Smart Bears don't talk to strangers: analysing privacy concerns and technical solutions in smart toys for children

Katerina Demetzou*, Leon Böck †, Obaida Hanteer**

** Business and Law Research Centre (OO&R), Radboud University, Netherlands, K.Demetzou@cs.ru.nl, †Technische Universität Darmstadt Telecooperation Lab, Germany, boeck@tk.tu-darmstadt.de, ** Data Science & Society Lab, IT University Of Copenhagen obha@itu.dk*

Keywords: smart toys, children, privacy & data protection, privacy enhancing technologies.

Abstract

The "Smart Bear" is a hypothetical connected-smart toy for children. While the functionalities it presents are appealing to both children and their parents, the privacy concerns that are raised should be taken into serious consideration. A big amount of personal data of the child (and probably of other uninformed minors and adults in physical vicinity) are processed and analysed, an accurate profile of the child is created and direct marketing practices would most probably take place. The toy could suddenly turn into a surveillance device, while malicious third parties might hack the device and proceed to activities that would even threaten the child's physical and/or mental health. Data minimisation and privacy enhancing technologies are suggested, that would, if not completely alleviate, at least diminish the risks presented. Cybersecurity measures constitute a necessary condition for the alleviation of privacy concerns. This paper concludes that while a zero privacy risk "Smart Bear" is currently not possible, a privacy-considerate "Smart Bear" is not that hard to achieve.

1 Introduction

This article builds on a use case, named the "Smart Bear", a connected smart toy targeted towards children. According to a JRC report from 2017 [1], the turnover of the connected toys on the market is expected to reach €10 billion by 2020. As is the case with the Internet of things revolution, the Internet of Toys comes with benefits but also with major concerns. Privacy, security as well as ethical concerns are raised. In this paper we will focus on the relevant privacy issues that might emerge by the use of this smart toy. In order to make the most out of the benefits presented by the "child-connected smart toy" interaction, these concerns should be raised in this early stage of the toy industry and measures that alleviate them should be proposed and become best practices. Via the example of the "Smart Bear" use case we try to address the aforementioned issues. Thus, we first describe the functionality of the toy (section 2) and the relevant provisions of the data protection legal framework (section 3). We proceed on to presenting the identified risks (section 4) as well as on to suggesting technical measures to mitigate these risks (section 4). We conclude by identifying to what extent a privacy friendly "Smart Bear" is indeed possible and to what extent the suggested privacy enhancing measures will alleviate the concerns raised.

2 Smart Bear functionalities

The "Smart Bear" is a toy targeted to children 4-10 years old, that aims at mentally stimulating the child, helping her discover her identity and develop her mind, learning cause and effect, exploring relationships and practicing skills that she will need as an adult. It is a physical toy, enhanced with a processor and the ability to connect to the internet. Furthermore, it contains sensors such as microphones or tactile sensors that process personal data of the child and of others in physical vicinity (parents, any other third person –minor or adult-). It is thus a "connected toy" meaning that it *"can connect to Internet-based platforms or to other devices to enable data collection, processing or sharing through a computer server"* [2]. It is also a "smart toy" given that via the variety of sensors embedded in it, it can simulate intelligence and interact with the child. We want to specifically point out that we consider that the toy is not equipped with a camera. We think that the additional security and privacy risks added by a camera are too severe to be considered in a privacy preserving scenario.

