# PolicyQA: A Reading Comprehension Dataset for Privacy Policies

**Wasi Uddin Ahmad**[*]
University of California, Los Angeles
wasiahmad@ucla.edu

**Jianfeng Chi**[*]
University of Virginia
jc6ub@virginia.edu

**Yuan Tian**
University of Virginia
yuant@virginia.edu

**Kai-Wei Chang**
University of California, Los Angeles
kwchang.cs@ucla.edu

## Abstract

Security and privacy policy documents are long and verbose. A question answering (QA) system can assist users in finding the information that is relevant and important to them. Prior studies in this domain frame the QA task as retrieving the most relevant text segment or a list of sentences from the policy document given a question. On the contrary, we argue that providing users with a short text span extracted from policy documents better helps them as it reduces the burden of searching the target information from a lengthy text segment. In this paper, we present PolicyQA, a dataset that contains 25,017 reading comprehension style examples curated from an existing corpus of 115 website privacy policies. PolicyQA provides 714 human-annotated questions written for a wide range of privacy practices. We present two strong neural baselines and rigorous analysis to reveal the advantages and challenges offered by PolicyQA.

Website: Amazon.com

Information You Give Us: We receive and store any information you enter on our Web site or give us in any other way. Click here to see ...

Question. How do you collect my information?

information you enter on our Web site

Promotional Offers: Sometimes we send offers to selected groups of Amazon.com customers on behalf of other businesses. When we do this, we do not give that business your name and address. If you do not want to receive such offers, ...

Question. Is my information shared with others?

we do not give that business your name and address

Table 1: Question-answer pairs that we collect from OPP-115 (Wilson et al., 2016a) dataset. The *Evidence* spans are highlighted in color and they are used to form the question-evidence pairs.

## 1 Introduction

Security and privacy policy documents describe how an entity collects, maintains, uses, and shares users' information. Users need to read the privacy policies of the websites they visit or the mobile applications they use and know about their privacy practices that are pertinent to them. However, prior works suggested that people do not read privacy policies because they are long and complicated (McDonald and Cranor, 2008), and confusing (Reidenberg et al., 2016). Hence, giving users access to a question answering system to search for answers from long and verbose policy documents can help them better understand their rights.

In recent years, we have witnessed noteworthy progress in developing question answering (QA) systems with a colossal effort to benchmark high-quality, large-scale datasets for a few application domains (e.g., Wikipedia, news articles). However, annotating large-scale QA datasets for domains such as security and privacy is challenging as it requires expert annotators (e.g., law students). Due to the difficulty of annotating policy documents at scale, the only available QA dataset is PrivacyQA (Ravichander et al., 2019) on privacy policies for 35 mobile applications.

An essential characteristic of policy documents is that they are well structured as they are written by following guidelines set by the policymakers. Besides, due to the homogeneous nature of different entities (e.g., Amazon, eBay), their privacy policies have a similar structure. Therefore, we can exploit the document structure (meta data) to form examples from existing corpora. In this paper, we

[*] Equal contribution.

|  | PolicyQA (This work) | PrivacyQA |
|---|---|---|
| Source | Website privacy policies | Mobile application privacy policies |
| # Policies | 115 | 35 |
| # Questions | 714 | 1,750 |
| # Annotations | 25,017 | 3,500 |
| Question annotator | Domain experts | Mechanical Turkers |
| Form of QA | Reading comprehension | Sentence selection |
| Answer type | A sequence of words | A list of sentences |

Table 2: Comparison of PolicyQA and PrivacyQA.

present PolicyQA, a reading comprehension style question answering dataset with 25,017 question-passage-evidence triples associated with text segments from privacy policy documents. PolicyQA consists of 714 questions on 115 website privacy policies and is curated from an existing corpus, OPP-115 (Wilson et al., 2016a). Table 1 presents a couple of examples from PolicyQA.

