Towards Cross-Domain Transferability of Text Generation Models for Legal Text

Vinayshekhar Bannihatti Kumar* **Kasturi Bhattacharjee*** **Rashmi Gangadharaiah**
AWS AI Labs
{vinayshk,kastb,rgangad}@amazon.com

Abstract

Legalese can often be filled with verbose domain-specific jargon which can make it challenging to understand and use for non-experts. Creating succinct summaries of legal documents often makes it easier for user comprehension. However, obtaining labeled data for every domain of legal text is challenging, which makes cross-domain transferability of text generation models for legal text, an important area of research. In this paper, we explore the ability of existing state-of-the-art T5 & BART-based summarization models to transfer across legal domains. We leverage publicly available datasets across four domains for this task, one of which is a new resource for summarizing privacy policies, that we curate and release for academic research. Our experiments demonstrate the low cross-domain transferability of these models, while also highlighting the benefits of combining different domains. Further, we compare the effectiveness of standard metrics for this task and illustrate the vast differences in their performance.

1 Introduction

Legalese is often perceived to be an expert language containing jargon-filled text, which makes it difficult for non-experts to comprehend (Kumar et al., 2019; Bannihatti Kumar et al., 2020; Obar and Oeldorf-Hirsch, 2020). However, owing to recent regulations (Voigt and Von dem Bussche, 2017; Moukad, 1979) there is a shift in paradigm to make legal documents more accessible to non-domain experts. Summarizing such documents is a vital step in this direction. A few examples include summarization over legislative bills (Kornilova and Eidelman, 2019; Zhang et al., 2020; Narayan et al., 2021) and legal contracts like terms of service (Manor and Li, 2019a; Jain et al., 2021; Shukla et al., 2022). However, obtaining annotated data for every domain of legal text for this task is expensive and often infeasible. Thus, exploring the ability of text generation models to transfer across multiple legal domains is of importance, particularly for low resource domains for which knowledge transfer from domains with large amounts of annotated data could be beneficial. While there has been research on various tasks and aspects of legal text, such as summarization (Jain et al., 2021; Kornilova and Eidelman, 2019; Zhang et al., 2020; Narayan et al., 2021), question answering (Ravichander et al., 2019; Keymanesh et al., 2021) and title generation in privacy policies (Gopinath et al., 2020), transferability of generative models across legal domains has remained relatively understudied.

In this work, we explore the cross-domain transferability of state-of-the-art text generation models across *four* distinctly different legal domains. We use standard summarization metrics to measure their degree of transferability. Further, we compare the effectiveness of such metrics at capturing the summarization capability of these models, and demonstrate the differences thereof. Further, since summarization datasets are not available in the privacy policy domain, we curate and release an annotated dataset for further research.

Contributions of our work:

- We measure the extent of cross domain transferability of T5 & BART-based summarization models on 4 different legal domains. Our experiments demonstrate the advantages of a multi-domain model over a single-domain one.
- We create a dataset for privacy policy summarization to enable further research in this area[1].
- We illustrate the shortcomings of BERTScore (Zhang et al., 2019) and

* Equal contribution

[1] https://github.com/awslabs/summarization-privacy-policies

Proceedings of the Natural Legal Language Processing Workshop 2022, pages 111 - 118

	Train	Dev	Test	Context # of chars			Summary # of chars		
				Min	Max	Avg	Min	Max	Avg
BillSum	15159	3032	3269	5004	19997	10319	65	4966	1193
JRC-Acquis (en)	2026	2242	328	50	888444	13603	6	2382	209
Legal contracts	369	36	41	44	3922	407	19	328	92
Privacy Policies	20000	2000	2000	6	9222	793	6	1689	64

Table 1: Dataset statistics for all datasets. We show the varying nature of each dataset with statistics on the number of characters in context and summary. As observed, the mean number of characters differs to a large extent across each dataset.

BARTScore (Yuan et al., 2021) on cross-domain transferability and demonstrate that traditional metrics like ROUGE-L & METEOR are better for such an assessment pertaining to the legal domain.