More precisely, the studied bear is provided with some basic features, that is the minimal design needed to meet the expected entertaining and educational requirements and to allow privacy concerns to be taken into consideration. The bear has three working modes: "**off**", "**on-offline**" and "**on-online**". When the *"**off**"* mode is selected, the bear is just like any other soft toy and does not consume energy as all of its functionalities and sensing capabilities are disabled. In the *"**on-offline**"* mode, the bear offers some *basic non-personalized learning* activities such as story tales, basic math, biology, or chemistry (or any other modules that can be installed in an agreement with the vendor while purchasing). Being an interactive friend, the bear is provided with microphones and speech recognition capabilities that would help listening to, understanding and accordingly reacting to the child's vocal inputs optimally. Some tactile sensors and temperature sensors might be added to support environment-dependent interaction (eg. saying 'thank you' when the bear is hugged). A mobile application developed by the vendors and controlled by the parents can be used to download/buy new modules and story tales and connects to the bear via a secure local connection media (Bluetooth, wired connection, or using

the home LAN) to manage the modules (add, delete, update, increase level etc.) periodically as the child progresses. The mobile application can also be used to provide a systematic feedback to the parents about the newly learned skills and the achieved goals. Through the app, the parents can choose to enable/disable some sensors and/or functionalities. A third working mode, the ***"on-online"*** mode, can be added to the bear after an agreement with the parents while purchasing. This mode can only be activated by the parents through the mobile app as it adds internet connectivity to the bear in order to support more personalized functionalities. When this mode is activated, the child inputs are processed in the cloud.

Given that the "Smart Bear" collects a big amount of personal data in order to perform its functions, it is of importance to refer to the most relevant provisions of the European legal framework on personal data protection.

3 The General Data Protection Regulation (GDPR) [3]

The right to privacy and the right to data protection are both fundamental rights, found respectively in articles 7 and 8 of the Charter of Fundamental Rights of the European Union (hereafter, the Charter). Following the adoption of the Lisbon Treaty in 2009, article 6 TEU grants the Charter the legal status of primary law in the European legal order. The General Data Protection Regulation (hereafter, the GDPR) which will be applicable from the 25th of May 2018, is the European legal framework that lays rules relating *"to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data"* [Article 1(1), GDPR]. Its legal basis is Article 16(1) of the Treaty on the Functioning of the European Union (TFEU). It should be noted that *"the right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality"* [Recital 4, GDPR].

A first point to be made is that the targeted consumers of this toy are children, which, according to Recital 75 of the GDPR, constitute a vulnerable group of data subjects [4]. The Working Party 29 mentions that children are *"not able to knowingly and thoughtfully oppose or consent to the processing of their data"* [5]. While Directive 95/46/EC (the predecessor of the GDPR) did not contain child-specific provisions, Recital 38 of the GDPR, acknowledges that *"Children merit specific protection with regard to their personal data, as they may be less aware of the risks, consequences and safeguards concerned and their rights in relation to the processing of personal data"*. More interestingly for the case of the "Smart Bear" is that *"such specific protection should, in particular, apply to the use of personal data of children for the purposes of marketing or creating personality or user profiles and the collection of personal data with regard to children when using services offered directly to a child"* [Recital 38, GDPR]. Furthermore, the GDPR underlines in Recital 75 that the processing of children's personal data may result in risks *to the rights and freedoms of natural persons.*

Our use case refers to a toy targeted to children of 4-10 years old. According to article 8 of the GDPR, *"Where the child is below the age of 16 years, such processing shall be lawful only if and to the extent that consent is given or authorized by the holder of the parental responsibility over the child"*. In the cases where consent should be obtained, the law requires that this should be an informed decision, meaning that the principle of transparency [Article 5(1)(a) GDPR] with regard to the processing of personal data should be respected.

When it comes to the provisions with regard to *profiling* the following points should be made. According to Recital 75 of the GDPR when the processing of personal data is done *"in order to create or use personal profiles"* then this could give rise to *"risk to the rights and freedoms of natural persons"*. The company that manufactures the "Smart Bear" can legitimately create the child's profile for the purpose of the proper functionality of the toy, as long as the data subject is informed about this practice. Recital 60 of the GDPR states that *" [...] the data subject should be informed of the existence of profiling and the consequences of such profiling"*.