In contrast to PrivacyQA (Ravichander et al., 2019) that focuses on extracting long text spans from policy documents, we argue that highlighting a shorter text span in the document facilitates the users to zoom into the policy and identify the target information quickly. To enable QA models to provide such short answers, PolicyQA provides examples with an average answer length of 13.5 words (in comparison, the PrivacyQA benchmark has examples with an average answer length of 139.6 words). We present a comparison between PrivacyQA and PolicyQA in Table 2.

In this work, we present two strong neural baseline models trained on PolicyQA and perform a thorough analysis to shed light on the advantages and challenges offered by the proposed dataset. The data and the implemented baseline models are made publicly available.[1]

## 2 Dataset

PolicyQA consists question-passage-answer triples, curated from OPP-115 (Wilson et al., 2016a). OPP-115 is a corpus of 115 website privacy policies (3,792 segments), manually annotated by skilled annotators following an annotation scheme predefined by domain experts. The annotation scheme is composed of 10 data practice categories (e.g., *First Party Collection/Use*, *Third Party Sharing/Collection*, *User Choice/Control* etc.). The data practices are further categorized into a set of practice attributes (e.g., *Personal Information Type*, *Purpose*, *User Type* etc.). Each practice attribute is associated with a predefined set of values. In the Appendix (in Table 9), we list all the attributes under the *First Party Collection/Use* category.

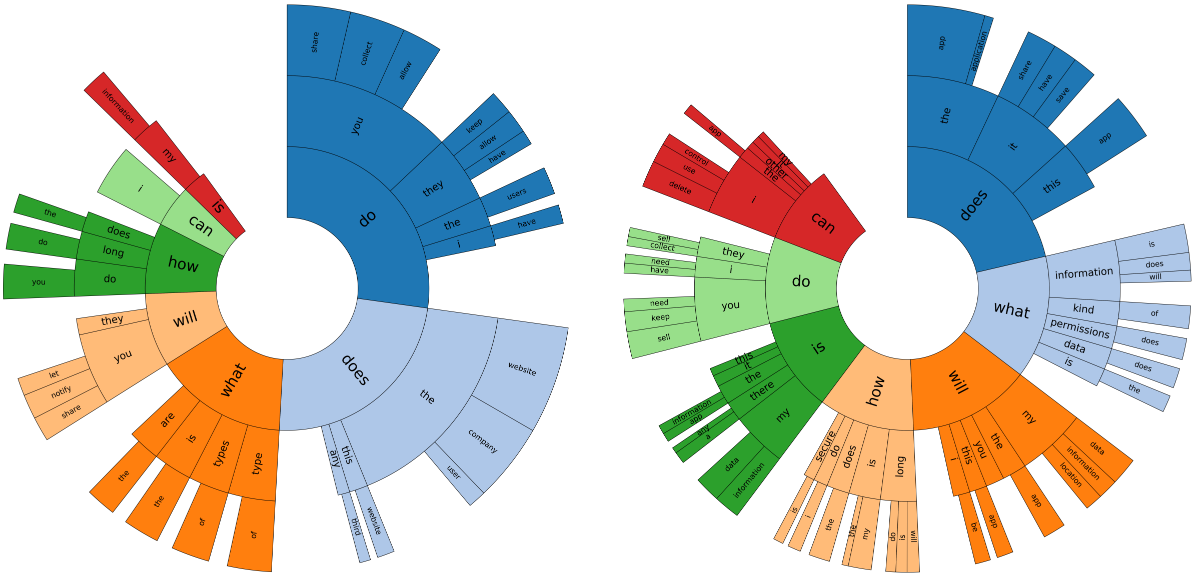
|  |
|---|
| "Practice": First Party Collection/Use<br>"Attribute": Purpose<br>"value": "Additional service/feature"<br>"startIndexInSegment": 360<br>"endIndexInSegment": 387<br>"selectedText": "responding to your requests" |
| "Practice": Third Party Sharing/Collection<br>"Attribute": Third Party Entity<br>"value": "Unnamed third party"<br>"startIndexInSegment": 573<br>"endIndexInSegment": 596<br>"selectedText": "Third-Party Advertisers" |

Table 3: Sample span annotations from OPP-115 associated with a segment of *Amazon.com* privacy policy.

