Towards Privacy-Friendly Smart Products

Kimberly García*‡, Zaira Zihlmann†§, Simon Mayer*¶, Aurelia Tamò-Larrieux *‖, Johannes Hooss***
*University of St.Gallen, Switzerland
†University of Lucerne, Switzerland
ORCID: ‡0000-0002-4971-2944, §0000-0002-3592-1606
¶0000-0001-6367-3454, ‖0000-0003-3404-7643, **johannes.hooss@student.unisg.ch

***Abstract*—Smart products, such as toy robots, must comply with multiple legal requirements of the countries they are sold and used in. Currently, compliance with the legal environment requires manually customizing products for different markets. In this paper, we explore a design approach for smart products that enforces compliance with aspects of the European Union's data protection principles within a product's firmware through a toy robot case study. To this end, we present an exchange between computer scientists and legal scholars that identified the relevant data flows, their processing needs, and the implementation decisions that could allow a device to operate while complying with the EU data protection law. By designing a data-minimizing toy robot, we show that the variety, amount, and quality of data that is exposed, processed, and stored outside a user's premises can be considerably reduced while preserving the device's functionality. In comparison with a robot designed using a traditional approach, in which 90% of the collected types of information are stored by the data controller or a remote service, our proposed design leads to the mandatory exposure of only 7 out of 15 collected types of information, all of which are legally required by the data controller to demonstrate consent. Moreover, our design is aligned with the Data Privacy Vocabulary, which enables the toy robot to cross geographic borders and seamlessly adjust its data processing activities to the local regulations.**

***Index Terms*—Data Protection, GDPR, Smart Products, Internet of Things, Data Privacy Vocabulary.**

I. Introduction

In a globalized world, products move across geographic regions, changing jurisdictions that make them subject to heterogeneous or even incompatible legal environments. This traditionally leads to the creation of multiple product variants that need to be customized and managed by the manufacturer to serve those different markets [1]. The European Union (EU) is an important marketplace for technology, including social robots, and EU manufacturers are required to ensure that their variant management complies with EU standards [2]. Among these is the General Data Protection Regulation (GDPR). The GDPR requires a specific legal ground to be fulfilled in order to process personal data in accordance with the fundamental principles of data protection law. Furthermore, the GDPR contains a norm on privacy by design and default, which mandates that Data Controllers (DCs) ensure, via technical and organizational measures, that they comply with the fundamental principles of data protection law (Article 25 GDPR).

However, the fundamental principles of data protection law are challenged by the data processing capabilities of toy robots. Such robots are equipped with various privacy-sensitive sensors, e.g., microphones and cameras, that are continuously processing personal data. The robots can easily be moved from one jurisdiction to another and can affect individuals, including vulnerable user groups such as children, in their private homes.

In light of this and the fact that implementing privacy-by-design requires a case-by-case analysis along with the tailoring of measures to a concrete use case [3], we have taken a toy robot as a use case to investigate how the legal environment of such a smart product can be reflected in its firmware implementation. We explore how this firmware can be designed to enable automatically enforcing its compliance with data protection principles by adapting, at run time, its data processing activities. This has the potential to not only increase the compliance of smart devices with their legal environment, but also to simplify variants management for products that are sold and used in multiple jurisdictions. While various data protection principles apply in our scenario, we focus on the implementation of a toy robot capable of complying with the consent requirements of data protection law, as with smart toys, consent seems to be manufacturers' preferred legal basis for processing personal data [4].

In the following, we introduce our proof-of-concept implementation of the toy robot firmware and compare it with a traditional software engineering approach.

II. Case Study: An Educational Toy Robot

Our toy robot was designed as a (mock) educational tool for young children. This robot roams private family premises, takes pictures of its environment every few seconds, and analyzes them to identify any known individuals (typically children) in its view. If an individual has been identified, the robot stops to perform an educationally valuable action, such as playing a song that motivates the identified individual to sing along to pre-selected personalized content. This content would be relevant for the children's age and current interests. Thus, to operate, the toy robot needs to record, process, and store several pieces of data, including personal data.