What is also of interest in our case is the legal provision on *direct marketing*. According to Recital 47 of the GDPR, this constitutes a legitimate interest of the data controller [6] (in this case the toy manufacturer) in the first place. However, this legitimate interest should not override *" [...] the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child"* [Article 6(1)(f) GDPR]. That requires a balancing exercise from the part of the data controller, the practice of which remains "a source of legal uncertainty" [7]. According to Article 21(2) GDPR, data subjects, therefore children as well, have the right to object to any profiling practice *"to the extent that it is related to such direct marketing"*. The data controller has to explicitly inform the data subject about this right [Recital 70 & Article 21(4) GDPR].

4 Privacy risks and security measures

In this section we will propose an approach for a privacy preserving "Smart Bear" by following the privacy by design approach by Gürses [8] and we will identify some privacy risks. For this we first introduce a straight forward implementation and then discuss the risks and how they can be resolved.

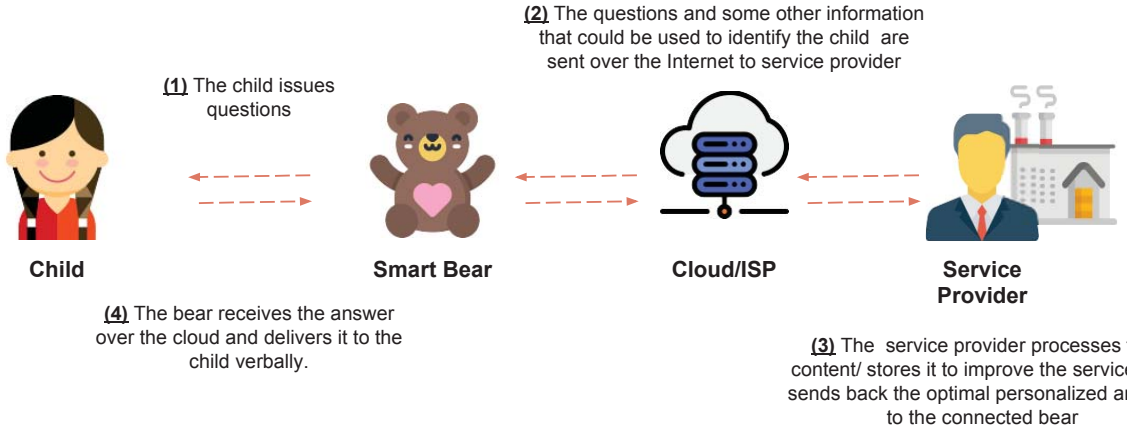

Figure (1) : straightforward implementation of the smart bear

4.1 Straight forward implementation

For the straight forward implementation of the Smart Bear we focus on the type of data collected and the way it is transmitted through different communication channels. The straight forward design is focused towards ease of implementation and cost savings. We will not focus in detail on how the functionality of the bear is realized but we want to point out that similar functionality already exists in other children toys or smart home devices such as Amazon's Alexa or Google Home.

Neither the ***"off"*** mode nor the ***"on-offline"*** mode require Internet connectivity and the data is directly processed on the Smart Bear. However, answering arbitrary questions would exceed the technical capabilities of a non-Internet connected device. Therefore, the ***"on-online"*** mode is required to activate the question answering feature of the toy.

Figure 1 depicts how the Smart Bear processes and answers arbitrary questions of the child. This is realized in a four step process. First, the child's question is processed by the Smart Bear. Secondly, the data is encrypted and transmitted over the internet to the server of the service provider. Afterwards, the service provider authenticates that the request is coming from a legitimate source to prevent others from using the resources illegitimately. The service provider then processes and stores the question. The answer is sent back encrypted to the Smart Bear which then, in step 4, verbally answers the child's question. The data stored by the service provider is used to improve the Artificial Intelligence (AI) that is used to answer the questions. By storing an ID together with all questions coming from a single Smart Bear, the service provider will be able to provide personalized answers for each customer.

