

Privacy and modern cars through a dual lens

1st Giampaolo Bella

*Dipartimento di Matematica e Informatica
Università degli Studi di Catania
Catania, Italy
giamp@dmf.unict.it*

3rd Marco De Vincenzi

*Istituto di Informatica e Telematica
Consiglio Nazionale delle Ricerche
Pisa, Italy
marco.devincenzi@iit.cnr.it*

2nd Pietro Biondi

*Dipartimento di Matematica e Informatica
Università degli Studi di Catania
Catania, Italy
pietro.biondi@phd.unict.it*

4th Giuseppe Tudisco

*Dipartimento di Matematica e Informatica
Università degli Studi di Catania
Catania, Italy
giuseppe.tudisco@studium.unict.it*

Abstract—Modern cars technologies are evolving quickly. They collect a variety of personal data and treat it on behalf of the car manufacturer to improve the drivers' experience. The precise terms of such a treatment are stated within the privacy policies accepted by the user when buying a car or through the infotainment system when it is first started.

This paper uses a double lens to assess people's privacy while they drive a car. The first approach is objective and studies the readability of privacy policies that comes with cars. We analyse the privacy policies of twelve car brands and apply well-known readability indices to evaluate the extent to which privacy policies are comprehensible by all drivers. The second approach targets drivers' opinions to extrapolate their privacy concerns and trust perceptions. We design a questionnaire to collect the opinions of 88 participants and draw essential statistics about them. Our combined findings indicate that privacy is insufficiently understood at present as an issue deriving from driving a car, hence future technologies should be tailored to make people more aware of the issue and to enable them to express their preferences.

Index Terms—automotive, privacy, drivers, cybersecurity

1. Introduction

Modern cars host highly developed technologies, such as infotainment systems and e-call boxes, routinely connected to the Internet. This increases the possible attack surface, and a number of examples of remotely hijacked cars exist. Cars may also collect drivers' (or passengers') personal data, hence privacy becomes a concern.

Intel estimates that a car can generate up to 4000 GB of data per day [1]. Thus, it is vitally important to understand and gather what types of (personal) data categories cars are collecting - and their manufacturers are treating - notably if these include special categories according to Regulation (EU) 2016/679, known as GDPR [2]. In addition, it is important to understand whether and to what extent drivers are aware of what and how much data they disclose to car manufacturers and how this data is

managed by them. In fact, there would be limited use in addressing a problem that drivers did not feel. Despite a few recent headlines on attacks to real cars [3], there is limited literature demonstrating how drivers feel about their privacy in their cars and what level of trust they pose e.g. in the interconnected infotainment systems that are becoming more and more common today, hence this objective.

The goal of this paper is to analyse privacy issues in modern cars through a dual lens. Through the first lens, we make an objective analysis of the availability and readability of the privacy policies the car makers provide to drivers as soon as they access to their service. To do this, we analyse the documents with respect to well-known readability indexes. The second lens is a subjective one. We constructed a ten-questions questionnaire in order to obtain the privacy concerns and trust perceptions of drivers with respect to modern cars. To do this, we statistically analysed the responses of 88 drivers. Our combined findings are that people's privacy concerns are rather low despite the moderate trust they generally express on their car's security and privacy measures, and this may be due to the opacity of the way privacy policies are written and presented. Therefore, it seems fair to claim that privacy is currently insufficiently understood as an issue deriving from driving a car.

The paper is structured as follows: Section 2 introduces the architecture of modern cars, Section 3 explains in detail the privacy issues of this field, Section 4 shows the study on privacy policies from an objective point of view, Section 5 describes the study from a subjective point of view, i.e. it analyses opinions from the point of view of privacy and trust of drivers with respect to modern cars, Section 6 comments on related work and conclusions.

2. Modern car architecture

Modern cars are computer on wheels. They have many different types of electronic control units (ECUs) on board that work together, thus enabling the complete operation of the car, they also provide and improve the usability and comfort of drivers and passengers.

In addition to control units, cars are composed of different sensors such as tyre pressure sensors, touch sensors that detect driver fatigue through grip, pulse or temperature sensors in the passenger compartment. Other sensors are installed above the roof of the car, such as radar or exterior cameras that take the car into the world of autonomous driving. In fact, modern cars tend to be increasingly connected to each other and to automotive infrastructures.