A. Proof of concept implementation

Our proof of concept uses a *Dexter Industries GoPiGo3* learning robot based on a Raspberry Pi 3 running *Raspbian for Robots* as operating system. To provide the robot with facial recognition and localization capabilities, a Raspberry Pi camera and an Adafruit GPS antenna were added to the assembly. The former allows the robot to customize the content it delivers, while the latter is used to self-localize to

comply with local regulations. On the software side, a Web application hosted locally was created to act as a first privacy friendly proxy between the robot and its users. Through this Web application, the user is informed about data processing and is able to provide consent. Moreover, online and offline image processing options were implemented to visually identify users. The online implementation uses Google AutoML Vision[1], which provides a straightforward way to process images in the cloud. It requires a minimum of 10 training images per user. However, to avoid overfitting and to get consistent predictions, we upload 30 images per user, which were augmented through affine transforms of the images, resulting in a total of 3000 images per user. For the offline processing option we used the Tensorflow implementation of the InceptionV3 neural network[2], removing the top neural network layer to train it with images of the users of our robot.

The implementation of our toy robot allows it to roam flat premises, capturing pictures and processing them locally or remotely every five seconds. When the robot identifies a user, using one of the implemented systems, with $>95\%$ accuracy, it plays the user's favorite song. If no user is recognized, the robot keeps roaming around and taking more pictures. To adhere to the data minimization principle, each picture the robot takes is replaced after 5 seconds by a new one.

III. What Might Be: Traditional Design of a Toy Robot

The following is a "what-might-be" analysis of a toy robot that has been designed following a conventional software engineering approach. As described in Section II-A, we have designed the educational toy robot ourselves to be in control of the data it exposes and their flows. However, products in the market collect similar types of personal data to provide similar services [5]. Thus, our robot use case can be considered a realistic scenario. Moreover, to stay as close as possible to a realistic, market-oriented system, we have adopted common practices from popular connected toys.

We begin with the toy robot setup process and show its flow of data[3]: After unpacking the toy robot, a user is asked to set up an online account with the robot's supplier (i.e., the DC according to GDPR), which requires the user to enter an email address and password as well as to complete a 5-min demographic survey (e.g., household members' names, genders, ages, and interests). The user then receives a confirmation email from the website. Subsequently, the user is prompted to enter their robot's serial number and assigns a name to it. Next, the user is asked to connect the robot to their home WiFi. To accomplish this, the user switches the robot to a special setup mode that allows the user to connect to the robot's WiFi. Once connected, the user is directed to the robot's locally hosted home page to enter the credentials for their home WiFi. Then, the robot connects to the home WiFi. The user then finalizes the setup through the Website by entering (for each individual to be recognized) a pseudo-identifier along with a set of pictures of the individual, and configures the robot to provide personalized responses for each recognized individual. This information is then linked to the household members and the training pictures and pseudo-identifiers are sent to a third-party facial recognition service for training (i.e., in our implementation Google Vision API).

Once the setup process has concluded, the robot starts roaming around the user's home. Every five seconds, the robot takes a new picture and uploads it to the DC's website via WiFi. The DC then forwards the picture to a facial recognition service, obtains the identity of the identified individual, and sends this information to the robot. Since the DC aims to become independent of the third-party facial recognition service, it retains all uploaded pictures to bootstrap and improve its own facial recognition algorithm. Finally, the robot performs an action based on the personalized configuration. Moreover, the robot uploads (to the DC) a daily usage report that includes, among others, data on the duration of usage, distance travelled, GPS location, number of classified individuals with classification reliability estimates, and system errors. This data is analyzed by the DC to identify future improvements to the device and the service, and to address further monetary exploitation opportunities. Table I provides an overview of the pieces of data that are processed during the robot setup process and in its subsequent operations. For instance, the table shows that the home network credentials (#7 in Table I) are the only piece of application-relevant data that does not leave the robot. Moreover, the business strategy of the DC is strongly reflected in the system design, since it shares as little data as possible with the Remote Service (RS) attempting to become independent of it. To achieve this, the DC collects as much information as possible from the user and stores it, ideally indefinitely, to keep improving its product and services, and to enable it to further monetize the acquired information.