In total, OPP-115 contains 23,000 data practices, 128,000 practice attributes, and 103,000 annotated text spans. Each text span belongs to a policy segment, and OPP-115 provides its character-level start and end indices. We provide an example in Table 3. We use the annotated spans, corresponding policy segments, and the associated {*Practice, Attribute, value*} triple to form PolicyQA examples. Next, we describe the question annotation process.

**Question annotations.** Two skilled annotators manually annotate the questions. During annotation, the annotators are provided with the triple {Practice, Attribute, value}, and the associated text span. For example, given the triple {*First Party Collection/Use*, *Personal Information Type*, *Contact*} and the associated text span "*name, address, telephone number, email address*", the annotators created questions, such as, (1) *What type of contact information does the company collect?*, (2) *Will you use my contact information?*, etc.

[1] https://github.com/wasiahmad/PolicyQA



(a) PolicyQA (This work)

(b) PrivacyQA

Figure 1: Distribution of trigram prefixes of questions in (a) PolicyQA and (b) PrivacyQA.

| Privacy Practice | Proportion | Example Question From PolicyQA |
|---|---|---|
| First Party Collection/Use | 44.4 % | why do you collect my data? |
| Third Party Sharing/Collection | 34.1 % | do they share my information with others? |
| Data Security | 2.2 % | do you use encryption to secure my data? |
| Data Retention | 1.7 % | how long they will keep my data? |
| User Access, Edit and Deletion | 3.1 % | will you let me access and edit my data? |
| User Choice/Control | 11.0 % | what use of information does the user choice apply to? |
| Policy Change | 1.9 % | how does the website notify about policy changes? |
| International and Specific Audiences | 1.5 % | what is the company's policy towards children? |
| Do Not Track | 0.1 % | do they honor the user's do not track preference? |

Table 4: OPP-115 categories to the questions from the PolicyQA dataset.

For a specific triple, the process is repeated for 5-10 randomly chosen samples to form a list of questions. We randomly assign a question from this list to the examples associated with the triple that were not chosen during the sampling process. In total, we considered 258 unique triples[2] and created 714 individual questions. In Table 4), we provide an example question for each practice category. Also, we compare the distribution of questions' trigram prefixes in PolicyQA (Figure 1a) with PrivacyQA (Figure 1b). It is important to note that, PolicyQA questions are written in a generic fashion to become applicable for text spans, associated with the same practice categories. Therefore, PolicyQA questions are less diverse than PrivacyQA.

We split OPP-115 into 75/20/20 policies to form training, validation, and test examples, respectively. Table 5 summarizes the data statistics.

## 3 Experiment

**Baselines.** PolicyQA frames the QA task as predicting the answer span that exists in the given policy segment. Hence, we consider two existing neural approaches from literature as baselines for PolicyQA. The first model is **BiDAF** (Seo et al., 2017) that uses a bi-directional attention flow mechanism to extract the evidence spans. The second baseline is based on **BERT** (Devlin et al., 2019) with two linear classifiers to predict the boundary of the evidence, as suggested in the original work.

**Implementation.** PolicyQA has a similar setting as SQuAD (Rajpurkar et al., 2016). Therefore, we pre-train the QA models using their default settings

[2] We exclude the text spans with practices labeled as "Other" and the values labeled as "Unspecified".

| Dataset | Train | Valid | Test |
|---|---|---|---|
| # Examples | 17,056 | 3,809 | 4,152 |
| # Policies | 75 | 20 | 20 |
| # Questions | 693 | 568 | 600 |
| # Passages | 2,137 | 574 | 497 |
| Avg. question length | 11.2 | 11.2 | 11.2 |
| Avg. passage length | 106.0 | 96.6 | 119.1 |
| Avg. answer length | 13.3 | 12.8 | 14.1 |