2.1. Communication domains

Modern cars appear to be increasingly complex and connected systems. An increasing number of entities receive and transmit data through the connected vehicle ecosystem. In particular, cars have several communication domains such as:

- **Vehicle-to-Vehicle (V2V)** communication includes a wireless network where cars exchange messages with information about what they are doing. This data includes speed, position, direction of travel, braking and loss of stability. The aim of V2V communication is to prevent accidents by allowing passing vehicles to exchange position and speed data via an ad hoc network.
- **Vehicle-to-Infrastructure (V2I)** communication is a communication model that allows vehicles to share information with the components that support a country's highway system. Such components include cameras, traffic lights, lane markers. Therefore, sensors can capture infrastructure data and provide travellers with real-time alerts on issues such as road conditions, traffic congestion and parking availability.
- **Intra-Vehicle communication (IV)** is the communication model in which the ECUs and sensors communicate with each other and exchange messages about the car's status. All this allows a car to function properly.
- **User-to-Vehicle communication (U2V)** concerns the inclusion of the smartphone, its data and functionalities (e.g. to allow easy hands-free dialling of the address book) via Bluetooth, Wi-Fi or USB mirroring.
- **Car maker and third-party services** need to collect data. The transmission of data to these services takes place through the use of a SIM card inside the car that provides connectivity to the car. In this scenario, the car transmits data both to the car maker's servers and to third-party services (e.g. for software updates or with satellites for geolocation), which are often accepted during the acceptance of privacy policies.
- **Emergency services** is one of the on board services. Thanks to the car's built-in SIM, the car can call the emergency services in case of a rescue need, such as a brutal impact that caused the airbags to deploy.

The information about a modern car, the data it handles and transmits, and the components from which it is made can be summarised in the Figure 1.

2.2. Data treatments

The car architectures are based on data and the latter are transmitted to the infrastructures via the communica-

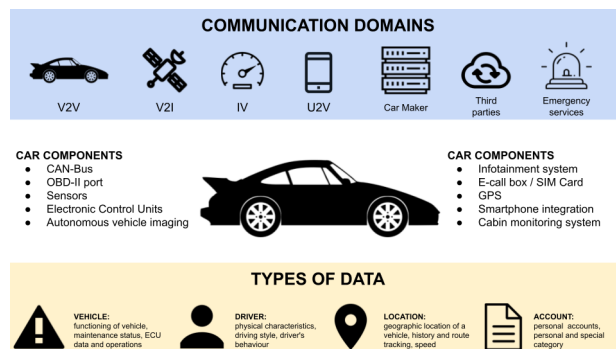


Figure 1. Infographics about data receivers, data type and car components.

tion domains. All these data can be classified into several types:

- **Vehicle Data** are all data concerning vehicle operation, maintenance status, mileage, tyre consumption, data from sensors.
- **Driver Data** relates to driver profiling. This can be done by sampling the physical characteristics or habits of the driver, such as, driving style of the vehicle, seat belt use, braking habits.
- **Location Data** includes data, such as the geographical position of a vehicle, route history and tracking, speed, direction of travel.
- **Account Data** contains all data relating to the driver's personal accounts.

Processing data provides a number of benefits to drivers. Vehicle sensors and internal components can produce data for diagnostic purposes to check the health of the vehicle. Predicting and preventing component failures can effectively improve driver safety. In addition, processing sensor data allows technicians to conduct remote diagnostics and, by analysing historical data, repair costs can be lower. The collection of data can also be useful in emergency situations. In fact, all cars sold in Europe since April 2018 must implement an emergency-call system, which is called eCall, as a measure to reduce fatalities caused by road accidents [4]. The eCall system is an electronic device that automatically alerts the emergency services in the event of a car accident. This system makes rescue operations faster and more effective and helps to save lives. When a severe impact is detected via the motion sensors, using the GPS and the vehicle's built-in SIM card for communication, data such as the exact position of the crashed car, direction of travel and type of impact are sent to the emergency services.

External service providers can use data to offer benefits to drivers, including economic benefits. These include usage-based car insurance, also known as pay-as-you-drive insurance. Insurers monitor the driving habits through a telematics device installed on the vehicle. By analysing data such as mileage, hard braking, rapid accelerations and cornering, the insurer is able to adapt the premium amount based on the driver's behaviour. The choice of a usage-based insurance over a traditional one has many potential benefits to the drivers. Thanks to the telematics device, drivers can monitor their driving habits and may get motivated to correct them because good drivers benefit from discounted premiums [5]. It is also

easier to investigate in car accidents by having access to the events that took place before the accident occurred.

