



AFRL-RI-RS-TR-2021-116

PERSONALIZED PRIVACY ASSISTANTS FOR BIG DATA & THE INTERNET OF THINGS

CARNEGIE MELLON UNIVERSITY

JULY 2021

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

**AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nations. Copies may be obtained from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-RI-RS-TR-2021-116 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /

STEVEN L. DRAGER
Work Unit Manager

/ S /

GREGORY J. HADYNSKI
Assistant Technical Advisor
Computing & Communications Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE*Form Approved*
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) JULY 2021			2. REPORT TYPE FINAL TECHNICAL REPORT		3. DATES COVERED (From - To) OCT 2015 – DEC 2020	
4. TITLE AND SUBTITLE PERSONALIZED PRIVACY ASSISTANTS FOR BIG DATA & THE INTERNET OF THINGS					5a. CONTRACT NUMBER FA8750-15-2-0277	
					5b. GRANT NUMBER N/A	
					5c. PROGRAM ELEMENT NUMBER 62303E	
6. AUTHOR(S) Norman Sadeh Alessandro Acquisti Lujo Bauer Matt Fredrikson Lorrie Cranor Anupam Datta Matthew Tschantz					5d. PROJECT NUMBER BRAN	
					5e. TASK NUMBER CM	
					5f. WORK UNIT NUMBER U1	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Carnegie Mellon University 5000 Forbes Avenue Pittsburgh, PA 15213					8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory/RITA 525 Brooks Road Rome NY 13441-4505 DARPA 675 North Randolph Street Arlington, VA 22203-2114					10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RI	
					11. SPONSOR/MONITOR'S REPORT NUMBER AFRL-RI-RS-TR-2021-116	
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT The "Personalized Privacy Assistants for Big Data and the Internet of Things (IoT)" project developed, demonstrated and evaluated novel techniques that empower end-users to effectively evaluate privacy policies and configure privacy settings in realistic mobile and IoT environments. This included addressing privacy challenges associated with Big Data and developing a privacy infrastructure that enables the automated discovery of IoT resources and their privacy practices. It also included work on privacy risks in Big Data machine learning pipelines. This report summarizes the research performed over the course of the project and major results.						
15. SUBJECT TERMS Privacy, Internet of Things, Mobile Apps, Big Data, Privacy Assistants, Usability, Artificial Intelligence, Trust, Transparency, Accountability, Explainability, Privacy Compliance, Machine Learning, Code Analysis, Privacy Nudging, Natural Language Processing						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	
a. REPORT	b. ABSTRACT	c. THIS PAGE			STEVEN L. DRAGER	
U	U	U	UU	103	19b. TELEPHONE NUMBER (Include area code) N/A	

Table of Contents

List of Figures.....	iv
Acknowledgments	vii
1. Summary.....	1
2. Introduction	3
3. Methods, Assumptions, and Procedures.....	4
3.1 Personalized Privacy Infrastructure for the Internet of Things and Personalized Privacy Assistants.....	5
3.2 Taxonomy of Privacy Dialog Primitives.....	8
3.2.1 Exploring How Privacy and Security Factor into IoT Device Purchase Behavior.....	9
3.2.2 An Expert Study to Determine IoT Privacy and Security Label Content	9
3.3 Modeling of User Privacy Preferences and Expectations	9
3.3.1 Mobile App Privacy Assistant Studies.....	10
3.3.2 Modeling People’s Privacy Expectations Across a Wide Array of IoT Scenarios.....	14
3.3.3 Informing the Design of Personalized Privacy Assistants.....	15
3.3.4 Understanding and Modeling People’s Privacy Attitudes Towards Video Analytics Technologies Across a Representative Cross-Section of Deployment Contexts.....	16
3.4 Nudging and Explanation.....	19
3.4.1 Mobile App Privacy Nudging - Studying the Impact of Framing on the Effectiveness of Privacy Nudges	20
3.4.2 Psychometrically Tailored Privacy Nudges	21
3.4.3 Transparency via Attribution in Vision Models and Internal Explanation for NLP models	23
3.4.4 Explanations for Mobile App Privacy Permission Recommendations.....	24
3.4.5 Transparency of Privacy Choices on the Web and Opt-Out Easy Browser Add-On Tool.....	26
3.5 Transparency and Compliance Analysis.....	27
3.5.1 Data Flow Modeling and Analysis.....	27
3.5.2 IFTTT: Identification and Analysis of Vulnerabilities in IoT End-User Programming Environments	28
3.6 Privacy Risk in Machine Learning Pipelines	29
3.6.1 Proxy use	29
3.6.2 Overfitting and Whitebox Membership inference	30
4. Results and Discussion.....	32
4.1 Personalized Privacy Infrastructure for the Internet of Things and Personalized Privacy Assistants.....	32
4.1.1 The Implementation and Deployment of IoT Portal.....	35

4.1.2 The Implementation of the IoT Assistant App	37
4.2 Taxonomy of Privacy Dialog Primitives.....	40
4.2.1 Exploring How Privacy and Security Factor into IoT Device Purchase Behavior.....	40
4.2.2 An Expert Study to Determine IoT Privacy and Security Label Content	41
4.3 Modeling of User Privacy Preferences and Expectations	42
4.3.1 Mobile App Privacy Assistant Studies.....	42
4.3.2 Modeling People’s Privacy Expectations Across a Wide Array of IoT Scenarios.....	50
4.3.3 Informing the Design of Personalized Privacy Assistants.....	50
4.3.4 The Experience Sampling Study.....	53
4.4 Nudging and Explanation.....	57
4.4.1 Evaluating the Impact of Framing on the Effectiveness of Privacy Nudges.....	57
4.4.2 Psychometrically Tailored Privacy Nudges	57
4.4.3 Transparency via Attribution in Vision Models and Internal Explanation for NLP models	60
4.4.4 Explanations for Mobile App Privacy Permission Recommendations.....	62
4.4.5 Transparency of Privacy Choices on the Web and Opt-Out Easy Browser Add-On Tool.....	62
4.5 Transparency and Compliance Analysis.....	63
4.5.1 Data Flow Modeling and Analysis.....	63
4.5.2 Mobile App Privacy Compliance Analysis	64
4.5.3 IFTTT	71
4.6 Privacy Risk in Machine Learning Pipelines	73
4.6.1 Proxy Use.....	73
4.6.2 Whitebox Membership Inference in Non-parametric Models and Ensembles.....	74
4.7 Integration, Deployment, and Evaluation of Technologies and Tools	75
4.7.1 Integration of IoT Privacy Infrastructure in TA3 Systems	75
4.7.2 Mobile app privacy assistant.....	75
4.7.3 IoT Privacy Infrastructure.....	76
4.7.4 The IoT Assistant App and Video Obfuscation Application	76
4.7.5 MAPS.....	77
4.7.6 Opt-Out Easy.....	77
4.7.7 IoT Security and Privacy Nutrition Labels	78
5. Conclusions	81
6. Recommendations	83
7. References.....	84
7.1 Conference Papers and Journal Articles	84
7.2 Short Selection of Conference Presentations, Tutorials, Panels and other Dissemination Events	88

7.3 URLs.....	90
7.4 Technical Reports.....	91
7.5 Issued Patents	92
8. List of Symbols, Abbreviations, and Acronyms	93

List of Figures

Figure 1: IoT Privacy Infrastructure - High Level Architecture	8
Figure 2: Motivating users to reflect on their current permission settings and possibly modify them, as part of study that involved collecting and analyzing people’s mobile app permission preferences as part of their regular daily interactions with their regular daily interactions with their regular smartphones [LAS+16].	11
Figure 3: Example of analysis, where people’s mobile app permission preferences are analyzed using clustering techniques.	12
Figure 4: A mobile app permission assistant asks users a small number of questions to assign them to a cluster/profile. Permissions for which people in the given profile exhibit highly homogeneous preferences are used as a basis for recommending possible changes to the user’s current mobile app permission settings. The user can also access an explanation for each recommendation [LAS+16].	13
Figure 5: In-situ study of people’s privacy attitudes towards Video Analytics technologies across a number of deployment scenarios. Screenshots of the study app and the web survey used for the evening review [ZFD+20a].	18
Figure 6: Studying the impact of framing on the effectiveness of privacy nudges in the context of mobile app location privacy nudges.	20
Figure 7: Nine different ways of framing mobile app location privacy nudges, from mundane nudges to nudges that place more emphasis on how location information might be used and associated privacy risks.	21
Figure 8: Two types of input attributions for an image classification model.	23
Figure 9: Slice of data pathways in an LSTM language model.	24
Figure 10: Uses (a,b,d) and non-uses (c) of sensitive information in decision trees.	27
Figure 11: The relationships between variables required for three different notions of proxy. The arrows represent causation; dashed lines, correlation.	29
Figure 12: Homepage of the IoT Privacy Infrastructure portal available at iotprivacy.io	32
Figure 13: The IoT Assistant mobile app, as available in the Apple iOS app store.	33
Figure 14: Screenshots of the IoT Assistant app showing resources deployed and publicized on the UC Irvine campus.	34
Figure 15: Screenshots of the IoT Assistant app showing resources deployed and publicized on the Carnegie Mellon University campus. 4.1.1 The Implementation and Deployment of IoT Portal.	35

Figure 16: IoT PI Dashboard from which users can check the status of their resources, registries, and templates. 36

Figure 17: Screenshots of the IoT Assistant app showing resource descriptions published in Miami, Washington DC, Boston and the Manhattan Financial District in New York. The app is currently configured to only display up to 50 IoT resources around a given location. Over 100,000 IoT resource descriptions have been published and can be discovered using the IoT Assistant app..... 37

Figure 18: Using the IoT Assistant app to discover nearby IoT resources, including their data collection and use practices and any privacy settings they make available. Notification settings allow users to configure their app to only be notified about IoT data collection practices they care about and at the frequency they find most useful. 38

Figure 19: Defining privacy controls (or privacy choices) when describing an IoT resource and its data practice via the IoT Portal. 39

Figure 20: Screenshots of Privacy Options made accessible to IoT Assistant users as they discover a particular IoT resource - in this case a Bluetooth location tracking device. Users typically need to authenticate with the organization controlling the IoT resource to manage the privacy options made available by that resource. 40

Figure 21: Clustering like-minded users based on their decisions to grant/deny permissions to mobile apps in different categories..... 43

Figure 22: In each cluster, the number of deny/allow decisions can be used as a basis for making recommendations to either grant/deny a given permission to a given category of apps. When the difference between the grant and deny decisions is not sufficiently strong or when there is not sufficient data, no recommendation is made. ... 44

Figure 23: Comparing F-1 scores for predictions with profiles that differentiate between the purpose for which apps in a given category request a permission and profiles that do not differentiate between the two..... 44

Figure 24: Screenshots of the personalized privacy assistant piloted in our initial study. 45

Figure 25: Number of recommendations made by the privacy assistant app that were accepted/rejected by each individual participant in the study (total of 27 participants who received recommendations). 46

Figure 26: Plotting the total number of user interactions: comparing predictive models that include permission purpose and models that do not. 48

Figure 27: Scatterplot showing recommendation accuracy versus the average number of user interactions. Each k represents the number of profiles the population is divided into. 49

Figure 28: Summary of collected responses organized around 16 different purposes. The bottom row shows the aggregated preferences across different purposes.	54
Figure 29: Profiles associated with a 6-cluster model.	56
Figure 30: Evaluating the effectiveness of framing in mobile app privacy nudges. Cells highlighted in purple are statistically significant.	57
Figure 31: Input space, output contour and classification (color), and attributions (vectors).	61
Figure 32: Subject-verb number agreement pathways in a 2-layer LSTM.	62
Figure 33: Password guessability under gender and age demographics.	63
Figure 34: Decision Tree Model with Proxy Use and Repaired Model.	64
Figure 35: MAPS: Screenshot of MAPS mobile compliance analysis system showing bulk analysis functionality.	66
Figure 36: MAPS: Screenshot of MAPS system showing different tabs that enable users to find details about individual apps, including detailed analysis results (“potential compliance issues”, “policy analysis”, and “static app analysis”).	67
Figure 37: MAPS: Looking at an app’s privacy policy.	68
Figure 38: MAPS: Summary of potential privacy compliance issues for a given app. ...	69
Figure 39: MAPS: Automated analysis of a privacy policy.	70
Figure 40: MAPS: Static analysis of a mobile app’s code.	71
Figure 41: Violating recipes, broken down by violation type.	72
Figure 42: White-box Shadow Model Attack for Membership Inference.	74
Figure 43: Membership inference attack success vs. number of estimators in ensembles.	74
Figure 44: Screenshots of Mobile App Privacy Assistant deployed in Google Play Store for rooted Android phones.	76
Figure 45: The User Interface of Opt-Out Easy.	78
Figure 46: Primary layer of IoT Security and Privacy Nutrition Label.	79
Figure 47: Secondary layer of IoT Security and Privacy Nutrition Label.	80

Acknowledgments

The investigators would like to thank the Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory (AFRL) for their support of the research reported herein. Special thanks to the DARPA Brandeis Program Managers, Dr. John Lanchbury, Mr. David Gunning, Mr. Jeremy Epstein, and Dr. Joshua Baron, to Mr. Timothy Bond, and to the team at AFRL and in particular Mr. Carl Thomas and Mr. Steven Drager. Their insightful input and feedback, their pragmatic and highly effective approach to organizing the overall initiative and managing individual projects, and their continued support over the years made a big difference. We would also like to take this opportunity to thank our many collaborators in the DARPA Brandeis initiative and in particular the Testbed for IoT-based Privacy-preserving Pervasive Spaces (TIPPERS) team at University of California Irvine (UCI) led by Prof. Sharad Mehrotra.

1. Summary

The Personalized Privacy Assistant (PPA) Project was launched to develop, demonstrate and evaluate novel technologies that empower end-users to effectively evaluate privacy policies and configure privacy settings in mobile and Internet of Things (IoT) contexts. This includes addressing privacy challenges associated with Big Data and developing a privacy infrastructure that enables the automated discovery of IoT resources and their privacy practices, including any available privacy choices. Work in the project was organized around the following themes:

- Personalized Privacy Infrastructure for the Internet of Things, including Personalized Privacy Assistants
- Taxonomy of Privacy Dialog Primitives
- Modeling of User Privacy Preferences and Expectations
- Transparency and Explanation
- Privacy Risk in Machine Learning Pipelines

Work in this project produced a number of new technologies, tools, models, and findings that have already influenced and are expected to continue influencing privacy practices. This includes the deployment of a publicly accessible privacy infrastructure for the Internet of Things to help publicize the presence of IoT systems and their data collection and use practices, including any available privacy choices. This includes the availability of an “IoT Assistant” app for people to discover these IoT systems and their data practices, including accessing any privacy choices they make available - the IoT Assistant app is available in both the iOS and Android app stores. This also includes the release of a personalized privacy assistant app for rooted Android phones to help people manage their mobile app permissions, the development of a mobile app privacy compliance tool (“MAPS”) that has been used to automatically analyze over 1 million Android apps for potential privacy compliance issues, the launch of the “Opt-Out Easy” browser extension to help people discover opt-out choices buried deep in the text of privacy policies (tool available in the Chrome and Firefox stores). In addition to these tools, research conducted in the project helped develop rich models of people’s privacy preferences and expectations across a variety of mobile app and IoT contexts, including the development of predictive models using machine learning techniques. These techniques have been shown to have the potential of drastically reducing user burden when it comes to helping users manage the myriad of privacy choices they need to make in mobile and IoT environments. This research has included exploring tradeoffs between user burden and control when it comes to configuring these technologies, including identifying differences in the way different people feel about relying on automated recommendations and even delegating some of their privacy decisions to predictive models. Research on privacy nudges has demonstrated that nudging can be

configured to be particularly effective at helping people make privacy decisions they are less likely to regret. Privacy nutrition labels have been developed to help simplify the presentation of key data practices to users when it comes to evaluating different IoT devices. Work on transparency has also yielded techniques to identify and analyze data flows and to help evaluate privacy risk in machine learning pipelines. In addition to the above, research conducted in this project has influenced ongoing privacy discussions, with findings presented at high profile events such as the Federal Trade Commission (FTC) Privacy Conference, the International Conference of Data Protection and Privacy Commissioners, events organized by the International Association of Privacy Professionals and many other venues, while also receiving significant exposure in the press. Most recently, the project also designed and evaluated a number of possible icons and textual descriptions and its recommended design was eventually adopted as the official recommendation for the new California Consumer Protection Act (CCPA).

2.Introduction

The Internet of Things and Big Data are making it impractical for people to keep up with the many different ways in which their data can potentially be collected and processed. What is needed is a new, more scalable paradigm that empowers users to regain appropriate control over the collection and handling of their data. This project introduced and researched **personalized privacy assistants** capable of helping users when it comes to selectively informing them about data practices, nudging them to carefully reflect on their privacy decisions and generally assisting them in managing their privacy by learning models of their preferences and expectations. Through targeted interactions, privacy assistants are intended to help their users better appreciate the ramifications associated with the processing of their data, and empower them to control such processing in an intuitive and effective manner. This includes selectively alerting users about practices they may not expect and/or feel comfortable with, recommending the selection of privacy settings and/or confirming with users other privacy choices, refining models of their preferences, and occasionally nudging them to carefully (re)consider their privacy decisions. Ultimately, by learning our preferences and expectations, privacy assistants will help their users more effectively manage their privacy across a wide range of devices and contexts without placing unrealistic burden on them.

Work in this project was organized around several closely related themes or tasks:

- Personalized Privacy Infrastructure for the Internet of Things, including Personalized Privacy Assistants
- Taxonomy of Privacy Dialog Primitives
- Modeling of User Privacy Preferences and Expectations
- Nudging and Explanation
- Transparency and Compliance Analysis
- Privacy Risk in Machine Learning Pipelines

The following summarizes the methods, assumptions and procedures used in this project. This is followed by a discussion of major results and their implications. Concluding remarks and recommendations are provided at the end of this report.

3. Methods, Assumptions, and Procedures

Work in this project combined user-centered design methodologies, human subject studies, software engineering, machine learning, code analysis and other relevant methodologies as further discussed below. User-centered design methodologies were used to understand users and their tasks in different contexts, to inform the design of technologies that include our IoT Privacy Infrastructure, our IoT privacy Assistant, our mobile app privacy assistant, our Opt-Out Easy browser extension and our privacy nutrition labels. Human subject studies were conducted to develop models of people's privacy preferences and expectations, to evaluate and refine different technologies (e.g., mobile app privacy assistant, Opt-Out Easy browser extension, privacy recommendations, design of privacy logos and textual descriptions for the California Consumer Privacy Act), interventions (e.g., privacy nudges). Software engineering methodologies guided the design and development of software artifacts such as our IoT Privacy Infrastructure, our MAPS Mobile App Privacy Compliance tool, or our Opt-Out Easy browser extension. Machine learning models were developed for people's privacy preferences and expectations as well as to help automatically extract disclosures from the text of privacy policies. Static and dynamic code analysis techniques were developed in support of our work on transparency, including our development of techniques and tools to automatically analyze mobile apps for potential privacy compliance issues. These methods and our work are further described below.

Research conducted as part of this project was motivated by the assumption that people generally care about their privacy, though often subject to diverse and complex privacy preferences and expectations. Another key assumption was that, while people care about their privacy, they struggle to effectively manage it as the amount of information they would need to familiarize themselves with and the number of decisions they would have to make is simply impractical. To make things worse, privacy is typically a secondary task with users often focused on other primary tasks as they are faced with privacy decisions. These assumptions - which had been documented already in prior research, including research by the PIs - were confirmed by several studies conducted in this project. Through work conducted as part of this project, including the development of a privacy infrastructure for mobile and IoT contexts and the development of privacy assistants to help users manage their privacy across these contexts, the PIs were able to demonstrate that it is possible to overcome some of the usability challenges associated with managing one's privacy.

Work in the project was organized around several tasks, as discussed below.

3.1 Personalized Privacy Infrastructure for the Internet of Things and Personalized Privacy Assistants

As we interact with an increasingly diverse set of sensing technologies, it becomes difficult to keep up with the many different ways in which data about us is collected and used. Research has shown that while people generally care about their privacy, they feel they have little awareness of, let alone control over, the collection and use of their data. This situation is the motivation for this project's work on personalized privacy assistants and on a privacy infrastructure for the Internet of Things. The objective is to empower people to regain control over their privacy in a world where the emergence of mobile technologies, the Internet of Things and Big Data is making it increasingly difficult for people to understand and control how data about them is being collected and used.

Within our IoT Privacy Infrastructure (or IoTPI) [DDS+18], which is now available to the public at large (iotprivacy.io), IoT sensors, devices, systems, services, and applications are modeled as "IoT resources" that collect data in a given area. The infrastructure enables people, whether owners or contributors from the general public, to describe IoT resources and their data practices (e.g., what data they collect, how they handle that data, for what purpose, and what types of privacy choices are made available to people whose data can possibly be captured). All users are required to be authenticated and are required to take responsibility for the resource listings they publish. Registries enable organizations and groups of people to curate collections of related resource descriptions such as resources in a mall, in an office building, on a campus, in a given neighborhood, etc. Registries can have owners (e.g., a mall operators, universities, neighborhood organizations) as well as administrators (e.g. personnel working for a mall operator or a university, or administrator in charge of a registry for a given neighborhood or office building). Resource listings include descriptions of the data being collected, the purpose for which collection is taking place, information about retention, analytics, anonymization/de-identification and more. It also includes information about the area where data collection takes place and a radius within which presence of the resource will be publicized (discoverability of the resource).

