) availability of medical records , and ( 3 ) ability to perform data abstraction . in addition , investigators targeted enrollment of geographically diverse patients to limit the potential bias associated with differences in practice patterns and to account for variation in disease prevalence ( eg , endemic mycoses ) that could alter management decisions . patients were identified by querying databases ( eg , billing and scheduling systems ) using the international classification of diseases , ninth revision , clinical modification codes for pn ( 793.1 , 786.6 , 518.89 , 519.8 , 519.9 ) to ensure homogeneity in patient identification and inclusion . manual chart abstraction was then used to identify those who met the criteria . to minimize selection bias , the sites were not permitted to use additional codes during database query to identify patients . to ensure a systematic sample , inclusion criteria included age 40 years and 89 years at the time of nodule finding , presentation to a pulmonologist , nodule size 8 to 20 mm , and definitive diagnosis ascertained by tissue diagnosis or radiographic follow - up for 2 years . exclusion criteria included chest ct scan performed &gt; 60 days prior to the initial visit , prior diagnosis of any cancer within 2 years of nodule detection , or incomplete chart data . patients were categorized into three groups by the most invasive procedure performed during management , as follows : surveillance ( serial imaging ) , biopsy ( ct scan - guided transthoracic needle aspir... </p>
Gold	<p>background : pulmonary nodules ( pns ) are a common reason for referral to pulmonologists . the majority of data for the evaluation and management of pns is derived from studies performed in academic medical centers . little is known about the prevalence and diagnosis of pns , the use of diagnostic testing , or the management of pns by community pulmonologists.methods:this multicenter observational record review evaluated 377 patients aged 40 to 89 years referred to 18 geographically diverse community pulmonary practices for intermediate pns ( 8 - 20 mm ) . study measures included the prevalence of malignancy , procedure / test use , and nodule pretest probability of malignancy as calculated by two previously validated models . the relationship between calculated pretest probability and management decisions was evaluated.result:the prevalence of malignancy was 25% ( n = 94 ) , nearly one - half of the patients ( 46% , n = 175 ) had surveillance alone . biopsy was performed on 125 patients ( 33.2% ) , a total of 77 patients ( 20.4% ) underwent surgery , of whom 35% ( n = 27 ) had benign disease . pet scan was used in 141 patients ( 37% ) . the false - positive rate for pet scan was 39% ( 95% ci , 27.1%-52.1% ) . pretest probability of malignancy calculations showed that 9.5% ( n = 36 ) were at a low risk , 79.6% ( n = 300 ) were at a moderate risk , and 10.8% ( n = 41 ) were at a high risk of malignancy . the rate of surgical resection was similar among the three groups ( 17% , 21% , 17% , respectively ; p = .69).conclusions : a substantial fraction of intermediate - sized nodules referred to pulmonologists ultimately prove to be lung cancer . despite advances in imaging and nonsurgical biopsy techniques , invasive sampling of low - risk nodules and surgical resection of benign nodules remain common , suggesting a lack of adherence to guidelines for the management of pns .</p>
Model	<p>objectivethe aim of this study was to document the prevalence of cancer ( 25% ) in patients with intermediate - sized pulmonary nodules referred to community pulmonologists.materials and methodsthis was a multicenter , community - based , observational study of patients with intermediate - sized pulmonary nodule who were referred to pulmonologists. patients were categorized into three groups by the most invasive procedure performed during management , as follows : surveillance ( serial imaging ) , biopsy ( ct - guided transthoracic needle aspiration [ tna ] or bronchoscopy ) , or surgery ( including mediastinoscopy , video - assisted thorascopic surgery , and/or thoracotomy ) . patients were categorized into three groups by the most invasive procedure performed during management , as follows : surveillance ( serial imaging ) , biopsy ( ct - guided transthoracic needle aspiration [ tna ] or bronchoscopy ) , or surgery ( including mediastinoscopy , video - assisted thorascopic surgery , and/or thoracotomy ) . patients were categorized into three groups by the most invasive procedure performed during management , as follows : surveillance ( serial imaging ) , biopsy ( ct - guided transthoracic needle aspiration [ tna ] or bronchoscopy ) , or surgery ( including mediastinoscopy , video - assisted thorascopic surgery , and/or thoracotomy ) . multivariate logistic regression was performed to identify factors associated with the use of an invasive diagnostic procedure.resultof the 377 patients included, 283 ( 75% ) had a nodule that was benign, and 94 ( 25% ) had a malignant nodule. the overall accuracy of pet scanning was 74%, with a false - positive ( fp ) rate of 39% and a false - negative ( fn ) rate of 9%. the overall accuracy of pet scanning was 74%, with a false - positive ( fp ) rate of 39% and a false - negative ( fn ) rate of 9%. nodules measuring &gt; 11 to 15 mm ( n = 48 ) had fn and fp rates of 9% and 36%, respectively.conclusionsthe prevalence of cancer in patients with intermediate - sized nodules was 25%. the rate of surgical resection for benign disease varied from 9% to 23% in screening trials and surgical series.</p>