3. Privacy Issues

Nowadays, when we think of privacy breaches, we immediately consider smartphone applications or online services that leak users' personal information.

In addition to smartphone and personal devices, also car cybersecurity is closely related to driver privacy, as modern cars collect a wide variety of personal data from drivers as a fundamental basis for the user experience. However, it is not easy to understand how cars exploit this data and, in particular, to what extent drivers understand the relevant implications, including the fact that car manufacturers are required to handle personal data in full compliance with relevant regulations. If car manufacturers collect data from a user in the European Union, this data is subject to the GDPR. Therefore, regardless of the geographical location of the servers that actually store the data, the company must ensure an adequate level of protection and privacy as required by the legislation.

The infotainment system also interacts directly with the driver and passengers by providing them with various functions, but these require data from users in order to properly work. By synchronising personal devices with the infotainment system, users are able to stream multimedia files, make calls using the car's speakers, read e-mails and send text messages. They can also surf the web using the system's browser. The infotainment system can provide directions and traffic information using the vehicle's real-time location. Compared to sensors, this is the main component that makes use of the personal data of its users. In addition, there is a variety of data concerning the personal data of drivers and passengers, including music preferences, favourite places, contact list, text messages and call logs. In particular, we note that this group may include "special categories of personal data", i.e. data on health, political and religious orientation, biometric data, ethnic origin, as defined by Article 9 of the GDPR. For example, the car manufacturer may infer certain health conditions of the driver by intersecting historical data from seat weight sensors with the chosen interior temperature, seat back adjustment, car position or heart rate via sensors on the car steering [6]. Remarkably, this data set can also include the driver's financial data to pay for fuel and parking [7].

Furthermore, through vehicle geolocation we may be able to obtain information that is repeated over time and thus derive a possible habit, hobby or routine of a driver. For example, we may notice that every Sunday at a certain time the car is near a church, so we could deduce that the driver goes to church every Sunday and through the information obtained in an open way from the network we are able to deduce also the information regarding the type of religion that the driver professes. Another example is recognising the driver through the seat sensor, which automatically adjusts the previously stored position based on weight. These technologies can also monitor eye movement to detect a driver's attention in order to determine whether a driver is falling asleep at the wheel [8]. It is clear that all these sensors allow for an

almost perfect identification of a driver and this increases the risk to the driver's privacy.

Finally, we can conclude that nowadays, more and more experiments show that cars can also manage the driver's personal data, often transmitted in the infotainment network, such as driving style [9], location history [10] and also more general data such as cabin preferences, music preferences, and credit card details [11]. However, the literature shows how the infotainment system can provide an entry point for attackers. Some examples involve exploiting vulnerabilities in the infotainment of a General Motors car to steal data from the remote system [12]. Additionally, a few years ago, researchers discovered a number of vulnerabilities that, when combined, allowed them to remotely hack a Tesla Model S [13].

4. Through the first lens: objective point of view

Car makers collect and manage drivers' data. As soon as a driver uses one of the car maker's service, she has to provide the consensus by signing a *privacy policy*, where each car maker should declare the type of data that it collects from the driver and from the car. In particular, following Article 12 of the GDPR, the controller of the data, in this case the car company, has to provide information to the user "*in a concise, transparent, intelligible and easily accessible form, using clear and plain language*" [2]. Our main target is to understand how easy it is for people to understand their data privacy in cars and the compliance with regulation like GDPR. In this section we present our analysis of privacy policies of twelve car makers and we consider a set of readability indexes evaluating the policy with respect to each of them.

4.1. Privacy policy collection

As a first step, we consider twelve car makers company: the top ten most famous car companies in Europe according to [14] (Audi, BMW, Ford, Mercedes, Opel, Peugeot, Renault, Skoda, Toyota, Volkswagen), Tesla, since it is the car company that most equips its cars with advanced technology, and KIA since we have a vehicle that we use for our cyber-security activity research. All the car manufacturers analysed in this work are shown in Table 2.