The IoT privacy infrastructure provides support for each resource to publicize the privacy choices made available to the public, such as opt-in/opt-out choices for the collection and use of their information, request for their data to be deleted and more. This includes support for data subject rights such as those mandated by regulations that include CCPA, CPRA, the European Union (EU) General Data Privacy Regulations (GDPR), Children's On-line Privacy Protection Act (COPPA), and more. At the same time, the IoTPI is regulation agnostic. This means that while it has the fields and

functionality necessary to accommodate the requirements associated with a rich set of regulations, it doesn't impose these requirements on any particular resource description. Instead, it is up to resource owners to make sure their descriptions comply with applicable regulatory requirements. Resource descriptions also include links to privacy policies and are not intended to act as substitutes for these policies (or "privacy policies"). Instead, they provide the basis for more succinct privacy notices in the form of privacy nutrition labels that focus on those types of data practices that people are most likely interested in learning about and would most likely want to be notified about.

As discussed above, resource descriptions can be entered by resource owners and administrators but they can also be entered by contributors from the general public. All users are required to be authenticated before they can create and publish resource descriptions. Many resource description fields include values such as "unspecified" to accommodate situations where a resource owner or a contributor does not have the answer to some questions (e.g., Information about retention practices or who data might be shared with, when the resource description is provided by a contributor who doesn't have access to this information). On the other hand, people are not allowed to publish a resource description until they explicitly indicate that they take responsibility for their description and that the description is accurate to the best of their knowledge.

The Privacy Infrastructure has been designed to make it as easy as possible for resource owners and contributors to enter and publish resource descriptions. In particular, it includes functionality to create and use resource templates that provide pre-filled descriptions of typical resources (e.g., typical video surveillance system in a mall or typical IoT device such as a Ring doorbell), saving people the trouble to research and/or enter details that can be pre-filled for them. This includes functionality that enables vendors to provide their own templates, where they take responsibility for details provided in their templates.

A first fully-fledged version of the infrastructure was deployed at Carnegie Mellon University (CMU) and UCI in 2016. This infrastructure was detailed in our 2018 Institute of Electrical and Electronics Engineers (IEEE) Pervasive Computing journal article, "Personalized Privacy Assistants for the Internet of Things: Providing Users with Notice and Choice" [DDS+18]. This included the development of first versions of our IoT Assistant app, which uses the infrastructure to enable users to discover IoT resources around them, including their data practices and any available privacy choices offered by these IoT resources.

With new personnel joining the team in summer 2018, we updated the codebase and re-implemented the infrastructure to improve scalability, reliability, and usability. This

included migrating to the cloud, using Amazon Web Services, which significantly improved scalability and reliability. Second, we completely re-implemented the IoT Assistant app using Flutter, a relatively new cross-platform app development framework that supports native apps in both iOS and Android, allowing our app to offer a significantly enhanced user experience. This work included multiple design iterations, from evaluating early prototypes to collecting feedback from users downloading the app from the iOS and Android Google Play app stores. Significant effort was also put into the development of functionality to support IoT resource templates, richer administration functionality, and support for a rich collection of APIs enabling IoT Assistant users to access privacy choices made available by IoT resources they discover. This also included multiple redesign and refinement of the IoT portal through which resource owners and contributors define and manage their resource descriptions as well as resource registries and resource templates.

A high-level depiction of the Infrastructure's architecture is shown in Figure 1. The infrastructure includes several key components:

- The IoT Web Portal (IoT Portal): This website enables owners and system administrators to describe IoT resources (including specifying the area where they can be discovered), their data collection and use practices and privacy options
- IoT Resource Registries (Registries), as already discussed earlier
- The IoT Assistant app (IoT Assistant): This is a cross-platform mobile app that enables users to discover nearby IoT resources and their data practices, receive personalized notifications about nearby IoT resources based on a rich collection of criteria they can personalize (e.g., type of data being collected, whether they data is personally identifiable, but also frequency of notifications).
- Privacy Option Management (POM): This is a system that helps users take advantage of privacy options made available by IoT resources they discover.

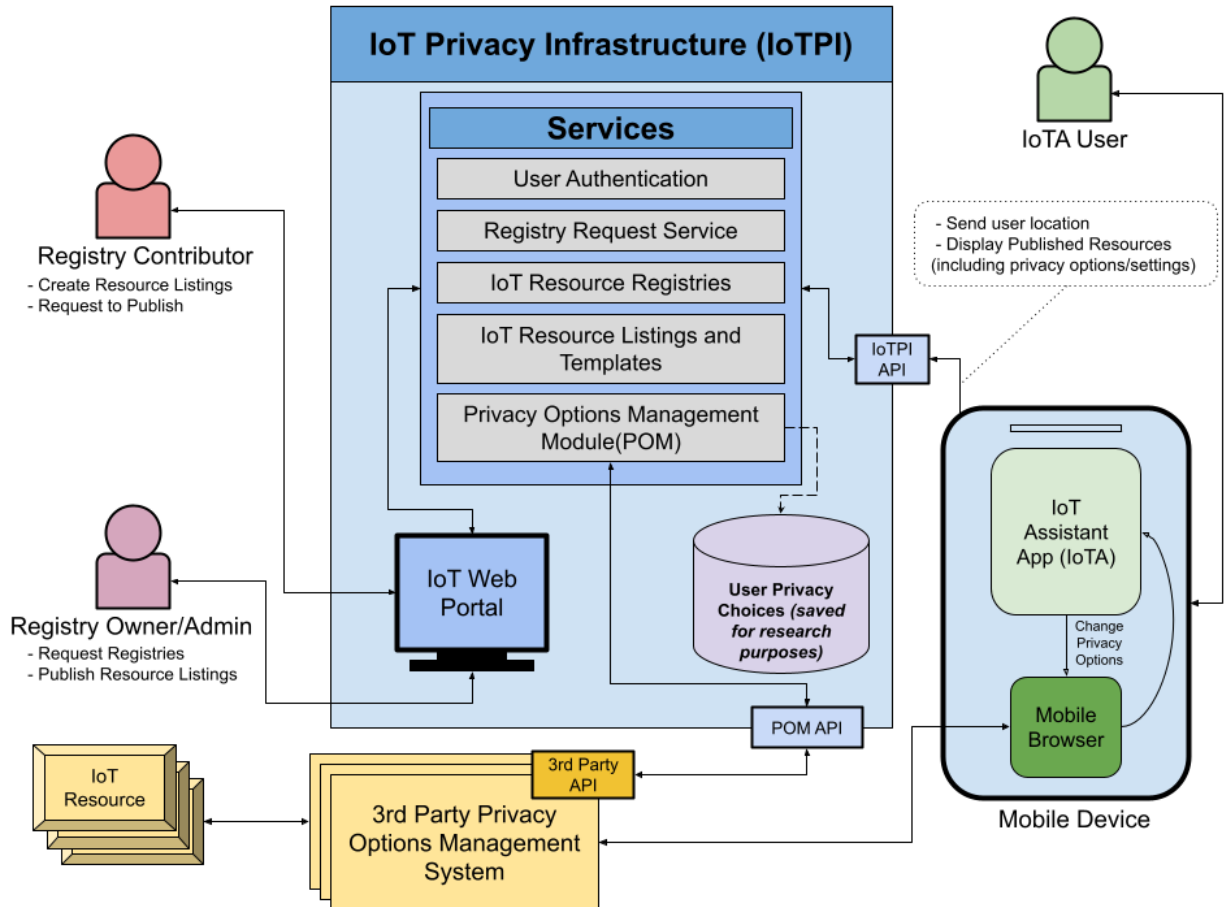


Figure 1: IoT Privacy Infrastructure - High Level Architecture

3.2 Taxonomy of Privacy Dialog Primitives

An important part of effectively communicating with users and motivating them to carefully consider privacy decisions at hand involves developing and evaluating the effectiveness of different dialog primitives. As part of this work, which complements our work on privacy nudging (summarized in another section below), research in this project focused on the evaluation of different types of recommendations, evaluating the impact of privacy recommendations attributed to privacy experts versus privacy recommendations attributed to friends [EDB+18]. Another important dimension of this work has revolved around the development of privacy nutrition labels. The idea here is to identify a small number of standardized privacy disclosures that can help inform users about key privacy issues they would most likely want to know about [EDA+19, EAC+20, EDA+21]. This research is further detailed below.

3.2.1 Exploring How Privacy and Security Factor into IoT Device Purchase Behavior

We conducted in-depth semi-structured interviews with 24 participants who had purchased at least one IoT device (smart home device or wearable) [EDA+21]. We explored interviewees' understanding of privacy and security issues associated with IoT devices and factors they considered when purchasing their device. At the end of each interview, we displayed prototype IoT security and privacy labels that we developed, and discussed them with interviewees. Finally, we conducted a 200-participant Mechanical Turk (MTurk) survey to probe the influence of privacy and security information when making IoT purchase decisions.

3.2.2 An Expert Study to Determine IoT Privacy and Security Label Content

Given consumers' scarce attention, presenting them with the most relevant security and privacy information in the most digestible form is crucial. To determine the most important information to include on IoT privacy and security labels, we solicited the opinions of privacy and security experts [EAC+20].

We conducted interviews and surveys with 22 privacy and security experts. To get different perspectives, we recruited experts from industry, academia, government, and non-governmental organizations (NGOs). We also ensured that these experts come from different backgrounds related to IoT (software, hardware, and policy). We used the iterative Delphi methodology to develop a consensus among the experts around important factors and an understanding of their reasons for or against including each factor.

3.3 Modeling of User Privacy Preferences and Expectations

People's privacy preferences and expectations have been shown to be complex and diverse. An important part of helping people manage their privacy requires developing a deeper understanding of these preferences and expectations, including those contextual attributes that influence them and how preferences and expectations vary across different individuals. This information can in turn be used to identify effective ways of notifying people about those data practices they care to be informed about without overwhelming them with large numbers of details they do not care about. This information can also be used to identify those privacy settings and options that should be made available to users, the granularity and expressiveness of these settings, looking at the level of control users wish to have as well as the burden placed on users by different possible settings/choices. Another important dimension of work in this area

has involved the development of predictive models and different possible ways of using these models to mitigate user burden while empowering users to retain the level of control they desire.

Our work in this area has included looking at the management of mobile app permissions on smartphones (e.g.[LAS+16,SZF+20]) as well as looking at a diverse range of IoT environments and technologies (e.g.,[EBH+17,PDY+17,ZFB+21,ZFD+20a,ZFD+20b]) , as summarized below.

3.3.1 Mobile App Privacy Assistant Studies

Over the years the number of mobile app permissions smartphone users have to manage has become unrealistically high. A typical smartphone user with 50 apps on their phone and an average of 3 permissions per app would have to configure 150 permission settings to ensure that the apps on their phone align with their privacy concerns. As part of this research, we have shown that using machine learning techniques it is possible to develop predictive models of people's privacy preferences that can help reduce the number of mobile app permission settings people need to manually configure. This research involved collecting detailed information about people's mobile app permission preferences (Figure 2) and using the resulting data to develop predictive models (Figure 3) [LAS+16]. In the example depicted in Figure 3, a small number of "profiles" based on seven different clusters is used to provide mobile app permission recommendations to users. This is done by first assigning users to a particular cluster based on their answers to a small number of questions, then relying on those mobile app permissions for which people in the cluster have been found to have highly similar preferences [LAS+16]. This work however did not stop with just the collection and analysis of people's privacy preferences, but also involved the development and launch of an actual personalized privacy assistant for mobile app permissions, which was evaluated with Android users as they went about their regular daily activities (Figure 4). Encouraged by the successful pilot of this technology and positive reactions from people who piloted it, the project went on and released a version of the personalized assistant in the Android Google Play store for use by people with rooted Android phones.

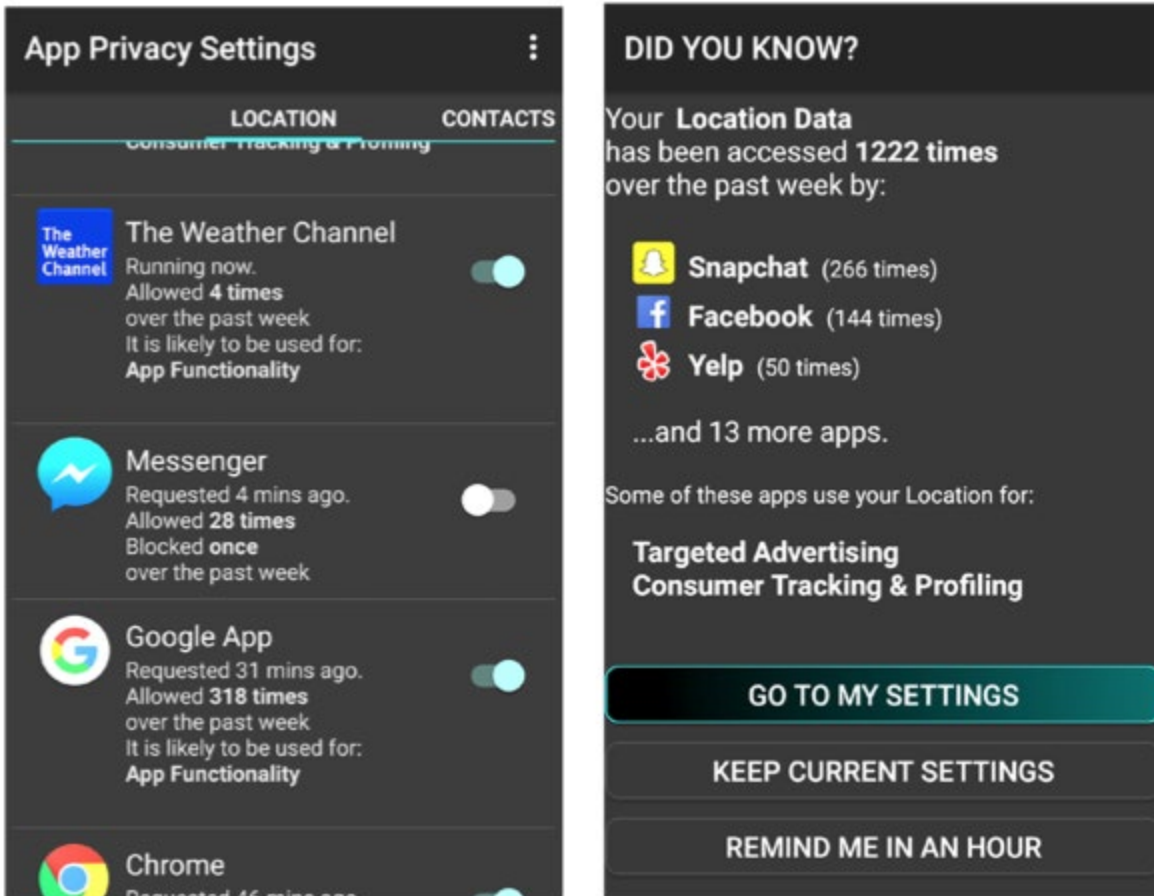


Figure 2: Motivating users to reflect on their current permission settings and possibly modify them, as part of study that involved collecting and analyzing people's mobile app permission preferences as part of their regular daily interactions with their regular daily interactions with their regular smartphones [LAS+16].

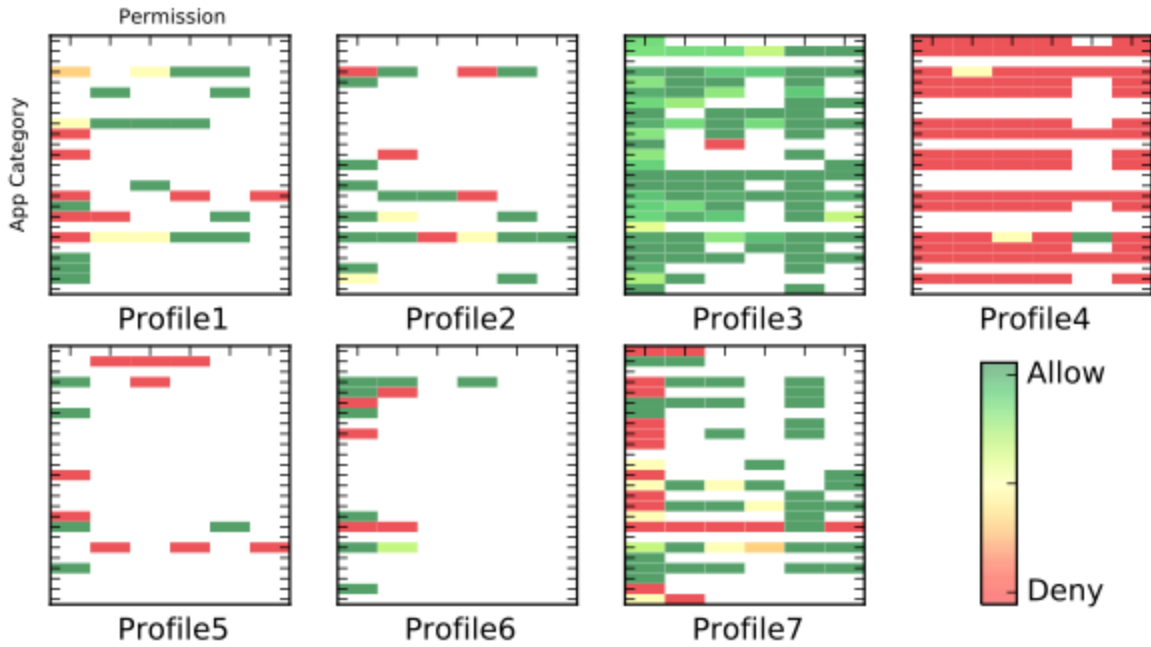


Figure 3: Example of analysis, where people’s mobile app permission preferences are analyzed using clustering techniques.

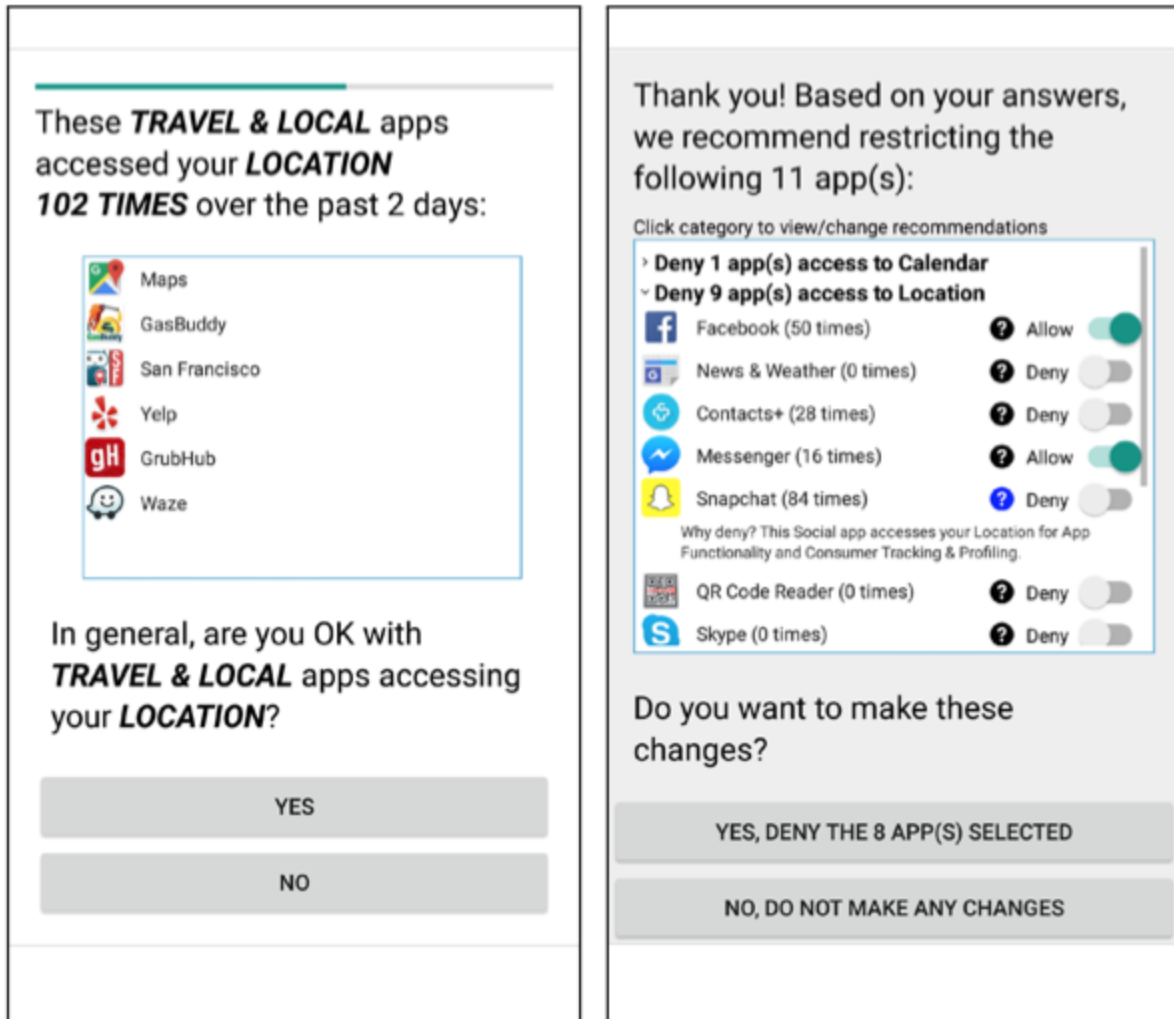


Figure 4: A mobile app permission assistant asks users a small number of questions to assign them to a cluster/profile. Permissions for which people in the given profile exhibit highly homogeneous preferences are used as a basis for recommending possible changes to the user's current mobile app permission settings. The user can also access an explanation for each recommendation [LAS+16].