Metrics	<p>Rouge1: 45.58, Rouge2: 9.56, RougeL: 18.37, RougeLsum: 38.1, Summary length (tokens): 470</p>
Top 75% example (Sorted by rougeL)	
Document	<p>a total of 1,217 dead birds were shipped at 4c to the tropical medicine institute " pedro kouri " and identified by ornithology experts . brain , heart , and kidneys were removed and tested for wnv by using reverse transcription polymerase chain reaction ( rt - pcr ) ( 12 ) . briefly , rna was extracted by using the qiamp viral rna kit ( qiagen , inc . , valencia , ca , usa ) . primers wn212 ( 5'-ttgttgcttcctcttgcttgcttctt-3' ) and wn619c ( 5'-cagccgacagacatggcattcaatc-3' ) were used to detect viral rna . a second rt - pcr with primers wn9483 ( 5'-caactacggcctcaaacattccac-3' ) and wn97943 ( 5'-ggaacgtgcgtccaaatcacatc-3' ) was performed on the same rna preparation . serum specimens from horses in havana and havana province were tested for antibodies to wnv by using a competitive enzyme - linked immunosorbent assay ( elisa ) with monoclonal antibodies 2.1112 , g and 086c-1 as described by blitvich <i>et al .</i> we tested 210 serum specimens from horses collected as part of an infectious anemia study . the immunoglobulin m ( igm ) test was not performed because horses were never suspected of having wnv and did not have any history of suspected viral encephalitis or other illness or symptoms . an inhibition value &gt; 30% was used as the diagnostic criterion to identify flavivirus antibody ( table 1 ) . * wnv , west nile virus ; elisa , enzyme - linked immunosorbent assay ; pnt , plaque reduction neutralization test ; slev , saint louis encephalitis virus . the cuban health ministry and medical services division conducted surveillance for encephalitis of unknown origin in patients &gt; 30 years of age . serum and cerebrospinal fluid specimens were shipped at 4c to the tropical medicine institute " pedro kouri . " human sera were screened for wnv igm and igg by using commercial igm and igg elisa kits ( focus technologies , cypress , ca , usa ) according to manufacturer 's instructions . hemagglutination - inhibition ( hi ) tests were also undertaken with wnv and saint louis encephalitis virus ( slev ) antigen ( 14 ) . reactive serum samples were further tested ( pnt ) with wnv ( ny99 , ontario , canada , 2001 isolate ) , slev ( purton strain , american type culture collection catalog no . vt-1265 ) , and dengue virus ( dengue 2 , ng - c strain ) . pnt was performed to confirm wnv - specific antibody and was carried out as described previously ( 15 ) by using a neutral red double - overlay procedure . ho .</p>
Gold	<p>a surveillance system to detect west nile virus ( wnv ) was established in cuba in 2002 . wnv infection was confirmed by serologic assays in 4 asymptomatic horses and 3 humans with encephalitis in 2003 and 2004 . these results are the first reported evidence of wnv activity in cuba .</p>
Model	<p>we report the first evidence of antibodies to west nile virus ( wnv ) in horses and humans in cuba. antibodies were detected by plaque reduction neutralization test in 10 of 1,217 dead birds tested. these findings provide evidence that wnv and slev may co - circulate in cuba.</p>
Metrics	<p>Rouge1: 51.06, Rouge2: 17.39, RougeL: 31.91, RougeLsum: 46.81, Summary length (tokens): 64</p>
Top 95% example (Sorted by rougeL)	
Document	<p>intra - articular injections of corticosteroids have been used for several decades in the management of inflammatory and degenerative joint conditions when first - line conservative therapies such as rest , ice , and anti - inflammatory medications fail to provide adequate symptom relief . based in part on this long history of successful utilization coupled with the findings of several randomized controlled trials , consensus statements and meta - analyses have concluded that intra - articular corticosteroid injections provide short - term patient benefit and clinical efficacy for chronic knee pain.13 more recently , various injectable hyaluronic acid agents have become commercially available and have enjoyed widespread clinical acceptance as an effective treatment for knee osteoarthritis . these agents are indicated for the treatment of the pain associated with osteoarthritis of the knee in patients who have failed to respond adequately to conservative nonpharmacologic therapy and simple analgesics , eg , acetaminophen . traditionally , intra - articular injections have been performed using anatomical landmarks to identify the correct trajectory for needle placement . however , different anatomical - guided injection techniques have yielded inconsistent intra - articular needle positioning due , in large part , to the fact that the physician can not directly visualize the area of interest , and variations in anatomy are common . incorrect needle placement has been partially attributed to variable clinical outcomes.410 furthermore , inaccurate corticosteroid injections in the knee , for example , may