Social Fairness, Accountability and Transparency of the Data Economy: Using Machine Learning to Combat the Emptiness of Privacy Policies

Przemysław Pałka  
Fellow in Private Law, Yale Law School  
przemyslaw.palka@yale.edu

Roozbeh Yousefzadeh  
Postdoctoral Fellow, Yale University  
roozbeh.yousefzadeh@yale.edu

Abstract

Online consumers do not read privacy policies. These policies outline the ways in which consumer data will be collected and used. Certain common usages - personalization of ads, prices or content - have long been identified as potential venues of algorithmic unfairness, lack of accountability and transparency. In this paper we draw the reader’s attention to a related problem, which we call social unfairness and non-transparency. We demonstrate that privacy policies of online apps and services are written in an essentially "empty" language - full of qualifiers like "we might," abstract terms like "third parties," and open-ended phrases like "among others." As a result, even a consumer with expert knowledge is unable to predict what will be done with what her data, and the imbalance of rights and duties strongly favors the data holders, to the detriment of consumers.

We argue that machine learning tools can be used to combat the "emptiness" of language of privacy policies, by increasing the text-processing power of individual consumers and consumer organizations. Further, once the social practice changes and policies start containing more actual information, machine learning can be used to process these documents with high speed and accuracy. To substantiate this claim, we review the existing literature and conduct out own experiment. We have tagged 15 privacy policies of services in competitive markets (jogging, food delivery and dating apps) and demonstrate that information derived from topic modeling in conjunction with a classification model allows one to detect a rigidly defined type of information with accuracy as high as 90 percent. This finding, given the small size of the initial sample, provides good prospects for future research. Hopefully, more and more applications empowering individual consumers and civil society will soon be able to follow.

1 Introduction

Online consumers are bound by countless contracts they neither read nor understand [4,32]. Every website we visit and app we use makes us accept its terms of service (“ToS”) and privacy policy (“PP”). These documents determine the rules governing consumer data collection, sharing and usage. Through the “acceptance” of ToS and PPs consumers not only agree to their data being gathered and analyzed, but also consent to being subjected to data-driven, machine learning-powered commercial practices. These include personalization of targeted ads [13,51], of displayed content [11] or of prices (“price discrimination”) [6]. These areas have long been recognized as venues of potential algorithmic unfairness and discrimination [21,32]. To tackle the technical side of the problem, numerous contributions aimed at increasing the fairness, accountability and transparency (“FAT”) of algorithms have been made in the recent years [11,24,38]. One should note, however, that the unfairness, non-transparency and lack of accountability in question stem not only from biased data
and opacity of algorithms; they result also from the unfair consumer contracts constituting the basis for data collection and usage. As the data used to train consumer-facing algorithms stems from the consumer activity online, to ensure overall fairness and transparency of these systems we need to guarantee the socio-legal fairness on top the algorithmic one. These two, though in practice often occurring jointly, in theory are independent of one another (see table below). And machine learning can be deployed to combat the socio-legal unfairness [25, 26].

Table 1: Algorithmic & socio-legal unfairness

<table>
<thead>
<tr>
<th>Legally fair data practices</th>
<th>Legally unfair data practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithmically fair data practices</td>
<td></td>
</tr>
<tr>
<td>Individuals fully understand and accept what data about them is being collected and how it will be used; the collection and usage occurs in a way ensuring that data is not biased (or de-biased) and trained algorithms deployed in a way ensuring algorithmic fairness.</td>
<td>Individuals do not understand or do not accept what data about them is being collected and how it will be used; yet the collection and usage occurs in a way ensuring that data is not biased (or de-biased) and trained algorithms deployed in a way ensuring algorithmic fairness.</td>
</tr>
<tr>
<td>Algorithmically unfair data practices</td>
<td></td>
</tr>
<tr>
<td>Individuals fully understand and accept what data about them is being collected and how it will be used; the collection and usage occurs in a way replicating or creating bias, and trained algorithms are used in an unfair, discriminatory manner.</td>
<td>Individuals do not understand or do not accept what data about them is being collected and how it will be used; and on top of that the collection and usage occurs in a way replicating or creating bias, and trained algorithms are used in an unfair, discriminatory manner.</td>
</tr>
</tbody>
</table>