While the number of mobile app permission settings a typical smartphone user has to configure is already unmanageable, research has also shown that available settings are overly simplistic and fail to capture key attributes influencing people's comfort when it comes to granting different apps access to different permissions. As part of research aimed at empowering people to control permission settings that are better aligned with their privacy preferences while avoiding a further increase in user burden, our project explored the possibility of giving users more expressive mobile app permissions while using machine learning to help mitigate user burden.

To what extent might these stronger predictive models help mitigate the greater user burden that would otherwise be associated with the configuration of more expressive privacy and security settings? Specifically, we focused on answering this question in the context of mobile app permissions, comparing models with permission settings that take the purpose of permissions into account versus models that do not. We compiled quantitative results aimed at evaluating this trade-off between accuracy and user burden across a number of parameter configurations.

Our first goal was to create a large corpus of user preferences about a variety of app permissions, for a variety of Android apps. The purpose of this corpus was to perform data mining which could potentially reveal insights into common factors along which user preferences would align. These patterns are indicative of the potential for improving predictive power.

Having analyzed this corpus to find statistically significant factors, our second goal was to determine how to use this predictive power to improve recommendation models for preferences, leveraging profiles that incorporate a combination of supervised and unsupervised machine learning (agglomerative hierarchical clusters and conditional inference trees). We also sought to create empirically derived guidance for the design of systems which employ these machine learning techniques to improve permissions management systems. We empirically determined the number of questions required to successfully profile users and count the instances where additional user input is required to make strong predictions. We measured the differences in efficiency and accuracy between the models which consider purpose and those which do not. Using machine learning, our approach demonstrated that it is possible to improve the expressiveness of mobile app permissions model without trading off accuracy for user burden and vice versa [SFZ+20].

3.3.2 Modeling People's Privacy Expectations Across a Wide Array of IoT Scenarios

As more complex IoT scenarios become possible, many other factors may play a role in determining individuals' privacy preferences. While some may feel comfortable with their location being tracked for the purpose of traffic prediction, others may want to limit tracking to their work commute or to certain days or hours in the day. Some may consent only if they are assured that their location data is retained and used in an anonymized form, etc.

We conducted a large-scale online vignette study to identify the contribution of different factors (such as the type of data, retention time, purpose of data collection, and location

of data collection) in promoting or inhibiting individuals' self-professed comfort levels [EBH+17]. We also studied the factors that trigger a desire for notifications about data collection. Our research identified which aspects of data collection or use by various IoT devices are most likely to cause discomfort, how realistic participants think these scenarios are, and which aspects they would like to be made aware of.

3.3.3 Informing the Design of Personalized Privacy Assistants

The increasing deployment of various IoT technologies in the daily environment creates new ways through which personal data is collected and used. People often have little awareness of, and even less control over, the data collection and use practices associated with IoT technologies. Personalized Privacy Assistants could address this issue by helping people discover and, when available, control the data collection practices of nearby IoT resources. This however requires understanding users' perceptions and acceptance of different possible configurations of PPAs, particularly in the IoT context. Without solid research evidence, it is challenging to design usable PPAs that match potential users' privacy needs and comfort levels.

To shed light on this issue, we conducted a qualitative interview study with 17 participants to explore user perceptions towards PPAs for IoT technologies [CFP+20]. Specifically, we focused on understanding how users would respond to different possible PPA implementations that may leverage machine learning-based predictions to assist users with privacy decisions. In the interviews, we asked participants about their opinions for three increasingly more autonomous PPA implementations: notification PPA provides end users with awareness and control over data requests; recommendation PPA gives users recommendations on how to respond to individual data requests; and auto PPA makes autonomous data sharing decisions for the user.

We recruited participants in the greater Pittsburgh area through both digital and non-digital channels. We invited potential participants to complete a short screening survey including questions about demographics and experiences with technology. From the pool of potential participants who responded to the screening survey, we selected a relatively diverse sample of participants for in-person interviews. Out of our 17 participants, eight self-identified as male, eight as female, and one as non-binary. Our participants had a mean age of 39 years (min = 22, max = 68) and were fairly well educated (13 had some form of higher education). Most of them were employed (eight full-time, four part-time). Only three participants self-identified as students, and four reported working with or studying in areas related to technology or security. Overall, the participants had a good, but incomplete, understanding of IoT technologies.

All interviews were audio recorded and transcribed into text for in-depth qualitative data analysis. Specifically, two researchers reviewed the passages related to participants' opinions about the different implementations of the PPA and their interaction preferences. Each researcher summarized these passages into higher-level concepts, and they discussed those concepts together. After noticing a high level of consistency in the concepts identified, we opted to forego formalizing a codebook and iterative coding process, instead resolving the few conflicts that occurred. After a comprehensive qualitative analysis on the coded data, we identified benefits and issues associated with each implementation. The findings from this qualitative interview study are described in Subsection 4.3.3.

3.3.4 Understanding and Modeling People's Privacy Attitudes Towards Video Analytics Technologies Across a Representative Cross-Section of Deployment Contexts

Over the past few years, facial recognition, a type of artificial intelligence (AI)-enabled video analytics technology, has become increasingly accurate with recent advances in deep learning and computer vision. The growing ubiquity of video analytics is contributing to the collection and inference of vast amounts of personal information, including people's whereabouts, their activities, whom they are with, and information about their mood, health, and behavior. Unfortunately, such data collection and usage often take place without people's awareness or consent. While video analytics technologies arguably have many potentially beneficial uses (e.g., law enforcement, authentication, mental health, advanced user interfaces), their broad deployment raises important privacy questions and has prompted increased scrutiny from both privacy advocates and regulators. New regulations such as the European Union's GDPR and the California Consumer Privacy Act mandate specific disclosure and choice requirements that apply to the deployment of video analytics technologies. While these regulations are important steps towards providing data subjects with more transparency and control over their data, they do not specify how people should be notified about the presence of video analytics, or how to effectively empower them to exercise their opt-in or opt-out rights. This includes addressing questions such as when to notify users, what to notify them about, how often to notify them, how to effectively capture their choices, and more.

Our research here aimed to address these issues by developing a more comprehensive understanding of how people feel about video analytics deployments in different contexts, looking both at the extent to which they expect to encounter them at venues they visit as part of their everyday activities and at how comfortable they are with the presence of such technologies across a range of realistic scenarios [ZFD+20a, ZFB+21,

ZFS21]. Our study focused on understanding people's privacy expectations and preferences. This included looking for possible social norms that might extend to a larger population, or alternatively identifying differences in how people respond to various deployment scenarios. The second set of questions was motivated by recent technical advances, namely (1) the development of real-time face denaturing functionality that enables video analytics software to only be applied to people who provide consent, and (2) the development of a privacy infrastructure for the Internet of Things that enables entities deploying video analytics software to publicize their data practices and allow data subjects to opt in or out of having their footage analyzed and/or shared. Using this functionality, it becomes possible to notify people in real-time as they approach areas where video analytics technologies are deployed and allow them to selectively opt in or out - as required by regulations such as GDPR or CCPA. Because expecting people to manually opt in or out of video analytics each time they come near this type of functionality would entail an unrealistically high number of privacy decisions, we use our data to explore the feasibility of developing predictive models that could assist users with their privacy decisions - with users able to review recommendations from the predictive models.

To answer these research questions, we designed an experience sampling study to collect people's responses to a variety of video analytics deployments (or "scenarios") in the context of their regular everyday activities. Context has been shown to play an important role in influencing people's privacy attitudes and decisions. Studying people's privacy attitudes through online surveys is often limited because participants answer questions about hypothetical scenarios and often lack context to provide meaningful answers. Accordingly, the experience sampling method has been repeatedly used in clinical trials, psychological experiments, and human-computer interaction (HCI) studies, yielding more ecologically valid behavior. This enables us to engage and survey participants in a timely and ecologically valid manner as they go about their normal daily lives. Participants are prompted to answer questions about plausible video analytics scenarios that could occur at the location in which they are actually situated (Figure 5). By systematically exploring more concrete scenarios in actual settings associated with people's day-to-day activities, we are able to elicit significantly richer reactions from participants and develop more nuanced models of their awareness, comfort level, and notification preferences pertaining to different deployment scenarios.

The scenarios considered in our in-situ study were informed by an extensive survey of news articles about real-world deployments of video analytics in a variety of different contexts (e.g., surveillance), marketing, authentication, employee performance evaluation, and church attendance tracking. These scenarios provided the basis for the identification of a set of relevant contextual attributes which were randomly manipulated

and matched against the different types of venues our subjects visited. We employed a factorial study design and developed a taxonomy that captured a representative set of attributes one might expect to influence individuals' privacy attitudes, including the purpose for which the data is collected, what type of facial recognition analyses are involved, retention of raw footage and analysis results, whether sharing with other entities are involved, and the types of places participants visited.



Figure 5: In-situ study of people's privacy attitudes towards Video Analytics technologies across a number of deployment scenarios. Screenshots of the study app and the web survey used for the evening review [ZFD+20a].

The 10-day study comprised the following five stages. First, eligible participants completed the consent forms for this study and downloaded the study app from the Google Play Store. Upon installing the app, participants completed a pre-study survey about their perceived knowledge level, comfort level, and notification preference with regard to facial recognition.

Second, participants were instructed to go about their regular daily activities. The study app collected participants' Global Positioning System locations via their smartphones. As they visited places for which we had one or more plausible deployment scenarios, the app would send them a push notification, prompting them to complete a short survey on a facial recognition scenario pertaining to their location (Figure 5). Third, on the days participants received push notifications via the app, they also received an email in the evening to answer a daily summary web survey. This web survey showed participants the places they visited when they received notifications, probed reasons for their in-situ answers, and asked a few additional questions. Fourth, after completing 10 days of evening reviews, participants concluded the study by filling out a post-study survey administered via Qualtrics. This survey contained free-response questions about their attitudes on facial recognition, the 10-item Internet Users' Information Privacy Concerns (IUIPC) scale on privacy concerns, as well as additional demographic questions like income, education level, and marital status. The last stage is optional: participants who indicated they were willing to be interviewed in their post-study survey may be invited to an online semi-structured interview. The interview contained questions about study validity, perceptions of scenarios, and clarifications with regard to their earlier responses. The findings from this study are described in Subsection 4.3.4.

3.4 Nudging and Explanation

Nudges have emerged within the privacy and security literature as an effective means of affecting, and possibly assisting, user behavior. Nudges work by modifying the structure of choices to encourage certain behaviors without altering economic incentives. Recent research has applied nudges to diverse privacy and security scenarios where users face hurdles in the decision making process, and where those hurdles can result in negative outcomes.

3.4.1 Mobile App Privacy Nudging - Studying the Impact of Framing on the Effectiveness of Privacy Nudges

As part of our prior research, we had already demonstrated how using privacy nudges, namely interventions designed to get people to reflect more carefully on their privacy decisions, it is possible to motivate users to visit privacy settings such as mobile app permissions and adjust them to that they are better aligned with one's privacy preferences (e.g., [ASS+15]). Encouraged by these results, we undertook to design a study aimed at exploring the impact of framing on the effectiveness of privacy nudges, using mobile app privacy nudges as our domain of investigation [Alm+17].

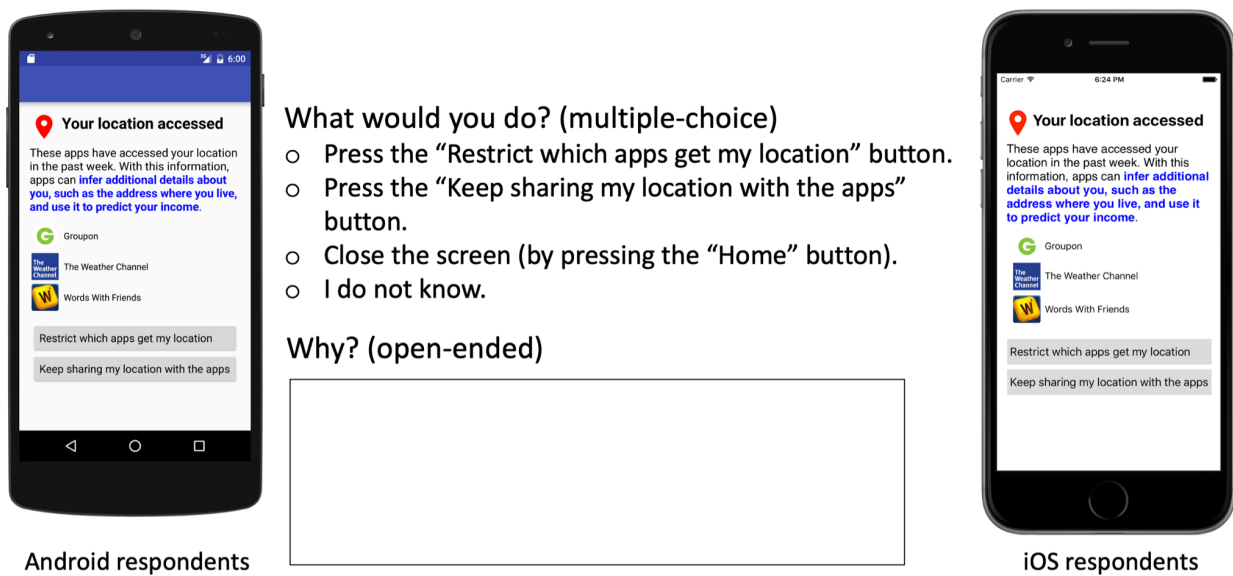


Figure 6: Studying the impact of framing on the effectiveness of privacy nudges in the context of mobile app location privacy nudges.

Condition	Wording
(1) Baseline	"These apps have accessed your location in the past week."
(2) Frequency	"These apps have accessed your location 1,865 times in the past week."
(3) Background	"These apps have accessed your location in the past week although you did not use them. "
(4) Purposes	"These apps have accessed your location in the past week for purposes not related to the apps' main function. "
(5) Purposes + Example	"These apps have accessed your location in the past week for purposes not related to the apps' main function, such as location-based advertising. "
(6) Inferences	"These apps have accessed your location in the past week. With this information, apps can infer additional details about you. "
(7) Inferences + Example	"These apps have accessed your location in the past week. With this information, apps can infer additional details about you, such as the address where you live. "
(8) Predictions	"These apps have accessed your location in the past week. With this information, apps can infer additional details about you, such as the address where you live, and use it to predict your income. "
(9) Predictions + Implications	"These apps have accessed your location in the past week. With this information, apps can infer additional details about you, such as the address where you live, and use it to predict your income. Knowing your income can affect prices and discounts you see in ads. "

Figure 7: Nine different ways of framing mobile app location privacy nudges, from mundane nudges to nudges that place more emphasis on how location information might be used and associated privacy risks.

Specifically, we focused on a number of different ways of framing a nudge designed to motivate users to revisit location permission access to their mobile apps. This involved manipulating the text shown to users in privacy nudges dealing with access by mobile apps to their location information (Figure 6). Figure 7 shows the different types of framing considered in our study.

3.4.2 Psychometrically Tailored Privacy Nudges

While effective, many of the applications of nudges within the privacy literature have focused on 'one-size-fits-all' approaches, where a certain behavioral intervention is applied to a diverse set of individuals, with no tailoring of the nudge based on characteristics unique to each individual.

To address this gap in the literature, we investigated psychometrically tailored nudges in the context of privacy behavior [WAS+19]. We conducted three online experiments that attempted to identify effects for tailored nudges on data disclosure choices. Across the three studies, participants completed surveys in which they were presented with a hypothetical (Study 1) or real (Studies 2 and 3) disclosure choice. Each disclosure choice was paired with a nudge designed to encourage participants to either allow or prohibit the disclosure of potentially sensitive information. After recording participants' disclosure choices, we measured participants along a variety of psychometric scales

designed to capture individual differences in decision making and personality. We examined whether changes in the differences in disclosure rates could be predicted by the measured psychometric variables. We modeled this relationship using logistic regression where the disclosure choice made by participants was our dependent variable.

Our examination of tailored privacy nudges took place within the context of data disclosure decisions. To facilitate this, each of our studies began by presenting participants with a hypothetical or real disclosure choice and asked them to make a 'Yes' or 'No' decision on whether or not they wished to disclose their data. We inserted nudges into these scenarios by modifying the choice text to elicit differences in disclosure rates between groups of participants. We employed two types of nudges across our three studies: 'framing' nudges and 'social norms' nudges. The first of these modified the choice text to leverage framing effects, while the second introduced additional information to the choice to establish social norms which might affect participants' disclosure behavior. In Study 1, we tested only the framing nudge. In Study 2 and Study 3, we tested both the framing and social norms nudges.

For each nudge, we created two different wordings or 'variants' of the nudge text. Relative to each other, these variants created the desired psychological effects we wish to leverage within our nudges. For example, we create the framing nudge for Study 1 by asking one group of participants whether they wished to 'Opt-In' to a service while asking the other group whether they wished to 'Opt-Out'. The framing effect exists only in the contrast between the two wordings. We applied a similar method to create the social norms nudge used in Study 2 and Study 3. We treated each nudge 'variant' as an experimental condition within our study. Participants were randomly assigned to a single nudge variant in a between-subjects design. For Study 1, this translated into two experimental conditions across one nudge. For Study 2 and Study 3, this translated into four experimental conditions across two nudges. Across the three studies, participants were only assigned to one disclosure decision that contains a single nudge variant.

Following the decision task, we asked all participants questions from psychometric inventories designed to measure their decision making and personality traits. We selected inventories for our studies that either relate to the cognitive effect being leveraged in the nudge or have been examined in other studies exploring psychometrically targeted nudges.

3.4.3 Transparency via Attribution in Vision Models and Internal Explanation for NLP models

Machine learnt models, especially deep models, are largely black-boxes resistant to human inspection. This, together with their increasing application to impactful scenarios, leads to problems such as lack of trust, potential violation of ethical norms, privacy violations, and others. While we cannot address the full breadth of these problems with all of their forms, we investigate two common types of machine learning models alongside common forms of explanations for their operation. We investigate attributions in vision models and internal explanations for natural language processing models (NLP).

Input Attribution in Vision Models.

Deep vision models such as convolutional neural networks are among the most difficult to explain. Instead of analyzing the internals of such models, popular forms of explanations are input attributions which quantify how much impact each input pixel has on a model output. Two such input attributions can be seen in Figure 8 (Two types of input attributions for an image classification model).

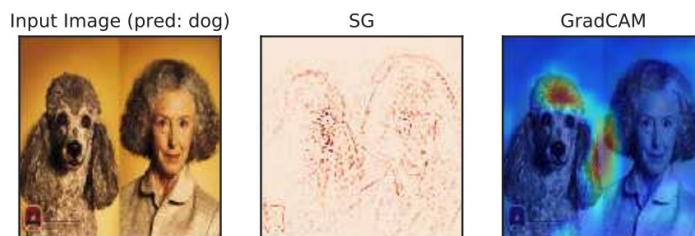


Figure 8: Two types of input attributions for an image classification model.

A variety of such attribution methods have already been proposed, each with differing design goals. As part of our work, we investigated these methods in terms of the mode of interpretation that is meant to occur [WMD+20]; that is, an input attribution may be interpreted differently by different people or in different contexts. In a related work, [WWD+20], we generalized a class of attribution methods while relating instances to the interpretation criteria. Additionally, we initiated a study of the relationship between robustness of models (against adversarial perturbations) and input attributions [WWR+20].

Internal Explanations in Natural Language Processing Models.

Unlike vision models, NLP models often have recurring intentional structures as for example in long short term memory (LSTM) models that contain elements meant to keep track of global input information and other elements meant to track local inputs. Explaining such models' operation can thus benefit from highlighting model components involved in various predictions tasks. Exercising NLP models on restricted tasks is not trivial however, as many models must incorporate a large amount of grammatical and syntactic information from the entire language being processed. In [LML+20] we investigate these problems with a novel analysis tool which is capable of pointing out paths throughout neural networks that are most responsible for conveying a concept of interest. For example, Figure 9 shows a (slice of data pathways in an LSTM language model) most responsible for that model's ability to correctly incorporate subject-verb number agreement in the English language.

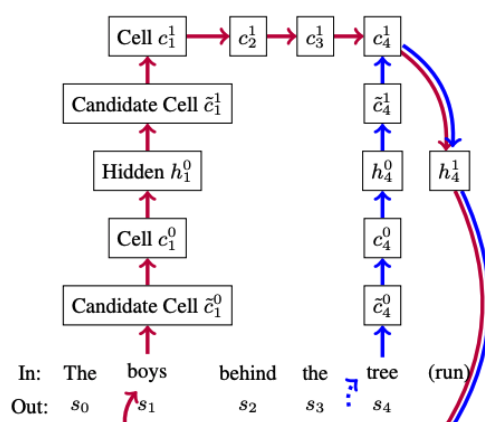


Figure 9: Slice of data pathways in an LSTM language model

3.4.4 Explanations for Mobile App Privacy Permission Recommendations

As automated recommendation and decision systems have grown in complexity, interest in explaining the resulting decisions has also grown. While multiple goals exist for explaining recommendations generated by AI systems, explaining how these systems function may increase trust in the system leading to higher adoption rates for the recommendations among users. Several studies in the privacy and security literature have focused on explanations with this goal in mind. This same focus may be applied to mobile privacy recommendations produced by privacy assistants. These recommendation systems suggest changes to users' privacy settings for individual apps such that apps' behavior regarding data access better conform with users' privacy preferences. This study seeks to examine whether explanations for mobile privacy

recommendations can increase adoption rates and what types of explanations may be most effective.

To answer these questions, we designed a study to examine the effectiveness of explanations for recommendations produced by a mobile privacy assistant for Android devices. The privacy assistant works by assigning each user to a privacy 'profile' based on their answers to several privacy preference questions. Based on the assigned profile and the specific apps the user has installed on their device, the privacy assistant generates a list of changes the user might make to their privacy settings for particular apps. These recommendations are based on user data collected as part of a previous study. For our study, we select a single 'deny' recommendation - where the privacy assistant recommends that the user change their permission settings to deny an app access to a particular type of data. Due to changes in the Android permission model introduced in Android version 10, we limit participation in our study to Android version 9 and lower.

