Vehicle Data Collection: A Privacy Policy Analysis and Comparison

Chiara Bodei[1][a], Gianpiero Costantino[2][b], Marco De Vincenzi[2][c], Ilaria Matteucci[2][d] and Anna Monreale[1][e]

[1]*Dipartimento di Informatica, Università di Pisa, Largo Bruno Pontecorvo 3, Pisa, Italy*

[2]*IIT, Consiglio Nazionale delle Ricerche, Via Giuseppe Moruzzi 1, Pisa, Italy*

Keywords: Automotive, Privacy Policy, Regulation, GDPR, Readability.

Abstract: In recent years, data can be considered the new fuel for road vehicle functionalities like driver-assistance systems or customized services. Therefore, the carmakers with their phone apps, synced with the infotainment system, can collect information from the drivers and vehicles to be processed inside or outside the car. In this context, we analyze different carmakers' privacy policies to define their readability and compliance with the EU General Data Protection Regulation, and provide analysis of carmakers' data collection. Besides, for the first time, we compare the most significant privacy regulations in automotive. Finally, we create an interactive dashboard to compare the different carmakers' policies and provide users with an efficient instrument to understand some relevant privacy aspects like which data the carmakers declare to collect. We find that carmakers could collect a large number of users and vehicle data, but, in some cases, the privacy policies seem to be quite challenging to read and do not provide some information like how collected data are protected or stored.

1 INTRODUCTION

Today, road vehicles can collect more data than we can imagine. People are usually aware that smartphones can gather several data, while few realize how much data cars collect during driving and when the phone is synced with the infotainment system (Vioreanu, 2022). The collected information is usually used by the carmakers to provide drivers with different features like Advanced Driver-Assistance Systems (ADAS) or personalized services, sending data to the carmaker servers to be processed. Nevertheless, the ownership of data can be controversial because, for example, Travelers United, a consumer advocacy group in Washington (USA) highlighted the problems of car companies refusing to share vehicle data with the car owners, even if they generated them (Leocha, 2022). Today, the documents, that can be used to clarify this situation, are the privacy policies of the different carmakers' apps that can be installed on the users' phones. In this context, we study the privacy policies of sixteen different carmakers to answer some privacy questions, understand the readability and provide a complete overview of the collected data, using an interactive dashboard. Our main contribution is to provide for the first time, dedicated only to automotive, a readability analysis and discussion of privacy policies, considering our cars not just vehicles, but nodes of a connected network.

The study of privacy policies has been a trending topic in the last years and people should be aware of the collected data employed by carmakers, so our work can contribute to aware drivers on possible privacy issues. The findings of our analysis underline the large quantity of data that carmakers declare to collect, but also the low readability of the privacy policies, which can be in contrast with some privacy regulation requirements.

Following the NIST definition, as privacy, we mean *the right of a party to maintain control over and confidentiality of information about itself*. However, the definition and perception of privacy can suffer cross-cultural differences, especially during interactions with new technologies (Li, 2022). For this reason, to define privacy, we provide also the indications of the privacy regulations that can be applied in the automotive context: in Section 3, we describe the Eu-

[a] https://orcid.org/0000-0002-0586-9333
[b] https://orcid.org/0000-0002-2900-262X
[c] https://orcid.org/0000-0002-2706-2936
[d] https://orcid.org/0000-0002-5936-8470
[e] https://orcid.org/0000-0001-8541-0284

Bodei, C., Costantino, G., De Vincenzi, M., Matteucci, I. and Monreale, A.
Vehicle Data Collection: A Privacy Policy Analysis and Comparison.
DOI: 10.5220/0011779500003405
In *Proceedings of the 9th International Conference on Information Systems Security and Privacy (ICISSP 2023)*, pages 626-633
ISBN: 978-989-758-624-8; ISSN: 2184-4356

ropean General Data Protection Regulation (GDPR) (European Parliament and Council of the European Union, 2016), the Chinese Provisions on the Management of Automotive Data Security (PMADS) (Provisions, 2022), and the American California Consumer Privacy Act (CCPA) (California State Legislature, 2018).