2 Datasets

In order to study cross-domain transferability of generative models, we select four summarization datasets consisting of legal text of varying domains, each of which is described below.

BillSum: This dataset (Kornilova and Eidelman, 2019) consists of US congressional bills collected over a 25 year time-period (103rd-115th sessions of US Congress) ranging from 1993-2018 & summarized by the respective legislative counsel.

JRC-Acquis (en): This dataset introduced by Steinberger et al. (2006) is composed of the contents, political objectives of treaties, legislation, declarations, etc. pertaining to the member states of the EU. We focus on the English subset of the corpus for this paper. The task here is to summarize the paragraphs of the documents using their titles.

Legal contracts: Curated by Manor and Li (2019b), this dataset is composed of unilateral legal contracts such as terms of service, terms of use and licensing agreements. Instead of summarizing the entire document as a whole, manually curated summaries of each section are provided.

Privacy Policies: Privacy policies are legal documents that disclose ways in which a company collects and manages their user data. Each section of the privacy policy discusses various facets of user data management. While there has been work done to *summarize sections* of privacy policies (Gopinath et al., 2020), there is no open sourced dataset available for this task.

Privacy Policies Dataset Creation: We leverage the ∼1M English language privacy policy dataset (Amos et al., 2021) in order to **create & release** a dataset for section summarization. To that end, we sample a subset of 20K privacy policies at random, from which we randomly select 24K sections. The dataset created for section summarization consist of <Body, Title> pairs extracted from these sections. For more details, please refer to Appendix A.1.

The statistics of the train/dev/test split for these datasets is shown in Table 1. Table 4 (Appendix) contains examples from each dataset, thereby highlighting the domain differences between them.

3 Methodology & Experiments

We leverage pretrained seq2seq Transformer-based text generation models such as BART (Lewis et al., 2020) and T5 (Raffel et al., 2020) for our experiments. In order to measure cross-domain transferability of generative models for the four domains discussed in Section 2, we first conduct experiments in the ***single-domain*** setting, in which each seq2seq model is fine-tuned with the <context, summary> pairs from the training split of a *single* domain-specific dataset, and subsequently used to generate summaries for the test splits of each dataset, both in and out-of-domain. Cross-domain performance of these models helps determine their transferability across different domains.

Further, in order to compare with a scenario in which the text generation model learns from all domains and is thereby able to incorporate the domain differences during generation, we propose a ***multi-domain*** setting, in which we fine-tune the model with training data of all domains, and generate summaries for each of the four datasets. For more details on experimental settings, please refer to Appendix A.3. Standard metrics such as **ROUGE-L** (Lin, 2004), **METEOR** (Banerjee and Lavie, 2005), **BERTScore** (Zhang et al., 2019) and **BARTScore** (Yuan et al., 2021) are reported to measure model performance.

4 Results

In this section, we report quantitative (Table 3) and qualitative results (Table 2) of the single and multi-domain text generation models.

Single-domain: As is evident from the single domain results, cross-domain model performance is low for both BART-base & T5-base models, across all reported metrics. For instance, a BART-base

Model	Generated Summary	Reference Summary
BART-base *single-domain*, FT: **Legal contracts**, Test: **JRC-Acquis (en)**	the European Economic Community should have a journal of its own.	*Decision creating the 'Official Journal of the European Communities'*
BART-base *multi-domain*, FT: **all domains**, Test: **JRC-Acquis (en)**	58/578/EEC: Council Decision of 15 September 1958 on the creation of the Official Journal of the European Communities	
T5-base *single-domain*, FT: **JRC-Acquis (en)** Test: **Privacy Policies**	Non-members - We do not collect any Personal Data about you - however, we may automatically collect information about your visits, such as browsing patterns - to analyse, manage and develop	*WHAT DATA DO WE COLLECT AND HOW?*
T5-base *multi-domain*, FT: **all domains** Test: **Privacy Policies**	Personal Data we collect	

Table 2: Summaries generated by single & multi-domain T5 & BART based generative models. **FT** represents the data (domain) the model was *fine-tuned* on.