In this paper we focus on one aspect of the socio-legal unfairness of ToS and PPs, namely the problem of emptiness of privacy policies. By "emptiness" we mean both: 1) consumer’s inability to predict what companies will do with her data and 2) blanket permissions for companies to collect essentially any data they want and use it for very broad ranges of purposes. We demonstrate how even an expert reader, with time and resources, will not be able to fully comprehend what data will be collected about her and how it will be used, based on the privacy policies currently used by companies (in the summer of 2020). As case studies, we analyze 15 privacy policies of apps in competitive markets (jogging, food deliver and dating apps). The privacy policies under analysis contain numerous qualifiers like “we might,” “sometimes;” abstract and general terms like “business partners,” “third parties,” “improve our services;” paired with phrases indicating open-endedness of enumerations, like “including” or “for example.’ As a result, the problem consumers potentially trying to read these privacy policies face is not only that they are long and complicated [2, 56, 59] but also that they are empty. The information is simply not there. Hence, consumers cannot be deemed to make informed choices about the services they use and data practices they accept, and the overall practice cannot be deemed socially fair or transparent. This state of affairs persists despite directly violating the data protection laws in the European Union (the GDPR), and arguably violating the spirit of the “notice and choice” model in the United States (see below, Section 2: Legal Background). As has been argued for in the literature, merely changing the law is insufficient to modify the companies’ practice - what is necessary is to empower consumers and their organisations with tools increasing their factual data processing capacities [25, 54]. This is where machine learning itself becomes useful.

First, we argue that machine-learning-powered tools can be used to automatically detect unclear, open-ended and conditional statements. This will be helpful for consumer organizations (often with limited resources) wishing to raise social awareness, rebuke companies, and/or potentially notify the regulators. We argue that exposing the scale of the problem is necessary to increase the social discontent, and that this discontent, properly channeled, can be a tool in the quest of having corporations become much more transparent about their data practices. In this paper we present initial quantitative evidence regarding the scale of emptiness of privacy policies, discuss the challenges in developing these kinds of systems, and argue for changes in both the contractual practice and in the law.
Second, we argue that once the privacy policies actually do contain exhaustive information about data practices (the information about the details all types of data collected, all entities with whom it is shared, and all purposes for which it is used) machine learning can be employed to build tools making sense of this data. Indeed, from the consumers’ point of view, all the ToS and PPs she accepts are themselves “big data” – too much to read and make sense of by a human mind \[34\]. However, just as corporations use machine learning to extract value from enormous data collections, consumers and civil society could have access to the tools enabling them to do the same. As a matter of fact, several projects aimed at building such tools already exist \[10, 5, 27, 14\]. In this paper we provide an additional piece of evidence that, as long as the task is clearly defined and the piece of information is present in the text, existing ML approaches can be a useful tool for automated text classification and information retrieval. We conduct an experiment on 15 privacy policies of online services in competitive sectors (jogging apps, food delivery apps and dating apps) and show that information derived from topic modeling in conjunction with a classification model allows one to detect a rigidly defined type of information with accuracy as high as 90 percent.

As a civil society, we are not far from the world in which every consumer can have an app on their phone, automatically analyzing contents of privacy policies and preparing personalized disclosures \[10\]. We are not far from the world in which every consumer advocacy organization has tools enabling it to scan, summarize and/or evaluate thousands of privacy policies. What we are missing to get to that world is not (only) technology; it is the lack of data. The problem is that privacy policies, as of today, are essentially empty.

2 Legal Background

In this section we discuss: (i) what we mean by formal and material unfairness of consumer contracts, including of privacy policies; (ii) the legal status of the “emptiness” problem. We further explain how the notion of "emptiness" is formalized for the purposes of measurement, and what types of clauses in PPs we are trying to teach the classifiers to detect.

2.1 Unfair Consumer Contracts and Privacy Policies in the EU and the US

Consumer contracts can be materially unfair or formally unfair. The former concerns the contents of contractual clauses; the question of whether the rights and duties stemming from a contract unfairly favor one party (the business) over the other (the consumer). The latter concerns the form of the contract and the circumstances of its conclusion; the questions of whether consumer could have understood what she agreed to, given the language used, amount of time she had, urgency of the matter, etc. This general distinction and categories have different specific legal meanings in different jurisdictions but can be useful as analytical tool, not least because, in some form, they can usually be found in various legal systems.