IV. Legal Considerations for Data Collection and Processing

To design a toy robot that is compliant with the GDPR, there are several legal considerations that should be addressed in its design and implementation. In the following, we focus on the characteristics of consent.

A. Consent should be Timely

While most of the consent requirements demand a case-by-case analysis, the timing requirement triggers a straightforward engineering implication, since consent must be obtained prior to the data processing activity [6]. In the case of smart products, it is thus key that 'set-up-notices' are provided to the user upon initial use of the device [7]. While the focus hereinafter is on the process of obtaining consent, it is important to bear in mind that the data subject has the right to withdraw consent. The data subject must be informed thereof

[1] https://cloud.google.com/vision/automl/docs

[2] https://www.tensorflow.org/api_docs/python/tf/keras/applications/inception_v3/InceptionV3

[3] This process is aligned with the setup of popular connected toys such as the Tonies ecosystem, see https://tonies.de/

TABLE I
OVERVIEW OF DATA PROCESSED BY A TOY ROBOT DESIGNED WITH A TRADITIONAL SOFTWARE ENGINEERING APPROACH WHERE DATA LOCATION IS *Local* (GREEN), *Data Controller, DC* (ORANGE), OR *Remote Service, RS* (RED).

#	Data item	Type	Location	Comments
1	User Email Address	String (UTF-8)	DC	The email address is recorded by the DC to enable the user to log in to its portal.
2	Robot Password	String (UTF-8)	DC	The password is recorded by the DC to enable the user to log in to its portal.
3	Demographic Survey	String (UTF-8)	DC	The demographic data is recorded by the DC.
4	Confirmation	Boolean	DC	The confirmation is recorded by the DC to complete the sign-up process (safeguarding from spamming).
5	Serial Number	String (UTF-8)	Local + DC	The serial number is recorded by the DC to link a specific robot to a specific user profile in the portal.
6	Robot Name	String (UTF-8)	DC	The robot name is recorded by the DC for user experience purposes.
7	Home Network Credentials	2 Strings (UTF-8)	Local	The home network credentials are stored locally on the robot.
8	Pseudo-Identifier (per Data Subject)	String (UTF-8)	Local + DC + RS	The DS Identifier is used across all system components.
9	Training Images (per DS)	jpg Files	DC + RS	The training images are used to train the third-party facial recognition service and are stored by the DC to establish its in-house facial recognition service.
10	Robot Response (per DS)	Any (e.g., mp3)	Local + DC	The desired response of the robot to each recognized individual is stored locally and by the DC.
11	Recorded Picture (every five seconds)	jpg File	Local + DC + RS	The recorded images are sent to the third-party facial recognition service and are stored by the DC to establish and improve its in-house facial recognition service.
12	Usage Report (every day)	Any (e.g., Strings)	Local + DC	The daily usage report is used for improvements to the device and service and for further monetary exploitation.

prior to giving consent and withdrawing consent should be as simple as giving it [6].

B. Consent should be Freely Given

Following Recital 42 of the GDPR, consent will not be regarded as freely given when the data subject is unable to refuse or withdraw it without detriment. Moreover, consent should be obtained separately from the general terms and conditions [8].

C. Consent should be Specific and Unambiguous

The specificity criterion prohibits blanket consent and demands the granting of consent linked to a particular processing purpose [9] [10]. When a service involves different processing operations for multiple purposes, the data subject should have the possibility to express its consent separately for each data processing operation and should not be forced to agree to consent in an all-or-nothing decision [8]. To provide for such granular consent, the options for consent should be distinctive and thereby allow separate consent for different purposes and types of processing [6]. This can be accomplished through separate forms or by ticking individual check-boxes [8]. These check-boxes must not be pre-ticked, as consent needs to be unambiguous, i.e., given through an active motion or declaration by the user [11].