Table 5: Statistics of the PolicyQA dataset.

| Fine-tuning | SQuAD Pre-training | Valid | | Test | |
|---|---|---|---|---|---|
| | | EM | F1 | EM | F1 |
| BiDAF | | | | | |
| ✗ | ✗ | 25.1 | 52.3 | 22.0 | 48.0 |
| ✗ | ✓ | 26.7 | 53.7 | 23.3 | 49.5 |
| ✓ | ✗ | 27.9 | 57.2 | 24.4 | 52.8 |
| BERT-base | | | | | |
| ✗ | ✗ | 30.5 | 59.4 | 28.1 | 55.6 |
| ✗ | ✓ | 30.5 | 60.2 | 28.0 | 56.2 |
| ✓ | ✗ | **32.8** | 60.9 | 28.6 | **56.6** |
| ✓ | ✓ | 32.7 | **61.2** | **29.5** | **56.6** |

Table 6: Performance of baselines on PolicyQA. The bold face values indicate the best performances.

on the SQuAD dataset. Besides, we consider leveraging unlabeled privacy policies in fine-tuning the models, as noted below.

- **Fine-tuning.** We train word embeddings using *fastText* (Bojanowski et al., 2017) based on a corpus of 130,000 privacy policies (137M words) collected from apps on the Google Play Store.[3] These word embeddings are used as fixed word representations in BiDAF while training on PolicyQA. Similarly, to adapt BERT to the privacy domain, we first fine-tune BERT using masked language modeling (Devlin et al., 2019) based on the privacy policies and then train on PolicyQA.

- **No fine-tuning.** In this setting, we use the publicly available *fastText* (Bojanowski et al., 2017) embeddings with BiDAF, and the BERT model is not fine-tuned on those privacy policies.

We adopt the default model architecture and optimization setup for the baseline methods. We detail the hyper-parameters in Appendix (in Table 10).

**Evaluation.** Following Rajpurkar et al. (2016), we use *exact match* (EM) and *F1 score* to evaluate the model's accuracy.

| BERT Size | Valid | | Test | |
|---|---|---|---|---|
| | EM | F1 | EM | F1 |
| Tiny | 21.0 | 47.1 | 15.5 | 39.9 |
| Mini | 26.5 | 55.2 | 22.8 | 49.8 |
| Small | 28.4 | 57.2 | 24.6 | 52.3 |
| Medium | **31.1** | 59.1 | 25.2 | 53.5 |
| Base | 30.5 | **59.4** | **28.1** | **55.6** |

Table 7: Performance of different sized QA models.

| | $\lvert\text{ans}\rvert$ | EM | F1 |
|---|---|---|---|
| Third Party Sharing/Collection | 9.3 | 35.0 | 60.2 |
| First Party Collection/Use | 10.1 | 28.3 | 55.7 |
| Data Retention | 10.6 | 29.1 | 55.9 |
| User Choice/Control | 11.0 | 24.3 | 53.2 |
| User Access, Edit and Deletion | 12.2 | 21.6 | 51.5 |
| Policy Change | 14.6 | 43.4 | 67.7 |
| Do Not Track | 30.9 | 37.5 | 69.2 |
| Data Security | 34.6 | 24.4 | 54.3 |
| Intl. and Specific Audiences | 52.8 | 5.3 | 43.1 |

Table 8: Test performance breakdown of BERT-base model for privacy practice categories, sorted by the average answer length as indicated by $\lvert\text{ans}\rvert$.

