I. INTRODUCTION

Years of research examining “smart” Internet of Things (IoT) devices have found that developers of these devices have poor privacy practices [1], [2]. This includes collection of personal information without notice or consent [3]–[5] and transmission of data to online advertisers and data brokers [6].

Automobiles are the next frontier for the “datafication” of consumer devices. To the best of our knowledge, all recent model year vehicles available in major market are connected cars: they include always-on internet connections, collect and transmit data about the vehicle and the driver, and incorporate companion smartphone apps.

A recent report from the Mozilla Foundation highlights privacy concerns around connected cars [7]. This report drew on automakers’ privacy policies to identify the different types of data that automakers claim to collect and discloses around data selling and sharing to third-parties. The practices revealed in the report are very concerning, in part because vehicle ownership is a de-facto requirement for modern life and car owners have little—if any—ability to opt-out of vehicle’s data collection. Recent reporting revealed that automakers are sharing driving data with insurance companies and that this is causing real-world harms to vehicle owners [8].

In this study we propose to use data subject access requests issued under the CCPA and GDPR to investigate the privacy implications of connected cars. The main shortcoming of the Mozilla study is that the report relies on disclosures from privacy policies, which may under- and over-disclose data collection and sharing practices. For example, studies of privacy policies from websites and mobile apps have found that they sometimes contain vague language that permits all data to be collected and shared [9]—thus revealing nothing about actual collection and sharing practices—or they fail to disclose all practices.

Our goal is to obtain at least one data report from every major auto manufacturer to examine the type and granularity of data being collected, as well as compare the provided data to the manufacturer’s privacy disclosures. Additionally, we will examine the types of data that are not included in the reports to uncover potential violations of data subject access rules.

II. RELATED WORK

Related studies have begun to investigate connected cars, including the readability of their privacy policies [10], [11], the legality of data collection from cars [12]–[15], and the security of vehicle data [16], [17]. Other studies focus on using vehicle data to train self-driving cars or traffic safety applications [18].

III. METHODOLOGY

We plan to perform a crowdsourced audit of data collection by connected auto manufacturers. We will recruit multiple vehicle owners to request the data for their own vehicle—across brands, manufacturers, and recent model years—from the automakers and third-parties that are known to receive vehicle data (e.g., Experian [8]) and donate this data to us. Since some manufacturers only honor data requests from locations with comprehensive data privacy legislation (e.g., California and Europe) [19], we will focus on recruiting participants in these locations. We do not anticipate needing a representative population for this study, and we recognize that the data in question is very sensitive, so we are planning to directly recruit participants from friends, family, and colleagues.

It would be challenging to collect a data report from an instance of every single car make, model, and options package produced between the year 2018 and 2024, so our goal is more modest: to get at least one report for each brand of car discussed in the Mozilla Foundation report [7]. Further, it may be unnecessary to gather data reports from each make and model of car from a given manufacturer because, to save on development costs, connected cars often share an underlying telematics platform. For example, Lexus is a sub-brand of Toyota, and we expect that cars from both brands will share much of the same data collection infrastructure.

Once we have data reports from participants, we plan to compare their content across vehicles and manufacturers. We will also compare them to (1) automakers’ stated privacy policies and (2) data made available to third parties by automakers to identify discrepancies. Given that the number of automakers is small, we do not anticipate the need for automation to review the privacy policies [20]. If we are able to recruit multiple participants who own vehicles from a given automaker, this will enable us to assess the consistency of data reports within that automaker. Finally, if we are able to recruit participants in the U.S. and Europe, this may facilitate trans-continental comparisons of data collection by connected cars.

REFERENCES