D. Consent should be Informed

In order to give informed consent, the data subject should, according to Recital 42, 'be aware at least of the identity of the controller and the purposes of the processing for which the personal data are intended.' To ensure that the information provided in the consent form can be understood by laypersons, it should be presented in an intelligible and easily accessible way [7], the language used should be clear and plain and should preferably be the local language [12].

E. Consent should be Explicit

The processing of special categories of data, such as biometric data, requires not only unambiguous but explicit consent [6]. In our use case, facial recognition technology processes data in order to uniquely identify the respective individual. Hence biometric data is being processed [13], requiring explicit consent. Explicit consent means expressed consent, i.e., the data subject must be requested to agree to a particular use of the data and must actively reply to the question in the affirmative [10]. The clearest way to ascertain that consent is explicit would be to obtain express confirmation of consent in a written and signed statement [11]. However, it is also possible to gain explicit consent by other means, for instance by having the data subject sign his or her name beneath the statement or tick a box next to it [8].

F. Further Requirements on Consent

Because the DC bears the burden of proof that valid consent was obtained, the consent process needs to be documented by the DC [6]. Thus, it is often recommended that the DC uses a double opt-in procedure when obtaining consent. Double opt-in requires that users first declare consent, e.g., by ticking a

checkbox or entering their email address. The data subject then receives a verification hyperlink via email or another electronic messaging service. By following the verification link, the data subject re-confirms consent [9] [14]. Scholars have advised to store the declaration of consent together with the name of the data subject or another reliable identifier (email address, etc.) and the time of the consent ('timestamp') [9]. The obligation to provide proof of consent exists for as long as the data processing activity in question persists [6].

With Article 8 of the GDPR requiring specific protection for children with respect to their personal data, parental consent must be obtained if connected devices, such as toys that target children, process children's personal data [14] [15]. Where the toy is offered directly to a child, Article 8(1) states that the child's consent is only valid if the child is at least 16 years old. However, Article 8(1) allows member states to set a lower age threshold, provided that it is not below 13 years. Therefore, the age threshold is fragmented throughout the EU and service providers must comply with the different age thresholds of the member states [14]. Even if Article 8 does not demand the controller to verify the age of the child, it is implicitly required [11] [14]. If the user indicates that he or she is above the respective age threshold, the controller should carry out an appropriate verification process to check if the indicated information is true [11]. When choosing an age verification mechanism, the DC should keep in mind that the privacy of children and other users can be put at risk by requiring the collection of additional personal data [15]. Thus, in some low-risk situations, it may be sufficient to require a user to indicate his year of birth or to declare that she is above a certain age [11].

Where the child is below the age threshold, Article 8(2) requires that 'the controller shall make reasonable efforts to verify that consent is given or authorised by the holder of parental responsibility over the child.' Verification of parental responsibility via email might be sufficient in low-risk cases, whereas in high-risk cases the DC should require more evidence in order to verify and retain the information [11]. In this respect, various measures can be taken, e.g., transmission of a document signed by the holder of parental responsibility or provision of a copy of a parent's government-issued identity card [15]. When deciding on a measure, the DC must avoid that the chosen technology leads to an excessive collection of personal data [11].

V. Data-Privacy-Compliance-First Design of a Toy Robot

To design and implement a toy robot that considers data protection law, the data flows identified on the previous analysis are augmented with a legal layer that guides the data processing practices to comply with GDPR. Thus, the user first unpacks the robot, starts its controller, and connects it to his or her computer. At start-up, the robot activates its on-board GPS module and acquires its physical location, i.e., a GPS coordinate. From this location, the robot computes its current logical location (i.e., a country code). The determined country code is stored on the device. The robot furthermore launches a local Web server. At this point, the robot does not have its own Internet connection, thus the user is prompted for the credentials of the user's home wireless network (i.e., SSID and password) so the robot can communicate with remote services. Next, the user directs a Web browser to the (fixed) URL that is connected to the local Web server to access the robot's welcome and configuration dashboard.