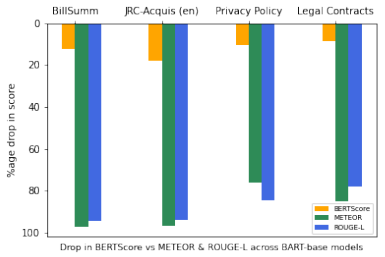
Model	Data for fine-tuning	Test Set	ROUGE-L	METEOR	BERTScore	BARTScore
		Single-domain				
BART-base	JRC-Acquis (en)	JRC-Acquis (en)	0.769	0.756	0.954	-1.771
		Legal contracts	0.099	0.077	0.830	-4.653
		BillSum	0.104	0.059	0.821	-3.855
		Privacy Policies	0.098	0.061	0.825	-6.138
	Legal contracts	Legal contracts	0.358	0.368	0.899	-3.266
		JRC-Acquis (en)	0.166	0.113	0.831	-4.967
		BillSum	0.065	0.036	0.820	-3.861
		Privacy Policies	0.116	0.085	0.833	-6.002
	BillSum	BillSum	0.343	*0.292*	0.883	-2.850
		JRC-Acquis (en)	0.21	0.308	0.839	-3.978
		Legal contracts	0.150	0.258	0.850	-3.900
		Privacy Policies	0.080	0.121	0.810	-5.480
	Privacy Policies	Privacy Policies	0.500	0.480	0.904	-4.140
		JRC-Acquis (en)	0.05	0.0264	0.788	-5.334
		Legal contracts	0.085	0.059	0.823	-4.410
		BillSum	0.020	0.009	0.778	-4.067
T5-base	JRC-Acquis (en)	JRC-Acquis (en)	0.756	0.756	0.955	-1.818
		Legal contracts	0.135	0.149	0.849	-4.077
		BillSum	0.161	0.102	0.842	-3.539
		Privacy Policies	0.133	0.116	0.829	-5.669
	Legal contracts	Legal contracts	0.277	0.307	0.885	-3.597
		JRC-Acquis (en)	0.210	0.165	0.839	-4.729
		BillSum	0.139	0.089	0.839	-3.651
		Privacy Policies	0.132	0.106	0.834	-5.893
	BillSum	BillSum	0.380	0.316	0.887	-2.752
		JRC-Acquis (en)	0.233	0.312	0.839	-3.954
		Legal contracts	0.159	0.262	0.856	-3.720
		Privacy Policies	0.09	0.131	0.817	-5.430
	Privacy Policies	Privacy Policies	0.456	0.450	0.897	-4.340
		JRC-Acquis (en)	0.113	0.054	0.794	-5.26
		Legal contracts	0.075	0.040	0.820	-4.510
		BillSum	0.062	0.020	0.800	-3.840
		Multi-domain				
BART-base	*All Domains Combined*	BillSum	**0.355**	**0.302**	**0.886**	**-2.817**
		JRC-Acquis (en)	**0.794**	**0.784**	**0.959**	**-1.628**
		Legal contracts	**0.387**	**0.396**	**0.902**	**-3.008**
		Privacy Policies	**0.513**	**0.503**	**0.907**	**-4.075**
T5-base	*All Domains Combined*	BillSum	**0.386**	**0.316**	**0.889**	**-2.743**
		JRC-Acquis (en)	**0.792**	**0.795**	**0.962**	**-1.603**
		Legal contracts	**0.351**	**0.388**	**0.898**	**-3.219**
		Privacy Policies	**0.497**	**0.484**	**0.903**	**-4.168**

Table 3: Model performance for single & multi-domain scenarios with **BART-base** & **T5-base** models across datasets.