### 3.1 Results and Analysis

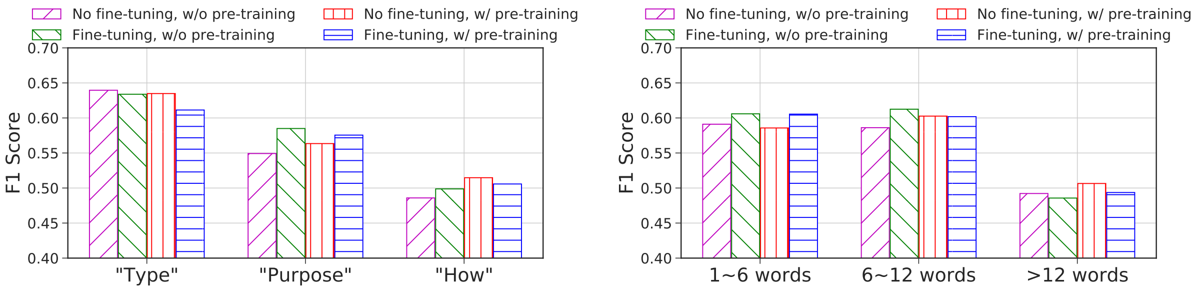
The experimental results are presented in Table 6. Overall, the BERT-base methods outperform the BiDAF models by 6.1% and 7.6% in terms of EM and F1 score (on the test split), respectively.

**Impact of fine-tuning.** Table 6 demonstrates that the fine-tuning step improves the downstream task performance. For example, BERT-base performance is improved by 0.5% and 1.0% EM and F1 score, respectively, on the test split. This result encourages to train/fine-tune BERT on a larger collection of security and privacy documents.

**Impact of SQuAD pre-training.** Given a small number of training examples, it is challenging to train deep neural models. Hence, we pre-train the extractive QA models on SQuAD (Rajpurkar et al., 2016) and then fine-tune on PolicyQA. The additional pre-training step improves performance. For example, in *no fine-tuning* setting, BiDAF, and BERT-base improve the performance by 1.5% and 0.6% F1 score, respectively (on the test split).

**Impact of model size.** We experiment with different sized BERT models (Turc et al., 2019) and the results in Table 7 shows that the performance improves with increased model size. The results also indicate that PolicyQA is a challenging dataset, and hence, a larger model performs better.

[3] We thank the authors of (Harkous et al., 2018) for sharing the 130,000 privacy policies.



(a)

(b)

Figure 2: BERT-base model's performance on (a) the three most frequent attributes of "First Party Collection/Use" and "Third Party Sharing/Collection" practice categories, and (b) questions with different answer lengths.

**Analysis.** We breakdown the test performance of the BERT-base method to examine the model performance across practice categories. The results are presented in Table 8. We see the model performs comparably on the three most frequent categories (comprise 89.5% of the total examples).

We further analyze the performance on questions associated with (1) the top three frequent attributes for the two most frequent practice categories, and (2) different answer lengths. The results are presented in Figure 2a and 2b. Our findings are (1) shorter evidence spans (e.g., evidence spans for *Personal Information Type* questions) are easier to extract than longer spans; and (2) SQuAD pretraining helps more in extracting shorter evidence spans. Leveraging diverse extractive QA resources may reduce the length bias and boost the QA performance on privacy policies.

## 4 Related Work

The *Usable Privacy Project* (Sadeh et al., 2013) has made several attempts to automate the analysis of privacy policies (Wilson et al., 2016a; Zimmeck et al., 2019). Noteworthy works include identification of policy segments commenting on specific data practices (Wilson et al., 2016b), extraction of Opt-Out choices, and their provisions in policy text (Sathyendra et al., 2016; Mysore Sathyendra et al., 2017), and others (Bhatia and Breaux, 2015; Bhatia et al., 2016). Kaur et al. (2018) used a keyword-based technique to compare online privacy policies. Natural language processing (NLP) techniques such as text alignment (Liu et al., 2014; Ramanath et al., 2014), text classification (Harkous et al., 2018; Zimmeck et al., 2019; Wilson et al., 2016a) and question answering (Shvartzshanider et al., 2018; Harkous et al., 2018; Ravichander et al., 2019) has been studied in prior works to facilitate privacy policy analysis.