4.2 Privacy and data protection concerns

The straight forward approach of a Smart Bear explained in the previous section introduces several privacy risks. When the Smart Bear is in the ***"off"*** mode, it does not transmit or record any data. While this setting is the least privacy intrusive, we still have to consider cases where personal data can be obtained and the Smart Bear's data storage can be analyzed (for example in the case where the Smart Bear is stolen).

The ***"on-offline"*** mode already increases the risk of leaking personal information. When transmitting data to the parent's smartphone or when receiving updates or installing new lessons, it could happen that this data are observed by an attacker eavesdropping on the communication between the Smart Bear and the smartphone. Furthermore, we have to consider that sensitive data is also stored on the smartphone. As many smartphones do not encrypt their stored data by default, data could also be illegally accessed, by hacking or stealing the smartphone, and then used to manipulate and access the Smart Bear functionality.

The ***"on-online"*** mode introduces the most severe privacy risks. Data in the form of questions asked by children is transmitted over the internet to be processed in the cloud. Such questions could contain several personal information such as names, locations or interests.

We can, thus, identify two elements of key importance to the functionality of the toy; firstly, the collection of a big amount of personal and non-personal data of the child and its environment and secondly, the connection to the Internet (in the case of the ***"on-online"*** mode). These elements also constitute the two basic sources of the following privacy risks.

- The Smart Bear is collecting a **broad range of personal data of the child** in order to offer personalized services and respond to the child's specific needs. Almost all everyday activities of the child are processed along with her thoughts and her biometric data (voice, fingerprints, etc.) which are tracked, recorded and analyzed. Given the quantity of data processed, a big amount of which could be sensitive data, and taken into account the vulnerability of the data subject (the child), the risks in the case of not proper security measures and not full transparency as to the data's use, are quite high. There is an important concern that these personal data might be used for purposes different than just for the

functionality of the Smart Bear. Additionally, if full transparency is not achieved, then the risk that the control of the individual (in this case of the child and the parent) will not be guaranteed, is high. Therefore, there is clearly a data protection risk as to the way these data will be manipulated and the control that the interested parties will have over their use.

- Apart from the processing of the child's personal data, the Smart Bear could process **personal data of other people** (be it other minors or adults) that exist within the environment of the child. In this case, the processing of their personal data will be done without their knowledge. This comes in direct contrast with the substance of the fundamental right to data protection, the very foundation of which is that the data subject has knowledge that her data is processed. According to Recital 60 of the GDPR *"The principles of fair and transparent processing require that the data subject be informed of the existence of the processing operation and its purposes"*. This is the starting point for all the rights enshrined in the data protection legal framework to be meaningfully exercised (eg. the right to access, the right to have one's data rectified etc). Again, we identify a data protection risk which refers, in this case, not to the child but to other data subjects in physical vicinity.
- In order for the Smart Bear to offer personalized services to the child it needs to create an **accurate profile** and understanding of the child's needs. The risk in this case is that the child will be categorized according to a profile created based on machine learning and their preferences and future behavior will be predicted accordingly [9]. This risk of the data subject being "trapped" in such a profile is significantly high in the case of children which are still in the very process of developing their personality, defining their preferences and creating the basis of their character. Where direct marketing also takes place, that heightens the risk to the child's right to privacy and right to development even more [10].
- **Privacy and surveillance risk**: One further concern that is highly linked to the security measures applied, relates to the Smart Bear being used as a surveillance device. First and foremost there is a data security risk, that is, the toy being hacked and used by third malicious parties as a way to getting access to the child's personal data, to eavesdrop on their conversations, to manipulate the functionalities of the toy in inappropriate ways etc. Blackmailing, kidnapping or pedophilic interests are just some examples of risks to be taken into serious consideration. Secondly, there is a risk that parents will use the toy to check on their children. Even though it is an understandable practice for safety reasons, it raises both a privacy but also an ethical concern. It should be noted that children own a fundamental right to privacy that applies also against their parents. That also raises privacy issues for other people (minors, parents, teachers etc) that are in the environment of the child and are also checked by the child's parents.