For example, in the European Union, both material and formal unfairness are regulated by the Directive 93/13 on Unfair Terms in Consumer Contracts\[3\]. The Directive defines a contractual term as materially unfair if "contrary to the requirement of good faith, it causes a significant imbalance in the parties’ rights and obligations arising under the contract, to the detriment of the consumer." (art. 3.1.) Such terms are invalid ex lege. Regarding the form, the Directive stipulates that "terms must always be drafted in plain, intelligible language. Where there is doubt about the meaning of a term, the interpretation most favourable to the consumer shall prevail." (art. 5) Moreover, to combat the effects of formal unfairness, another legal act - Directive 2011/83 on Consumer Rights\[2\] - gives consumers "14 days to withdraw from a distance or off-premises contract, without giving any reason." (art. 9).

In the United States these matters are governed not by legislation, but by the common law unconscionability doctrine. A contract will be considered unconscionable if it has been concluded in circumstances depriving one party of meaningful choice, and when its terms unreasonably favor the other party \[28\]. According to § 2–302 of the Uniform Commercial Code: "If the court as a matter of law finds the contract or any clause of the contract to have been unconscionable at the time it was


made the court may refuse to enforce the contract, or it may enforce the remainder of the contract without the unconscionable clause, or it may so limit the application of any unconscionable clause as to avoid any unconscionable result." Arguably, both as a matter of doctrine as a matter of practice, the threshold of nonenforceability is higher in the United States than in the European Union; but both legal systems share the commitments to procedural and substantive fairness in consumer contracts. This holds across various sectors of the economy.

Privacy policies are a peculiar type of contracts given that - as a matter of law - they play several roles at the same time. In other words: legally speaking, they are more than just contracts. On the one hand, they are transparency tools, providing information to consumers. Using them might be required by certain privacy/data protection laws, e.g. the GDPR in the EU (art.12-14), the CCPA in California or the Federal COPPA, or by the consumer protection agencies, for example the Federal Trade Commission in the US. If one looks at privacy policies from this perspective, not having one (or using one that does not meet the substantive or formal standards) would be a violation of an administrative requirement, or an unfair/deceptive trade practice, respectively. On the other hand, privacy policies are contracts containing promises made by companies to consumers, in exchange for rights of companies to collect and use data in the ways in which they describe in PPs. Admittedly, looking at privacy policies as contracts comes much more naturally to the American lawyers than to Europeans. However, even in the EU the realization of the promise-making function of the PPs is slowly taking hold.

Consequently, a comprehensive account of what a "fair" privacy policy looks like would need to account for both the contractual doctrine, the consumer law and the relevant privacy/data protection laws. The answer, arguably, differs across jurisdictions, not only between the US and the EU, but also between various states of each Union. However, on the general level, a privacy policy should be "fair" both materially (by not unreasonably/significantly favoring one party over the other) and formally (by being written in a way that enables a consumer to understand what she is about to agree to, and to exercise a meaningful choice regarding whether to do so or not.) We argue that, as an empirical matter in 2020, many privacy policies could be deemed unfair along both dimensions. In this paper we concentrate the pervasiveness of the "empty" language, exemplified by some types of blanket permissions creating significant imbalance of power, to the detriment of consumer.

2.2 Features of Unfairness Analyzed in this Study

Emptiness of privacy policies cuts across both formal and material contractual unfairness. Consider the following example, from the privacy policy of a jogging app Strava:

"We may also engage service providers to collect information about your use of the Services over time on our behalf so that we or they may promote Strava or display information that may be relevant to your interests on the Services or other websites or services."

First, the conditional verb "may" indicates that Strava believes to have a right to engage in such practices, but provides no way of knowing whether they will actually do so, unlike verbs "will" or "do" would (see below). Second, the clause is full of general and abstract terms, with ambiguous meaning. What exactly is "information about my use of Services?" Who are "service providers?" What is meant by "information that may be relevant to my interests?" What "other websites or services?" Notice how, based solely on this one sentence, the company acquires a right to share essentially any information gathered in connection with consumers’ "use of the services" with any other company engaging in personalisation of ads, of search results, displayed content, etc. There is no way for a consumer to know what she agrees to (procedural unfairness) and the blanket right given to the company clearly creates an imbalance of power (potentially material unfairness).

