# Poster: Ontology Enabled Chatbot for Applying Privacy by Design in IoT Systems

Lamya Alkhariji
alkharijiLa@cardiff.ac.uk
Cardiff University
United Kingdom

Suparna De
s.de@surrey.ac.uk
University of Surrey
United Kingdom

Omer Rana
ranaof@cardiff.ac.uk
Cardiff University
United Kingdom

Charith Perera
pererac@cardiff.ac.uk
Cardiff University
United Kingdom

## ABSTRACT

Our aim is to create a personal assistant, a chatbot, that can answer queries from software developers regarding Privacy by Design (PbD) methods and applications throughout the design phase of IoT system development. We used semantic web technologies to model the PARROT Ontology that includes knowledge underlying PbD measurements, their intersections with privacy patterns, IoT system needs, and the privacy patterns that should be applied across IoT systems. To determine the PARROT ontology's requirements, a collection of real-world IoT use cases were aided by a series of workshops to gather Competency Questions (CQs) from researchers and software engineers, resulting in 81 selected CQs. In a user study, the PARROT ontology was able to answer up to 58% of software developers' privacy-related issues.

## CCS CONCEPTS

• **Security and privacy → Privacy protections**; **Privacy protections**; • **Software and its engineering** → *Designing software*; • **Theory of computation → Semantics and reasoning**.

## KEYWORDS

privacy by design, ontology, chatbot, personal assistant

**ACM Reference Format:**
Lamya Alkhariji, Suparna De, Omer Rana, and Charith Perera. 2022. Poster: Ontology Enabled Chatbot for Applying Privacy by Design in IoT Systems. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22), November 7–11, 2022, Los Angeles, CA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3548606.3563504

## 1 INTRODUCTION

With the widespread use of Internet of Things (IoT) technology, preserving individual privacy has become a rising challenge that demands immediate attention. Various authorities have shown an increased interest in the establishment and enforcement of privacy laws and practices recently. Many software engineers are concerned that these guidelines are theoretical, requiring field specialists to interpret them into more practical representations. While this is viable for huge corporations, such tailored techniques can be overwhelming for small and medium-sized businesses (SMEs).

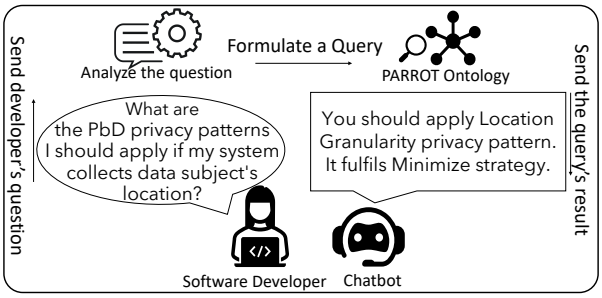


**Figure 1: Chatbot System Example**

Privacy by Design (PbD) is a concept that promotes thinking about data protection throughout the system design phase and may provide a realistic solution to preserve the data subject's privacy, according to [6]. Various organizations and scholars have suggested multiple PbD measurements at various levels of abstraction, including principles, guidelines, strategies, and privacy patterns, as described in our previous paper [1]. However, to design an efficient IoT system, the software developer must first evaluate which PbD principles are best suited to the system in question, which adds substantial effort to the development process. This study proposes a personal assistant solution to assist software developers in obtaining the necessary assistance to ensure that their systems meet the PbD concept. In other words, we introduce a chatbot that is enhanced with the PARROT ontology which models IoT system needs and binds them to the relevant PbD measurements. The questions of this paper are as follows:

(1) What are the common PbD questions that software engineers have when designing an IoT application?
(2) What information need to be modeled in a privacy knowledge ontology to answer these questions?
(3) How much proportion of software developers' questions can a chatbot that is enhanced with the ontology answer?


*CCS '22, November 7–11, 2022, Los Angeles, CA, USA*

ACM ISBN 978-1-4503-9450-5/22/11.
https://doi.org/10.1145/3548606.3563504

## 2 MOTIVATION

Nora is developing an IoT system considering user privacy. She searches about the required PbD practices in order to understand them and then applies them in her system. She finds many resources and documents that describe ways to protect the privacy of users, but she is confused by the large number of available documents and their variations. For example, Cavoukian's principle, "Proactive, not Reactive; Preventative not Remedial" that she finds easy to understand, but these are vague in application. Looking at another document, Hoepman's strategies, she is unsure whether she needs to apply all of the strategies. On looking at the first strategy, "Minimise", she approaches her system with the intent of minimising data. This becomes somewhat confusing where she seeks further explanations of this strategy. She discovers the privacy patterns which she finds applicable. However, it offers so many patterns that she is not sure which ones best explain the "Minimise" strategy. She goes back to her system having struggles to find the appropriate practices to deploy.