Then, we download the privacy policy by using two different channels: during the installation of the respective app or from the company website and contacting the customer service of the car maker. The aim is being sure to analyse the most recent privacy policy of the considered car maker. Hence, we verify the correctness of the selected privacy policies by comparing the two of them and considering the one retrieved contacting the customer services and those one obtaining by using each car maker's app. The customer services were contacted using the social network "Facebook". However, we received the required documents in six cases over twelve. Hence, it is necessary to contact via email or by phone another call center. The result of this double check is that the privacy policies downloaded from the app, are the same sent from the customer services. So, we can confirm that for a user it is

quite easy to access and obtain the most updated privacy policy document. However, it is sometimes quite difficult to download the document and, without any particular translation, it is only possible to read it online, but this does not affect the possibility of being informed.

4.2. Policy readability analysis

Once we have established that the driver can access the privacy policies quite easily, we analysed whether the content is easy to be understood. A text analysis on privacy policy documents (written in English) is performed by using “textstat”, a Python library to calculate statistics from text, that allows also to compute readability indexes [15]. From the analysis, several metrics, such as, the *number of words* or the *number of sentences* are extracted. We consider also other metrics. In fact, one of the most important features of a text, especially of a privacy policy, is the *readability*, that is the document quality of being easy and enjoyable to read [16]. By the word “easy” we mean that the policy must be in an easily accessible form with clear and simple language.

To understand if a text is readable, several indexes can be calculated. In this specific case, using the same Python library *textstat*, we calculate the **Coleman-Liau** index, the **SMOG** index, the **Automated Readability** Index and the **Flesch Reading Ease** Index. The first three indexes 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 or more, that is graduated level. The 13th grade or above is considered university level. Table 1 shows an approximate comparison between the index scores and the US education level [17].

TABLE 1. TABLE COMPARING SCORES AND EDUCATION LEVELS [17].

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

The **Coleman-Liau** Index (CLI) [18] depends on the complexity of the words, measured from the number of letters, and the complexity of the sentences.

The **SMOG** Index [19], acronym for “Simple Measure of Gobbledygook” uses the polysyllables (words of 3 or more syllables) in a certain number of sentences (at least 30).

Despite the similarity with the previous indexes, the **Automated Readability** Index (ARI) [20] takes into consideration also the number of characters, in addition to the number of words and sentences.

The fourth is the **Flesch Reading Ease** Index (FREI) [21], that differs from all the three previous indexes because it outputs a score instead of the school grade. The score ranges from 0 to 100 and the lower value indicates a text extremely difficult to read. It uses the number of words, sentences and also the number of syllables.

To combine the results obtained by the previous indexes into a single metric, the last column of Table

TABLE 2. METRICS ESTABLISHED FOR EACH PRIVACY POLICY ORDERED ACCORDING TO THE GIDR

Privacy Policies Metrics						
Company	Number of Words	Number of Sentences	CLI	SMOG	ARI	FREI
Ford	9744	1327	8.7	10.0	5.1	58.3
Peugeot	2151	437	9.0	9.0	6.0	47.7
Kia	22043	3096	9.6	10.4	5.8	49.8
Skoda	5831	860	9.9	9.9	6.1	49.7
Mercedes	8591	1387	10.0	10.0	6.0	44.0
Opel	2438	323	11.0	10.0	7.0	46.6
Audi	13661	1410	10.4	11.1	6.8	49.4
BMW	991	119	11.1	10.5	7.1	49.1
Tesla	13224	1453	10.8	11.1	7.0	49.5
Volkswagen	11742	1206	12.2	11.8	8.3	42.1
Toyota	3279	263	11.7	13.2	8.8	43.6
Renault	2568	94	12.7	17.3	15.9	37.3

2 presents a *General Index of Difficulty of Readability* (GIDR) that is calculated by combining the previous four indexes: the lowest value “0” indicates the most readable privacy policy, while the value “100” the most difficult among the selected documents. GIDR is formulated like Equation 1 and it is built starting from a harmonic mean. To merge the obtained indexes, we combined with a simple average the three indexes (CLI, ARI, SMOG) that have the same comparable scale (the US school grade). Then, we merged this mean with the FREI values. Due to the different scales of the mean and the FREI values, we chose the harmonic mean because it allows us to find division relationships between fractions without worrying about common denominators. The obtained value was normalised in a range between 0 and 100 using the Equation 2. In Table 2, we summarise all findings obtained by analysing the twelve privacy policies.