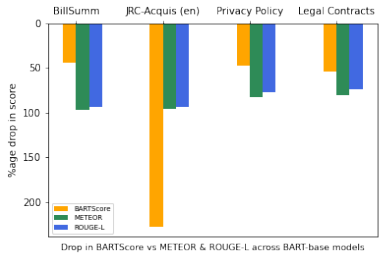
model trained on **JRC-Acquis (en)** yields 0.769 ROUGE-L score for text of the *same domain*, while achieving a much lower ROUGE-L score of 0.104 on *a different domain* (**BillSum**). A similar behavior is observed for **T5-base** as well. Here, a T5-base model trained on **Privacy policy** is able to obtain a METEOR score of 0.45 on a test set of the same domain, while a model trained on **Legal contracts** achieves 0.106 METEOR score on the same Privacy policy test set. Thus text generation models are observed to yield low cross-domain transferability for legal text.

Multi-domain: We observe the multi-domain T5 & BART models to yield better performance across each domain, in comparison to the single-domain setting. For instance, the METEOR score for the *best single-domain T5-base* model for **Privacy Policy** test set is 0.45, while the multi-domain T5 model is able to achieve 0.484 on the same test set. Similarly, the multi-domain BART-base model yields a ROUGE-L score of 0.794 on the **JRC-Acquis (en)** dataset, for which the corresponding best single-domain model performance is 0.769. This illustrates that it helps the model to learn from the domain differences of these datasets.

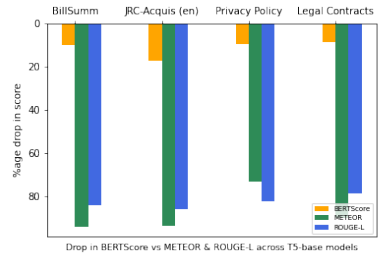
Comparing summarization metrics: An interesting observation is that the percentage drop in performance between the best and worst performing models, reflected via BERTScore & BARTScore is significantly less as compared to that obtained using other metrics such as METEOR ROUGE-L, for $14/16$ settings considered in this study. For instance, in Figures 1a & 1c, we consider, for each dataset, the best & the worst performing models



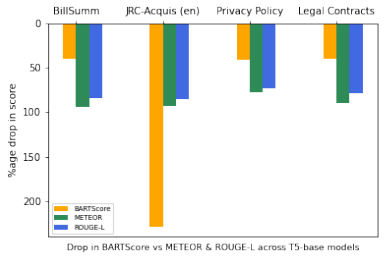
(a) BERTScore vs METEOR & ROUGE-L for BART models



(b) BARTScore vs METEOR & ROUGE-L for BART models



(c) BERTScore vs METEOR & ROUGE-L for T5 models



(d) BARTScore vs METEOR & ROUGE-L for T5 models

Figure 1: Percentage drop in BERTScore & BARTScore as compared to METEOR & ROUGE-L. Figures 1b & 1a provide a metrics comparison for the multi-domain model & the worst performing single-domain model w.r.t BARTScore & BertScore respectively. Similarly, Figures 1d & 1c illustrate the comparison for the corresponding T5-base counterparts.

based on BERTScore. For each pair of models, we plot the percentage drop in performance for BERTScore, and the corresponding METEOR & ROUGE-L scores. We observe that text overlap metrics like ROUGE-L and METEOR exhibit a significant drop in performance when compared to BERTScore. A similar trend is observed for BARTScore as well, which captures a lower drop in performance for 3 out of 4 datasets (Figures 1b & 1d). This illustrates that perhaps not all metrics are equally capable of capturing model performance adequately for summarization. Furthermore, upon manual investigation, we observe that the deterioration of quality of generated summaries is better reflected by ROUGE-L & METEOR, when compared to BERTScore and BARTScore.

Qualitative results Table 2 demonstrates model-generated summaries for a few of the single & multi-domain models. As is evident, the multi-domain models generate summaries that are closer to the reference, in each case.