Based on the concerns mentioned above, we suggest that data minimisation and privacy enhancing technologies be applied so as to provide a privacy preserving "on-online" mode for the Smart Bear.

4.3 Data minimisation

"Data minimisation" is a data processing principle mentioned in Article 5(1c) of the GDPR that obliges the data controller to process only personal data that are *"adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed"*. In order to minimize the collected data, we need to consider exactly what data is required to achieve the service provider's goals. This also comes in line with another data processing principle, namely the *"purpose limitation"* principle, according to which *"personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes"* [Article 5(1)(b) GDPR]. In our use-case of the Smart Bear the provider has two major goals/purposes: 1) Storing query data to improve the smart interaction functionality of the Smart Bear and 2) Linking successive questions to provide better answers. In the following paragraphs we will look at each of the goals individually and discuss a solution to minimise the data required for the provider.

The manufacturer needs to collect personalized data to optimize the smart answering functionality of the Smart Bear. We argue that the privacy risk created, could be greatly minimised by removing personal IDs from the data. While this deteriorates the personalization of the provided answers, we argue that this can be overcome to some degree by grouping the users. As an example, the age of the child could be used to classify sets of users. Based on the age groups, the Smart Bear will adapt its answers to consider the evolution stages of the children. This may be sufficient to provide good answers while storing the requests without any personal identifiers. While this removes a direct link between stored queries and users, there may still be information such as names or locations that may reveal information about the users. To overcome this issue, we recommend that language processing tools are used to identify names and specific locations and replace them with generic placeholders before the question is sent over the Internet. This will allow service