$$GIDR = \left(\frac{1}{\frac{CLI+SMOG+ARI}{3}} + \frac{1}{100 - FREI} \right)^{-1} \quad (1)$$

$$GIDR_n = 100 \times \frac{GIDR - \min(GIDR)}{\max(GIDR) - \min(GIDR)} \quad (2)$$

As a result, Table 2 shows that the number of words is a relative parameter to establish whether a policy is readable or not. It is not possible to identify any particular cluster, because, for instance, car makers belonging to the same group, like Volkswagen-Audi-Skoda or Peugeot-Opel, have different GIDR values. Instead, it is possible to note the concentration of the most car makers (9 over 12) in the first quartile of the GIDR value distribution, meaning a similar readability. However, inside this last set, Ford has a significant lower value than the others, resulting the most readable text. This difference seems to be determined by the FREI value, that, with respect to the others three readability indexes, is the only considering the number of syllables. This situation means that Ford could be the easiest readable text, because it uses shorter words with respect to others, and especially with respect to Renault.

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. In our analysis, the SMOG and CLI indexes show values close

to 11, that correspond to one of the last years of high school. The ARI index, probably due to its equation, has slightly lower values, on average at a middle/high school level, but directly proportional and in line with the other two indexes, SMOG and CLI.

To summarise our analysis, it can be stated that the reading and the comprehension of the twelve privacy policy documents requires, 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. This situation, for example in Europe, can affect the possibility of people to be informed in a concise, transparent and easily form as stated by GDPR [2]. In fact, according to [22], in 2020 in Europe only 35.9% of people aged 25–54 has tertiary educational attainment and 21.9% among people aged 54–74.

5. Through the second lens: subjective point of view

In this section, conversely with respect to the previous one, we show a subjective analysis of drivers' privacy concerns and trust perceptions. The following sections show the questionnaire we used for the study and the analysis of the findings obtained.

5.1. Questionnaire approach and structure

To explore the issue of privacy in modern cars, we decided to conduct a study of drivers using a questionnaire to ask about their concerns about privacy and perceptions of trust in modern cars. Specifically, the questionnaire is divided into three parts. The first part deals with basic information about the respondents, including essential demographic data, including for the purpose of confirming that they are over 18 years old, and a check on the reliability of their answers. The second part deals with respondents' privacy concerns and the third part with their perceptions of trust, so these two parts contain the key questions for our study.

A 7-point balanced Likert scale is used for most of the questions in the questionnaire. This is the most commonly used approach to scale responses in a research survey. Moreover, the range captures the intensity of their feelings for a given item. As such, likert scales have found application in psychology and social sciences, statistics, business and marketing [23].

In more detail, in Appendix A, we show the core questions of the questionnaire and the type of answers allowed. In particular, the questionnaire begins with a question (Q0) that asks participants how many hours a week they spend driving their vehicle. This information can be interesting because an individual who spends a lot of time driving is likely to have a greater knowledge of the features provided by their car compared to an occasional driver. The next question (Q1) asks participants to evaluate their knowledge about modern cars. Then question Q2 asks respondents whether or not they agree that modern vehicles are similar and comparable to modern computers. Question Q3 introduces the part related to the collection and processing of personal data. The question asks participants to select the kind of data they think modern cars

collect. Then question Q4 asks participants whether or not they agree that personal data collected by a modern car about its driver is necessary for the full functioning of the car. This question allows us to understand if drivers believe that, in order to be able to use all the features provided by their vehicle without any sort of limitation, they need to provide their personal data. Question Q5 asks whether or not respondents think it is necessary to transmit the personal data collected over the Internet. In relation to the previous question, participants may agree to provide their personal data to obtain additional features (e.g. statistics about driving style and vehicle usage). Question Q6 asks participants whether or not they agree that a modern vehicle safeguards its driver's life. This question introduces the survey part about the trust that drivers pose in their car. Question Q7 asks respondents whether or not they agree that a modern car protects its driver's personal data better than it safeguards its driver's life. Then with question Q8 participants are asked if they agree that a modern car processes the personal data it collects about its driver in a legitimate way that is consistent with the relevant regulations (e.g. *GDPR*). Question Q9 asks participants whether they agree that a modern car carries out a systematic and extensive evaluation of the personal data it collects about its driver on the basis of automated processing in order to evaluate personal aspects. In fact, according to the current legislation (art. 22 of the *GDPR*) these processes must be properly declared and explicit consent is required for the proposed purposes. In addition, (art. 32 of) the *GDPR* requires the use of adequate security measures to protect the rights and freedoms of data subjects. We ask participants about it with the last question (Q10) where they are asked whether or not they agree that the personal data a modern car collects about its driver is protected by suitable technology when the car transmits data over the Internet.