Among the question answering (QA) methods, Harkous et al. (2018) framed the task as retrieving the most relevant policy segment as an answer, while Ravichander et al. (2019) presented a dataset and models to answer questions with a list of sentences. In comparison to the prior QA approaches, we encourage developing QA systems capable of providing precise answers by using PolicyQA.

## 5 Conclusion

This work aims to utilize an existing resource to build a question answering (QA) dataset that can contribute to the development of QA systems in the security and privacy domain that can have a sizeable real-word impact. We propose PolicyQA, a corpus of privacy policy questions curated from 115 annotated website privacy policies. We present two strong neural baseline methods and provide thorough ablation analysis to reveal important considerations that affect answer span prediction. In our future work, we want to explore how transfer learning can benefit question answering for the security and privacy domain.

## Acknowledgments

This work was supported in part by National Science Foundation Grant OAC 1920462.

## References

Jaspreet Bhatia and Travis D Breaux. 2015. Towards an information type lexicon for privacy policies. In *2015 IEEE eighth international workshop on requirements engineering and law (RELAW)*, pages 19–24. IEEE.

Jaspreet Bhatia, Morgan C Evans, Sudarshan Wadkar, and Travis D Breaux. 2016. Automated extraction of regulated information types using hyponymy relations. In *2016 IEEE 24th International Requirements Engineering Conference Workshops (REW)*, pages 19–25. IEEE.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G Shin, and Karl Aberer. 2018. Polisis: Automated analysis and presentation of privacy policies using deep learning. In *27th {USENIX} Security Symposium ({USENIX} Security 18)*, pages 531–548.

Jasmin Kaur, Rozita A Dara, Charlie Obimbo, Fei Song, and Karen Menard. 2018. A comprehensive keyword analysis of online privacy policies. *Information Security Journal: A Global Perspective*, 27(5-6):260–275.

Fei Liu, Rohan Ramanath, Norman Sadeh, and Noah A. Smith. 2014. A step towards usable privacy policy: Automatic alignment of privacy statements. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 884–894, Dublin, Ireland.

Aleecia M McDonald and Lorrie Faith Cranor. 2008. The cost of reading privacy policies. *Isjlp*, 4:543.

Kanthashree Mysore Sathyendra, Shomir Wilson, Florian Schaub, Sebastian Zimmeck, and Norman Sadeh. 2017. Identifying the provision of choices in privacy policy text. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2774–2779.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. SQuAD: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Austin, Texas. Association for Computational Linguistics.

Rohan Ramanath, Fei Liu, Norman Sadeh, and Noah A. Smith. 2014. Unsupervised alignment of privacy policies using hidden Markov models. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 605–610.

Abhilasha Ravichander, Alan W Black, Shomir Wilson, Thomas Norton, and Norman Sadeh. 2019. Question answering for privacy policies: Combining computational and legal perspectives. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4947–4958.

Joel R Reidenberg, Jaspreet Bhatia, Travis D Breaux, and Thomas B Norton. 2016. Ambiguity in privacy policies and the impact of regulation. *The Journal of Legal Studies*, 45(S2):S163–S190.

Norman Sadeh, Alessandro Acquisti, Travis D Breaux, Lorrie Faith Cranor, Aleecia M McDonald, Joel R Reidenberg, Noah A Smith, Fei Liu, N Cameron Russell, Florian Schaub, et al. 2013. The usable privacy policy project. *Technical report, Technical Report, CMU-ISR-13-119*.

Kanthashree Mysore Sathyendra, Florian Schaub, Shomir Wilson, and Norman Sadeh. 2016. Automatic extraction of opt-out choices from privacy policies. In *2016 AAAI Fall Symposium Series*.

Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. 2017. Bidirectional attention flow for machine comprehension. In *International Conference on Learning Representations3*.

Yan Shvartzshanider, Ananth Balashankar, Thomas Wies, and Lakshminarayanan Subramanian. 2018. RECIPE: Applying open domain question answering to privacy policies. In *Proceedings of the Workshop on Machine Reading for Question Answering*, pages 71–77, Melbourne, Australia. Association for Computational Linguistics.