5 Conclusion & Future Work

In this paper, we study the cross-domain transferability of neural text generation models across four different domains of legal text. We consider seq2seq model architectures such as BART & T5 and fine-tune them on datasets of specific domains. Based on standard generation metrics such as ROUGE, METEOR, BERTScore & BARTScore, we find such models to show a drop in performance for cross-domain settings. Further, our experiments demonstrate the benefits of combining different domains to train models for such tasks. Moreover, we observe some metrics to be more effective at capturing the differences in predicted and ground-truth summaries. We also curate & release a dataset on title generation for privacy policies for further research in this direction. In the future, we wish to explore text generation specific to legal text for low resource scenarios including zero and few-shot settings.

References

Ryan Amos, Gunes Acar, Eli Lucherini, Mihir Kshirsagar, Arvind Narayanan, and Jonathan Mayer. 2021. Privacy policies over time: Curation and analysis of a million-document dataset. In *Proceedings of the Web Conference 2021*, pages 2165–2176.

Satanjeev Banerjee and Alon Lavie. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics.

Vinayshekhar Bannihatti Kumar, Roger Iyengar, Namita Nisal, Yuanyuan Feng, Hana Habib, Peter Story, Sushain Cherivirala, Margaret Hagan, Lorrie Cranor, Shomir Wilson, et al. 2020. Finding a choice in a haystack: Automatic extraction of opt-out statements from privacy policy text. In *Proceedings of The Web Conference 2020*, pages 1943–1954.

Abhijith Athreya Mysore Gopinath, Vinayshekhar Bannihatti Kumar, Shomir Wilson, and Norman Sadeh. 2020. Automatic section title generation to improve the readability of privacy policies.

Deepali Jain, Malaya Dutta Borah, and Anupam Biswas. 2021. Summarization of legal documents: Where are we now and the way forward. *Computer Science Review*, 40:100388.

Moniba Keymanesh, Micha Elsner, and Srinivasan Parthasarathy. 2021. Privacy policy question answering assistant: A query-guided extractive summarization approach. *arXiv preprint arXiv:2109.14638*.

Anastassia Kornilova and Vlad Eidelman. 2019. Billsum: A corpus for automatic summarization of us legislation. *arXiv preprint arXiv:1910.00523*.

Vinayshekhar Bannihatti Kumar, Abhilasha Ravichander, Peter Story, and Norman Sadeh. 2019. Quantifying the effect of in-domain distributed word representations: A study of privacy policies. In *AAAI Spring Symposium on Privacy-Enhancing Artificial Intelligence and Language Technologies*.

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.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Laura Manor and Junyi Jessy Li. 2019a. Plain english summarization of contracts. *arXiv preprint arXiv:1906.00424*.

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

Rosemary Moukad. 1979. New york's plain english law. *Fordham Urb. LJ*, 8:451.

Shashi Narayan, Yao Zhao, Joshua Maynez, Gonçalo Simões, Vitaly Nikolaev, and Ryan McDonald. 2021. Planning with learned entity prompts for abstractive summarization. *Transactions of the Association for Computational Linguistics*, 9:1475–1492.

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

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(140):1–67.

Abhilasha Ravichander, Alan W Black, Shomir Wilson, Thomas Norton, and Norman Sadeh. 2019. Question answering for privacy policies: Combining computational and legal perspectives. *arXiv preprint arXiv:1911.00841*.

Bharti Shukla, Sonam Gupta, Arun Kumar Yadav, and Divakar Yadav. 2022. Text summarization of legal documents using reinforcement learning: A study. In *Intelligent Sustainable Systems*, pages 403–414. Springer.

Ralf Steinberger, Bruno Pouliquen, Anna Widiger, Camelia Ignat, Tomaz Erjavec, Dan Tufis, and Dániel Varga. 2006. The jrc-acquis: A multilingual aligned parallel corpus with 20+ languages. *arXiv preprint cs/0609058*.

Paul Voigt and Axel Von dem Bussche. 2017. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, 10(3152676):10–5555.