The paper is structured as follows: Section 2 describes the related work, while Section 3 is the legal background about privacy regulations. Section 4 is the readability analysis of the privacy policies using four different indexes. Section 5 contains an investigation of the policies concerning Articles 9 and 10 of the GDPR, which defines some special categories of data that should be treated properly. Section 6 describes the dynamic dashboard with the comparison among the carmakers' privacy policies. Section 7 reports the findings and possible future works.

2 RELATED WORK

In recent years, attention to privacy risk perception and the analysis of privacy policies have been raised.

In (Fabian et al., 2017), the authors present a large-scale study on the readability of nearly 50,000 privacy policies of English websites. From this work, we inherit some of the applied readability indexes, but our analysis is limited only to the automotive privacy policies. Another significant work (Lawson et al., 2015) on the privacy of connected vehicles was released in 2015, when the Canadian Freedom of Information and Privacy Association (FIPA) wrote a yearlong study on privacy, consumer choice, and vehicle technology. This work is a complete document with a detailed description of every aspect and concern about emerging connected vehicles, but, it is mainly focused on the Canadian audience and it does not deal directly with the privacy policies documents. However, this guide provides a specific indication for our work, because it states that, during a purchasing decision, *"it would remain unrealistic to expect the average car purchaser to be able to review and compare the privacy policies of various carmakers, dealers and other relevant service providers."*. Our work can address this issue because the dashboard compares the privacy policies of different carmakers. Another valuable work in the automotive field is the Pesé survey (Pesé, 2019), where authors describe the automotive privacy attacks and define a privacy score, quantifying the risk associated with each vehicular sensor and the related attack, but without directly analyzing the privacy policies. Another relevant work is (Zaeem et al., 2020), which states that privacy policies can be lengthy and hard to comprehend. To address this problem, researchers have utilized machine learning to devise tools that automatically summarize online privacy policies for web pages. In our work, with the reading analysis, we verify the assumption that the documents could be challenging to understand, and we provide an intuitive instrument to compare carmakers' privacy policies and show the differences.

Regarding the comparison among the privacy regulation in automotive, significant work is (Michael Tan and Thomas Kahl, 2022), where authors compare PMADS and the GDPR with a specific focus on the automotive industry. In our work, we add the CCPA to consider also another significant area for the automotive industry like the California and USA.

3 LEGAL BACKGROUND

Even if we do not address directly the legal aspects of data collection, we need to identify some legal requirements and possible constraints which can help us to compare the privacy policies. We choose the European GDPR, the Chinese PMADS, and the American CCPA, which, can be considered three of the most significant legal documents for the protection of personal data in automotive.

The GDPR, effective in May 2018, is a legal framework that sets guidelines for the collection and processing of personal information for companies and organizations that handle information of European Union (EU) citizens. The EU has drafted another document that can be used to regulate data privacy: the more automotive-related Guidelines 1/2020 (EDPB, 2020) on processing personal data in the context of connected vehicles and mobility-related applications, written by the European Data Protection Board and published in early 2020. In particular, the guidelines define the connected vehicles as "terminal equipment" just like a computer, a smartphone, or a smart TV and identify three special categories of data: location, biometrics, and offenses.

The PMADS is a Chinese regulation that was issued on August 2021 and entered into force on October 2021. It aims to regulate vehicle data processing activities to protect the rights and interests of individuals and organizations. It distinguishes between *personal data*, which includes any information that could infer a person's identity or behavior, and *important data*, which includes data that may endanger national security, for example, in military areas.

The other relevant regulation is the CCPA, which is an advanced state statute to protect privacy rights in California State and contains the broadest defini-

tion of *personal information* like any information that identifies or is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household.