Through the configuration dashboard, the user accesses the consent forms—*consent is timely*, which are made available in the local language by default (based on the determined country code). The system also informs users that they may withdraw consent or modify their consent settings at any time via the robot's configuration page—*consent is informed*. When entering this configuration page for the first time, the user is prompted to change the system's password to ensure that the locally stored data is secure. Next, the user is prompted to enter pseudo-identifiers of the data subjects (i.e., of the individuals that shall be recognized by the system). The user is encouraged to enter pseudonyms or nicknames instead of the real names of the data subjects. Then, for each named data subject the following *consent process* is executed:

- The subject is prompted to enter its date of birth; the user is instructed that the real date of birth is required. The parental consent requirement is then based on the entered date of birth and the local jurisdiction regarding the legal protection age of the data subject.
- An additional form is displayed later in the process if parental consent needs to be obtained.
- The user provides facial recognition consent, stating that he or she is informed about the purpose of data processing—*consent is specific and unambiguous*. The toy robot data processing is restricted to facial recognition capabilities. Other processing capabilities (e.g., audio for voice recognition) would need to be consented to by the data subject through a consent process similar to the facial recognition consent process, but using a separate from. To ensure *freely given consent*, the user is asked to choose between three options regarding the processing of the pictures the robot takes: *No processing*, *offline processing*, and *online processing*. Short statements inform the user about the consequences of selecting either of the options, including that the accuracy of *offline processing* (i.e., on the robot) is generally inferior to *online* facial recognition, and that *no processing* of data will provide the subjects with generic-non personalized content.
- After selecting one of the processing options, the user is prompted again to affirm the online or offline processing of images taken by the robot.
- The robot then ask the user for his or her email address, it then sends an email containing a hyperlink to verify the consent settings. This hyperlink contains the serial number of the robot together with a hash of the data subject's consent choices. The user clicks the received hyperlink to confirm the consent settings—*consent is*

explicit. Finally, user consent settings are stored locally on the robot as well as on the DC's back-end, allowing the DC to demonstrate consent of the data subject. This concludes the consent process for each data subject.

In case a user wants to update their consent settings later on, the previous process needs to be repeated. If the data subject opts for online or offline processing of images, the system finally prompts the user to upload training images to the robot, for each to-be-recognized data subject. Depending on the data subjects' (collective) choices, the robot then either uses these images to train a local classification model or uploads them to an online image classification service. When all data subjects agree to using an online classification service, the user is redirected to the welcome page which displays a 'Start' button.

Once configured, the robot roams the user's premises. In case the user selected offline processing of images, the robot takes a new picture every five seconds and uses its local classification model, which returns the classified category (i.e., the name of the data subject that was configured before). If the user selected online processing of images, new pictures are instead submitted to an online classifier which returns the category. In both cases, the robot's response is to perform a personalized action for the identified individual (e.g., playing a music file). If the user gave no permission to process recorded images, the robot does not take pictures and does not activate the camera-relevant features of its software and hardware.

A. Data Privacy Vocabulary

The W3C Data Privacy Vocabulary (DPV)[4] was used as a foundation to design and implement our toy robot that is compliant with GDPR. DPV provides a machine-readable representation of terms relevant to personal data handling, in adherence to the GDPR. This vocabulary is a specification that (if broadly standardized) could become a strong basis for software engineers and designers to better understand the GDPR. So they can, together with legal experts, making implementation decisions that best suit a device's operations while complying with the law. Moreover, since the machine-readable representation of this vocabulary is an ontology, it could be seamlessly linked to other legal regulations; for example geographic, opening the possibility of having a robot that remains compliant even when moving within jurisdictions. On this basis, the robot could for instance autonomously request local restrictions regarding parental consent from a Web service to obtain consent from users in a compliant way.

