Privacy and modern cars through a dual lens

Giampaolo Bella*, Pietro Biondi*, Marco De Vincenzi♦, Giuseppe Tudisco♦
giamp@dmi.unict.it, pietro.biondi@phd.unict.it, marco.devincenzi@iit.cnr.it, giuseppe.tudisco@studium.unict.it

* Dipartimento di Matematica e Informatica, Università di Catania, Catania, Italy
♦ Istituto di Informatica e Telematica, Consiglio Nazionale delle Ricerche, Pisa, Italy
Automotive cybersecurity

COMMUNICATION DOMAINS

V2V V2I IV U2V Car Maker Third parties Emergency services

CAR COMPONENTS
- CAN-Bus
- OBD-II port
- Sensors
- Electronic Control Units
- Autonomous vehicle imaging

CAR COMPONENTS
- Infotainment system
- E-call box / SIM Card
- GPS
- Smartphone integration
- Cabin monitoring system

TYPES OF DATA

VEHICLE: functioning of vehicle, maintenance status, ECU data and operations

DRIVER: physical characteristics, driving style, driver's behaviour

LOCATION: geographic location of a vehicle, history and route tracking, speed

ACCOUNT: personal accounts, personal and special category

Bella, Biondi, De Vincenzi, Tudisco
Art 12. GDPR “The controller of the data, has to provide information to the user in a concise, transparent, intelligible and easily accessible form, using clear and plain language”

Focus: Understand if the privacy policies are compliant with article 12 of the GDPR

Approach: Analysis of the privacy policies of twelve car manufacturers and study of readability indices evaluating the policy

Steps:

1. Privacy policy collection
2. Policy readability analysis
Privacy policy collection

Twelve carmakers company: the top ten most famous car companies in Europe plus Tesla and KIA.

We download the privacy policy by using two different channels:

- During the installation of the respective app
- From the company website and contacting the customer service
Policy readability analysis
Textstat, a Python library to calculate statistics from text, that allows also to compute readability indexes.

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.

| Score/Grade | Education Level       |
|-------------|-----------------------|
| 1-4         | Elementary School     |
| 5-8         | Middle School         |
| 9-12        | High School           |
| 13-16       | Undergraduate         |
| 17+         | Graduate              |
• **CLI**: depends on the complexity of the words, measured from the number of letters, and the complexity of the sentences.

• **SMOG**: uses the polysyllables (words of 3 or more syllables) in a certain number of sentences (at least 30).

• **ARI**: takes into consideration also the number of characters, in addition to the number of words and sentences.

• **FREI**: that differs from all the three previous indexes because it outputs a score instead of the school grade.

• **GIDR**: 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.
| Company      | Number of Words | Number of Sentences | CLI  | SMOG | ARI  | FREI  | GIDR |
|--------------|-----------------|---------------------|------|------|------|-------|------|
| Ford         | 9744            | 1327                | 8.7  | 10.0 | 5.1  | 58.3  | 0.0  |
| Peugeot      | 2151            | 437                 | 9.0  | 9.0  | 6.0  | 47.7  | 3.6  |
| Kia          | 22043           | 3096                | 9.6  | 10.4 | 5.8  | 49.8  | 10.7 |
| Skoda        | 5831            | 860                 | 9.9  | 9.9  | 6.1  | 49.7  | 12.5 |
| Mercedes     | 8591            | 1387                | 10.0 | 10.0 | 6.0  | 44.0  | 14.3 |
| Opel         | 2438            | 323                 | 11.0 | 10.0 | 7.0  | 46.6  | 21.4 |
| Audi         | 13661           | 1410                | 10.4 | 11.1 | 6.8  | 49.4  | 23.2 |
| BMW          | 991             | 119                 | 11.1 | 10.5 | 7.1  | 49.1  | 25.0 |
| Tesla        | 13224           | 1453                | 10.8 | 11.1 | 7.0  | 49.5  | 25.0 |
| Volkswagen   | 11742           | 1206                | 12.2 | 11.8 | 8.3  | 42.1  | 42.9 |
| Toyota       | 3279            | 263                 | 11.7 | 13.2 | 8.8  | 43.6  | 48.2 |
| Renault      | 2568            | 94                  | 12.7 | 17.3 | 15.9 | 37.3  | 100.0|
Questionnaire design:
- Basic information
- Capturing concerns on privacy
- Capturing perceptions of trust

Participant
- Choice of the subjects
- Sample of 88 respondents

Essential sample statistics
Explanation of results and presentation in tables
A 7-point balanced Likert scale is used for most of the questions in the questionnaire.

Simplifying the analysis: grouping the 7 levels of agreement into 3 categories

| Levels of agreement       | Categories      |
|---------------------------|-----------------|
| Strongly agree            | Agreeing        |
| Agree                     |                 |
| Somewhat agree            |                 |
| Neither agree nor disagree| Undecided       |
| Somewhat disagree         |                 |
| Disagree                  | Disagreeing     |
| Strongly disagree         |                 |
● **Q1:** evaluates the driver's knowledge on modern cars.
● **Q2:** asks respondents whether or not they agree that modern cars are similar to modern computers.
● **Q3:** select all the categories of data they think a car collects.
● **Q4:** asks whether they think it is necessary to collect personal data to achieve full vehicle functionality.
● **Q5:** asks whether personal data collected by a modern car about its driver needs to be transmitted over the internet.
● **Q6:** asks whether participants agree that a modern vehicle safeguards the life of its driver.
● **Q7:** asks participants whether a modern car protects its driver's personal data better than its driver's life.
● **Q8:** asks whether the data collected from the vehicle is legitimately processed according to the relevant regulations.
● **Q9:** asks if participants believe that the personal data collected is systematically analysed and evaluated using automated processes (including profiling).
● **Q10:** asks whether the participants feel that the data transmitted over the Internet are protected by adequate technologies.
## Analysis of findings

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

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

- Reading and understanding the twelve privacy policy documents requires 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.

- The interviewed sample feel quite informed about modern vehicles.

- Collection of personal data, the participants seem to be equally divided.

- Q9 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.

- Regarding the transmission of collected data just a few of the participants think it is truly necessary.
Privacy and modern cars through a dual lens

Thank you for your attention

Giampaolo Bella*, Pietro Biondi*, Marco De Vincenzi†, Giuseppe Tudisco*

giamp@dmi.unict.it, pietro.biondi@phd.unict.it, marco.devincenzi@iit.cnr.it, giuseppe.tudisco@studium.unict.it

* Dipartimento di Matematica e Informatica, Università di Catania, Catania, Italy
† Istituto di Informatica e Telematica, Consiglio Nazionale delle Ricerche, Pisa, Italy