Following these regulations, it is possible to define legal requirements and a framework, which underlines the importance of a deeper analysis of an emerging problem like user privacy in connected vehicles. In particular, all the regulations identify and classify the data according to their sensibility and importance. Different types of data can require different privacy levels, however, as shown in (Sardianos et al., 2018), data, which can be classified with low privacy requirements like the location, can be used anyway to infer sensitive information as health status.

In Table 1, we report a comparison among the three regulations. In particular, we can notice that the GDPR was the first issued regulation and may have been a guideline for car manufacturers and automotive companies in recent years. Other countries like USA and China have followed the EU regulation and may even overtake European supervisory authorities in data protection in the future (Michael Tan and Thomas Kahl, 2022). The main differences among the three documents are the definitions of personal and special data. While the GDPR and the PMADS distinguish among the two categories, in the CCPA, all data are classified as personal. The PMADS identifies as special data, information that can influence public security, while GDPR seems to be more person-centered.

To conclude, the comparison of the different regulations allows us to categorize the different data and to understand the levels of privacy required for every piece of information, collected by the carmakers, in the different territories.

4 PRIVACY POLICIES READABILITY

We start our readability analysis using an updated version[1] of the database as (Bodei et al., 2020), so in July 2021, we selected the top fifteen best-selling carmakers in Europe in 2020 (Statistics, 2021) and collected their mobile app privacy policies. Besides, we added Tesla, a carmaker with advanced technologies like autopilot, which requires a large quantity of data to be developed. We decided to choose the best-selling car brands in Europe, where the regulation GDPR is legally binding. For this reason, in Section 5, we focus on compliance only with European regulation.

In Table 2, we report the names of the apps from which are downloaded the privacy policies. We consider readability as the quality of being easy and pleasant to read (University Press Cambridge, 2012). In particular, the word "easy" should indicate which data are collected and how are processed. This information should be easily accessible in clear and plain language (European Parliament and Council of the European Union, 2016). To understand whether a text is readable, several indexes can be used, for example, in our analysis, we refer to the Coleman-Liau Index (CLI), the Simple Measure of Gobbledygook (SMOG) index, the Automated Readability Index (ARI), and the Flesch Reading Ease Index (FREI). These indexes compute the readability referring to a general group of readers without considering factors like age or gender, however, until today, they can be considered some of the most significant readability indexes (Fabian et al., 2017). The first three use the U.S. school grade to label a text as difficult or easy to read. The U.S. schools have different grades, starting from 1 to 17 which is the graduated level. The 13th grade or above is considered university level. Table 3 shows an approximate comparison between scores and the US education level (Derguech et al., 2018). As far as we know, the three indexes, CLI, SMOG, and ARI are considered three relevant indexes to evaluate text readability. They have been used since the 1960s/70s to define the scholastic level necessary for the comprehension of a text, starting from different values and coefficients as defined in the respective equations.

The Coleman-Liau Index (CLI) (Coleman and Liau, 1975) is a specific test to understand text readability and results in the U.S. grade level. This index (Equation 1), is based on the complexity of the words, measured by the number of letters, and the complexity of sentences, measured by the number of words in a sentence, multiplied by some defined coefficients.

$$CLI = 0.0588 \times L - 0.296 \times S - 15.8 \qquad (1)$$

where L is the average number of letters per 100 words and S is the average number of sentences per 100 words.

The SMOG Index (McLaughlin, 1969) shows the U.S. grade level necessary to understand the text. In its formula, Equation 2, it uses the polysyllables (words of 3 or more syllables) in a certain number of sentences (at least 30).

$$SMOG = 1.0430 \times \sqrt{P \times \frac{30}{S}} + 3.1291 \qquad (2)$$

where P is the number of polysyllables and S is the number of sentences.

[1]We consider the same database entries but with updated values

Table 1: Comparison of the three main legal documents related to data privacy and applicable in automotive.