---

4 California Consumer Privacy Act of 2018, Assembly Bill No. 375 CHAPTER 55 An act to add Title 1.81.5 (commencing with Section 1798.100) to Part 4 of Division 3 of the Civil Code, relating to privacy.
To formalize the notion of "emptiness" in a way grounded in law, we relied on the European Guidelines on Transparency used by the European Data Protection Board. This document specifies how a privacy policy should be formulated in order to be compliant with the GDPR. We chose it for two reasons: (i) it directly applies to the companies directing their activities at consumers in the EU, whether domestic or foreign (thereby applying to many American online services as well); (ii) it contains the most specified instruction that allowed us to tag privacy policies in an unambiguous manner. We should note that what follows below does not directly apply in the United States, but contains a type of a standard the American lawmaker could, one day, consider. The EDPD writes the following:

The information should be concrete and definitive; it should not be phrased in abstract or ambivalent terms or leave room for different interpretations. In particular the purposes of, and legal basis for, processing the personal data should be clear.

Poor Practice Examples

The following phrases are not sufficiently clear as to the purposes of processing:

- “We may use your personal data to develop new services” (as it is unclear what the “services” are or how the data will help develop them);
- “We may use your personal data for research purposes (as it is unclear what kind of “research” this refers to); and
- “We may use your personal data to offer personalised services” (as it is unclear what the “personalisation” entails).

Good Practice Examples

- “We will retain your shopping history and use details of the products you have previously purchased to make suggestions to you for other products which we believe you will also be interested in” (it is clear that what types of data will be processed, that the data subject will be subject to targeted advertisements for products and that their data will be used to enable this);
- “We will retain and evaluate information on your recent visits to our website and how you move around different sections of our website for analytics purposes to understand how people use our website so that we can make it more intuitive” (it is clear what type of data will be processed and the type of analysis which the controller is going to undertake); and
- “We will keep a record of the articles on our website that you have clicked on and use that information to target advertising on this website to you that is relevant to your interests, which we have identified based on articles you have read” (it is clear what the personalisation entails and how the interests attributed to the data subject have been identified).

Language qualifiers such as “may”, “might”, “some”, “often” and “possible” should also be avoided.

Note how these guidelines pertain solely to the language of the privacy policy, regardless of what the contents of the clauses are. In the following section we explain how we have transformed these guidelines into a tagging instruction, as well as what the results of our initial empirical study have been.

Further, given the omnipresence of empty language in the privacy policies, we chose three specific kinds of clauses that we distinguished based on content of the created entitlements: clauses giving companies rights to sell data as elements of corporate transactions; rights to combine the collected data with data from other sources; and rights to use and share the collected data in order to vindicate private rights of the companies or others. In the next section we discuss the tagging instruction and the examples of each.

3 Corpus

In this study, we analyzed 15 privacy policies of apps in three competitive markets:

---

7 Article 29 Working Party Guidelines on transparency under Regulation 2016/679, adopted on 29 November 2017, as last Revised and Adopted on 11 April 2018, 17/EN WP260 rev.01
1. Jogging: Strava Running and Cycling, Adidas Running, Runkeeper (Asics), Map My Run (Under Armour), and Nike+ Run Club;
2. Food Delivery: Grubhub (also Seamless), Uber Eats, DoorDash (also Caviar), Instacart, and Postmates;

We chose these documents for two material reasons. First, these are all competitive markets, where consumers have a reasonable choice between various options, and so one could expect that some level of competition regarding data practices exists or could materialize. Second, the kind of data generated by these apps can be very sensitive. For example, jogging apps allow users to upload data about their height, weight and age, and further track their athletic performance. This kind of information, once paired with info about consumers’ dietary choices, can constitute a pretty detailed view of one’s health-related habits. Hence, one could imagine that questions about with whom this data will be shared, or with what data from other sources it will be combined, are of interests to consumers. This is not even to mention the type of sensitive information collected by dating apps.