Weizhe Yuan, Graham Neubig, and Pengfei Liu. 2021. Bartscore: Evaluating generated text as text generation. *Advances in Neural Information Processing Systems*, 34:27263–27277.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. 2020. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In *International Conference on Machine Learning*, pages 11328–11339. PMLR.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

A Appendix

A.1 Privacy policy Dataset creation

In Algorithm 1 we describe the algorithm that we used for the creation of <Context, Summary> pairs in the case of privacy policies. Table 6 contains a few examples of the data. The data is further split into train/dev/test splits as specified in Table 1.

Algorithm 1 Algorithm for creation of privacy policy summarization corpus

$N \leftarrow NumberOfPolicies$
$i \leftarrow 0$
$title \leftarrow$ ""
$runningContent \leftarrow$ ""
$samples \leftarrow []$
while $i < N$ **do**
  $lines \leftarrow Policy_i$
  if $lines[0]$ is '#' or '**' **then** samples.append(<Body, Title>)
    $title \leftarrow lines[i]$
    $runningContent \leftarrow$ ""
  else
    $runningContent += lines[i]$
  end if
end while

A.2 Data Sources & Examples

The specific sources for obtaining the datasets are as follows:

1. BillSum: We obtained the data from HuggingFace's Datasets library.

2. JRC-Acquis (en):The data was downloaded from this link.

3. Legal contracts: The data was downloaded from this link.

In Table 4, we show different samples from all the datasets considered in this paper. We can see from the samples below that the 4 domains considered in this paper vary widely from one another.

A.3 Experimental Details

BART-base & **T5-base** model checkpoints are initialized from HuggingFace for each of the experiments and fine-tuned using the provided script. For the single-domain experiments, we conduct Hyper-parameter optimization(HPO) using the dev set of the corresponding dataset. In case of the multi-domain experiments, we use a dev set built by combining the dev splits of each dataset for HPO. The following are the best hyper-parameters for each of the models.

- **T5 & BART-base Single-domain (JRC-Acquis (en)):**learning rate: 5e-05, batch size: 32, optimizer: Adam, number of epochs: 3.0
- **BART-base Single-domain (Legal contracts):**learning rate: 5e-05, batch size: 32, optimizer: Adam, number of epochs: 3.0
- **T5-base Single-domain (Legal contracts):**learning rate: 5e-05, batch size: 16, optimizer: Adam, number of epochs: 3.0
- **T5 & BART-base Single-domain (BillSum) & Single-domain (Privacy Policies):**learning rate: 5e-05, batch size: 12, optimizer: Adam, number of epochs: 3.0
- **BART-base Multi-domain:** learning rate: 8e-05, batch size: 64, optimizer: Adam, number of epochs: 5.0
- **T5-base Multi-domain:** learning rate: 8e-05, batch size: 16, optimizer: Adam, epsilon=1e-08, number of epochs: 5.0

A.4 Qualitative Examples

In Table 5 we show the different summaries that our model produces along with the reference summaries.