Topic	GDPR	PMADS	CCPA
Enforcement	2018	2021	2020
Who it protects	EU citizens	Mainland territory of the People's Republic of China	California citizens
Definition of Personal Data	*Article 4* Any information to identify a person: • Name • Identification number • Location • Physical data • Economic data • Cultural data	*Article 3* • Information that could infer an identity	*1798.140-15* Information that identifies a consumer: • Name • Alias • Address • Internet protocol / Email address • Driver's license number • Geolocation • ...
Definition of Special Data	*Articles 9 and 10* Data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data, or data concerning a natural person's sex life or sexual orientation. Personal data relating to criminal convictions and offenses or related security measures.	*Articles 3 and 10* Data on the flow of people and traffic in military administrative areas, national defense science and industrial units or other units that involve state secrets [...] data on the operation of automobile charging networks, data on types and traffic volume, etc., audiovisual data of individuals' faces, voices, and license plates, etc., outside the vehicle. Drivers' biometric data such as fingerprints, voiceprint, facial images, and heart rhythm can be collected.	[Not provided, all data included in personal data category]
Data Anonymization	Yes	Yes	Yes [Deidentification (1798.140-8)]
Right to Deletion	Yes	Yes	Yes

Table 2: Privacy Policies' App.

Company	App Name
Audi	myAudi
BMW	My BMW
Citroen	My Citroen
Fiat	Uconnect LIVE
Ford	FordPass
Hyundai	Bluelink Europe
Kia	Kia UVO (UVO Connect)
Mercedes	Mercedes Me
Opel	myOpel
Peugeot	myPeugeot
Renault/Dacia	MY Renault
Skoda	MySkoda (Skoda Connect)
Tesla	Tesla
Toyota	MyT
Volkswagen	We Connect

Table 3: Table comparing scores and education levels (Derguech et al., 2018).

Score/Grade	Education Level
1-4	Elementary School
5-8	Middle School
9-12	High School
13-16	Undergraduate
17+	Graduate

The Automated Readability Index (ARI) (Senter and Smith, 1967) measures the readability of a text and returns the U.S. grade level to understand the text. With respect to the other, this one also takes into account the number of characters (Equation 3).

$$ARI = 4.71 \times \frac{C}{W} + 0.5\frac{W}{S} - 21.43 \qquad (3)$$

where C is the number of characters (letters and numbers), W is the number of words, and S is the number of sentences.

The fourth is the Flesch Reading Ease Index (FREI) (Flesch, 1981) which differs from all three previous indexes because it outputs a score instead of a school grade. The score starts from 0 to 100 and the lowest value indicates a text extremely difficult to read. The formula, Equation 4, uses the number of words, sentences, and also syllables.

$$FREI = 206.835 - 1.015 \times \frac{W}{S} - 84.6 \times \frac{Sy}{W} \qquad (4)$$

where W is the number of words, S is the number of sentences, and, Sy is the number of syllables.

Table 4: FREI index interpretation.

Score	Interpretation
100-90	Very easy
90-80	Easy
80-60	Fairly easy
60-40	Fairly difficult
40-30	Difficult
30-10	Very difficult

Table 5 summarizes all the results obtained by analyzing the privacy policy documents using the Python library *readability* [2]. The library enables the calculation of different metrics and our Python program outputs a table containing the metrics of the text in addition to the number of words.

As a result, Table 5 shows that the number of words is not a significant parameter to establish whether a policy is readable or not. It is not pos-

[2]https://pypi.org/project/readability/

Table 5: The calculated metrics for each privacy policy.

Privacy Policies Metrics					
Company	**Number of Words**	**CLI**	**SMOG**	**ARI**	**FREI**
Audi	16005	11.74	18.56	13.38	36.86
BMW	3380	11.08	14.36	12.75	43.33
Citroen	2623	10.76	16.45	12.15	41.69
Fiat	821	10.23	12.04	10.23	48.18
Ford	12185	12.44	13.82	13.86	39.65
Hyundai	5808	11.27	17.62	10.57	46.65
Kia	22663	10.71	14.51	9.96	47.12
Mercedes	7376	13.13	16.11	13.14	31.89
Opel	2460	11.60	14.84	12.74	39.44
Peugeot	2135	10.86	16.28	13.03	39.69
Renault/Dacia	4155	12.69	17.88	15.22	34.13
Skoda	1496	11.23	16.81	13.34	36.28
Tesla	6657	13.15	17.28	17.78	27.12
Toyota	6503	12.60	16.11	14.69	33.96
Volkswagen (VW)	15313	12.58	16.61	14.84	32.47