3.1 General Emptiness of Language

Upon this corpus, we set off to tag the privacy policies along two dimensions: general emptiness of language, and three specific issues. The first task that we set out to realize was to verify whether privacy policies under consideration contain instances of such language. In particular, we specified three kinds of tags, based on the EDPD Guidliness cited in the previous section:

- <qua>…</qua> – “qualifiers” – for sentences including words rendering future behavior of the data controller uncertain. Those include:
  - “we may/might/can” instead of “we will/are/do”
  - “sometimes/possibly”
- <op>…</op> – “open ended” – for sentences including phrases indicating open-endedness of the catalogue. For example:
  - “including”
  - “for example”/”such as”
- <ab>…</ab> – “abstract” – contain general and abstract terms referring to an unspecified category of entities, data or processing purposes. Those include:
  - “business partners”/”third parties”/”advertising partners”
  - “usage data”
  - “improve our services”/”advertising purposes”

Consider, as an example, a clause from the Grubhub privacy policy: We may also work with third-party partners to employ technologies, including statistical modeling tools, that permit us to recognize and contact you across multiple devices.

This one clause exemplifies all three types of problems defined above. First, the conditional verb makes it impossible to know whether the company actually does this or not. Second, it contains an abstract term of “third-party partners,” without specifying who those are. And third, it provides an example of “technologies” used by those unidentified partners, leaving a door open for other “similar” activities.

After having tagged three privacy policies (Strava, Nike, Under Armour) we realized that the pervasiveness of the "empty" language far exceeds our expectations. In the corpus of 796 sentences, 120 were tagged as “qualified,” 209 as “open-ended” and 265 as “abstract.” This amounts to 15 percent, 26 percent and 33 percent, respectively (note that some sentences where tagged with two or three tags). As these language was detected in the critical clauses stipulating what type of data would be collected, how will it be used, and with whom it will be shared, the policies were deemed completely empty. Moreover, it seemed that with this size of corpus - 15 policies - given the heterogeneity of the tagged sentences and their sheer amount, the ML-classifier won’t be able to pick them up at any

---

8 For a press report on this subject, see: Bridget Read, How Much Do Your Dating Apps Know About You?, VOGUE, September 27, 2017, available at: https://www.vogue.com/article/dating-apps-privacy
satisfactory rate. Hence, at this stage of the research, we decided to simply report the quantitative result and moved to the second task.

Figure 1 shows the share of tagged sentences in the privacy policies.

![Figure 1: Percentage of sentences tagged with issues explained in Section 3.1. Note that some sentences are tagged with multiple tags, but the figure does not demonstrate the overlap of the tags. The untagged (yellow) piece has no overlap with other pieces.](image)

### 3.2 Specific Kinds of Blanket Rights for Companies

To provide further evidence that machine learning can be used to successfully retrieve information from the privacy policies (and thereby aid in increasing the socio-legal fairness) as long as the task is clear and the information is actually there, we moved on to creating a second set of tags, chosen because a very limited amount of clauses ended up fitting these categories. We created the following tags:

- **<tab>…</tab>** – “transaction/acquisition/bankruptcy” – clauses giving companies rights to share/use personal data as an element of “corporate transactions” – sales of assets, merger/acquisition, restructuring/reorganization, dissolution/bankruptcy etc. Examples:
  - **Hinge**: We may transfer your information if we are involved, whether in whole or in part, in a merger, sale, acquisition, divestiture, restructuring, reorganization, dissolution, bankruptcy or other change of ownership or control.
  - **Postmates**: [We may share your information with third parties in the following cases:] while negotiating or in relation to a business transaction, such as a merger, sale of assets, or bankruptcy;

- **<pre>…</pre>** – “private rights enforcement” – clauses giving companies rights to share/use personal data to enforce their private rights, or rights of third parties, when not mandated by law (so not following a subpoena/responding to a warrant, but just a private action). Examples:
  - **Strava**: We may also retain, preserve or disclose your information if we determine that this is reasonably necessary or appropriate (…) to prevent or detect violations of our Terms of Service or fraud or abuse of Strava or its members, or to protect our operations or our property or other legal rights, including by disclosure to our legal counsel and other consultants and third parties in connection with actual or potential litigation.
  - **Grubhub**: We disclose personal information to third parties, such as legal advisors and law enforcement, as required by law or subpoena or if we reasonably believe that such action is necessary to (…) enforce our Terms of Use or to protect the security or integrity of our Services; (c) detect, suppress or prevent fraud or reduce credit risk and collect debts owed to us; and/or (d) exercise or protect the rights, property, or personal safety of Grubhub, our visitors, or others.