Using the platform of the Android mobile privacy assistant, we test two different types of explanations. In both cases, the explanations were designed to explain how the privacy assistant works with the goal of increasing recommendation adoption rates. In the first explanation type, we used information about the 'group' of users used to train the recommendation model within the privacy assistant. These explanations make users aware of this 'group' data and similarity of their privacy preferences to those users from the training data which share the same privacy profile assignment. With the second type of explanation, our goal was to add additional information to the first explanation to make the functionality of the privacy assistant more salient to users. Our hypothesis is that a greater understanding of how the privacy assistant operates may increase the rate at which users adopt its recommendations. For the second type of explanation, we combine the 'group data' from the first explanation type with inferences about the recommender derived using 'quantitative input influence' methods. In the context of the privacy assistant, quantitative input influence can be applied to determine which of the privacy questions users answered was most influential in their final privacy profile assignment. This therefore relates a decision made by users (their answers to the privacy questions) to the final recommendation.

Our study design tests the recommendations of the privacy assistant and the two explanation types using four experimental conditions. In the first condition, participants do not receive a recommendation or explanation. Participants in this condition are presented with a recommendation screen in the privacy assistant app that contains only the instructions on how they can access and change their permission settings if they wish. Participants assigned to the second experimental condition are shown a

recommendation when one is available but are not provided with any explanation. The third and fourth experimental conditions contain recommendations and explanations. The third condition presents participants with a recommendation and 'group' explanation while the fourth condition presents participants with a recommendation along with a 'group + QII' recommendation. By comparing the number of attempts by participants to change their permission settings between conditions 1 and 2, we can assess the effectiveness of the recommendations. By comparing the number of attempts between condition 2 and conditions 3 and 4, we can determine the impact of the explanations. Finally, by comparing conditions 3 and 4, we can examine the effectiveness of the two different types of explanations.

While the recommendation and explanation a participant might receive depends in part on their randomly assigned experimental condition, the study procedure is the same for all participants. After consenting to the study, participants are instructed on how to download and install the Android privacy assistant from the Google Play Store. Once installed, participants open the app, upon which the privacy assistant categorizes each app on the participants device. Following the categorization step, the participant is presented with up to five privacy questions. Their answers to these questions, along with the currently installed apps and their permission settings are used to generate a single deny recommendation. Depending on the participant's assigned privacy profile and installed apps, a deny recommendation may not always be available. Based on whether a recommendation is available and the participant's randomly assigned condition, the participant may be shown a recommendation and explanation. The participant is then given the opportunity to view the recommendation/explanation and make any changes to their privacy settings. Once they have completed this step, participants are instructed on how to uninstall the privacy assistant and are redirected to an exit survey on Qualtrics. Within this survey, participants are asked questions designed to measure the quality of the recommendation and explanation along with their trust in the privacy assistant as a whole.

3.4.5 Transparency of Privacy Choices on the Web and Opt-Out Easy Browser Add-On Tool

Another key problem around web data privacy is the lack of transparency for privacy choices. Even though many websites provide privacy choices to users, such as opting in or out of certain web data collection and use practices, it is very difficult for average web users to find these choices and configure the settings according to their privacy preferences. Website privacy policies sometimes provide users the option to opt-out of certain collections and uses of their personal data. Unfortunately, many privacy policies bury these instructions deep in their text, and few web users have the time or skill necessary to discover them.

To address this issue, our work focused on using natural language processing techniques to automatically detect opt-out choices in privacy policy text. Our overall approach for extracting and classifying opt-out choices combines heuristics to identify commonly found opt-out hyperlinks with supervised machine learning to automatically identify less conspicuous instances. Built upon this work, we also developed and successfully evaluated Opt-Out Easy, a web browser extension designed to present available opt-out choices to users as they browse the web [KIN+20].

3.5 Transparency and Compliance Analysis

3.5.1 Data Flow Modeling and Analysis

The use of sensitive data whether directly or indirectly in systems including machine learnt models is common cause of privacy violations. Defining use, however, is not always easy especially when the apparent privacy violations occur sufficiently downstream from the sensitive information use that causes them, or when sensitive information is derived from non-sensitive data. We investigate two aspects of the complexity of use: 1) information theoretic analyses of secrets (i.e., sensitive information) derived by potentially non-sensitive or vulnerable data (i.e., demographics), and 2) indirect uses by association (i.e., proxies) in machine learnt models.

Both of these cases point out difficulties in reasoning about flow of sensitive information that lead to privacy violations but require different approaches. We take an information theoretic approach in [AMH17] to analyze the simple flow of data from secret picking strategy to the secret itself and what it implies about its vulnerability to being guessed. When dealing with machine learnt models, on the other hand, we focus on the syntax and semantics of common models (see for example proxy uses in simple decision trees in Figure 10 (Uses and non-uses of sensitive information in decision trees) from [DFK+17]).

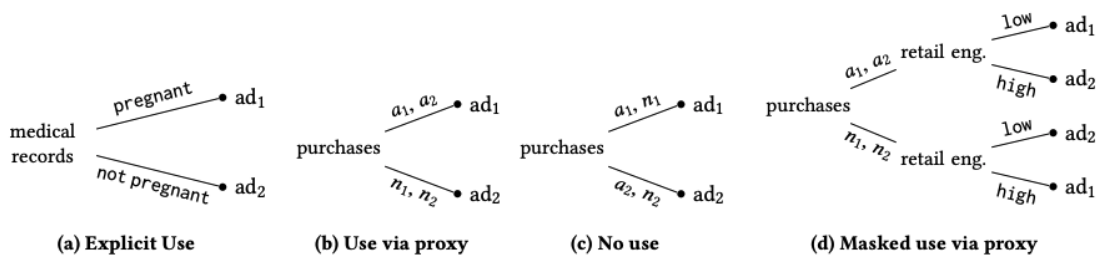


Figure 10: Uses (a,b,d) and non-uses (c) of sensitive information in decision trees.

3.5.2 IFTTT: Identification and Analysis of Vulnerabilities in IoT End-User Programming Environments

The use of end-user programming, such as if-this-then-that (IFTTT), is becoming increasingly common. Services like IFTTT allow users to easily create new functionality by connecting arbitrary Internet-of-Things devices and online services using simple if-then rules, commonly known as applets or recipes. However, such convenience at times comes at the cost of security and privacy risks for end users. Security and privacy risks could arise from rules that inadvertently leak private information, rules that can be triggered from an untrusted source but execute a potentially dangerous action, or rules that chain together unexpectedly (where the action executed by one rule serves as a trigger for another). For example, a user could write a rule to open the window if the temperature rises above a certain threshold. This rule enables an attacker who can affect the temperature of the user's house to cause the window to open and hence give him access to the house. The attacker could affect the temperature of the house in different ways: for example, if there is an outside fuse box, the attacker could flip the breaker and turn off the air conditioning; if the air conditioning has an exhaust to the outside, the attacker could cover it. To give another example, to conveniently manage email attachments, a user could write a rule that uploads all attachments from newly received emails to her OneDrive folder. Later, if the user receives an email with a malicious attachment and uses OneDrive to sync multiple devices, then the malicious attachment would automatically be copied to multiple devices, increasing the likelihood that the user will mistakenly execute the malicious program.

To gain an in-depth understanding of the extent to which users may be creating rules that expose them to potential security and privacy risks, we carried out two investigations: one that examines potential risks through an information-flow analysis of all published IFTTT rules [SAB+17]; and one that examines users' actual IFTTT rules in the context in which they are used [CSK+20].

In the first investigation, we examined a set of 19,323 unique published IFTTT rules—most rules are published so that they can be reused—collected by Ur et al. [UPB+16]. Based on our manual inspection of the rules, we defined an information-flow lattice consisting of labels that specify the secrecy and integrity levels of rules' triggers and actions. We analyzed individual rules based on this information-flow model to identify rules that were violations of integrity or secrecy. We also manually examined a random sample of these violating rules, we categorized the potential harms to users resulting from these rules.

In the second investigation, we study the risks of real-world use of IFTTT by collecting and analyzing 732 applets (each of which implements one or more rules) installed by 28

participants and participants' responses to several survey questions. Survey questions addressed the context in which the applets are used (e.g., who cloud storage documents are shared with), participants' understanding and perception of secrecy and integrity risks (e.g., if they had considered certain risks when setting up rules, if they had experienced any harms, and if they believed certain risks were possible for a particular rule), and how they would react to specific violations identified in prior work. We used the automated information-flow-based analysis developed as part of our first investigation to identify applets that have secrecy or integrity violations. We then examined the applets manually, in more detail, considering context such as their titles. Through this examination we determine to what extent the violations identified by automated analyses correspond to real risks and harms, and we also identify new harms that can be caused by IFTTT rules but that are beyond the scope of current automated analyses.

3.6 Privacy Risk in Machine Learning Pipelines

3.6.1 Proxy use

We studied notions of proxy use that consider why a model uses a correlated feature. Our study assumed a classifier produced by an ML algorithm running on a training data set, both of which are known to the investigator. It assumes that a putative proxy targets some protected attribute, such as medical status and that the proxy may be used by the classifier to reach some decision. For this setting, we developed a systematic range of proxy notions. Here we show three such notions of proxy, each being more restrictive (Figure 11).

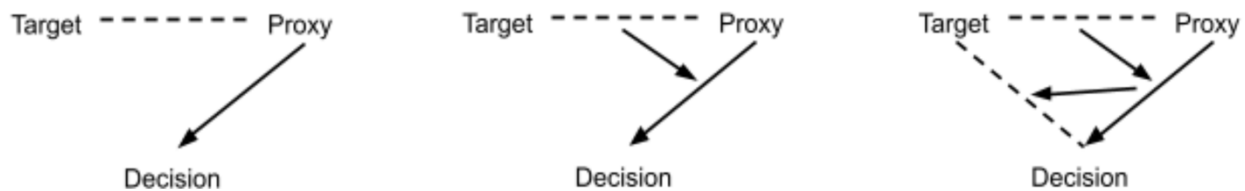


Figure 11: The relationships between variables required for three different notions of proxy. The arrows represent causation; dashed lines, correlation.

They vary in terms of which causal and correlational conditions imply them. For example, the third, strongest notion of proxy, requires that a correlation between the targeted variable and the proxy causes the classifier to use the proxy in a way that introduces a correlation between target and the classifier's decision.

Unlike preexisting notions that use first-order forms of causation and correlation, many of the newly proposed notions depend upon second-order forms of causation and correlation. Whereas first-order forms look at individuals, second-order ones look at sets of individuals, such as training data sets. These second-order forms can describe how machine learning systems use data sets to select models and how relations between individual-level (first-order) variables in such data sets can affect the learning process. A goal of this work was to put such second-order forms of causation onto a firm footing, including comparing it to the distinction between triggering and structuring causes found in prior work.

Another goal was to relate these forms of proxies to those found in prior legal analyses of proxy discrimination, including a definition of “unintentional proxy discrimination” that looks for statistical independence between the proxy and the correct decision given the target. The final goal was providing practical auditing approaches for different notions of proxy.

3.6.2 Overfitting and Whitebox Membership Inference

A machine learning model is said to overfit its training data when its predictive accuracy on unseen test sets diverges from the accuracy observed during training. Recent work by Fredrikson demonstrates an explicit connection between overfitting and privacy risk by characterizing the vulnerability of models to certain privacy attacks as a function of the model’s generalization error. Intuitively, this connection can be thought of in terms of the model “remembering too much” about the training data. This view implicitly assumes that the model’s greater accuracy on training data stems from it learning properties specific to training instances, but that are not representative of the general distribution. Privacy risk then follows when an attacker is able to make inferences about the training data by decoding this behavior through black-box queries.

The blackbox view of overfitting and privacy is a promising step forward in our understanding of causes of privacy risk in machine learning pipelines, but it does not explain all of the risk associated with releasing trained models. Notably, some models exhibit provably good generalization despite encoding considerable information specific to the training data in their structure and parameters. For example, both k-Nearest Neighbor and Support Vector Machine models consist of some or all training instances encoded in explicit form though their black-box behavior is known to generalize both in theory and in practice. Both of these examples are blatantly non-private for white-box attackers. Further, less blatantly problematic models have been shown to be vulnerable to leaking training data in cases of adversarial control over the training process. It remains unknown, however, whether benignly-trained models encode specific information about their training data in a way that is susceptible to adversarial inference.

We explore the privacy risk of widely-used learning algorithms to white-box adversaries who gain access to the already trained models. In particular, we will extend the game-based formulation of membership and attribute inference attacks developed in our prior work to account for white-box adversaries, and attempt to derive bounds that characterize these adversaries' advantage in terms of structural properties of the model and training data. Our approach for deriving lower bounds is constructive, and is based on identifying statistical assumptions about the training data and the resulting model properties that lead to privacy risk (i.e., adversarial advantage) under novel white-box attacks. We will then validate the practical significance of these bounds by measuring privacy risk empirically, mounting the corresponding attacks on models trained using real data.

4.Results and Discussion

The following summarizes key results of our research and discusses their implications.

4.1 Personalized Privacy Infrastructure for the Internet of Things and Personalized Privacy Assistants

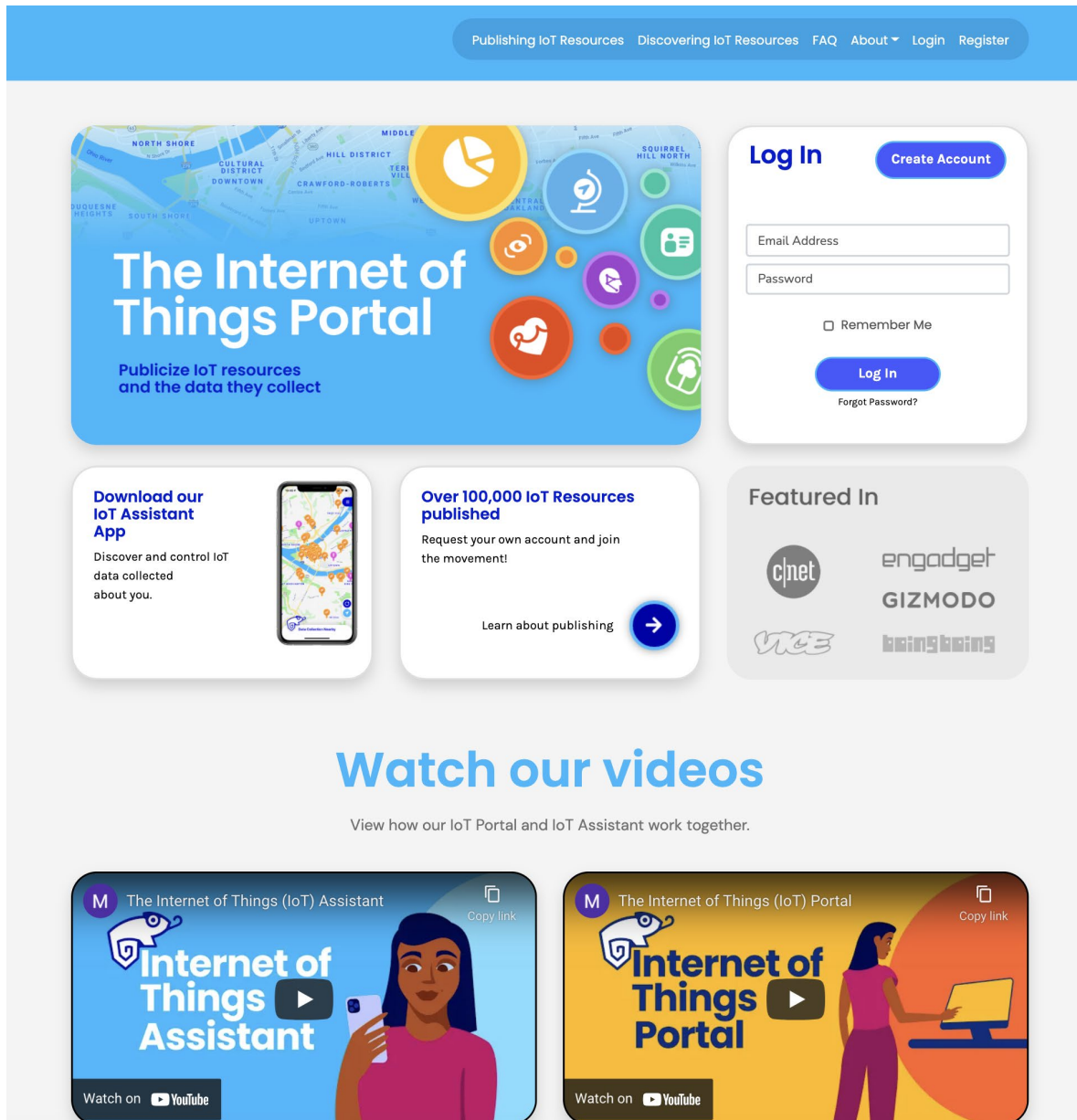


Figure 12: Homepage of the IoT Privacy Infrastructure portal available at iotprivacy.io

As part of our research, we deployed, evaluated and refined successive versions of our IoT Privacy Infrastructure and associated IoT Assistant app. The IoT Privacy Infrastructure has been made available to the public (Figure 12) and supports the creation and publication of resource descriptions and resource templates as well as the management of registries in the United States of America, Canada, the European Union, Australia, New Zealand, Iceland, Liechtenstein, Norway, Switzerland, and the United Kingdom.

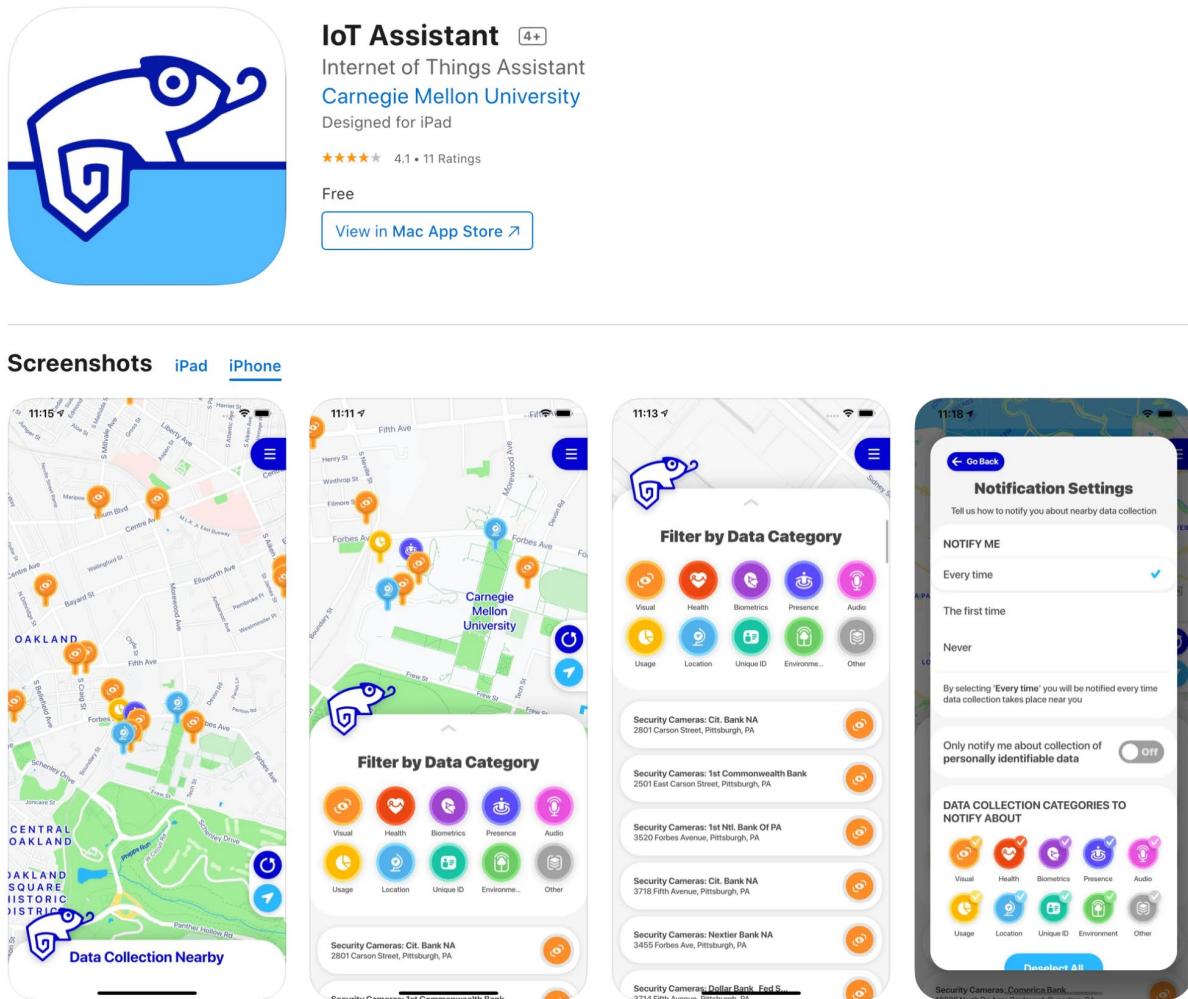


Figure 13: The IoT Assistant mobile app, as available in the Apple iOS app store.

The IoT Assistant app is available in both the Apple iOS App Store [IoT AiOS] (see Figure 13) and the Android Google Play Store [IoT AGoogle].