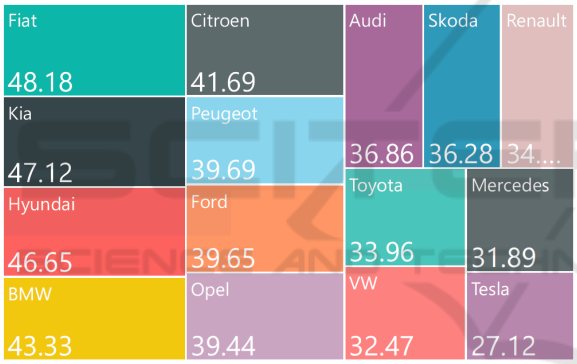

Figure 1: Treemap of the FREI readability index. Bigger rectangles represent an easier text, while smaller rectangles a more difficult text to read.

sible to identify any particular set, because, for instance, carmakers belonging to the same group, like Volkswagen-Audi-Skoda have different metric values. We can only retrieve similar indexes for Opel and Peugeot, belonging to the same industrial group, and that share almost the same privacy policy document.

Referring again to Table 5, it can be observed that all privacy policy documents require at least a high school/university level of education to be completely understood. In particular, the CLI index shows values close to or above 11, which correspond to the last years of high school, while SMOG shows higher values near the university level. To summarize our findings, the reading and comprehension of the privacy policies require, on average, a high level of education equal to the last years of high school or the first years of university to be comprehensible in every part.

5 GDPR SPECIAL CATEGORIES

Processing the privacy policies, we search for the data special categories belonging to Articles 9 and 10 of the GDPR, which requires a defined treatment different from generic personal data, identified in Article 4. As a preliminary step, we report Table 6, where collected data by the carmakers are classified with the related keywords, retrieved from the privacy policies. This classification allows users to understand in an easier way which data categories each carmaker declares to collect and we use this work as a baseline for Articles 9 and 10 categories.

Article 9 states how to process some special categories of personal data. In particular, in Article 9, in addition to Table 6 elements, we can identify other special categories as *"personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data intended to uniquely identify a natural person, data concerning health or a natural person's sex life or sexual orientation"* (European Parliament and Council of the European Union, 2016). Besides, Article 9 states that these categories can be processed only under certain conditions like a vital interest of the data subject or public interest. In addition to the special categories identified by Article 9, Article 10 identifies another special data category related to criminal convictions and offenses that can only be registered under the control of the official authority.

The categories considered in Article 9 and Article 10 are data that have been classified as special, however, we can infer the same information starting from the generic personal data as in Table 6. For example, if some geolocation data, usually composed of coordinates and time logs, show that, on a specific day of the week, a person goes to a place of worship at the same time as there is a celebration, it can probably reveal the religious beliefs of the person. The same assumption can be applied to every special data category, also using other information that a vehicle collects like "voice and messages". This shows how the processing of data by car companies is a really sensitive activity.

To verify if carmakers declare to collect Article 9 and Article 10 special categories, we design and develop a Python tool, which is based on the NLTK library (Loper and Bird, 2002). The tool is fed with the data categories and keywords, as defined in Table 6, and it can automatically identify the possible categories collected in each privacy policy. Once fed with the keywords, the tool takes as input the privacy policy text. Then, it finds all the nouns and adjectives

Table 6: Automotive data categories with related keywords (Bodei et al., 2020).