- **<agg>…</agg>** - “aggregation/combination” – clauses giving companies rights to aggregate/combine their consumers’ data with data from other sources. Examples:
  - **Adidas**: We may also combine this information with other information we collect as you interact with our brand across apps, social media and marketing messages we send you. Aggregating data allows adidas to update and correct the information contained in our database and to provide you with product recommendations and special offers.
– **Uber:** Uber may combine the data collected from these sources with other data in its possession.

We used these instruction to tag the 15 privacy policies mentioned above. As a result, we created a corpus of 420 sentences out of 3,967 sentences in the 15 documents. The set was tagged by two people independently, and then a cross-comparison, aimed at ensuring consistency, was run. The word clouds below illustrate the words occurring in each training set.

![Word Clouds](image)

Figure 2: Word cloud for the (a) Sharing of Data in Corporate Transactions, (b) Use of Data for Private Rights Enforcement, and (c) Rights to Aggregate and Combine Data tags, described in Section 3.2. We can see clear distinction between the common words for each tag, demonstrating that each tag is targeting a narrow and specific issue. At the same time, we see that all the privacy policies have issues concerning these tags.

In the following section, we discuss the methodology used to train classification models to detect these terms, as well as discuss the results of the experiment aimed at verifying how well the machine will fare regarding the clause detection.

### 4 Machine Learning Methodology and Experiments

We first consider previous machine learning approaches with respect to privacy policies. Then, we analyze all the 15 privacy policies with respect to the three issues explained in Section 3.2.

#### 4.1 Previous ML Approaches about Privacy Policies

Machine learning has been deployed before with respect to privacy policies. Notably, Wilson et al. annotated a combination of 115 privacy policies and investigated the potential of annotating them automatically with tools such as support vector machines, logistic regression, etc.

Sadeh et al. suggested a method to semi-automatically extract key features from privacy policies and to effectively communicate them to users. Tesfay et al. also proposed using machine learning to summarize lengthy privacy policies into condensed notes, and Costante et al. proposed a method to label the privacy policies with various privacy categories. These methods intend to allow the users to make informed decisions about privacy policies, however, as we saw in numerous examples, privacy policies are written mostly in vague and open-ended languages and moreover, they often require the users to avoid using the services, unless they agree with the entirety of privacy policies. So, summarizing such privacy policies for individual users does not necessarily help users on the individual-level.

Here, we first use machine learning, in an unsupervised way, in order to study the systematic deficiencies in the privacy policies. Once we understand the specific patterns in the data, we extend our analysis in a supervised way and show that machine learning can be advantageous in automatic detection of of particular kinds of clauses in privacy policies.
4.2 Bigrams

Clearly, the sequence of words matter in text analysis. Using $n$-gram analysis we can study the common sequences of different ($n$) size that words appear together. Here, we consider bigrams which corresponds to word sequences of size 2. Figure 3 shows the bigram derived from the combination of texts that are tagged with <tab>, <pre>, and <agg>.

![Bigram Diagram]

Figure 3: Bigram derived from all the three tags of (1) Sharing of Data in Corporate Transactions, (2) Use of Data for Private Rights Enforcement, and (3) Rights to Aggregate and Combine Data.

This analysis provides the information as to which combination of words are common in the tagged sentences. For example, we see that words "combine" and "information" frequently appear as a sequence in the tagged sentences. Similarly, words "merge" and "sale" appear together frequently.

4.3 Topic Modeling

In natural language processing, topic modeling is an unsupervised statistical approach that aims to find abstract topics in a set of documents by finding latent structures in them [44]. In other words, this method analyzes the documents and automatically identifies clusters of words that are related and tend to appear together. From this perspective, topic modeling can be viewed as a clustering method, and each of the identified clusters in the documents would correspond to a topic [12][19]. For example, one could extract the top topic models from the publications of a major newspaper. This could lead to topics such as Politics, Economy, Science, Art, etc.