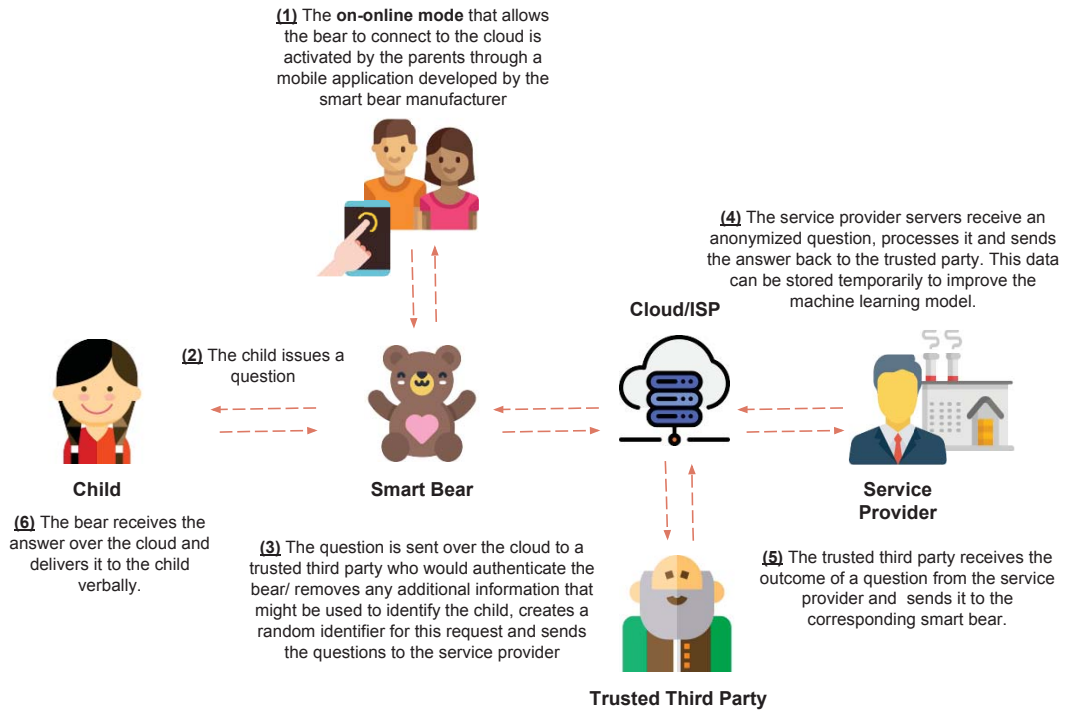

Figure (2) : Privacy considerate implementation of the smart bear

providers to still analyze the general type of questions while alleviating the risk that a dishonest provider will try to link the queries to the customers. Lastly, we suggest that data is only stored for a limited amount of time. Once the service provider has used past questions to improve the model of the artificial intelligence system, the original data should be discarded. This can be achieved through incremental learning techniques such as Learn++ [11]. Therefore, raw data may only be stored for short periods of time in between optimizations.

The second goal of the Smart Bear service provider is in conflict with the removal of identifiers from the stored queries. If the service provider wants to link consecutive queries there needs to be a way to link them. To do this, we propose that random short lived identifiers are used. This could be realized by setting the same identifier to questions that are sent within a predefined time window. If another question is sent after this time window is exceeded, a new identifier will be generated to avoid linkage between unrelated requests. This allows the service provider to link related requests while limiting the linkability of requests to a minimum.

4.4 Identification and mitigation of privacy risks

While data minimisation is an important step for reducing privacy risks, we also need to consider that internal and external parties may actively try to undermine the security and privacy of the Smart Bear. Therefore, we need to identify the potential attackers against the system. In our use-case we identified malicious system providers, malicious ISPs and external actors as potential attackers. In the following paragraphs, we will discuss each attacker, their motivation, capabilities and goals in more detail.

A malicious system provider may try to link questions to customers allowing them to increase their revenue through targeted advertisements. This could be achieved by storing the IP address of the incoming queries to identify the customers based on that. Furthermore, users could potentially be tracked if their IPs change by correlating similarities between the questions sent. This puts the users' and especially the child's privacy at risk. An obvious approach to anonymize the traffic would be the use of Tor [12] or Mix networks [13]. However, this would prevent the service provider from identifying legitimate requests. Therefore, we suggest the use of a Trusted Third Party (TTP). Any questions forwarded to the service provider will be processed through the TTP. This removes the possibility for the service provider to identify customers based on their IP address, while the TTP can verify that requests are coming from legitimate customers.

A malicious ISP can potentially invade a user's privacy by observing usage patterns of the Smart Bear. Even though traffic that is transmitted over the Internet is encrypted, an ISP can identify traffic going to the TTP based on non-encrypted Metadata. This would allow an ISP to track the location of the Smart Bear and therefore the location of the child (however, there are no clear motivations for ISPs to do so). Furthermore, we want to point out that similar data on adults can already be collected by monitoring cell phone activities or other connected devices. Nevertheless, a possible approach to anonymize the endpoints of the communication would be the use of anonymization services such as Tor [12] or Mix Networks [13].

Lastly, we want to consider possible motivations and attack vectors of external parties. The case of My Doll Cayla [14] highlights that external parties may not only try to invade the

privacy of the child but to also introduce security risks by obtaining access to the Smart Bear to observe and even talk to the child. As our case study is mainly focused on the privacy concerns, we will not go into greater detail regarding security mechanisms to prevent hackers from easily obtaining access to the Smart Bear, the parents' smartphone or even the infrastructure of the service provider. However, some of the approaches we suggested to protect the privacy of the child such as encryption, or authentication also provide protection against third parties that intend to harm the security of the child.

We identify two major ways in which external parties could endanger the privacy of the child. The first is to obtain access to the Smart Bear's functionality or to the data transmitted by the Smart Bear. The use of strong cryptographic protocols can prevent external parties from accessing the data transmitted over the Internet or local wireless networks. Another way of gaining access to the Smart Bear could be to impersonate the smartphone of the parents that is used to manage the Smart Bears functionality. To prevent this, strong authentication is required, such that deters anyone with the correct application from obtaining access to the Smart Bear. One possible approach could be that the smartphone is connected via wired connection, such as USB, to the Smart Bear and provide a password to authenticate a legitimate user. If the password is correct the Smart Bear and the smartphone could exchange cryptographic keys through which they communicate and authenticate each other. This would prevent external parties from easily obtaining management access to the Smart Bear. While there are multiple other motives for external parties, we argue that many of these risks can be prevented with proper security engineering.