Category	Keywords	Category	Keywords
Personally Identifiable Information	• Name • Surname • Address • Date of birth • Mobile number • Email address • License plate number	**Geolocation**	• Position • GPS time • Speed • Directions • Traffic • Departure and destination name • Estimated travel time • Point-of-interest searching (POI)
Driver's Phone	• IP address • MAC address • OS version • Browser Information	**Financial**	• Customer ID • Credit card number • Purchasing • Financial data for payments • Fuel costs
Offences and Violations	• Speeding • Information on car accident • Information on airbag usage • Vehicle security systems usage	**Driver's Behavior**	• Driving style • Travels statistics • Steering movements • Accelerator and brake usage
Vehicle Status	• Vehicle Identification Number (VIN) • Engine status • ECUs status • Oil level • Tyre pressure • Automatic maintenance requests • Maintenance history	**Surrounding vehicle environment**	• Detected signs and lanes • Environment • Static and dynamic objects near the car • Side distance from near objects • Climate • Light influx
Voice and Messages	• Emergency call • Voice controls • To perform voice recognition • Messages and chat with call center	**App Usage**	• Behavior • Logs • Time • Duration

of the entire text. It compares the found grammar elements with the database of keywords to find a possible correspondence and it outputs for each privacy policy a list of sentences, containing each word, that can define the collection of a category by the carmaker.

We use this tool to find the collection of the special categories, defined in the GDPR. In particular, we perform a keyword search with the tool using the words contained in the special categories identified in these two articles. For example, we looked for the word "religious" in the texts to verify whether it was perhaps used to declare a collection of this kind of data. From our search, we did not identify sensitive words, like "racial", "ethnic", "political", and others inside the privacy policies of the sixteen carmakers. Nevertheless, this does not indicate that carmakers do not declare the collection of special data categories. Car companies may use a paraphrase of a sentence with synonyms or antonyms to indicate a special category. Due to this fact, we decided to perform a second step in the privacy policy documents, but this time considering synonyms or antonyms of each special category keyword belonging to *Articles 9* and *10* of the GDPR.

Table 7 reports some examples of our investigation made with the Python tool, enriched with synonyms or antonyms of each special category.

For example, our analysis says that in the *Mercedes* privacy policy we found that the word "gender" appears as a synonym of "sex", which is a word designing a special category in *Article 9*. After finding the word, we manually read the sentence where the word "gender" appears. The result is that the word "gender" is related to the usage of pronouns in the text and not to the collection of data related to sexual orientation. To conclude, after analyzing all the privacy policy documents, our opinion is that none of the car companies claims to directly collect special data regulated from *Articles 9* and *10* of GDPR.

6 THE ONLINE DASHBOARD

To provide the reader with a quick overview of carmakers' mobile app data collection and improve the comprehension of the associated privacy policies, we implement a dynamic dashboard, available online [3], based on the findings of our analysis. To build the dashboard, we use Microsoft Power BI, a business data analytics tool, which allows us to summarize data in an interactive dashboard. One of the main features of our dashboard is being easy to access at first glance because it is organized like a set composed of tables comparing different metrics. Another feature is the possibility to drill down in the graphs to have a smaller granularity to answer information questions.

As shown in Figure 2, the dashboard consists of a slicer and four main graphs. The upper horizontal slicer allows users to select and compare carmakers. The first upper stacked column chart represents the data category collected by each carmaker. This chart is based on data and categories of Table 6 and it allows us to compare the different carmakers and the quantity of data that they declare to collect. We can

[3]https://app.powerbi.com/view?r=eyJrIjoiYWU2ODg1NjQtNjQxOS00ZWVlLTk5YzUtNTkzYjg4NTJmYjNhIiwidCI6ImM3NDU2YjMxLWEyMjAtNDdmNS1iZTUyLTQ3MzgyODY3MGFhMSIsImMiOjh9

Table 7: Example of detection of special data category word with taxonomy in a privacy policy.