For computation, we use the Latent Dirichlet Allocation (LDA) [9] method which is one the common methods for discovering topic models. This method relies on a hierarchical Bayesian model and it considers the arrangements of words in the documents.

As our first step, we combine all the tagged text associated with our three tags and feed them as the input to the LDA algorithm. Figure 4 shows the first 3 topic models we obtain from the combined text. Interestingly, each obtained model corresponds to one of the tags. Note that the algorithm does not know that the combined text is associated with our three tags, rather it views them as a bunch of sentences. So, the fact that LDA extracts our main three topics from the combined text tells us that the issues/tags that we have defined have negligible overlap in their scope and keywords. Hence, each of our tags corresponds to a specific topic detectable by natural language processing. This is highly desirable from the machine learning perspective and also from the practical viewpoint, because analysis of each topic (e.g., sharing of users’ information in corporate transactions vs rights to aggregate and combine data) can possibly lead to a separate set of remedies for policy makers.

---

11 In other words, bigram is a special case of n-gram, where $n = 2$. 

---
Figure 4: Topic models learned from the bigrams of combined set of tagged portion of all privacy policies. Interestingly, the top 3 models are corresponding to each of the three specific tags we have considered in Section 3.2. Topic model 1 is clearly related to Rights to Aggregate and Combine Data <agg>. Topic model 2 is related to Rights to Sale Data in Corporate Transactions <tab>, and finally, topic model 3 is related to Use of Data for Private Rights Enforcement <pre>. This demonstrates each of these three tags are specific in scope and have negligible overlap, i.e., they are well-defined in mathematical language.

This separation holds up even we disregard the word sequences and compute the top three topic models using the bag of words for all the text associated with these three tags, as shown in Figure 5.

Figure 5: Top 3 topic models learned from the bag of words of the combined set of tagged portion of all privacy policies. The association of topic models with our tags still holds, when we disregard the word sequences.

The top keywords associated with each of our three tags are common among all the 15 privacy policies. Moreover, those keywords are rarely used for other purposes. For example, the keywords "Acquisition", "Merger", and "Bankruptcy" are highly correlated with the tag <tab> while they generally do not appear for text not associated with this tag. In the next section, we leverage these structural relationships to build a classification model that can label each sentence in privacy policies.

4.4 Classification Model for Automated Tagging

We propose a computational method to build a classification model and then use it to tag the sentences in unseen privacy policies. Specifically, we use 4 tags (<tab>,<pre>,<agg>,and untagged) to label all the 3,967 sentences in the 15 privacy policies. For training, we use only a subset of 12 privacy policies, and then test the model on the remaining 3 policies.

Our computational method can be summarized as the following:

1. Break each policy into individual sentences.
2. Label each sentence based on the tags associated with it. If there is no tag associated with it, label it as untagged.
3. Pick 12 privacy policies as the training set and the remaining 3 as the testing set.
4. Create a bag of words from the training set, save it as $\mathcal{B}$.
5. Compute the Term Frequency–Inverse Document Frequency (tf-idf) matrix for training and testing set, using the vocabulary in $\mathcal{B}$, derived from training set.
6. Train a multi-class error-correcting output codes (ECOC) model on the training set.
7. Label the sentences in the testing set.

We repeat this experiment for all the distinct combinations of choosing 3 policies out of the 1512 and we achieve nearly perfect (above 90%) accuracy in labeling all the sentences. The high accuracy seems promising, but we note that our dataset is relatively small. In order to establish that our approach is practical, we plan to test our model on more privacy policies in the near future.

In this model, we use the top-1 classification, so generally the model will label each sentence with only one of the tags. However, in cases where the score for the top-1 class is relatively close to the score for the top-2 class, we can interpret the model's classification as a double tag.

The tagging with our method described above can be followed up by further expert review of the tagged text to remove the false positives. This could be very beneficial in practice, if incidents of false negatives are negligible. In order to achieve that, we need to analyze a larger number of privacy policies, using more sophisticated tools such as tensor decomposition.

4.5 Tensor Decomposition

Here, we briefly mention tensor decomposition as a future direction of research to analyze the patterns in large number of privacy policies. Tensor decomposition has broad applications in machine learning, especially to extract patterns and structures in large datasets, including text [20, 18].