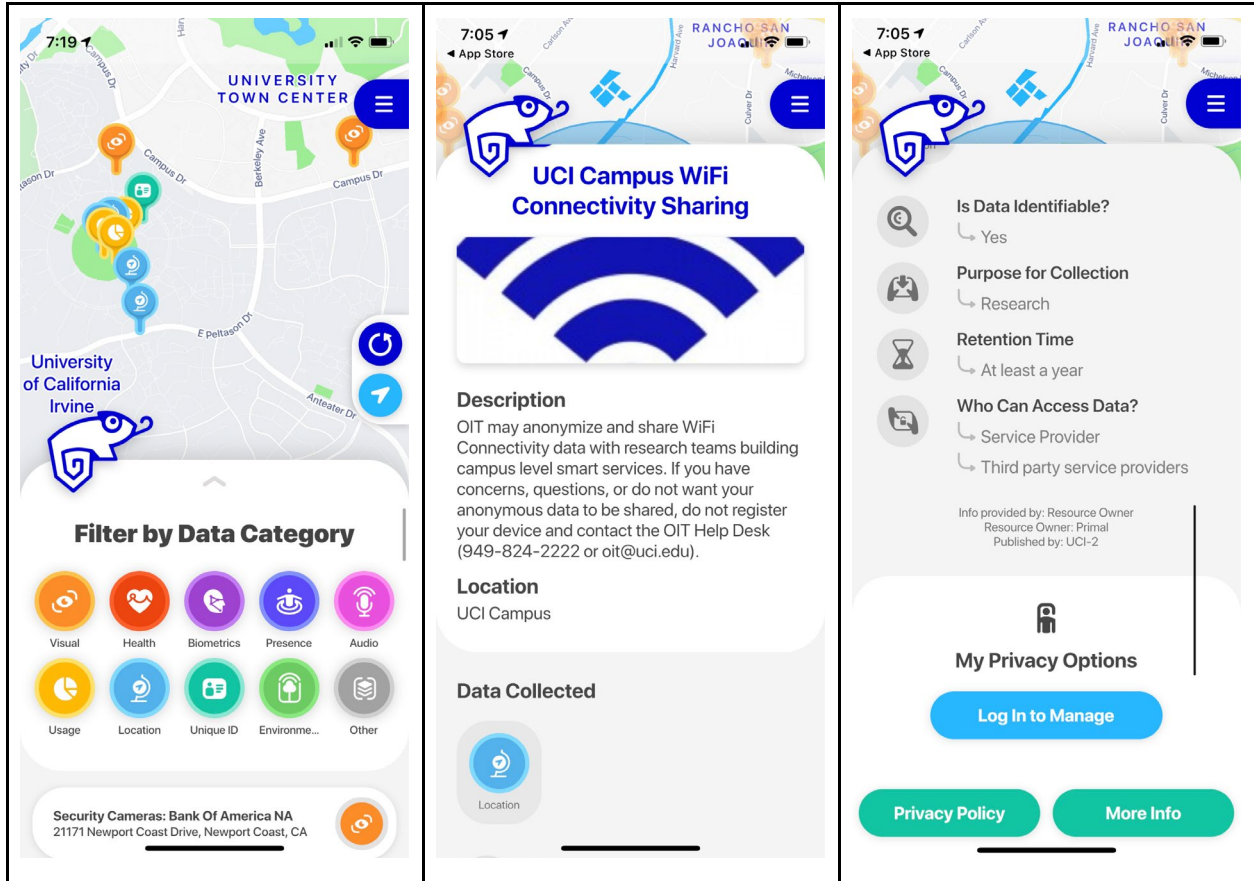


Figure 14: Screenshots of the IoT Assistant app showing resources deployed and publicized on the UC Irvine campus.

Early deployments of the infrastructure focused initially on the campus of the University of California at Irvine (Figure 14) and the campus at Carnegie Mellon University (Figure 15).

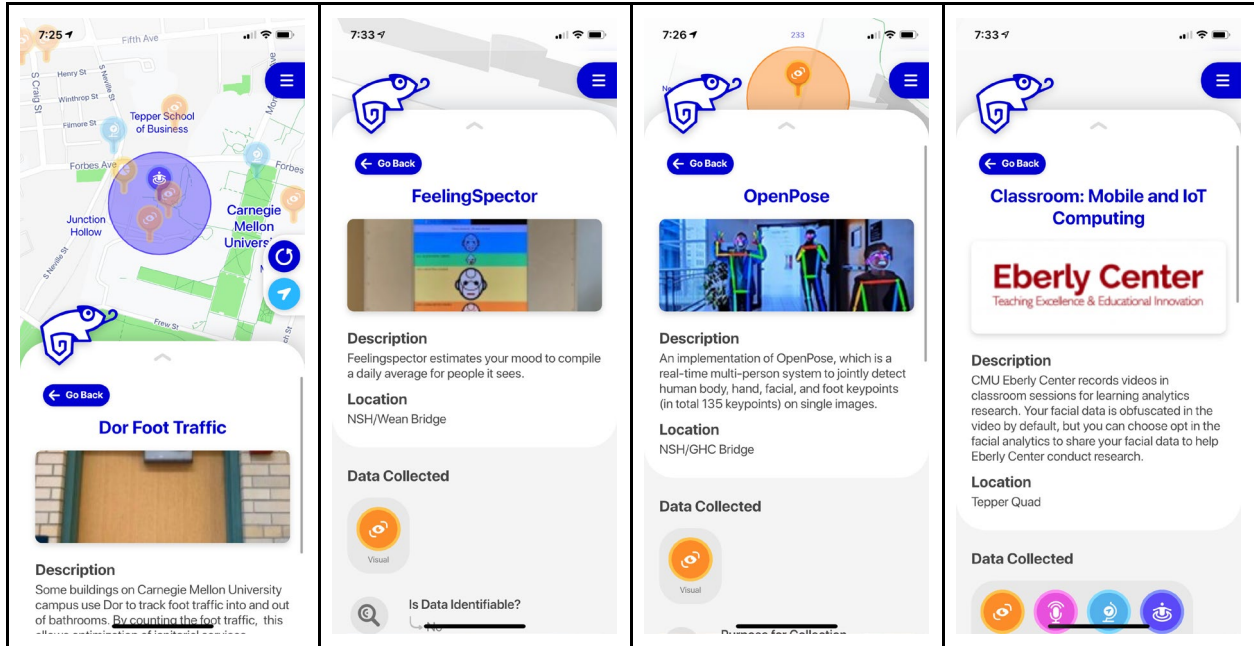


Figure 15: Screenshots of the IoT Assistant app showing resources deployed and publicized on the Carnegie Mellon University campus.

4.1.1 The Implementation and Deployment of IoT Portal

The most recent version of the IoT Portal is designed to support a wide range of stakeholders, including resource owners (e.g., mall operators, stores, restaurants, office building operators, universities, neighborhood organizations, municipalities as well as individuals such as home owners), contributors from the public at large, resource template contributors, including device manufacturers and system integrators, and registry owners.

The new IoT Portal features a completely redesigned user interface that provides a streamlined process for different stakeholders to create IoT resource listings and templates and to publish their IoT resources. As shown in Figure 16, a dashboard provides users with an overview of their resources, registries, and templates and easy access to functionality to manage each of these.

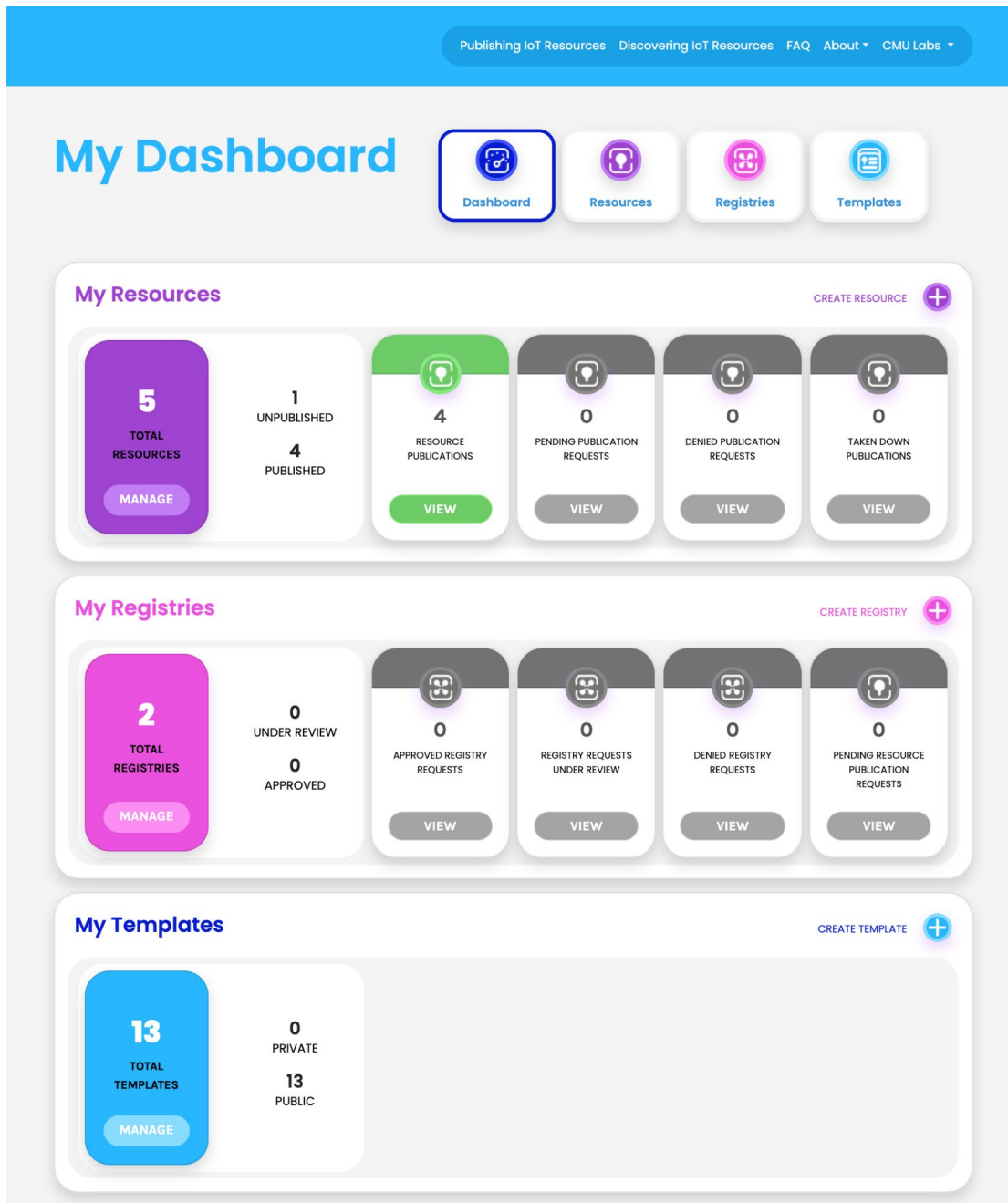


Figure 16: IoT PI Dashboard from which users can check the status of their resources, registries, and templates.

At the time of writing the IoT Privacy Infrastructure already hosts well over 100,000 IoT resource descriptions. Figure 17 provides screenshots of resource descriptions published in Miami, Washington DC, the Manhattan Financial District and Boston. The IoT Assistant app was downloaded by over 15,000 users within the first week of its release.

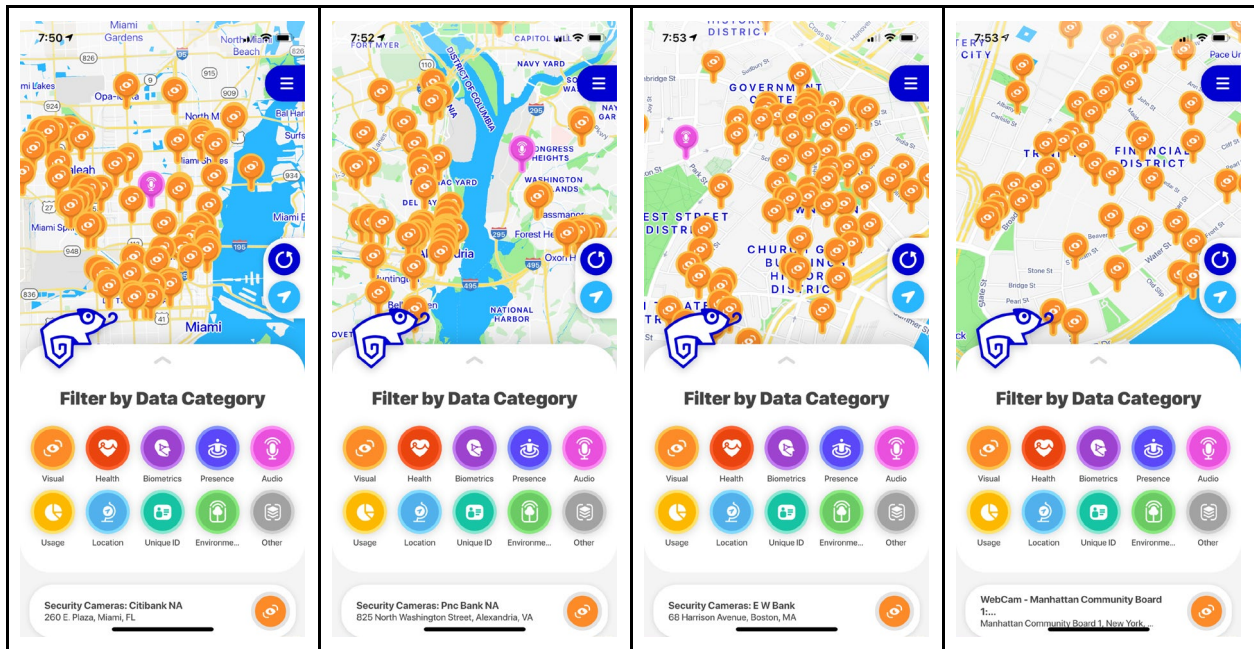


Figure 17: Screenshots of the IoT Assistant app showing resource descriptions published in Miami, Washington DC, Boston and the Manhattan Financial District in New York. The app is currently configured to only display up to 50 IoT resources around a given location. Over 100,000 IoT resource descriptions have been published and can be discovered using the IoT Assistant app.

4.1.2 The Implementation of the IoT Assistant App

The most recent release (1.2.8) of the IoT Assistant App (IoTAssistant) is available for download in more than 30 countries on Apple’s App Store and Google Play Store. The IoTAssistant app enables people to discover nearby IoT resources through a map-based interface (Figure 18, left screenshot). Each pin on the map is a published IoT resource description. Clicking a pin will take users to a screen with more information about the IoT resource and its data practices (Figure 18, middle screenshot). The IoTAssistant app supports customizable notification functionality (Figure 18, the right screenshot) that enables users to specify the types of data collections they want to be notified about and the frequency of notifications. In addition, the most recent version of the IoTAssistant app also supports the scanning of quick response (QR) codes that can be used to publicize nearby IoT resources. The QR codes can be automatically generated from the IoT portal, when defining resources. A user who does not have the IoTAssistant app and scans a QR code will be redirected to the app store, where he or she can download the IoTAssistant app before being redirected to the description of the IoT resource in the app, once the app has been downloaded.

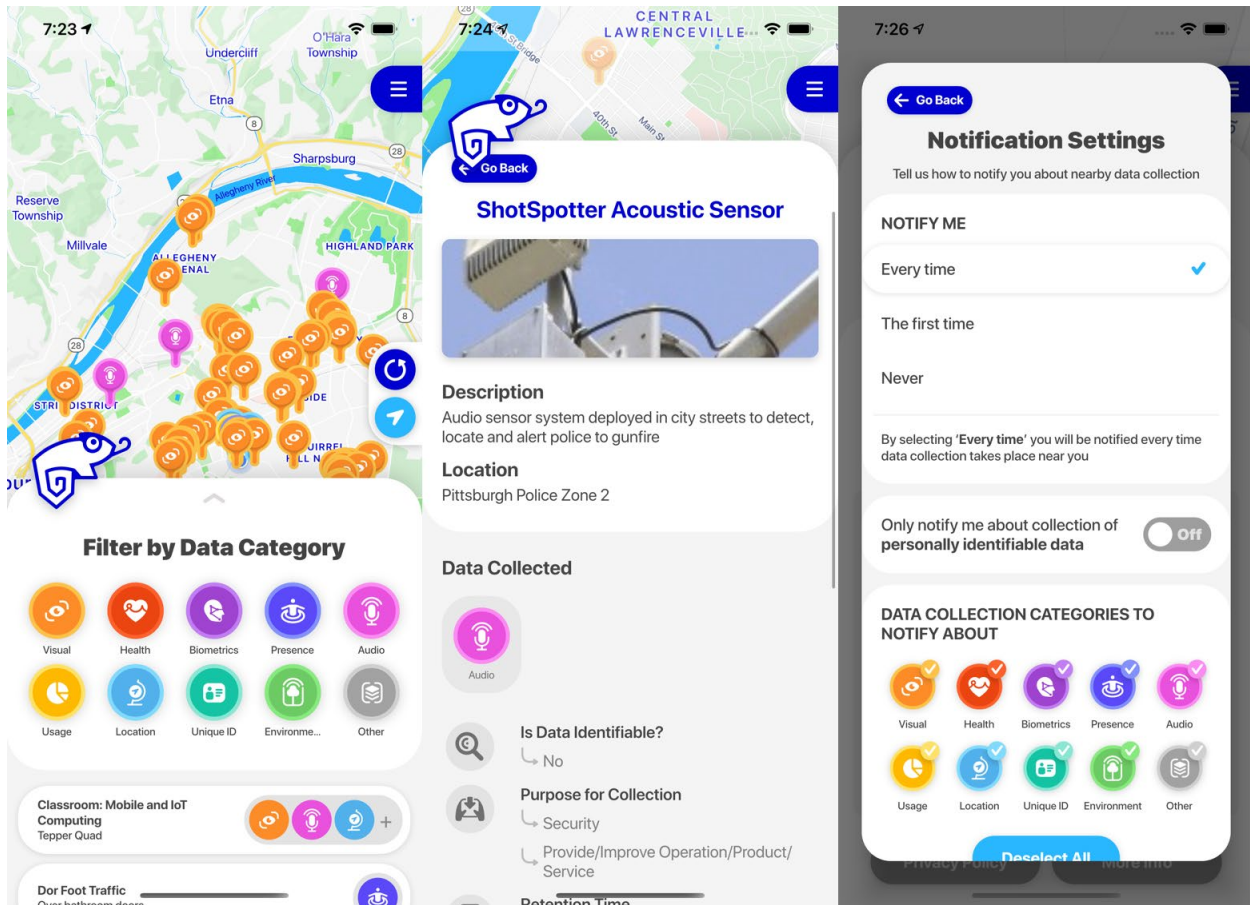


Figure 18: Using the IoT Assistant app to discover nearby IoT resources, including their data collection and use practices and any privacy settings they make available. Notification settings allow users to configure their app to only be notified about IoT data collection practices they care about and at the frequency they find most useful.

As shown in Figure 19, resource templates and resource descriptions can specify a rich set of privacy choices, which allow individual users to select from choices such as opting in or out of some data practices, request that their data be deleted, or request copy of their data, as required under regulations such as GDPR or CCPA. Figure 20 shows how these privacy choices can then be accessed by IoT Assistant users as they discover these resources. Accessing privacy options typically requires users to authenticate (Figure 20) - with authentication typically being with systems controlled by owners of the resources (i.e., we do not expect resource owners to rely on our own authentication system), though some privacy options can also be taken advantage of via email. Further details on the design principles that have driven the design of the IoT assistant privacy controls are discussed in our 2021 CHI paper “A Design Space for Privacy Choices: Towards Meaningful Privacy Control in the Internet of Things” [FYN21].

Publishing IoT Resources Discovering IoT Resources FAQ About Norman Sadeh ADMIN

Dashboard Resources Registries Templates

Create a New Resource

Search for a template to pre-fill the form

Fields marked with an asterisk * are required

STEP 1 Basic Information STEP 2 Location STEP 3 Data STEP 4 Data Practices **STEP 5 Privacy Options**

Previous Save as Draft Create Resource

This resource provides users the following privacy options
(select all that apply)

- Request Data Deletion
Accept user requests to delete their data that has been collected by this resource
- Request Copy of Data
Accept user requests to download their data that has been collected by this resource
- Allow/Deny Data Collection
Accept user requests to allow or deny data collection by this resource
- Allow/Deny Data Sharing
Accept user requests to allow or deny sharing of the data collected by this resource

Privacy Options Management URL

Privacy Options Status Check - API endpoint

Figure 19: Defining privacy controls (or privacy choices) when describing an IoT resource and its data practice via the IoT Portal.