Company privacy policy	Word in GDPR	Word detected in the privacy policy	Semantic analysis: sentence in the privacy policy	Collection of special data category
Mercedes	sex	gender	"To make this Policy easier to read, the text uses only the male forms of pronouns for natural persons. The words he, his, and him are always intended to include all individuals, regardless of **gender** identity."	No
Kia	orientation	preference	"Reset of account: Your account may be reset by setting the respective **preference** (e.g. in the UVO App)."	No
Peugeot	religious	religious	"Please note that you should not include sensitive data (such as information about racial or ethnic origin, political opinions, **religious** or philosophical beliefs, or health) in your message."	No

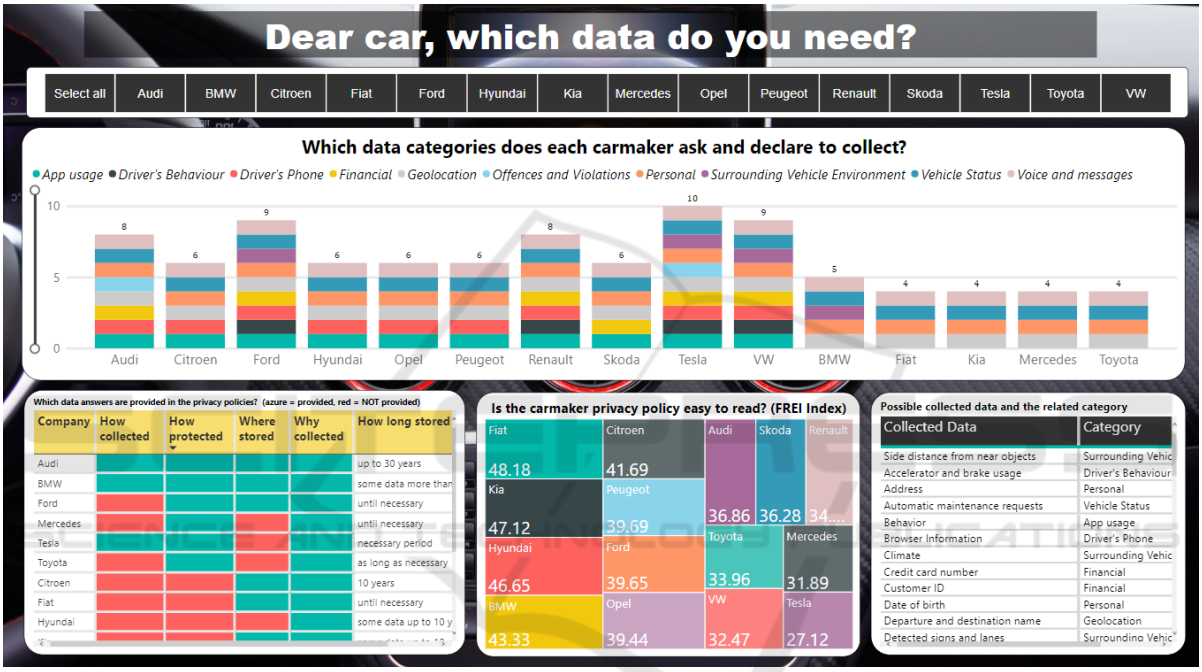

Figure 2: Dynamic dashboard to summarize our findings.

notice that Tesla seems the company which collects more data, while Fiat, Kia, Mercedes, and Toyota declare to collect only four data categories over ten. The bottom left table answers several questions about how carmakers declare to manage information like how they collect, protect, and store our data. An important element that we can retrieve from this table is how long our data are stored. We span from up to 30 years of Audi, to "*until necessary*" of some companies like Mercedes and Ford. The centered treemap represents the FREI readability index as shown in Figure 1. The last graph in the bottom-right corner shows the collected data with the respective category to which they belong.