5.2. Analysis of findings

We have submitted the questionnaire to friends and colleagues to refine the methodology of the questions and the analysis of the findings, and we also think it could be a valid sample to be translated into a larger sample through crowdsourcing. The result discussed in this chapter is based on a sample of 88 people who responded to the survey described above, moreover, we anticipate that these findings are very promising.

Starting with the first question, Q0 asks participants how many hours per week they spend driving their vehicle. The answers are shown in Table 3. It turns out that 80% of the participants do not drive more than 9 hours per week, this is probably due to the fact that many people have reduced their mobility by car because of the pandemic.

Then participants were asked to evaluate their knowledge about modern cars. Considering the values in Table 4, it can be affirmed that the interviewed sample considers itself knowledgeable about modern cars. Just a minority of participants (about 13%) think they are not sufficiently knowledgeable about modern cars. Moreover, 17% of participants think they have average knowledge while the rest of them (70%) are quite confident about their knowledge.

To simplify interpretation and at the same time make it more expressive, answers were classified into the agree-

TABLE 3. ANSWERS TO THE Q0

	Q0
3-6 hours	38
7-9 hours	32
10-12 hours	13
13-15 hours	1
16-20 hours	1
21+ hours	3

TABLE 4. ANSWERS TO THE Q1

	Q1
Knowledgeable about modern cars	70%
Average knowledge	17%
Not knowledgeable about modern cars	13%

TABLE 5. ANSWERS TO THE Q3

	Q3
Personal data about the driver	69
Public data about the driver	56
Public data not about the driver	44
Special categories of personal data about the driver	15
Financial data about the driver	15
No data at all	1

ing and disagreeing category, according to their level of agreement/disagreement. Participants that selected neither agree nor disagree are reported as undecided. All findings discussed below refer to the Table 6, except question 3 which refers to Table 5. Question 2 has a high rate of agreement, so we can say that the participants agree that a modern car is similar to a modern computer. From question 3 we note that the predominant categories according to respondents are: “personal data about the driver” (selected by 69%); “public data about the driver” (selected by 56%); “public data not about the driver” (selected by 44%). Few participants think that their vehicle collects more sensitive data belonging to the special categories of personal and financial data (both 15%). Just one participant thinks that modern cars do not collect any data at all. Question 4 shows that 32% of the participants agree with the statement above, moreover 35% of them are undecided and 33% of them disagree with the statement. It seems that participants are somehow equally distributed. From the answers to question 5 it can be seen that just 26% of the sample agrees to the transmission of data over the Internet, 26% of participants are undecided while 48% of them disagree with the statement. This means that the sample in general is not very convinced to send personal data over the Internet. Question 6 shows that 80% of the participants agree with the above statement and only 6% disagree with the statement while 14% of them are undecided. The answers to question 7 show that a large part of the sample is undecided on this statement (41%), 22% of the participants agree with the statement while 37% disagree. Question 8 shows that the 50% of the participants agree with this statement while the 34% are undecided and the rest of them (16%) disagree. Question 9 shows that 52% of participants agree with this statement, 31% of them are undecided and 17% disagree with the statement. The last question (Q10) shows that just a minority of them (about 23%) answered negatively to the question. The majority (51%), agree that their data is protected using appropriate methodologies and techniques, indicating a good perception of security and trust on the part of drivers.