Nora thus decides to use a personalized assistant tool. She draws the system she is designing in the tool's interface, and once she submits the DFD diagram, the tool returns it with annotations and comments about the privacy patterns required for each node in the diagram. She explores these comments, which help her ascertain what she needs to do to implement the appropriate patterns. Seeking further explanation, she uses a chatbot to ask questions about the meaning of these privacy patterns. The Figure 1 shows an example of a question that Nora could get. The question will be analysed, then, queried in the PARROT ontology. Finally, the chatbot will provide Nora with the query results.

## 3 METHODOLOGY

The progress of the current research was organised into four phases. These were information gathering, analysis, development, and evaluation. In this section, we will explain each phase in detail.

### 3.1 Gathering

The initial phase of this research was to employ six representative IoT use cases to gather the information needed to model the PARROT ontology as CQs. The CQs were gathered and asked by software engineers in workshops where they were provided with the use cases and prompt to ask questions about how to preserve privacy in the provided use cases designs. CQs that were out of scope or duplicated were eliminated, while others were adjusted to fit the needs of this study. This resulted in 81 valid CQs that were then utilized to build the ontology requirements. The answers to the retained CQs were determined and then constructed into a set of Privacy Patterns.

### 3.2 Analysis

The knowledge from the gathering phase was categorised and tagged in the analyse phase. The CQs were categorised into five types and 20 sub-types based on the issues presented in them. In addition, as determined by prior research [1], answers -which are sets of privacy patterns- were tagged utilising Hoepman's [5] eight tags. The categories and sub-types are described in the following:

- **Data Collection.** These CQs narrate the types of data that will be gathered within the system. This includes four sub-types: location, personal information, photo, and routine. An example of the latter type of CQ might be *What PbD patterns should I apply if my system stores a user's food intake information?*
- **Device:** These were CQs about a specified device, leading to the generation of four sub-types: mobile phone, camera, microphone, and reading sensor, potentially covering any sensor that could be used in the system.
- **Process:** These CQs considered the processes applied to data subject information across five sub-types: these were share, access, third-party, route, and profile.
- **Storage:** The CQs were around storing data subject information and periods appropriate to this which emerged two sub-types: cloud and local.
- **Dignity:** These are CQs that consider the regulations and procedures that ensure a data subject's dignity and right to privacy. This category thus contains four sub-types, advantage, agreement, notify, and control.

### 3.3 Development

The PARROT ontology was developed using a top-down approach. As a starting point, four existing ontologies, SKOS [1], GDPRtEXT [2], SSN [3], and SOSA[4] were reused, beside classes created to model the knowledge that was to be included in the PARROT ontology. The PbD measurements were organised in levels (Principles, Strategies, Guidelines, Patterns) depending on their abstractness and specificity. These PbD measurements and their connections were thus modelled together [1], along with the data set analysed in the previous phase. The resultant PARROT ontology includes 39 classes, three object properties, and 225 individuals. Readers are referred to the supplementary technical report [2] for a more detailed explanation of ontology classes with their modeled object properties and instances and it is also available online[5].

### 3.4 Evaluation

The content of the PARROT ontology was then evaluated using the Wizard of Oz [6] technique [4] via a user study, that is ethically approved, with participants who have various software engineering experience. In the user study participants played the software developer role and were provided with two use cases along with their descriptions and DFD diagrams, and they were prompted to ask a privacy expert about how to apply PbD measurements into those systems.

A total of 193 questions were taken from participants that were checked for validity, being ranked as valid, duplicated, and discarded for the questions that are out of scope. After filtration, we had 81 valid questions. These valid questions were then assigned to

[1] www.w3.org/TR/2008/WD-skos-reference-20080829/skos.html
[2] www.w3.org/community/dpvcg/wiki/Data_Protection_Ontology_by_Bartolini_et._al
[3] www.w3.org/TR/2017/REC-vocab-ssn-20171019/
[4] www.w3.org/2015/spatial/wiki/SOSA_Ontology
[5] github.com/alkharijiLa/PARROT/blob/main/PARROT.owl
[6] A methodology that allows testing of a prototype of a system before its actual development by means of having a person simulate the interface.

the categories. Finally, after sorting and classifying the collected questions, these were answered using the PARROT ontology.