Similar to Table I, in which the data flows for a traditionally designed toy robot are documented, Table II shows the data processed by our implementation of the toy robot that follows a data-privacy-first approach. For each data item, its storage location (local or remote) is shown along with the mapping of the item to the DPV vocabulary and a comment. In comparison to the robot designed following a traditional approach, our proposed design clearly minimizes the variety, amount, and quality of data that leaves the premises of the user: Instead of storing over 90% of the collected types of information (some of which point towards large amounts of data, e.g. 'usage reports') the privacy-by-design robot uploads under 70% of the collected types of information. Moreover, the user is required to explicitly consent to the upload of three types of information and five of them are only collected optionally.

Regarding the information types that are processed locally, three of them (#1-3 in Table 2) are used to facilitate the consent process for users and to better secure the user's local data. These three items are either processed transiently or stored encrypted; the other two (#12 and #15 in Table 2) are optional in this design, although #15 is required to enable the robot to fulfill its application purpose. All the information types that are uploaded to the DC (#4-10 in Table 2), are required by the DC to demonstrate consent. Regarding the information types shared with the third-party service, the derived category (#14) has been decoupled from the data subjects' pseudo-identifiers and the usage of the service has been made optional, by outfitting the robot with an additional local (i.e., offline) facial recognition system. In case of such offline processing, the robot uses the training images to train its algorithms, but does not store them. If the user selects online processing, these images are uploaded to the DC and the remote classification service, but again neither stored locally nor by the DC. The data items used to personalize content for a specific user (#11,#13-#14) are always optional, given that a user can select to play pre-defined content according to the child's age.

VI. Limitations

The focus of this paper lies on the implementation of the consent requirements set out by the GDPR. It is important to keep in mind that consent constitutes an important legal basis but is also a rather controversial concept. A key issue in this respect is that data subjects, both adults and children, encounter problems to understand how their data are processed, what potential risks are involved, and how to counterbalance these risks against the advantages associated with the processing. Consequently, there are concerns about the ability of data subjects to provide freely given and informed consent [16]. Nonetheless, the GDPR upholds consent as one legal basis, yet has tightened up the requirements for establishing valid consent [10]. The way these tightened requirements are implemented in the consent process executed in our case study seems reasonable from a compliance point of view, as users are, for example, given the choice between three options for processing the images captured by the robot, informed about the consequences of choosing one of them, and have the possibility to change the chosen options via the configuration dashboard. However, from the user's experience perspective, this consent process could be burdensome as the user is asked to give consent multiple times, which is likely to result in spending time setting up and operating the toy robot. Consequently, additional research on privacy-friendly smart products should consider these experiences at the early stages of designing for legal compliance [17], and should, where applicable, enhance the remote processing of data

[4] https://dpvcg.github.io/dpv/

TABLE II

OVERVIEW OF DATA PROCESSED BY A TOY ROBOT DESIGNED TO CONSIDER DATA PRIVACY LAW WHERE DATA LOCATION IS *Local* (GREEN), *Data Controller, DC* (ORANGE), OR *Remote Service, RS* (RED).