In summary, the interviewed sample feel quite informed about modern vehicles. The majority of participants agrees that systems and technologies present in modern cars are increasingly similar to modern computers. Regarding the collection of personal data, the participants seem to be equally divided between those who agree, those

TABLE 6. QUESTIONNAIRE RESPONSES IN PERCENT

	Q2	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Agreeing	83%	32%	26%	80%	22%	50%	52%	51%
Disagreeing	6%	33%	48%	6%	41%	34%	31%	23%
Undecided	11%	35%	26%	14%	37%	16%	17%	26%

who disagree and those who neither agree nor disagree. Moreover, according to the sample, the data collection is more oriented towards public and personal data, with no interest in financial information or special categories of personal data. The level of agreements regarding the collection of personal data may be due to the fact that drivers think it is neither necessary nor useful. The answers of question 9 tell us that half of the sample thinks that their data is analysed and studied by the vehicle systems in order to evaluate some personal aspects. This statement could have increased the level of disagreement of data collection, showing that there are some privacy concerns. Regarding the transmission of collected data just a few of the participants think it is truly necessary. This result could suggest that drivers have some privacy concerns about their personal data. In fact, considering the answers of question 5, almost half of the interviewed sample does not want personal data to be transmitted on the Internet by the vehicle. This kind of drivers may think that they have not enough control on their personal data once they are transmitted. Thus, it seems that drivers demonstrate to have some risk perceptions regarding their data. Recent attacks against car manufacturers that targeted drivers’ personal data [24], [25] could have influenced the drivers in this decision.

6. Related work and conclusions

In 2014, Schoettle and Sivak [26] surveyed public opinions in Australia, the United States and the United Kingdom regarding connected vehicles. Their research noted that people (drivers as well as non-drivers) expressed a high level of concern about the safety of connected cars, which does not seem surprising on the basis of the novelty of the concept at the time.

In 2016, Derikx et al. [27] investigated whether drivers’ privacy concerns can be compensated by offering monetary benefits. They analysed the case of usage-based auto insurance services where the rate is tailored to driving behaviour and measured mileage and found out that drivers were willing to give up their privacy when offered a small financial compensation. We argue that the international research community may not have fully realised the necessity and relevance of such a compliance beyond its sheer legal urgency. This is confirmed by the scant literature focusing on protecting drivers’ data. The “CANDY” attack reconfirms how data can be stolen following security weaknesses, which derive, in this particular case, from optimistic network isolation assumptions made at application layer [28].

A few works emphasise the overarching problem of how to effectively transmit the contents of a lengthy policy to people. Those wishing to use a service routinely accept the terms of the service provider without fully understanding them. As a result, users are not actually informed [29]. Furthermore, the literature referred to above

While there is some general awareness that treating people's personal data is essential to people's privacy and, consequently, freedom today, this paper showed that awareness to be very limited in the automotive domain. Car drivers' privacy concerns are lower than we think they should, especially given the quantity and quality of personal data that cars collect and manufacturers treat. This cannot be justified in terms of drivers' overall trust, which is found to be comparatively low. As a possible reason, privacy policies are found not to reach drivers well, indicating that the entire privacy area aboard modern cars demands immediate attention.

This research has received funding from COSCA (CONceptualising Secure Cars) [30], a project supported by the European Union’s Horizon 2020 research and innovation programme under the NGI TRUST grant agreement no 825618.

- [1] Brian Krzanich, “Data is the New Oil in the Future of Automated Driving.” <https://newsroom.intel.com/editorials/krzanich-the-future-of-automated-driving/>, 2016.
- [2] European Union, “General Data Protection Regulation (EU Regulation 2016/679).” <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2016:119:FULL>, 2016.
- [3] C. Miller and C. Valasek, “A survey of remote automotive attack surfaces,” *Black Hat USA*, 2014.
- [4] European Commission, “eCall in all new cars from April 2018.” <https://ec.europa.eu/digital-single-market/en/news/ecall-all-new-cars-april-2018>, 2020.
- [5] National Association of Insurance Commissioners, “TELEMATICS/USAGE-BASED INSURANCE.” https://content.naic.org/cipr_topics/topic_telematicsusage_based_insurance.htm, 2020.
- [6] Z. O. Abu-Faraj, W. Al Chamaa, A. Al Hadchiti, Y. Sraj, and J. Tannous, “Design and development of a heart-attack detection steering wheel,” in *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–6, 2018.
- [7] S. Negru, “Connected cars and in-car payments: the road so far and the road ahead.” <https://tinyurl.com/Simona-Negru-cars>, 2019.
- [8] Future of Privacy Forum, “Personal data in your car.” <https://fpf.org/wp-content/uploads/2017/01/consumerguide.pdf>, 2020.
- [9] Miro Enev, Alex Takakuwa, Karl Koscher, and Tadayoshi Kohno, “Automobile Driver Fingerprinting.” https://petsymposium.org/2016/files/papers/Automobile_Driver_Fingerprinting.pdf, 2016.
- [10] M. L. Bernardi, M. Cimitile, F. Martinelli, and F. Mercaldo, “Driver and Path Detection through Time-Series Classification,” *Journal of Advanced Transportation*, vol. 2018, pp. 1–20, 2018.
- [11] E. Union, “Guidelines 1/2020 on processing personal data in the context of connected vehicles and mobility related applications.” https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_202001_connectedvehicles.pdf, 2020.
- [12] Jeff Crume, “OwnStar: Yet another car hack.” <https://insidetheinternetsecurity.wordpress.com/2015/08/05/ownstar-yet-another-car-hack/>, 2015.
- [13] L. Constantin, “Researchers hack Tesla Model S with remote attack.” <https://www.pcworld.com/article/3121999/researchers-demonstrate-remote-attack-against-tesla-model-s.html>, 2016.