Lastly, external parties could obtain access to the data stored on the Smart Bear or smartphone by stealing the device. To prevent this, we recommend to encrypt all stored data on the device.

5 Conclusion

This article used a use case scenario (the "Smart Bear" connected-smart toy) in order to first of all highlight the privacy risks and secondly to suggest appropriate data minimisation and privacy enhancing technologies that could potentially alleviate them. What should be noted in the first place is that a trade-off between the functionalities of the toy and privacy exists to some extent if we want to protect the privacy of the child. Smart connected toys introduce new risks to both privacy and security of children that should be taken well into account by parents when deciding on purchasing such a toy. However, we have to consider that the market for smart connected toys is growing. Therefore, we think it is not only a necessary addition to take privacy into consideration when developing such toys, but it should also be considered as an important feature of the connected toy. The technical measures suggested in this paper are necessary but not sufficient for the alleviation of privacy concerns. Legal safeguards provided by the GDPR must also apply. Data protection principles (Article 5 GDPR) must be followed by the data controller, data subjects' rights must be respected (Chapter III of the GDPR) and tools that contribute to an effective data protection (eg. DPIA Article 35 GDPR) must be used when the relevant legal conditions are met. While a zero privacy risk Smart Bear is currently not possible, a privacy-considerate Smart Bear is not that hard to achieve.

Acknowledgements

This use case was conceived and dealt with during the activities of the PiLab Summer School, June 18-23,2017, Berg-en-Dal, Nijmegen, The Netherlands; the name "Smart Bear" is a hypothetical name, used only for the purposes of dealing with a specific case study during the summer school.

References

[1]Chaudron S., Di Gioia R., et al "Kaleidoscope on the Internet of Toys – Safety, security, privacy and societal insights", EUR 28397 EN, doi: 10.2788/05383

[2] Future of Privacy Forum – Family Online Institute (FOSI), "Kids & the connected home: privacy in the age of connected dolls, talking dinosaurs and battling robots" (2016)

[3] European Parliament and Council (2016) Regulation (EU) 2016/679 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 4 May 2016, L119/1)

[4] According to the definition of Article 4(1) GDPR, a "data subject" is *"an identified or identifiable natural person"* to whom the personal data relate.

[5] Article 29 Working Party, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, WP 248 rev.01.

[6] According to the definition of Article 4(7) GDPR, "data controller" is the *"natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data."*

[7] Eva Lievens, Valerie Verdoodt. "Looking for needles in a haystack: Key issues affecting children's rights in the General Data Protection Regulation", Computer Law & Security Review: The International Journal of Technology Law and Practice (2017), doi:10/1016/j.clsr.2017.09.007

[8] Gürses, Seda, Carmela Troncoso, and Claudia Diaz. "Engineering privacy by design." (2011)

[9] Council of Europe (2010) Recommendation CM/Rec(2010)13 on the protection of individuals with regard to automatic processing of personal data in the context of profiling.

[10] Articles 16 and 6 United Nations Convention on the Rights of the Child.

[11] Polikar, Robi, et al. "Learn++: An incremental learning algorithm for supervised neural networks." IEEE transactions on systems, man, and cybernetics, part C (applications and reviews) 31.4 (2001): 497-508.

[12] Dingledine, Roger, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. Naval Research Lab Washington DC, 2004.

[13] Chaum, David L. "Untraceable electronic mail, return addresses, and digital pseudonyms." Communications of the ACM 24.2 (1981): 84-90.

[14] Bundesnetzagentur, Press Release “Bundesnetzagentur removes children's doll "Cayla" from the market”, Bonn, 17 February 2017.