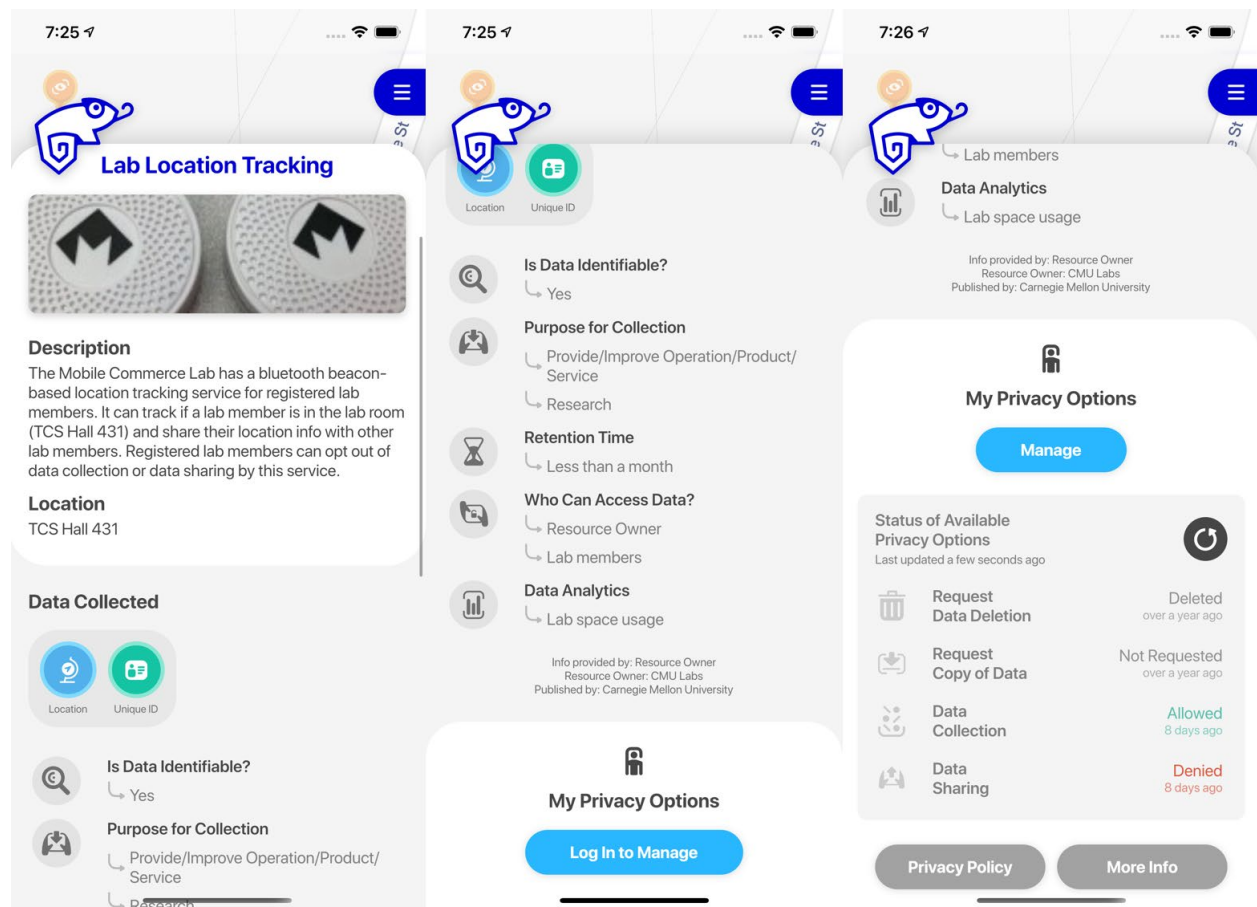


Figure 20: Screenshots of Privacy Options made accessible to IoT Assistant users as they discover a particular IoT resource - in this case a Bluetooth location tracking device. Users typically need to authenticate with the organization controlling the IoT resource to manage the privacy options made available by that resource.

4.2 Taxonomy of Privacy Dialog Primitives

The results of our research on taxonomies of privacy dialog primitives are detailed below with relevant discussion.

4.2.1 Exploring How Privacy and Security Factor into IoT Device Purchase Behavior

In the interview study exploring users' IoT device purchase behavior, we found that about half of our interviewees had limited and often incorrect knowledge about privacy and security and that this impacted their ability to make informed privacy and security decisions. In addition, most of our interviewees had not considered privacy and security

before purchasing IoT devices, but reported being concerned after the purchase. Those who were concerned about privacy and security at the pre-purchase evaluation stage reported difficulty finding useful information about device privacy and security.

We found that security and privacy are among the factors that people would consider in their future IoT device purchase decisions. Survey participants reported that security and privacy would have significantly more influence on their decisions to purchase a smart security camera than a smart thermostat or toothbrush, likely due to their perceptions of the sensitivity of the data those devices collect. Almost all interviewees acknowledged the importance of knowing privacy and security information related to IoT devices before making purchase decisions and said they would pay a small premium for such information to be provided, especially when purchasing a device that they perceive to collect more sensitive information (e.g., a smart camera capturing images).

Almost all interviewees found our prototype labels easy to understand and able to provide information they would consider in a purchase decision. We found that interviewees often focused on privacy and security choices, expert ratings, purpose of data collection, and the convenience of security mechanisms. From our findings, we distilled recommendations for the design of privacy and security labels that enable consumers to make informed IoT device purchase decisions. Our findings on consumers' interest in IoT nutrition labels, and ways to make them more useful, are important and timely contributions as policy makers debate new IoT privacy and security regulations.

4.2.2 An Expert Study to Determine IoT Privacy and Security Label Content

In the expert study to determine the content for IoT privacy and security nutrition label, we found that differences in opinions were driven less by fundamental differences in beliefs, but rather by differences in work experience and priorities. For example, some experts were more knowledgeable about specific security mechanisms, standards, or regulations, and prioritized factors related to their area of expertise or their organization's mission.

Most factors identified as important by experts are factors that they believe will inform consumers. Experts also identified some factors for inclusion that could inform experts only, mostly to be able to hold companies accountable.

A layered label includes a primary layer that presents the most important and glanceable content, followed by a secondary layer for additional information. In our study, we asked experts to specify the layer on which the information should be

included on the label. They mostly recommend putting only information that would be understandable and important to most consumers on the primary layer.

We designed a prototype layered label based on our expert elicitation study. We then conducted semi-structured interviews with 15 consumers of IoT devices (smart home devices or wearables) and presented our prototype to them. We show that all of our participants had a clear understanding of the information presented on the primary layer of the label. Although some of the factors on the secondary layer of the label were less understandable to participants who lacked privacy and security expertise, all of our participants reported that they still want such information to be included on the label mainly to be as informed as possible. In addition, consumer participants reported that having all the important privacy and security factors, even unfamiliar ones, on the label would help them easily search online to find more information.

4.3 Modeling of User Privacy Preferences and Expectations

4.3.1 Mobile App Privacy Assistant Studies

As part of our research on privacy assistants and the development of predictive models to help people manage their mobile app privacy assistants, we collected privacy preferences people have for granting/denying different types of permissions to different categories of mobile apps (see Subsection 3.3.1 for details). This data, which was collected from Android users as they used their regular phones as part of their regular activities while receiving a daily nudge motivating them to review their permission settings, was analyzed using clustering techniques. This research showed that even a small number of clusters can go a long way in predicting many of a user's privacy preferences. Figure 21 displays an example of a set of seven clusters obtained by combining users with similar preferences when it comes to granting/denying permissions to different categories of mobile apps [LAS16+]. Blue shades indicate a tendency to grant a given permission to a certain category of apps, whereas red shades indicate a tendency to deny. The intensity of the color indicates how pronounced the tendency is among people in a given cluster.

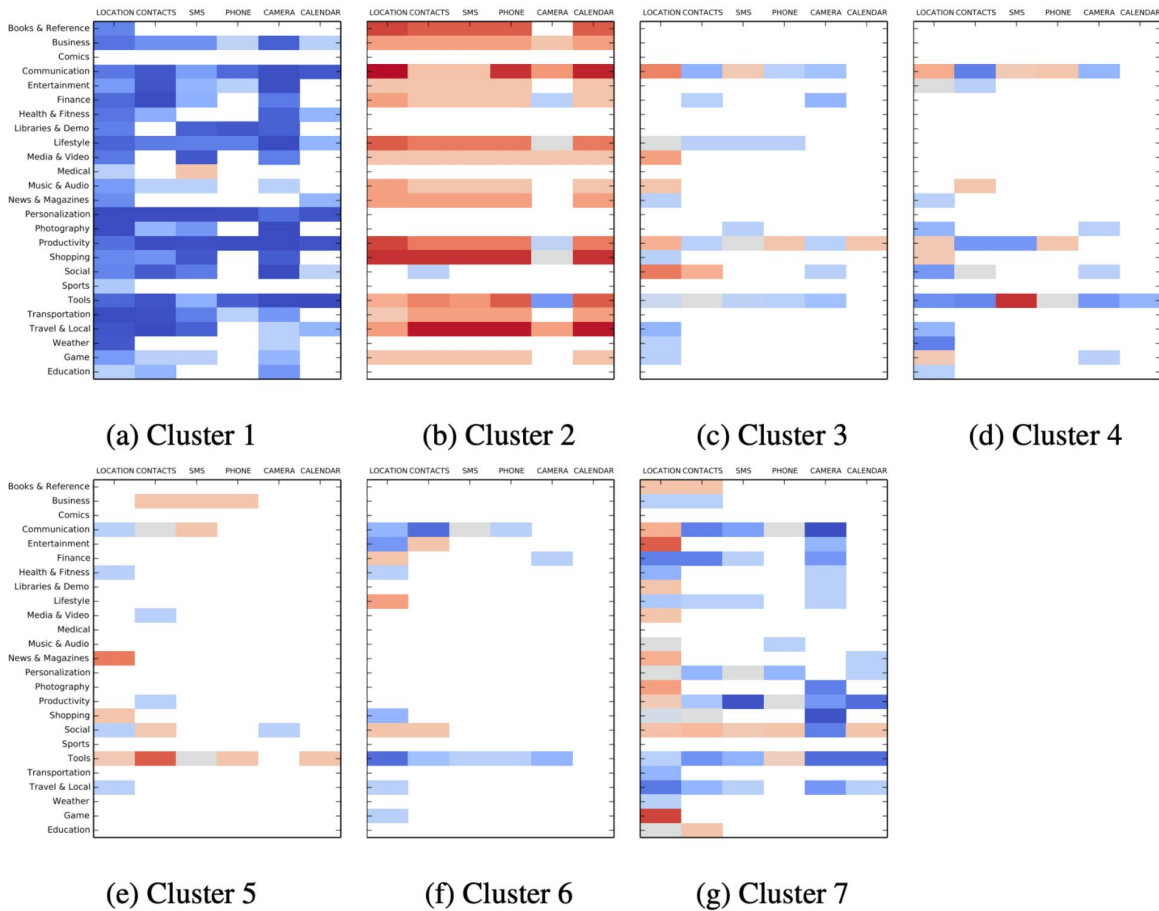


Figure 21: Clustering like-minded users based on their decisions to grant/deny permissions to mobile apps in different categories.

These clusters can in turn be used to predict people’s privacy preferences for denying/granting permissions to their apps by assigning them to clusters that best match their expressed preferences [LAS16+]. The end result is a model that can be used to recommend permission settings to users (Figure 22). Predictive power was also shown to increase if these models distinguish between the different purposes for which an app request access to a given permission (Figure 23). As can be seen, the profiles that distinguish between purposes (e.g., core functionality versus marketing versus advertising) are shown to have stronger predictive power.

Clusters and privacy preference profiles for granting/denying permissions to different mobile apps were used as the basis for the development and evaluation of a first mobile app privacy assistant designed to recommend mobile app permission settings to its users [LAS+16].

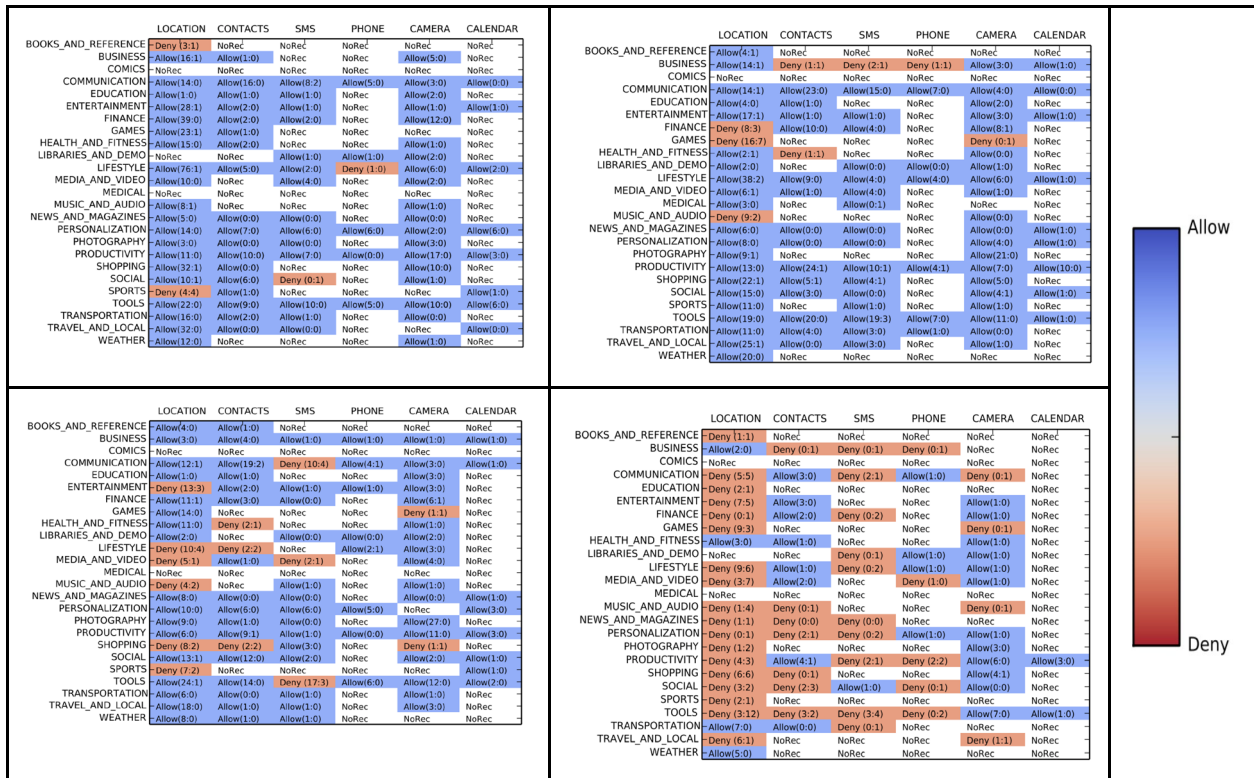


Figure 22: In each cluster, the number of deny/allow decisions can be used as a basis for making recommendations to either grant/deny a given permission to a given category of apps. When the difference between the grant and deny decisions is not sufficiently strong or when there is not sufficient data, no recommendation is made.

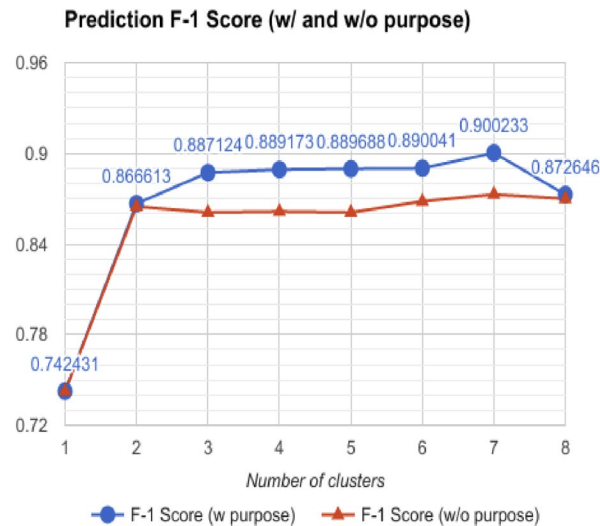


Figure 23: Comparing F-1 scores for predictions with profiles that differentiate between the purpose for which apps in a given category request a permission and profiles that do not differentiate between the two.

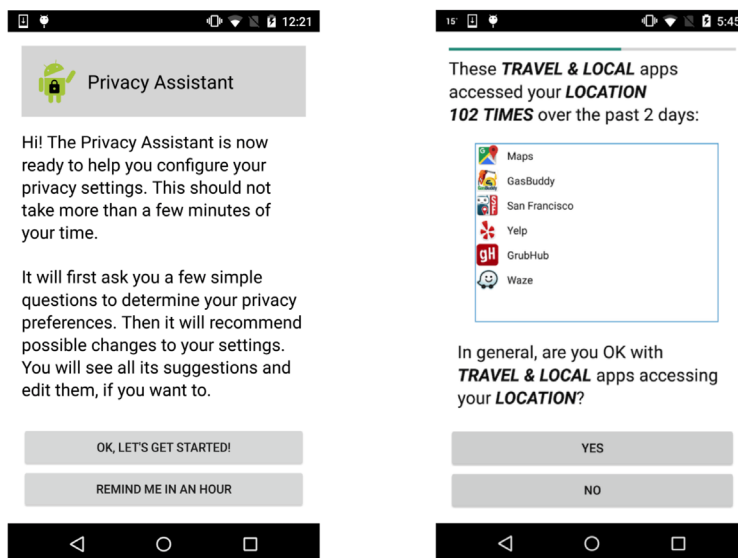


Figure 24: Screenshots of the personalized privacy assistant piloted in our initial study.

Figure 24 as well as Figures in Subsection 3.3.1 show screenshots of the first mobile app privacy assistant developed and piloted by our project [LAS+16, Liu19]. After users answer between 3 and 5 questions, they are placed in a cluster and receive permission recommendations based on the cluster to which they have been assigned. They can review the recommendations and decide whether or not to accept them, whether individually or in bulk. See Subsection 3.3.1 for additional screenshots of the privacy assistant app evaluated as part of the study reported in [LAS+16].

This privacy assistant started by collecting data about the app installed by a user on his/her smartphone and the permissions requested by these apps. It would then rely on decision trees to dynamically ask questions to users about their mobile app permission preferences (see screenshot in Subsection 3.3.1). Using these decision trees, users would be assigned to a cluster and its associated mobile app permission recommendation (“privacy profile”). These models would in turn be used to review the user’s current mobile app permissions and recommend possible changes in the form of recommendations to possibly deny permissions that were currently allowed (see screenshot in Subsection 3.3.1). The user could then review recommendations and decide to accept or reject them individually or in bulk.

The resulting privacy assistant was evaluated with 49 Android users on their rooted Android phones as part of their regular daily activities. Of the 49 participants, 27 received recommendations and 22 did not, as our model indicated that these participants were satisfied with their existing permission settings. Among the 249 recommendations received by these 27 participants, 196 recommendations were

accepted (Figure 25). Overall participants accepted 78.7% of all recommendations. Each row represents a different participant. [LAS+16].

To make sure that these recommendations truly reflected the participants' preferences, they were subjected to daily nudges for another 6 days prompting them to review their permissions, including those they had modified based on the assistant's recommendations. Participants only modified 10 of the 196 recommendations they had previously accepted. Detailed interviews and additional data collected as part of this study strongly suggested that participants truly benefited from the functionality and liked the recommendations, including the ability to review them and decide whether or not to accept them. Additional details on this work have been reported in [LAS+16, Liu19].

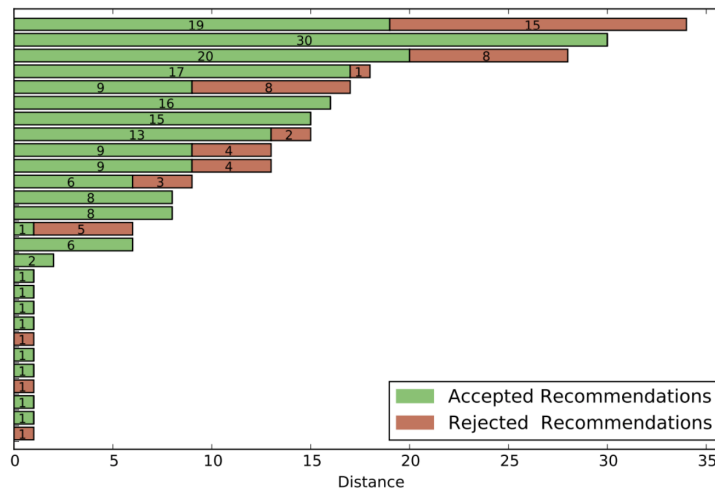


Figure 25: Number of recommendations made by the privacy assistant app that were accepted/rejected by each individual participant in the study (total of 27 participants who received recommendations).

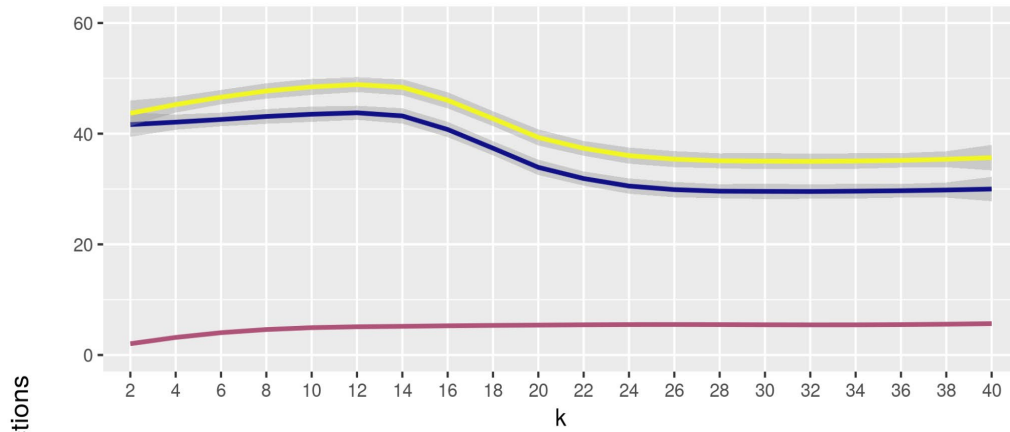
Encouraged by these results, our project developed an improved version of the privacy assistant, which was released on the Android Google Play Store for use on rooted Android phones, as further discussed in Section 4.7.2 and detailed in [Liu19].