7 CONCLUSION AND FUTURE WORK

In our study, we analyze sixteen privacy policies of different carmakers. We define their readability, we study the possibility that special GDPR data are collected, and, finally, we create a summary dashboard to compare the different policies and collected data categories. Besides, we provide a focus on the definition of privacy and data category in automotive comparing three different regulations. As our findings, we can state that carmakers' privacy policies need a high school level to be understood, confirming a general trend of difficulty also for privacy documents outside automotive. Another key question of our work is: *is the data collection compliant with the GDPR Articles*

9 and 10? The answer is apparently yes, because the carmakers declare that they do not collect sensitive or special categories of data. Despite this, the collection of different categories of data in a large quantity can lead an external subject to understand the behavior of any user and infer truly sensitive information. Moreover, several privacy policies are quite complex and some information such as how long data are stored and how data is protected are not so easy to find.

To conclude, we can state that the privacy policies of the carmakers are the main instrument to inform users about data processing, however, they need to be more readable to be compliant with the different regulations, providing also more answers to the most relevant privacy questions like where user data are stored.

ACKNOWLEDGMENTS

The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 883135 (E-Corridor). This work was partially supported by project SERICS (PE00000014) under the NRRP MUR program funded by the EU - NGEU.

REFERENCES

Bodei, C., Costantino, G., De Vincenzi, M., Monreale, A., and Matteucci, I. (2020). Privacy and security issues in vehicular ad hoc networks.

California State Legislature (2018). Ab-375, chau. privacy: personal information: businesses.

Coleman, M. and Liau, T. L. (1975). A computer readability formula designed for machine scoring. pages 283–284.

Derguech, W., Zainab, S. S., and D'Aquin, M. (2018). Assessing the readability of policy documents: The case of terms of use of online services. ICEGOV '18, page 247–256, New York, NY, USA. Association for Computing Machinery.

EDPB (2020). Guidelines 1/2020 on processing personal data in the context of connected vehicles and mobility related applications.

European Parliament and Council of the European Union (2016). Eu general data protection regulation (gdpr): Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation), oj 2016 l 119/1.

Fabian, B., Ermakova, T., and Lentz, T. (2017). Large-scale readability analysis of privacy policies. In Sheth, A. P., Ngonga, A., Wang, Y., Chang, E., Slezak, D., Franczyk, B., Alt, R., Tao, X., and Unland, R., editors, *Proceedings of the International Conference on Web Intelligence, Leipzig, Germany, August 23-26, 2017*, pages 18–25. ACM.

Flesch, R. (1981). How to write plain english.

Lawson, P., McPhail, B., and Lawton, E. (2015). The connected car: Who is in the driver's seat? a study on privacy and onboard vehicle telematics technology.

Leocha, C. (2022). What data does your car collect about your life? Accessed on October 28, 2022.

Li, Y. (2022). *Cross-Cultural Privacy Differences*, pages 267–292.

Loper, E. and Bird, S. (2002). NLTK: the natural language toolkit. *CoRR*.

McLaughlin, G. (1969). Smog grading a new readability formula. pages 639–646.

Michael Tan and Thomas Kahl (2022). New PRC Data Rules versus GDPR: . Accessed on November 5, 2022.

Pesé, M. (2019). Survey of automotive privacy regulations and privacy-related attacks.

Provisions (2022). Several provisions on the management of automobile data security. *https://www.chinajusticeobserver.com/law/x/ /provisions-on-the-administration-of-automative-data-security-20210816*.

Sardianos, C., Varlamis, I., and Bouras, G. (2018). Extracting user habits from google maps history logs. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 690–697.

Senter, R. and Smith, E. (1967). Automated readability index.

Statistics, C. S. (2021). Best-selling car brands in europe in 2020.

University Press Cambridge (2012). Cambridge advanced learner's dictionary & thesaurus.

Vioreanu, D. (2022). Connected cars collect more information on you than you imagine. Accessed on November 5, 2022.

Zaeem, R. N., Anya, S., Issa, A., Nimergood, J., Rogers, I., Shah, V., Srivastava, A., and Barber, K. S. (2020). Privacycheck's machine learning to digest privacy policies: Competitor analysis and usage patterns. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 291–298.