

Using Machine Learning to Enforce the CCPA: An Experimental Study

Jens Frankenreiter* Julian Nyarko‡ Dane Thorley⁺

Extended Abstract and Project Description

The CCPA provides one of the most ambitious regulatory approaches to privacy protections the U.S. has seen to date. However, many companies that fall under the CCPA allegedly fail to comply with its requirements. While the California Attorney General’s office plays an important part in ensuring enforcement, the CCPA also relies on the general public to ensure compliance.¹ In this study, we examine the feasibility of promoting said compliance using a field experiment that builds on recent advancements in computational linguistics and machine learning.

To that end, we are currently collecting, at regular intervals, the privacy policies of 1,614 randomly selected websites. We then train machine learning classifiers to automatically determine whether these privacy policies are compliant with specific requirements of the CCPA. The requirements we focus on are required notifications about the new privacy rights introduced in the CCPA, in particular the right to delete and the right to no-discrimination.

In a first step, our goal is to present descriptive evidence for the extent to which websites comply with the CCPA. The structure of our data as a panel allows us to investigate absolute levels of compliance as well as changes over time.

* Associate Professor of Law, Washington University in St. Louis School of Law.

‡ Assistant Professor of Law, Stanford Law School.

⁺ Associate Professor of Law, BYU Law School.

¹ CCPA Section 155(a), <https://oag.ca.gov/consumer-privacy-tool> [<https://perma.cc/3PDG-Z8J2>].

Then, in a second step, we are going to select a subset of companies that are non-compliant into an experimental study. The goal of this experimental study is to assess the extent to which simple, low-cost interventions can induce changes in the privacy policies in order to ensure compliance with the CCPA. Among the treatments we are currently contemplating and would be particularly thrilled to receive feedback on are:

- Doing nothing (control condition)
- Informing companies of their non-compliance
 - Should we inform companies that our notice acts as notice under the CCPA?
 - Should we stress the fact that we are researchers? Concerned citizens? Average Joe's?
- Informing companies of their non-compliance and suggesting specific language to remedy this non-compliance
- Inform companies that continued non-compliance will result in us notifying the California AG's office, which can result in enforcement actions
 - [We are in active conversation with the AG's office]