Further work on mobile app permissions and the effectiveness of machine learning when it comes to mitigating tradeoffs between expressiveness and user burden of privacy controls made available to users was also reported in [SFZ+20]. In this work, we administered a survey which collected participants' Android permissions preferences for a variety of apps under two conditions: one with purpose-specific permissions and another with permissions that extend across all possible purposes. We analyzed responses using logistic regression and machine learning. Our study sampled 5964

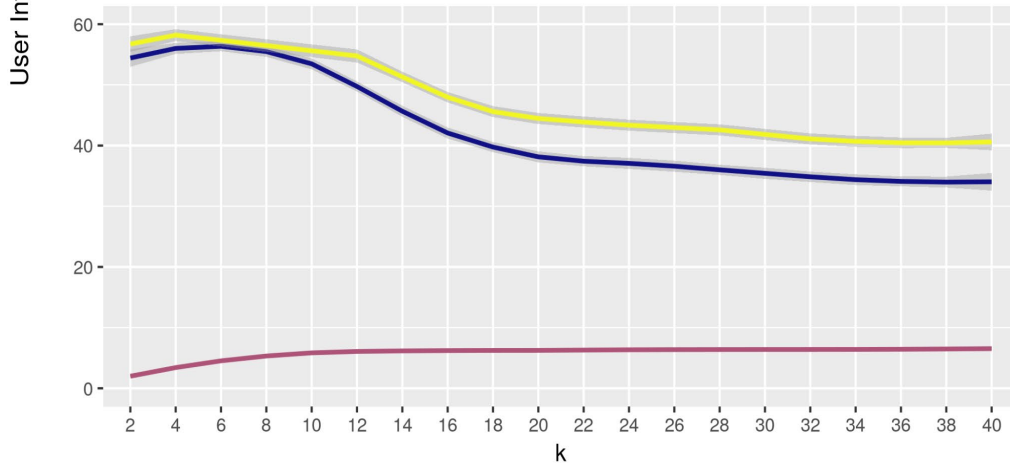
observations of preferences toward three sensitive Android app permissions (calendar, location, contacts), with user preferences across 108 apps, from a large sample of Android users (n=994) in the United States. Our aim was to discover whether machine learning could help mitigate the trade-off between accuracy and user burden when it comes to configuring Android app permissions. We applied machine learning techniques to evaluate if profile-based models could improve app permission management in terms of accuracy and user burden.

Our results (Figure 26 and 27) show that it is possible to select parameters that can improve accuracy, reduce user burden, or achieve both objectives simultaneously. This is due to the predictive power gained from models that incorporate additional factors such as the purpose for granting the permission. Specifically, Figure 26 plots the number of user interactions to cluster users (i.e., assigning them to particular clusters), as represented by the “profile” line, and the number of manual user interactions required to configure one’s permission settings when taking advantage of permission recommendations made using the cluster to which a user is assigned. The number of these latter interactions is represented by the “ask” line, which corresponds to the number of permissions where the user needs to be asked because the model does not have a recommendation. K is a parameter corresponding to the number of clusters used in a given model. Models that include purpose require significantly fewer user interactions overall (lower user burden). Results for larger values of K should not be considered, as they most likely involve some level of overfitting.

Efficiency (Purpose Specific Model, 36 apps, simulated)
Fewer user interactions is better.



Efficiency (Purpose Independent Model, 36 apps, simulated)



Interaction Type — ask — profile — sum

Figure 26: Plotting the total number of user interactions: comparing predictive models that include permission purpose and models that do not.

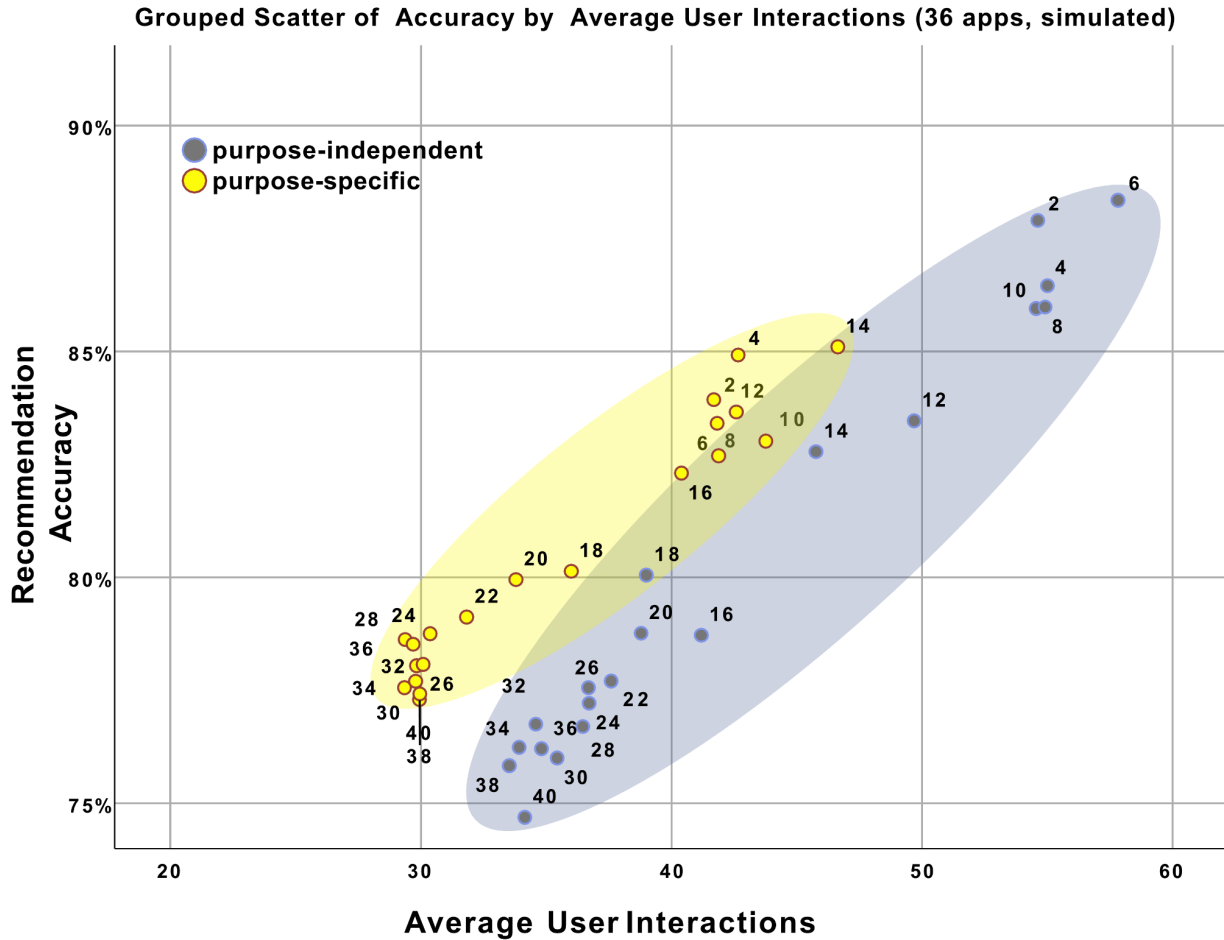


Figure 27: Scatterplot showing recommendation accuracy versus the average number of user interactions. Each k represents the number of profiles the population is divided into.

By default, adding additional factors to mobile app permissions would be expected to increase accuracy as well as user burden. For example, mobile app permissions could be designed to offer users the ability to select different permissions subject to the purpose for which a permission is being requested by an, but this would multiply the configuration burden by the number of specified purposes. Our results suggest that machine learning can help mitigate trade-offs between accuracy and user burden. In the context of models that take the purpose of permissions into account, our study suggests that it is possible to get the “best of both worlds”, namely doing a better job at accurately capturing people’s preferences while simultaneously reducing the number of decisions they have to make. This is accomplished using machine learning to assign users to privacy profiles and using these profiles to infer many permissions for each user.

These results argue for the introduction of purpose-specific permissions in mobile operating systems such as Android and iOS. They not only show that people's permission preferences are influenced by the purpose for which permissions are requested, but also suggest that, using machine learning, interfaces could be built to mitigate the increase in user burden that would otherwise result from the introduction of finer, purpose-specific permissions.

4.3.2 Modeling People's Privacy Expectations Across a Wide Array of IoT Scenarios

The results of our study outlined in Subsection 3.3.2 inform the design of more transparent IoT connected systems—we envision our results can be used to improve privacy notices for IoT devices, and develop more advanced personal privacy assistants. We show that individuals' comfort levels in a variety of IoT data collection scenarios are related to specific aspects of that data collection. Many of our findings are consistent with observations made in prior work, but our quantitative methodology and the scale of our experiment allows us to understand the effect of individual factors and their relative importance more precisely. Leveraging our qualitative and quantitative results, we advance explanations for many of the differences among these factors. We show that whether or not participants think the use of their data is beneficial to them has a profound influence on their comfort level. We also find that participants' desire for notification is closely related to whether or not they feel comfortable with data collection in a particular scenario.

4.3.3 Informing the Design of Personalized Privacy Assistants

From the qualitative interview study exploring people's perception towards Personalized Privacy Assistants, we developed a deep understanding of people's diverse preferences towards PPAs at different automation levels. As described in 3.3.3, we asked participants about their opinions for three increasingly more autonomous PPA implementations: Notification PPA, Recommendation PPA, and Auto PPA. Participants' opinions of PPA generally became less positive as automation increased, from Notification to Auto, but a few, showed a non-linear relationship between their opinions and the level of automation, as detailed below.

Automation Level 1: Notification PPA

Most participants (n=11) had a positive reaction to an implementation of PPA that could provide users with awareness of data collection around them. However, this reaction was almost always accompanied by a desire, and at times expectation, that the system would also provide them with control over these data collections. To prevent users from being overwhelmed by notifications, participants suggested only being notified about

new and unexpected devices, or specific device types, being able to define “known locations” to not be notified by things in trusted spaces, receiving batched notifications, and being able to set the frequency of notifications. These findings are reflected in the design of our IoT Assistant app.

Automation level 2: Recommendation PPA

Participants’ opinions were mostly positive—but there was a clear preference toward external recommendations (e.g., experts, manufacturers, friends) over recommendations based on past behavior. Overall, participants found this level of automation helpful and educational, and they appreciated that the tool could reduce their cognitive burden while augmenting their knowledge about what would be a good choice to make. We originally conceived and framed recommendations as based on users’ preferences. However, participants presented a range of opinions on possible sources. Six participants suggested offering recommendations from authoritative sources with no vested interests in the data, but one of them also saw the benefits of recommendations from device manufacturers, since they know the technology the best. Two of them suggested recommendations based on crowd-sourcing or user reviews. Regardless of the source, participants wanted transparent disclosure of that source. Four participants were aware that some sources of recommendations might have biases or even try to manipulate users. As such it seems preferable for Privacy Assistants to be offered by independent third parties.

Automation Level 3: Auto PPA

About a third of participants did not want a PPA to make decisions for them. Five participants did not want to yield control over their decisions. One said, “I don't like to be fully controlled by a device, you know?” Furthermore, three participants were unsure whether the technology could accurately predict their decisions and were thus hesitant to allow it to do so. One reason for this, as one of the three participants stated, is that we are not always consistent in our decisions: “I could change my mind. Nine times out of ten I'm going to go this way, but I've got a very good reason for that tenth time not to do that.” Among the other two-thirds of our participants, we observed positive opinions towards an Auto PPA. These positive opinions reflected an appreciation of the convenience of outsourcing this type of decision-making to a computerized system. One participant rejected the Notification PPA due to a fear of becoming overwhelmed, but said that the Auto PPA presented a valid tradeoff: “There I feel we're obtaining a utility value to a human individual and I would consider owning such an appliance, as part of the digital world.”

From the study results, we generated several design implications for PPAs from a user-centered perspective. First, PPAs should probably only offer automation of decisions or recommendations when they can achieve a high level of accuracy in their predictions. We noticed that participants in our interviews were not always excited about automating the information analysis process. Participants were concerned about potential biases from the sources of recommendations or incorrect suggestions for inferred recommendations. Based on our findings, a good design for this functionality would allow users to pick the types of recommendation sources that they would prefer, ranging from expert opinions (authoritative) to real users' opinions (crowd-sourced). It is worth noting that our mobile app privacy assistant ([LAS+16]) is consistent with this finding, as it offers recommendations, which users can review and reject, whether individually or in bulk. This is also consistent with participants in the evaluation of this mobile app permission assistant generally receiving very positive reviews from participants [LAS+16].

Second, PPAs should also ensure that users are not overwhelmed by notifications, not only because notifications can reduce individuals' ability to properly process the information being presented, but because notifications can make them anxious and resigned. One way to avoid this is to remove unnecessary notifications by incorporating a "trusted location" feature, such that users would not be notified about devices in those locations. For non-trusted locations, in lieu of potentially overwhelming individual notifications, the PPA could list devices once and request a decision, revisiting that decision periodically if new devices were added or the user's preferences changed. Another way to avoid unnecessary notifications is to specify data collection situations where users are always opposed to or always in favor of sharing. As already discussed these findings are now in part reflected already in the design of our IoT Assistant, which allows users to personalize their notifications settings.

Third, PPAs will benefit from a user-centered decision selection function. For Notification PPA, we observed a clear desire to have a tool that not only provides information about data collection, but that also collects and enforces users' preferences related to data collection. A next level of automation, Recommendation PPA, could implement users' predefined preferences in an automated way. For example, when prompted to make a decision, the user can choose to have the PPA "remember this decision," informing the system that they no longer want to be asked about that specific data collection. This could be based on data types, devices, companies, etc. or a combination of these variables. For Auto PPA, it is vital to provide an auditing mechanism where users are able to verify and adjust decisions made on their behalf. This mechanism was considered essential when discussing an autonomous PPA, serving as a tool to avoid perpetuating incorrect decisions. This mechanism would also

prove beneficial to users who choose a lower level of automation, as it would allow them to review and revise past decisions as their preferences evolve.

4.3.4 The Experience Sampling Study

We now turn our attention to summarizing key results of our study of people's privacy attitudes towards video analytics through the experience study summarized in Subsection 3.3.4.

When surveying participants' responses to facial recognition scenarios, we focused on four related questions: how surprised they were by the scenario presented to them (surprise level), how comfortable they were with the collection and use of their data as assumed in that scenario (comfort level), to what extent they would want to be notified about the deployment scenario at the location they visited (notification preference), and whether, if given a choice they would have allowed or denied the data practices described in that scenario at that particular location at the time they visited that location (allow/deny preference). Figure 28 provides a summary of collected responses organized around the 16 categories of scenarios (or "purposes"). As can be seen, people's responses vary for each scenario. In other words, "one size fits all" would fail to capture individuals' diverse preferences when presented with these scenarios. At the same time, some scenarios elicit more consistent responses from participants than others. For instance, generic surveillance scenarios appear to surprise participants the least and to elicit acceptance by the most. Yet, even in the presence of such scenarios, 60% of participants reported they would want to be notified at least the first time they encounter these scenarios at a given venue and over 35% indicated they would want to be notified each time. At the other end of the spectrum, scenarios involving facial recognition for the purpose of evaluating employee productivity or tracking attendance at venues elicited the greatest level of surprise and lowest level of comfort among our participants, with barely 20% reporting that, if given a chance, they would consent to the use of these technologies for the purpose of evaluating employee productivity. Similarly, participants expressed significant levels of surprise and discomfort with scenarios involving the use of facial recognition to make health and medical predictions or to track the attendance of individuals.

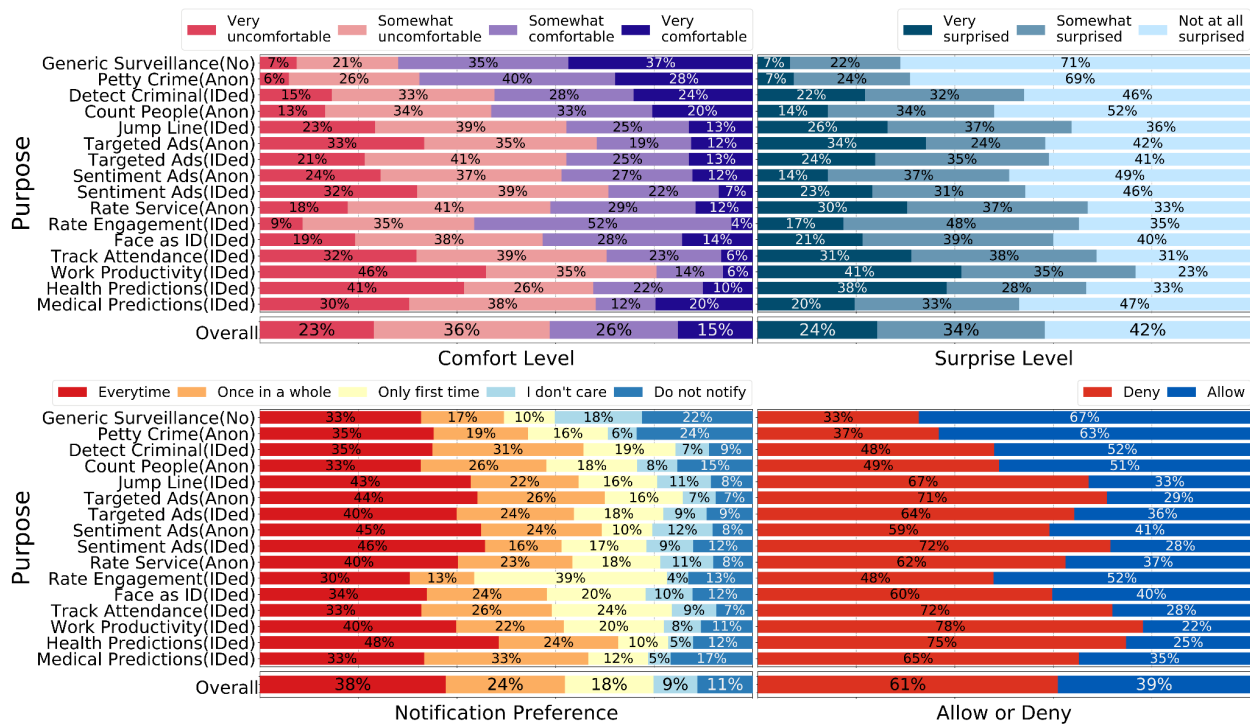


Figure 28: Summary of collected responses organized around 16 different purposes. The bottom row shows the aggregated preferences across different purposes.

We investigated whether people's decisions to allow or deny data collection have a relationship with the contextual attributes. We constructed our model using generalized linear mixed model (GLMM) regression, which is particularly useful for data analysis with repeated measures from each participant. Among all the attributes, we found that "purpose" exhibits the strongest correlation with the decision to allow or deny data practices associated with our scenarios. Some of the place type attributes were also found to have an influence on participants' allow or deny decisions. The number of days participants had been in the study also seemed to influence their allow/deny decisions. Our findings showed that people's privacy preferences are both diverse and complex. They depend on a number of contextual attributes such as the purpose for using video analytics. As such, our findings are another illustration of contextual integrity principles. The importance of purpose information identified in our study (i.e., for what purpose video analytics is being applied) is largely consistent with results reported in earlier publications.

A majority of the study participants reported increased awareness resulting from participation in the study. They did not realize facial recognition could be used for so many different purposes, at such a diverse set of venues, and with this level of sophistication. As participants grew more aware of possible video analytics

deployments, they gained a more grounded estimate of their knowledge level. Such increased awareness leads some participants to deliberate on facial recognition usages. Three interviewees described their deliberation on facial recognition usages as the study progressed. Our results clearly indicate that many people were taken by surprise when encountering a variety of video analytics scenarios considered in our study. While many expect surveillance cameras to be widely deployed, few are aware of other types of deployments such as deployments for targeted advertising, attendance, productivity, and more. These less expected scenarios are also those that generally seem to generate the greatest discomfort among participants and those for which, if given a chance, they would often opt out (or not opt in).

These results make a strong case for the adoption of more effective notification mechanisms than today's typical "this area under camera surveillance" signs. Not only are people likely to miss these signs, but even if they do not, these signs fail to disclose whether video analytics is being used, for what purpose, who has access to the footage and results, and more. Our study shows that many of these attributes have a significant impact on people's desire to be notified about deployments of video analytics. And obviously, these signs do not provide people with the ability to opt in or out of these practices. Our findings support new disclosure requirements under regulations like GDPR, which mandates the disclosure of this information at or before the point of collection. Our findings also demonstrate the urgent need to provide people with choices to decide whether or not to allow the collection and processing of their data, as our participants expressed diverse levels of comfort with these scenarios with many not feeling comfortable with at least some of them. Regulatory disclosure requirements help improve transparency of video analytics deployments. While some study participants grew more concerned about facial recognition, we observed others becoming more accepting of it as they learned about potential benefits of some deployments. These findings suggest that increased transparency and awareness would help data subjects make informed decisions.

These findings have contributed to informing the design of our privacy assistants and our IoT Privacy Infrastructure. Using our IoT Assistant, users can access opt-in or opt-out functionality made available by IoT resources to indicate whether they agree or not to the collection and processing of their data by these resources. Given the growing deployment of cameras, taking advantage of such functionality could still be hampered by the number of notifications and decisions a typical person would be confronted with each day when passing within range of cameras. As such the ability to customize one's notification preferences, as supported in our IoT Assistant app, can help alleviate this problem. Similarly, being able to configure default opt-in/opt-out decisions for different types of video analytics deployments could help significantly reduce user burden. Our

findings showing the diversity of people’s privacy and expectation preferences across different deployment scenarios further confirm the need for personalized privacy assistant functionality with different users relying on different configurations of notification and default opt-in/opt-out settings for different video analytics deployments (and more generally different IoT resource deployments). To keep user burden manageable, one needs configurations that allow users to automatically opt in or out of scenarios for which they have pretty definite preferences (e.g., “I want to opt out of any video analytics deployment that shares my data with insurance companies”). For other scenarios, users could be notified and prompted to make manual opt-in or -out decisions. Given how rich and diverse people's privacy preferences are, enabling users to manually specify each and every of their notification and opt-in/opt-out preferences would result in unacceptable burden on users. Instead (just as we had done for mobile app permissions), using machine learning to develop privacy preference profiles based on the data collected from this study, we have been able to show that it is possible to use simple clustering techniques to reduce the number of decisions users would likely have to make [ZFD+20b], as shown in Figure 29. In this figure, each cluster profile contains 3 columns: the left one displays the average mean value (deny=-1, allow=1), and the right column represents the cluster profile, where the blue color represents an allow decision, red means a deny, and white means no decision, either because not enough data points are available or for lack of a two-thirds majority. The middle column shows the variances, ranging from 0 to 1. The 3 numbers (D/A/T) in each entry in the right column represent the distribution of deny (“D”) and allow (“A”) collected for members of the cluster for the corresponding purpose, with T=D+A representing the total number of decisions collected for the given purpose from members of the cluster.