Domain	Context	Summary
BillSum	SECTION 1. SHORT TITLE.This Act may be cited as the "Taxpayer Transparency Act of 2013".. 2. REQUIREMENTS FOR PRINTED MATERIALS AND ADVERTISEMENTS BY FEDERAL AGENCIES.(a) Identification of Funding Sources.–Each communication funded a Federal agency for advertising or educational purposes shall state–(1) in the case of a printed communication, including mass mailings, signs, and billboards, that the communication is printed and published at taxpayer expense; and(2) in the case of a communication transmitted through radio, television, the Internet, or any means other than the means referred to in paragraph (1), that the communication is produced and disseminated at taxpayer expense..... (A) means any mailing or distribution of 499 or more newsletters, pamphlets, or other printed matter with substantially identical content, whether such matter is deposited singly or in bulk, or at the same time or different times; and(B) does not include any mailing–(i) in direct response to a communication from a person to whom the matter is mailed; or(ii) of a news release to the communications media.(e) Source of Funds.–The funds used by a Federal agency to carry this Act shall be derived from amounts made available to the agency advertising or other communications regarding the programs and of the agency.	Taxpayer Transparency Act of 2013 - Requires each communication funded by a federal agency for advertising or educational purposes to clearly state: (1) in the case of a printed communication, including mass mailings, signs, and billboards, that the communication is printed and published at taxpayer expense; and (2) in the case of a communication transmitted through radio, television, or the Internet, that the communication is produced and disseminated at taxpayer expense. Requires any such printed communication, including e-mails, to be of sufficient size to be clearly readable, to be set apart from the other contents of the communication, and to be printed with a reasonable degree of color contrast between the background and the printed statement. Exempts from such requirements: (1) information in or relating to a solicitation for offers for a federal contract or applications or submissions of a bid or proposal for a federal grant or other means of funding under a federal program; and (2) advertisements for employment opportunities, not including advertising materials developed for use in recruiting and retaining personnel for the Armed Forces.
JRC-Acquis (en)	2006/C 252/02) The Minister for Economic Affairs of the Kingdom of the Netherlands hereby gives notice that an application has been received for authorisation to prospect for hydrocarbons in block P1 as indicated on the map appended as Annex 3 to the Mining Regulation (Mijnbouwregeling) (Government Gazette (Staatscourant) 2002, No 245). With reference to Article 3(2) of Directive 94/22/EC of the European Parliament and of the Council of 30 May 1994 on the conditions for granting and using authorisations for the prospection, exploration and production of hydrocarbons and the publication required by Article 15 of the Mining Act (Mijnbouwwet) (Bulletin of Acts and Decrees (Staatsblad) 2002, No 542), the Minister for Economic Affairs hereby invites interested parties to submit an application for authorisation to prospect for hydrocarbons in block P1. The Minister for Economic Affairs is the competent authority for the granting of authorisations. The criteria, conditions and requirements referred to in Articles 5(1), 5(2) and 6(2) of the Directive are set out in the Mining Act (Bulletin of Acts and Decrees 2002, No 542). Applications may be submitted during the 13 weeks following the publication of this notice in the Official Journal of the European Union and should be sent to the Minister for Economic Affairs, for the attention of the Director for the Energy Market, Bezuidenhoutseweg 30, The Hague, Netherlands, and marked "personal". Applications submitted after the expiry of this period will not be considered. A decision on the applications will be taken not later than twelve months after this period has expired. Further information can be obtained from the following telephone number: (31-70) 379 72 98	Notice inviting applications for authorisation to prospect for hydrocarbons in block P1 of the Dutch continental shelf
Legal contracts	we may also automatically collect device specific information when you install access or use our services. this information may include information such as the hardware model operating system information app version app usage and debugging information browser information ip address and device identifiers.	the service may use tracking pixels web beacons browser fingerprinting and or device fingerprinting on users.
Privacy Policies	We take reasonable steps to ensure that personal information we process is accurate, complete, and current by using the most recent information provided to us.	Data Integrity

Table 4: Example of context and summary for the four datasets considered in this paper. For presentation purpose, we have reduced the size of context for BillSum.

Model	Generated Summary	Reference Summary
BART-base *single-domain* FT: **Legal contracts** Test: **JRC-Acquis (en)**	THE COMMISSION OF THE EUROPEAN COMMUNITIES, Having regard to the Treaty establishing the European Community, having regard to Council Regulation 90/425/EEC of 26 June 1990 concerning veterinary and zootechnical checks applicable in intra-Community trade.	98/168/EC: Commission Decision of 17 February 1998 amending Decision 93/70/EEC on codification for the message 'Animo' to include certain types of mammalian waste products (Text with EEA relevance)
BART-base *multi-domain* FT: **all domains** Test: **JRC-Acquis (en)**	98/766/EC: Commission Decision of 17 February 1998 amending Decision 93/70/EEC laying down the codes to be used for animals and animal products (Text with EEA relevance)	
T5-base *single-domain* FT: **JRC-Acquis (en)** Test: **Privacy Policies**	Complaints lodged in the summer of 2006 - Information Commissioner's Office - Data protection supervisory authority	Right to complain
T5-base *multi-domain* FT: **all domains** Test: **Privacy Policies**	Right to lodge a complaint	

Table 5: More Summaries generated by single & multi-domain T5 & BART based generative models. **FT** represents the data (domain) the model was fine-tuned on.