#	Data Item	Type	Location	DPV Relevant Concept	Comments
1	GPS Coordinate	3 Floating-point numbers	Local	dpv:GPSCoordinate	Since GPS is a self-localization system, the robot's location is not learned by any external entity. This data is stored transiently only.
2	Country Code	2 Characters (UTF-8)	Local	dvp:Country	The robot uses a reverse geo-coding algorithm that is implemented locally and therefore does not require any information to leave the device. This data is stored transiently only.
3	Robot Password	String (UTF-8)	Local	Not related to the individual, therefore no GDPR relevance	The robot password is stored locally, encrypted.
4	Pseudo-Identifier (per Data Subject)	String (UTF-8)	Local + DC	dpv:UserName	The DS Identifier is recorded by the DC as part of the consent process.
5	Date of Birth (per DS)	3 Integer numbers	Local + DC	dpv:age	The date of birth of each DS is required to enable the system to comply with its data protection obligations regarding parental consent.
6	Selected Option (per DS)	Short Integer number	Local + DC	dpv:PrivacyPreference	The selected processing option for images is recorded by the DC to enable it to demonstrate consent.
7	Affirmation (per DS)	Boolean	Local + DC	dpv:PrivacyPreference	The affirmation is recorded by the DC to enable it to demonstrate consent.
8	Email Address (per DS)	String (UTF-8)	DC	dvp:EmailAddress	The email address is recorded by the DC to enable it to demonstrate consent and to enable unambiguous consent.
9	Serial Number	String (UTF-8)	Local + DC	dpv:MACAddress	The serial number is recorded by the DC (together with the consent settings) to enable it to demonstrate consent.
10	Confirmation (per DS)	Boolean	DC	dpv:Consent	The confirmation is recorded by the DC to enable it to demonstrate consent.
11	Training Images (per DS)	jpg Files	Local or RS (depends on user setting)	dpv:Picture	In case of local classification, the training images are used to train the local classifier and are subsequently deleted. In case of the usage of an online service for classification, the images are transmitted to that service.
12	Home Network Credentials	2 Strings (UTF-8)	Local	dpv:Password	The home network credentials are stored locally on the robot and deleted whenever the user consent settings are modified and this information is not anymore required (i.e., when the user chooses a processing option different from online processing).
13	Recorded Picture (every five seconds)	jpg File	Local or RS (depends on user setting)	dvp:Picture	In case of local processing, recorded pictures are stored locally until they are successfully classified. Then they are deleted. In case of online processing, recorded pictures are transmitted to the remote service.
14	Derived Category (every five seconds)	Integer number	Local or RS (depends on user setting)	dpv:Identifying	In case the image classification process takes place using a local and locally trained model (our implementation is based on an Inception-v3 Convolutional Neural Network), the classified category is stored locally to trigger the robot response to that individual. In case the image classification takes place using an online service (in our implementation, the Google Vision API is used), the classified category is derived and transmitted by that service. The service thus has access to this information. The category is decoupled from the pseudo-identifier to ensure anonymity.
15	Robot Response (per DS)	Any (e.g., mp3)	Local	dpv:FavoriteMusic	The desired response of the robot to each recognized individual is stored locally.

with approaches such as federated learning [18], differential privacy [19], or secure multi-party computation [20] to balance the product's privacy/experience trade-off. The aim of our proof-of-concept is to demonstrate that even with a simple implementation — which is the one more likely to be selected by manufacturers of inexpensive articles — a more privacy-friendly product can be achieved by having engineers and legal experts collaborate in the design of smart products.

VII. Conclusion

In this paper, we present the use case of a toy robot's firmware that follows the GDPR data protection principles. Thus, we carefully analyzed the different data flows needed for the robot to provide its functionalities: recognizing children to play educational content tailored to their age and preferences. This analysis led us to implement variants of the functionality that require more or less data and different levels of data exposure (i.e., online vs. offline processing), enabling granular and specific consent of users to the handling and processing of their data. The consent forms were designed to be informative and easily understood. Additionally, our implementation uses the DPV vocabulary as a foundation to later integrate other regulations, and to enable the automatic compliance with laws when the robot is moved between jurisdictions.

As the toy robot shows, even when minimal data is required to provide a personalized service, e.g., playing educational content, it is undoubtedly necessary that designers and software engineers establish an interdisciplinary dialogue with legal professionals to ensure that functionality requirements are appropriately balanced with data privacy requirements. Although this incurs additional effort, our work shows that it can lead to a considerable reduction in the amount of personal data that is shared with the DC and other third parties. Furthermore, smart products designed in this way could flexibly adapt to a user's given consent, which could reduce the amount of data shared with the DC while preserving essential user experience aspects, albeit at lower quality.

Furthermore, the proposed design provides an opportunity to promote transparency. By adopting privacy-by-design and utilizing machine-readable law vocabularies, standardized and understandable documentation could be made available to end users to learn about the way their data is handled.