- [14] H. Bekker, "Q1/2019 Europe: Best-Selling Car Manufacturers and Brands." <https://www.best-selling-cars.com/europe/q1-2019-europe-best-selling-car-manufacturers-and-brands/>, 2019.
- [15] Shivam Bansal and Chaitanya Aggarwal, "Textstat 0.7.1." <https://pypi.org/project/textstat/>, 2021.
- [16] U. P. Cambridge, *Cambridge Advanced Learner's Dictionary & Thesaurus*. 2012.
- [17] W. Derguech, S. S. e. Zainab, and M. D'Aquin, "Assessing the readability of policy documents: The case of terms of use of online services," ICEGOV '18, (New York, NY, USA), p. 247–256, Association for Computing Machinery, 2018.
- [18] M. Coleman and T. L. Liao, "A computer readability formula designed for machine scoring," vol. 60, pp. 283–284, 1975.
- [19] G. McLaughlin, "Smog grading a new readability formula," vol. 12, no. 8, pp. 639–646, 1969.
- [20] R. Senter and E. Smith, *Automated Readability Index*. AMRL-TR-6620, 1967.
- [21] R. Flesch, *How to Write Plain English*. University of Cambridge, 2016.
- [22] Eurostat, "Educational attainment statistics." https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Educational_attainment_statistics, 2021.
- [23] Formplus, "The 4,5, and 7 Point Likert Scale + [Questionnaire Examples]." <https://www.formpl.us/blog/point-likert-scale>, 2020.
- [24] C. Cimpanu, "Toyota announces second security breach in the last five weeks." <https://www.zdnet.com/article/toyota-announces-second-security-breach-in-the-last-five-weeks/>, 2019.
- [25] T. Robinson, "BMW customer database for sale on dark web." <https://www.scmagazine.com/home/security-news/bmw-customer-database-for-sale-on-dark-web/>, 2020.
- [26] B. Schoettle and M. Sivak, "A survey of public opinion about connected vehicles in the U.S., the U.K., and Australia," in *2014 International Conference on Connected Vehicles and Expo (IC-CVE)*, pp. 687–692, IEEE, 11 2014.
- [27] S. Derikx, M. de Reuver, and M. Kroesen, "Can privacy concerns for insurance of connected cars be compensated?," *Electronic Markets*, vol. 26, pp. 73–81, Feb 2016.
- [28] G. Costantino et al., "CANDY: A Social Engineering Attack to Leak Information from Infotainment System," in *IEEE 87th VTC*, 2018.
- [29] R. Pardo and D. Le Métayer, "Analysis of privacy policies to enhance informed consent," in *Data and Applications Security and Privacy XXXIII*, Springer, 2019.
- [30] COSCA Team, "Conceptualising Secure CARS (COSCA) Website." <https://cosca-project.dmi.unict.it/>, 2020.

The core questionnaire questions are listed below.

0. How many hours a week do you drive a car?
- 3-6 hours; ○ 7-9 hour; ○ 10-12 hours; ○ 13-15 hours; ○ 16-20 hours; ○ 21+ hours
- 1) Are you knowledgeable about modern cars?
- Not at all □ – □ – □ – □ – □ – □ – □ Very knowledgeable about modern cars
- 2) How much do you agree with the following statement: a modern car is similar to a modern computer.
- Strongly disagree □ – □ – □ – □ – □ – □ – □ Strongly agree

- [illegible]