Context	Summary
We may share aggregated Non-Identifying Information and we may otherwise disclose Non-Identifying Information (including, without limitation, Hashed Information) to third parties. We do not share your Personal Information with third parties for those third parties' marketing purposes unless we first provide you with the opportunity to opt-in to or opt-out of such sharing. We may also share the information we have collected about you, including Personal Information, as disclosed at the time you provide your information, with your consent, as otherwise described in this Privacy Policy, or in the following circumstances	INFORMATION SHARING AND DISCLOSURE
You have the right at any time to access any Personal Data we hold about you, and where you feel the Personal Information that we hold is not correct, to request that the Personal Information is corrected.0 You also have the right to have your Personal Information deleted. All of the Personal Information, along with other data collected (as noted in the table above) is information that you can access, amend or delete by logging into your SOFTWARE112 Account. If you have any questions about accessing, correcting, amending, or deleting your information then you can contact us.	How can I Access, Amend, Correct and/or Delete my Personal Data?
Occasionally, at our discretion, we may include or offer third party products or services on our website. These third party sites have separate and independent privacy policies. We therefore have no responsibility or liability for the content and activities of these linked sites. Nonetheless, we seek to protect the integrity of our site and welcome any feedback about these sites.	Third party links
The information we collect from you will be used by Microsoft and its controlled subsidiaries and affiliates to enable the features you are using and provide the service(s) or carry out the transaction(s) you have requested or authorized.0 It may also be used to analyze and improve Microsoft products and services. In order to offer you a more consistent and personalized experience in your interactions with Microsoft, information collected through one Microsoft service may be combined with information obtained through other Microsoft services. We may also supplement the information we collect with information obtained from other companies. For example, we may use services from other companies that enable us to derive a general geographic area based on your IP address in order to customize certain services to your geographic area. Except as described in this statement, personal information you provide will not be transferred to third parties without your consent. We occasionally hire other companies to provide limited services on our behalf, such as packaging, sending and delivering purchases and other mailings, answering customer questions about products or services, processing event registration, or performing statistical analysis of our services. We will only provide those companies the personal information they need to deliver the service, and they are prohibited from using that information for any other purpose. Microsoft may access or disclose information about you, including the content of your communications, in order to: (a) comply with the law or respond to lawful requests or legal process; (b) protect the rights or property of Microsoft or our customers, including the enforcement of our agreements or policies governing your use of the services; or (c) act on a good faith belief that such access or disclosure is necessary to protect the personal safety of Microsoft employees, customers, or the public.0 We may also disclose personal information as part of a corporate transaction such as a merger or sale of assets. Information that is collected by or sent to Microsoft by WebPI may be stored and processed in the United States or any other country in which Microsoft or its affiliates, subsidiaries, or service providers maintain facilities. Microsoft abides by the safe harbor framework as set forth by the U.S. Department of Commerce regarding the collection, use, and retention of data from the European Union, the European Economic Area, and Switzerland.	Collection and Use of Your Information
You will find links to other websites on our websites to keep you really well informed. We do not have any influence upon the design and the content of these external websites.	Links to other websites
The advertisements diffused on our site are proposed by third companies. They may use data on users' visits to target content that may be of interest to them	Advertisements
Most of the content on this website is ours and subject to our copyright, but some of the content is owned by others. For instance where we link to other websites. You may: * use and enjoy the content for your own personal information purposes; and * share our posts on social media. If you want to use the content for any other purpose, please ask our permission first. You can contact us at info@thesouthafrican.com.	Content on this website

Table 6: Example of context and summary for the domain of privacy policies.