result in post - injection pain , crystal synovitis , hemarthrosis , joint sepsis , and steroid articular cartilage atrophy , as well as systemic effects , such as fluid retention or exacerbation of hypertension or diabetes mellitus.1 therefore , identification of methods and proper training to aid in correct needle placement during these procedures is warranted . various imaging modalities can be used to improve the accuracy of intra - articular injections , including fluoroscopy , computed tomography , and magnetic resonance imaging . however , musculoskeletal ultrasound is one of the most practical because it is rapid , safe , relatively inexpensive , emits no ionizing radiation , and can be performed in the outpatient clinical setting.11,12 ultrasound utilizes high - frequency sound waves to visualize soft tissues and bony structures and is a f... </p>
Gold	<p>intra - articular corticosteroid and hyaluronic acid injections provide short - term symptom amelioration for arthritic conditions involving structural damage or degenerative changes in the knee . conventional palpation - guided anatomical injections frequently result in inaccurate needle placement into extra - articular tissue and adjacent structures . the purpose of this review was to determine the effect of ultrasound guidance on the accuracy of needle placement , clinical outcomes , and cost - effectiveness in comparison with anatomical landmark - guided intra - articular large joint injections , with particular emphasis on the knee . a total of 13 relevant studies were identified ; five studied the knee , seven studied the shoulder , one used both the knee and shoulder , and none studied the hip . ultrasound was used in seven studies ; the remaining studies utilized air arthrography , fluoroscopy , magnetic resonance arthrography , or magnetic resonance imaging . across all studies ( using all imaging modalities and all joints ) , needle placement accuracy ranged from 63% to 100% with ultrasound and from 39% to 100% with conventional anatomical guidance . imaging guidance improved the accuracy of intra - articular injections of the knee ( 96.7% versus 81.0% , p &lt; 0.001 ) and shoulder ( 97.3% versus 65.4% , p &lt; 0.001 ) . imaging guidance improved the accuracy of intra - articular injections of the knee ( 96.7% versus 81.0% , p &lt; 0.001 ) and shoulder ( 97.3% versus 65.4% , p &lt; 0.001 ) . imaging guidance improved the accuracy of intra - articular injections of the knee ( 96.7% versus 81.0% , p &lt; 0.001 ) and shoulder ( 97.3% versus 65.4% , p &lt; 0.001 ) . in particular , ultrasound guidance of knee injections resulted in better accuracy than did anatomical guidance ( 95.8% versus 77.8% , p &lt; 0.001 ) . in particular , ultrasound guidance of knee injections resulted in better accuracy than did anatomical guidance ( 95.8% versus 77.8% , p &lt; 0.001 ) . conclusionthe use of imaging guidance , in particular ultrasound , improves the accuracy of intra - articular injections in large joints , including the knee . furthermore , accurate ultrasound - guided intra - articular knee injections improve clinical outcomes and lower health care costs .</p>
Model	<p>objectivethe purpose of this review was to determine the effect of ultrasound guidance on the accuracy of needle placement, clinical outcomes, and cost - effectiveness in comparison with anatomical landmark - guided intra - articular large joint injections , with particular emphasis on the knee.methods:the peer - reviewed literature was searched using medline and relevant bibliographies published in english language journals through december 31, 2011 that compared the accuracy of intra - articular injections with anatomical landmark - guided intra - articular large joint injections , with particular emphasis on the knee.results:a total of 13 studies met the entry criteria. needle placement accuracy ranged from 63% to 100% with ultrasound and from 39% to 100% with conventional anatomical guidance . imaging guidance improved the accuracy of intra - articular injections of the knee ( 96.7% versus 81.0%, p &lt; 0.001 ) and shoulder ( 97.3% versus 65.4%, p &lt; 0.001 ) . imaging guidance improved the accuracy of intra - articular injections of the knee ( 96.7% versus 81.0%, p &lt; 0.001 ) and shoulder ( 97.3% versus 65.4%, p &lt; 0.001 ) . imaging guidance improved the accuracy of intra - articular injections of the knee ( 96.7% versus 81.0%, p &lt; 0.001 ) and shoulder ( 97.3% versus 65.4%, p &lt; 0.001 ) . in particular , ultrasound guidance of knee injections resulted in better accuracy than did anatomical guidance ( 95.8% versus 77.8%, p &lt; 0.001 ) . in particular , ultrasound guidance of knee injections resulted in better accuracy than did anatomical guidance ( 95.8% versus 77.8%, p &lt; 0.001 ) . conclusionthe use of imaging guidance , in particular ultrasound , improves the accuracy of intra - articular injections in large joints , including the knee . furthermore , accurate ultrasound - guided intra - articular knee injections improve clinical outcomes and lower health care costs.</p>
Metrics	<p>Rouge1: 62.21, Rouge2: 43.74, RougeL: 48.51, RougeLsum: 58.7, Summary length (tokens): 464</p>

Table 16: Examples of the PubMed dataset using the model pubmed-4096-512 base diverse