References

[1] B. Avak, "Variant management of modular product families in the market phase," Ph.D. dissertation, ETH Zurich, 2006.

[2] European Commission, "Statement by executive vice-president margrethe vestager on the launch of a sector inquiry on the consumer internet of things," 16 July 2020. [Online]. Available: https://ec.europa.eu/commission/presscorner/api/files/document/print/en/speech_20_1367/SPEECH_20_1367_EN.pdf

[3] A. Tamò-Larrieux, *Designing for privacy and its legal framework*. Springer, 2018.

[4] A. McStay and G. Rosner, "Emotional artificial intelligence in children's toys and devices: Ethics, governance and practical remedies," *Big Data & Society*, vol. 8, no. 1, pp. 1–16, 2021.

[5] J. Lau, B. Zimmerman, and F. Schaub, "Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, pp. 1–31, 2018.

[6] European Data Protection Board, "Guidelines 05/2020 on consent under regulation 2016/679: Version 1.1."

[7] F. Schaub, R. Balebako, A. L. Durity, and L. F. Cranor, "A design space for effective privacy notices," in *The Cambridge handbook of consumer privacy*, E. Selinger, J. Polonetsky, and O. Tene, Eds. Cambridge: Cambridge University Press, 2018, pp. 365–393.

[8] Information Commissioner's Office, "Consent." [Online]. Available: https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/consent-1-0.pdf

[9] S. Dienst, "Lawful processing of personal data in companies under the general data protection regulation," in *New European general data protection regulation*, T. Kugler and D. Rücker, Eds. München and Baden-Baden and Oxford and München: C.H. Beck and Nomos and Hart and Verlag C.H. BECK, 2018.

[10] L. A. Bygrave and L. Tosoni, "Article 4(11). consent," in *The EU general data protection regulation (GDPR)*, C. Kuner, L. A. Bygrave, L. Drechsler, and C. Docksey, Eds. Oxford: Oxford University Press, 2020.

[11] Article 29 Working Party, "Guidelines on consent under regulation 2016/679': Wp 259 rev.01."

[12] J. Taeger, "Art. 7. Bedingung für die Einwilligung," in *DSGVO - BDSG*, ser. Kommunikation & Recht, J. Taeger and D. Gabel, Eds. Frankfurt am Main: Fachmedien Recht und Wirtschaft dfv Mediengruppe, 2019.

[13] Y. Welinder and A. Palmer, "Face recognition, real-time identification, and beyond," in *The Cambridge handbook of consumer privacy*, E. Selinger, J. Polonetsky, and O. Tene, Eds. Cambridge: Cambridge University Press, 2018, pp. 102–124.

[14] E. Kosta, "Article 8. conditions applicable to child's consent in relation to information society services," in *The EU general data protection regulation (GDPR)*, C. Kuner, L. A. Bygrave, L. Drechsler, and C. Docksey, Eds. Oxford: Oxford University Press, 2020.

[15] Centre for Information Policy Leadership, "Gdpr implementation in respect of children's data and consent." [Online]. Available: https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_white_paper_-_gdpr_implementation_in_respect_of_childrens_data_and_consent.pdf

[16] B. Custers, F. Dechesne, W. Pieters, B. Schermer, and S. van der Hof, "Consent and privacy," in *The Routledge handbook of the ethics of consent*, ser. Routledge handbooks in applied ethics, A. Müller and P. Schaber, Eds. Boca Raton, FL: Routledge an imprint of Taylor and Francis, 2018.

[17] O. Ayalon and E. Toch, "User-centered privacy-by-design: Evaluating the appropriateness of design prototypes," *International Journal of Human-Computer Studies*, vol. 154, 2021.

[18] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[19] P. C. M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman, "Local differential privacy for deep learning," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 5827–5842, 2019.

[20] A.-T. Tran, T.-D. Luong, J. Karnjana, and V.-N. Huynh, "An efficient approach for privacy preserving decentralized deep learning models based on secure multi-party computation," *Neurocomputing*, vol. 422, pp. 245–262, 2021.