We specifically plan to use Higher Order Generalized Singular Value Decomposition (HO-GSVD) which is a spectral decomposition method [23]. Using this method, one can decompose n number of (full column rank) matrices, \( D_i, i \in \{1, 2, \ldots, n\} \), as

\[
D_i = U_i \Sigma_i V, \forall i \in \{1, 2, \ldots, n\}, \tag{1}
\]

where, \( U_i \)'s, are normalized left basis vectors, \( \Sigma_i \)'s are the singular values, and \( V \) consists of normalized orthogonal right basis vectors [33, 35].

In our analysis, each \( D_i \) would contain the parts of privacy policies that are tagged with a specific topic/issue. For example, \( D_1 \) could refer to text associated with , and \( D_1 \) could refer to text associated with .

The \( V \) is common for the decomposition of all the datasets and it is orthogonal. Hence, the vectors in \( V \) can generally be viewed as the patterns present in all \( D_i \)'s. The singular values in the decomposition are positive scalars organized in diagonal elements of \( \Sigma_i \). Each \( D_i \) has its own set of singular values and the comparison of singular values for each \( D_i \) is the key to understanding which pattern is specific to each class/topic [8]. This way, we would able to extract specific patterns associated with each issue in the documents and then leverage those patterns to improve our classification models.

5 Discussion and related work

Legal scholarship devoted to using machine learning-powered tools to automate certain types of legal tasks, specifically flagging of clauses in contracts, is growing [41, 26, 30]. Similarly, several contributions calling for legal reforms necessary for the realization algorithmic fairness, accountability and transparency were made [8, 29, 1]. However, the amount of works devoted to using machine learning to advance social fairness, accountability and transparency is still relatively small [25]. This paper aims at being a little contribution to that field.

A lot remains to be done, of course. From the legal point of view, most importantly, formalizable standards of procedural and material unfairness of privacy policies need to be identified in other jurisdictions than on the level of the European Union. This is particularly important both on the federal level in the United States, where the standard is much less precise [49], on the level of (member) states of both Unions (where California, Nevada, France or Germany have pretty robust standards applicable to privacy policies). One this is done, we will be able to tag more privacy policies according to more

\[ \text{This means } \binom{15}{3} = 455 \text{ combinations, and we tested all of them.} \]

\[ \text{Further details about the computational approach and our code will be released at } \url{https://github.com/roozbeh-yz/privacy_policies} \]
legal standards and comment on both the accuracy of various ML approaches, and on the quality of legal standards in different jurisdictions (from the point of view of their formalizability).

Regarding our machine learning approach: Our preliminary results clearly show that we can use machine learning to automatically tag the suspicious parts of unseen privacy policies. Such automatic process would considerably facilitate the analysis of privacy policies and would make it practical to tackle this issue, despite its massive scale.

**To automate annotating privacy policies with focus on avoiding false negatives:** By defining narrow and specific issues in privacy policies, we were able to achieve very high accuracy in using machine learning. Our next focus will be to extend our analysis to a larger set of policies and investigate the generalization of our method in a larger domain. Our focus will be on avoiding false negatives in our model’s predictions, so that a human expert can review the tags and identify the troublesome parts of privacy policies efficiently. To avoid false negatives, it is essential to understand the underlying structures of text, and for that, we will use powerful mathematical tools such as tensor decomposition.

**To summarize the privacy policies:** There are many machine learning approaches for summarizing text and they perform reasonably well in many applications, including legal documents [22]. For privacy policies, however, the hollowness, vagueness, and lack of adequate information would make such approaches unpractical. We intend to explore that further in the future, and provide legal and policy recommendations.

6 Conclusions

Algorithmic Fairness, Accountability and Transparency need to go hand in hand with Social Fairness, Accountability and Transparency. Consumers need to be able to understand what data about them is collected, and how it is being used. As of today, they are unable to know that. As we have shown, the problem is not only the complexity of these documents - it’s first and foremost their emptiness. In this paper we outlined the theory behind combating it, and produced initial results suggesting that machine-learning powered approaches can be helpful here. Machine learning can be used both to help detect empty language, and to retrieve specific types of information from the PPs (once the information is already there). The results of the experiment conducted on a relatively small sample of PPs suggest that, when more data is generated, accuracy of these systems can grow.

References