## 4 RESULTS

The evaluation aimed to measure the extent to which the PARROT ontology can answer software developers' questions about privacy-preserving measures in the design of IoT systems. Above and beyond the three ranks declared earlier (i.e., valid, duplicated, and discarded), two additional ranks for **valid** questions were thus created: **missing** and **not available**. A **missing** question is a valid question that is not yet covered in the ontology; such cases thus need to be modeled and added to the PARROT ontology. An example was *Can I get the date of birth of the driver to check if he has a license?* A **Not available** question is also a valid question, but one for which there are insufficient privacy patterns to cover the issue raised. These questions should lead to the creation of a list of new privacy pattern suggestions for future work. For examples, *If we have an external copy or backup of the information, how can I keep this private?* Such a question require privacy patterns that are taken into account the rights of multiple people, including those who do not use the service directly and who thus have not provided consent for their data to be collected.

After sorting and classifying the collected questions, these were answered using the PARROT ontology. Of the 81 valid questions, only 45 questions were answered successfully. Overall, there were 14 **missing** questions, and 21 **not available** questions, as shown in Figure 2. The full table of questions, the accompanying analysis, and the relevant queries and answers can be found in [3].

## 5 CONCLUSION AND FUTURE WORK

The goal of this study was to see if the queries that software developers have about designing privacy-preserving IoT systems could be described and addressed by a chatbot using an ontology. As a result, the necessary knowledge that has to be represented to serve that function has been discovered, and the PARROT ontology established in this research encompasses two knowledge sectors. The first includes knowledge about IoT system devices and activities, as well as how these should be handled by such PbD measurements. This information was gathered by researchers using Competency Questions (CQs) developed from six real-world IoT systems. This research also provided a categorised framework of the resulting questions and their treatments that lays the groundwork for additional concerns about IoT systems to be addressed.

In user research, the content of the PARROT ontology was evaluated, and it was discovered that the present version of the PARROT ontology could answer 56% of the questions posed. This occurred in part because 18% of the questions were completely lacking from the ontology, necessitating modelling in future editions. However, due to a lack of PbD measurements, the remaining 26% of queries could not be addressed. In summary, the PARROT ontology includes most of the available knowledge necessary to aid software engineers in applying PbD measures to their system designs, as well as to explain PbD metrics to software engineers by demonstrating linkages to other measurements.

Future work will extract knowledge from a broader range of IoT use cases. It is also recommended that all PbD measures be linked to

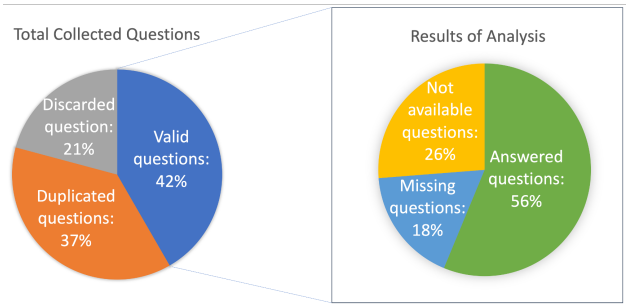


**Figure 2: The results of the study analysis, showing that the PARROT ontology was able to answer only 56% of the questions asked by participant.**

GDPR regulations to ensure that software engineers are fully aware of which rules they must follow throughout their systems. Finally, a chatbot interface for the PARROT ontology will be created in the near future to make it easier for software programmers to interact with and retrieve information.

## REFERENCES

[1] Lamya Alkhariji, Nada Alhirabi, Mansour Naser Alraja, Mahmoud Barhamgi, Omer Rana, and Charith Perera. 2020. Synthesising privacy by design knowledge towards explainable internet of things application designing in healthcare. *arXiv* (2020). arXiv:2011.03747

[2] Lamya Alkhariji, Suparna De, Omer Rana, and Charith Perera. 2022. PARROT Ontology Tech. Report. (2022). https://orca.cardiff.ac.uk/149337/

[3] Lamya Alkhariji, Suparna De, Omer Rana, and Charith Perera. 2023. Semantics-based privacy by design for Internet of Things applications. *Future Generation Computer Systems* 138 (2023), 280–295.

[4] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of oz studies-why and how. *International Conference on Intelligent User Interfaces, Proceedings IUI* Part F127502 (1993), 193–200.

[5] Jaap-Henk Hoepman. 2018. Privacy Design Strategies (The Little Blue Book). *Transactions of the Canadian Society for Mechanical Engineering* (2018), 30. https://www.cs.ru.nl/~jhh/publications/pds-booklet.pdf

[6] Richmond Y. Wong and Deirdre K. Mulligan. 2019. Bringing Design to the Privacy Table. (2019), 1–17.

Received 16 August 2022; revised 07 September 2022; accepted 18 September 2022