Iulia Turc, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Well-read students learn better: The impact of student initialization on knowledge distillation. *arXiv preprint arXiv:1908.08962*.

Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N. Cameron Russell, Thomas B. Norton, Eduard Hovy, Joel Reidenberg, and Norman Sadeh. 2016a. The creation and analysis of a website privacy policy corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1330–1340.

Shomir Wilson, Florian Schaub, Rohan Ramanath, Norman Sadeh, Fei Liu, Noah A Smith, and Frederick Liu. 2016b. Crowdsourcing annotations for websites' privacy policies: Can it really work? In *Proceedings of the 25th International Conference on World Wide Web*, pages 133–143. International World Wide Web Conferences Steering Committee.

Sebastian Zimmeck, Peter Story, Daniel Smullen, Abhilasha Ravichander, Ziqi Wang, Joel Reidenberg, N Cameron Russell, and Norman Sadeh. 2019. Maps: Scaling privacy compliance analysis to a million apps. *Proceedings on Privacy Enhancing Technologies*, 2019(3):66–86.

| Attribute | Values |
|---|---|
| Does/Does Not | Does; Does Not |
| Collection Mode | Explicit; Implicit; Unspecified |
| Action First-Party | Collect on website; Collect in mobile app; Collect on mobile website; Track user on other websites; Collect from user on other websites; Receive from other parts of company/affiliates; Receive from other service/third-party (unnamed); Receive from other service/third-party (named); Other; Unspecified |
| Identifiability | Identifiable; Aggregated or anonymized; Other; Unspecified |
| Personal Information Type | Financial; Health; Contact; Location; Demographic; Personal identifier; User online activities; User profile; Social media data; IP address and device IDs; Cookies and tracking elements; Computer information; Survey data; Generic personal information; Other; Unspecified |
| Purpose | Basic service/feature; Additional service/feature; Advertising; Marketing; Analytics/Research; Personalization/Customization; Service Operation and Security; Legal requirement; Merger/Acquisition; Other; Unspecified |
| User Type | User without account; User with account; Other; Unspecified |
| Choice Type | Dont use service/feature; Opt-in; Opt-out link; Opt-out via contacting company; First-party privacy controls; Third-party privacy controls; Browser/device privacy controls; Other; Unspecified |
| Choice Scope | Collection; Use; Both; Unspecified |

Table 9: The attributes and their values for the *First Party Collection/Use* data practice category. We do not consider the data practices associated with "Unspecified" values.

| Model | Hyper-parameter | Value | Model | Hyper-parameter | Value |
|---|---|---|---|---|---|
| BiDAF | dimension | 300 | BERT | $d_{model}$ | 768 |
| | rnn_type | LSTM | | num_heads | 12 |
| | num_layers | 1 | | num_layers | 12 |
| | hidden_size | 300 | | $d_{ff}$ | 3072 |
| | dropout | 0.2 | | dropout | 0.2 |
| | optimizer | Adam | | optimizer | BertAdam |
| | learning rate | 0.001 | | learning rate | 0.00003 |
| | batch size | 16 | | batch size | 16 |
| | epoch | 15 | | epoch | 5 |

Table 10: Hyper-parameters used in our experiments.

| Value | Example Question From PolicyQA |
|---|---|
| Collect on website | Do you collect my information on your website? |
| Collect in mobile app | Will you collect my data if I use your phone app? |
| Collect on mobile website | How do you collect data when I use my mobile? |
| Track user on other websites | Do they track users' activities on other websites? |
| Collect from user on other websites | Does the website collect my info on other websites? |
| Receive from other parts of company/affiliates | Do you collect my information from your affiliates? |
| Receive from other service/third-party (unnamed) | Does the website obtain my data from others? |
| Receive from other service/third-party (named) | Who provides you my data? |
| Other | How do you receive data from users? |

Table 11: Examples questions from PolicyQA for the "Action First-Party" attribute under the *First Party Collection/Use* data practice category.