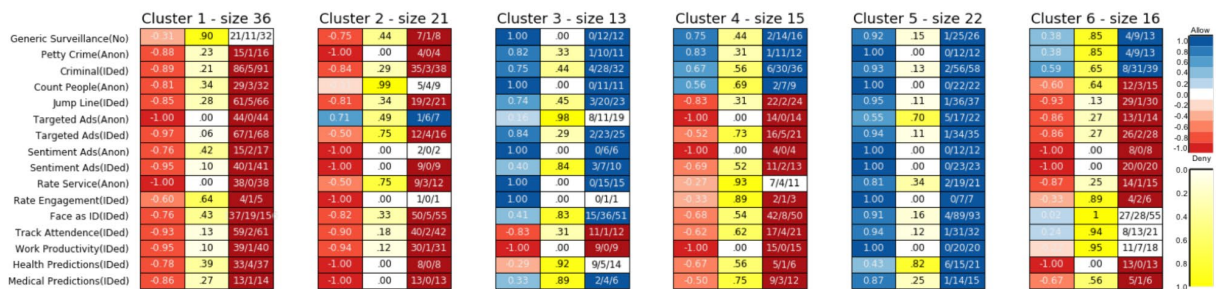


Figure 29: Profiles associated with a 6-cluster model.

Here again the idea is that these models could be used to recommend settings to users, who could review allow/deny recommendations and decide whether or not to accept them.

4.4 Nudging and Explanation

We now move on to discussing the main findings of our research on privacy nudges conducted as part of this project, focusing on findings of the studies introduced in Section 3.4.

4.4.1 Evaluating the Impact of Framing on the Effectiveness of Privacy Nudges

As reported in [Alm17], our work evaluating different types of framing in privacy nudges shown to mobile app users has demonstrated that framing nudges in a way that more clearly reveals the privacy risks associated with one’s privacy decisions tends to be more effective (see Figure 30).

	N	Adjust Settings	Keep Settings	Close Screen	Don’t Know
(1) Baseline	96	56 (58.33%)	33 (34.38%)	7 (7.29%)	0 (0%)
(2) Frequency	97	67 (69.07%)	22 (22.68%)	8 (8.25%)	0 (0%)
(3) Background	99	75 (75.76%)	20 (20.2%)	4 (4.04%)	0 (0%)
(4) Purposes	97	75 (77.3%)	15 (15.5%)	3 (3.1%)	4 (4.12%)
(5) Purposes + Example	94	78 (83%)	15 (16%)	0 (0%)	1 (1%)
(6) Inferences	94	58 (61.7%)	34 (36.17%)	2 (2.13%)	0 (0%)
(7) Inferences + Example	92	65 (70.65%)	21 (22.83%)	6 (6.52%)	0 (0%)
(8) Predictions	96	69 (71.88%)	18 (18.75%)	7 (7.29%)	2 (2.08%)
(9) Predictions + Implications	96	74 (77.08%)	15 (15.63%)	5 (5.21%)	2 (2.08%)

Figure 30: Evaluating the effectiveness of framing in mobile app privacy nudges. Cells highlighted in purple are statistically significant.

4.4.2 Psychometrically Tailored Privacy Nudges

The results and discussion of the three studies on tailored privacy nudges outlined in 3.4.2 are detailed below.

With Study 1, our goal was to identify whether psychometric variables could be used to infer differences in disclosure rates for a single nudge. We recruited a sample of 200 participants from Mechanical Turk to complete Study 1, out of which we conducted our analysis on a final sample of 143 participants.

Across the two experimental conditions in Study 1, 55% (78) of participants chose to allow the data disclosure within the hypothetical IoT scenario. We measured the size of the framing effect by comparing the likelihood of participants to allow the data disclosure between nudge conditions. Of the 143 participants considered in our analysis, 64% (46) of those assigned to the 'Opt-In' condition chose to allow the disclosure while only 44% (32) assigned to the 'Opt-Out' condition chose to do the same. The difference in disclosure rates between these conditions was significant with a p-value of $p=0.02291$ when tested using Chi-squared, suggesting a significant main effect of the nudge.

We used logistic regression models to conduct a difference-in-differences analysis of the response data. In each regression, the coefficient of interest to our research question was the interaction between the psychometric variable scores and the nudge condition assigned to the participant. This allowed us to see whether the effect of the assigned nudge on disclosure likelihood varied significantly with the psychometric variables.

While none of the interaction coefficients we examined were significant, the direction of the coefficients made intuitive sense. For both the Resistance to Framing and Log-ADR variable scores, the sign of the interaction coefficient was negative, indicating that a higher decision competence for these variables may translate into a decrease in the effectiveness of the framing nudge on disclosure behavior.

In Study 2, we sought to determine whether the null result observed in Study 1 was robust by focusing on the potential limitations of Study 1 that may have contributed to our null results. We addressed these issues in part by switching from hypothetical to real disclosure choices and including an additional nudge within our study design. We additionally expanded upon the selection of psychometric variables to capture a broader range of psychometric traits.

For Study 2, we recruited 1,200 participants from Mechanical Turk. Of those recruited, 1,198 completed the initial survey and 966 completed the followup survey and passed the minimum threshold for attention check questions. We tested multiple logistic regression models in Study 2 as part of our exploratory analysis to examine the relationship between disclosure choice, our nudges, and the measured psychometric variables. For each model, our dependent variable was participants' disclosure choice. Our explanatory variables were the assigned nudge condition and the relevant psychometric variable score. Across the two nudges and 15 psychometric variables, we constructed 30 sets of regression models. We used the IUIPC score and demographic variables as controls within our regressions. The interaction term between the assigned

nudge condition and psychometric variable score was the primary coefficient of interest to our research question.

Out of the regression models that we created, we identified 3 pairs of psychometric variables and nudge conditions with potentially significant interaction terms. These included the Recognizing Social Norms score and social norms nudge, the Big Five Extraversion score and framing nudge, and the Big Five Conscientiousness score and framing nudge.

Of the potentially significant effects, the interaction between the Recognizing Social Norms score and the social norms nudge was both the most intuitive and had the strongest effect. The direction of the coefficient implied that as a participant is more likely to recognize social norms, they may be more likely to be influenced by a social norms nudge. While less intuitive, the effects between the two Big Five traits and the framing nudge could also be interpreted. The direction of the regression coefficient for Big Five Extraversion score and the framing nudge condition suggested that those who are more extroverted may be more likely to be influenced by a framing nudge. Likewise, the direction of the coefficient for the Big Five Conscientiousness score and framing nudge condition suggest that participants who are less conscientious may be more likely to be influenced by a framing nudge.

Study 3 built upon the results of Study 2 by attempting to replicate the three potentially significant effects we identified using a separate sample. We recruited 2,000 participants to complete our replication study using Mechanical Turk. During recruitment, we excluded participants that had previously participated in Study 1 or Study 2. Out of this total, 1,996 participants provided complete responses and passed the minimum threshold for attention check questions.

We observed similar effect sizes on both the framing and social norms nudges compared to Study 2. We tested the same logistic regression models that we used in Study 2 for the three potentially significant pairs of psychometric variables and nudge conditions (Recognizing Social Norms score and social norms nudge, the Big Five Extraversion score and framing nudge, and the Big Five Conscientiousness score and framing nudge). Overall, we failed to replicate the effects from Study 2. For the three logistic models, the interaction terms were either not significant, weakly significant (at the $p < 0.1$ level), or lost significance when controls were added.

Whereas the interaction term for the Recognizing Social Norms score and social norms nudge pair was the strongest of the three effects in Study 2, the same effect was only significant at the $p < 0.1$ level when control variables were excluded from the regression

in Study 3. When the control variables were added, the coefficient for the interaction term was no longer significant. For the Big Five Conscientiousness score and framing nudge pair, the direction of the interaction coefficient changed from negative to positive. Again, the interaction term for this variable and condition pair was only weakly significant for the regression model without control variables. When control variables were added, this term lost significance.

Overall, the results of our three studies indicated that effects for tailored privacy nudges are difficult to identify with consistency. Although Study 2 identified three potentially significant effects, our replication of these effects in Study 3 found them to be either fragile or non-existent. This result suggests that tailored privacy nudges at the scale of our studies may not be practical in application.

4.4.3 Transparency via Attribution in Vision Models and Internal Explanation for NLP models

Input Attribution in Vision Models

We have analyzed input attribution methods in terms of interpretability along one of several criteria [WMD+20]. The criteria: sufficiency, necessity, and proportionality, organize the various motivations presented in literature, and correspond to a measurable test with which an attribution can be evaluated. Sufficiency, for example, indicates that an attribution should highlight image elements which, were they to be presented alone, would be “sufficient” for the explained classification. Evaluating many popular attribution methods, we grouped and ranked them for their most appropriate interpretation. Some criteria are mutually exclusive, however, and our results indicate that there is no best method for each interpretation; additional work is necessary. In a parallel work [WWD+20], we generalized a class of attribution methods parameterized by the interpretation criterion.

In [WWR+20] we also investigated the relationship between adversarial perturbations and input attributions. Most input attributions methods incorporate gradients of classification scores in terms of inputs in some way. Methods for finding adversarial perturbations typically operate on the very same gradient by iteratively adjusting an image until a model makes a wrong prediction. For example, consider Figure 31 where the output contour of a simple model is visualized over the input space with several inputs (dots) highlighting their attributions which in this case are vectors (in the same way gradients are). These same vectors also point towards adversarial examples in some cases. Given the relationship, making models robust to adversarial perturbations must also impact attributions which are to be used as explanations. In this work we also demonstrate new techniques for making explanations robust.

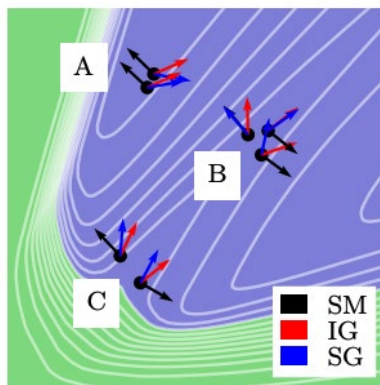


Figure 31: Input space, output contour and classification (color), and attributions (vectors).

Internal Explanations in Natural Language Processing Models

As part of [LML+20], we described a novel internal explanation methodology for analyzing neural models. Named influence paths, they highlight pathways in models that are most responsible for its handling of a specified concept. We applied influence paths to understanding standard LSTM language models on a variety of features of the English language such as subject-verb number agreement. The technology lets us discover how such concepts are implemented in LSTMs and how errors can arise when confusing nouns interact in the pathways. For example, Figure 32 (Subject-verb number agreement pathways in a 2-layer LSTM.) demonstrates the pathways involved in conveying the grammatical number of a subject in a sentence intersecting the pathway from a noun that causes an error in the predicting the grammatical number of a subsequent verb.

Additional research in this area, looking at feature-wide bias implications can also be found in [LBF+2018].

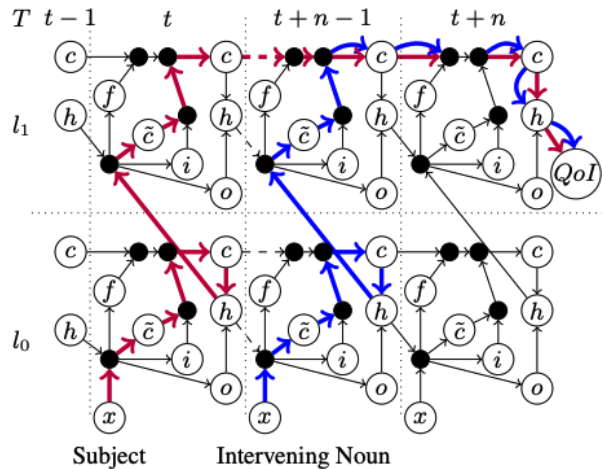


Figure 32: Subject-verb number agreement pathways in a 2-layer LSTM.

4.4.4 Explanations for Mobile App Privacy Permission Recommendations

While use of our explanation functionality made available in the mobile app privacy assistant we piloted proved somewhat limited, anecdotal evidence collected as part of interviews with participants suggest that such functionality can help people become more comfortable with automated recommendations by the assistants [LAS+16].

4.4.5 Transparency of Privacy Choices on the Web and Opt-Out Easy Browser Add-On Tool

The results of our work to increase the transparency of privacy choices on the web were published in our full paper at the 2020 Web Conference, “Finding a Choice in a Haystack: Automatic Extraction of Opt-Out Statements from Privacy Policy Text.” [KIN+20].

In this paper, we describe the technology behind Opt-Out Easy. First, we created two corpora of opt-out choices, which enabled the training of classifiers to identify opt-outs in privacy policies. Our overall approach for extracting and classifying opt-out choices combines heuristics to identify commonly found opt-out hyperlinks with supervised machine learning to automatically identify less conspicuous instances. Our approach achieves a precision of 0.93 and a recall of 0.9. Our paper also includes a usability evaluation of Opt-Out Easy. Finally, we present results of a large-scale analysis of opt-outs found in the text of thousands of the most popular websites. Our results suggest that many privacy policies do not have opt-out links, but that popular websites are more

likely to include them. Advertising opt-outs are by far the most common type of opt-out link, accounting for 60% of the opt-outs we detected.

An evaluation of our Opt-Out Easy browser extension also suggest that it is quite effective at helping users identify and take advantage of available opt-out choices [KIN+20].

4.5 Transparency and Compliance Analysis

4.5.1 Data Flow Modeling and Analysis

In [AMH17] we explored information theoretic definitions of privacy or information security in contexts where sensitive secrets are picked in a manner potentially influenced by non-sensitive features. For example, one's passwords are impacted by one's gender or age due to a variety of factors such as familiarity with certain topics or using one's birthday as part of a password (see for example Figure 33). In this work we decomposed an information theoretic notion of vulnerability of such secrets in terms of their inherent guessability and the guessability of the method with which the secrets are picked. The decomposition offers a more complete account vulnerability in realistic scenarios were secrets cannot be assumed to be picked perfectly and without undue influence from such things as demographics.

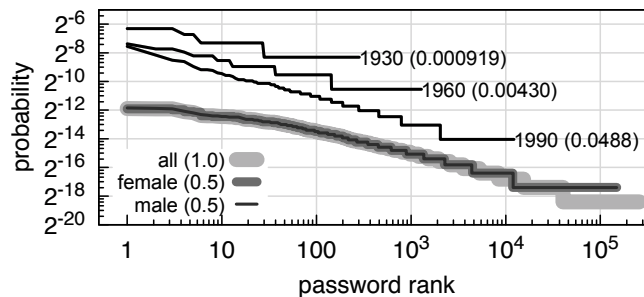


Figure 33: Password guessability under gender and age demographics.

In [DFK+17] we described a definition of proxy uses of private information in machine learnt models as model components having both sufficient association with a sensitive attribute and sufficient influence on the outcome of the model. This quantitative definition allowed us to define algorithms for detecting proxy uses in a large class of models as well as to perturb or repair these models so as to reduce the significance of association or influence thereby clearing models of proxy use. In the example shown in Figure 34, a decision tree model's nodes are placed on an association vs. influence graph with a forbidden region shown in the dark area in the top right. A repair algorithm

then produces a new model indicated by + marks that has no subcomponents in the forbidden region, i.e., has no proxy use of the strength indicated by that region.

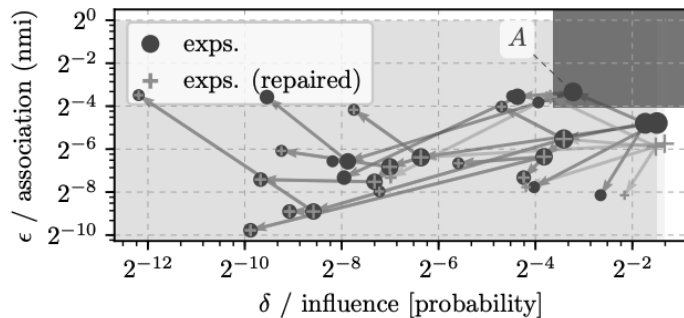


Figure 34: Decision Tree Model with Proxy Use and Repaired Model.

4.5.2 Mobile App Privacy Compliance Analysis

The app economy is largely reliant on data collection as its primary revenue model. To comply with legal requirements, app developers are often obligated to notify users of their privacy practices in privacy policies. However, prior research has suggested that many developers are not accurately disclosing their apps' privacy practices. Evaluating discrepancies between apps' code and privacy policies enables the identification of potential compliance issues. However, identifying these discrepancies manually is too time consuming to be scalable, showing the need for automated solutions.

Our approach is an automated analysis pipeline for identifying such discrepancies. We start by automatically retrieving mobile apps from the Google Play Store and automatically identifying and downloading the apps' privacy policies. Next, we use static code analysis to determine apps' privacy-relevant practices. Then, we use machine learning techniques to analyze apps' privacy policies, determining which practices are described and which are absent. Finally, we compare app behavior to the practices described in privacy policies, flagging discrepancies as potential compliance issues.

We published early work in this area at NDSS 2017, "Automated Analysis of Privacy Requirements for Mobile Apps" [ZWZ+16]. Our analysis of 17,991 free Android apps showed the viability of combining machine learning-based privacy policy analysis with static code analysis of apps. Results suggest that 71% of apps that lack a privacy policy should have one. Also, for 9,050 apps that have a policy, we find many instances of potential inconsistencies between what the app policy seems to state and what the code of the app appears to do. In particular, as many as 41% of

these apps could be collecting location information and 17% could be sharing such with third parties without disclosing so in their policies. Overall, each app exhibited a mean of 1.83 potential privacy requirement inconsistencies.

We refined our techniques, as detailed in a PoPETS 2019 paper titled “MAPS: Scaling Privacy Compliance Analysis to a Million Apps” [ZSS+19]. In particular, we increased the granularity of our analysis techniques, allowing us to distinguish between different types of information collection (e.g., fine- and coarse-grained location access). We also introduced technical improvements to make our analyses more robust and scalable. The result is our Mobile App Privacy System (MAPS) for automatically comparing apps’ code and privacy policies. We also published a related paper, “Natural Language Processing for Mobile App Privacy Compliance” [SZR+19], in which we give more details about our machine learning techniques.

Our automated pipeline enables the analysis of large populations of apps. In its first application, we conducted a privacy evaluation for a set of over one million Android apps from the Google Play Store. We found broad evidence of potential non-compliance. Many apps did not have privacy policies to begin with, and policies that did exist were often silent on the practices performed by apps. For example, 12.1% of apps had at least one location-related potential compliance issue. We hope that our extensive analysis will motivate app stores, government regulators, and app developers to more effectively review apps for potential compliance issues.

A related contribution was our creation of the APP-350 corpus. In order to train our machine learning classifiers, we needed a dataset of privacy policies annotated with different types of privacy-relevant behavior. For example, “first-party access of location data” or “third-party access of phone numbers.” We annotated the privacy policies of 350 mobile apps, and have made this dataset available to the community (<https://usableprivacy.org/data>).

MAPS provides a user interface (see Figure 35, 36, 37, 38, 39, and 40) that enables users at regulatory agencies to filter analysis results according to a number of criteria (e.g., Play Store category). This functionality has been piloted with several regulatory agencies. For example, we interacted with personnel at the FTC to focus on potential compliance issues under COPPA [SZR+19] and also used our app to help a large European manufacturer of electronic health devices analyze several of their mobile apps for potential compliance issues just before the EU GDPR came into effect.

Reports

Past Reports

Analyses

x Large Analysis #3

Potential Compliance Issues

All practices

All types

All specificities

[Refresh](#)

[fewer filters ^](#)

Static Analysis Practices

All

Installs

All

Third Parties

All

Content Ratings

All

Statuses

x Completed
x

Categories

All

Developers

All

Latest per app

Columns Export Page

Search:

Application	Static Analysis	Policy Analyses	Potential Compliance Issues
Chants D'Esperance with Tunes	5 practices performed	1 policy	8 potential issues
Classical music for baby	4 practices performed	1 policy	8 potential issues
Kids Learn Phonics: ABC Songs & Preschool Rhymes.	6 practices performed	1 policy	11 potential issues
Kids Top Nursery Rhymes Videos - Offline Learning	5 practices performed	1 policy	10 potential issues
Storybook Rhymes Volume 1	5 practices performed	0 policies	10 potential issues
Hymns and Praise with Tunes	5 practices performed	1 policy	8 potential issues
Bible en français Louis Segond	4 practices performed	0 policies	8 potential issues
Worship and Praise Lyrics	5 practices performed	1 policy	8 potential issues
LDS Music	4 practices performed	2 policies	16 potential issues
Chants D'Esperance	2 practices performed	0 policies	4 potential issues

Showing 1 to 10 of 1,035,853 entries (filtered from 1,049,790 total entries)

Figure 35: MAPS: Screenshot of MAPS mobile compliance analysis system showing bulk analysis functionality.

App Analysis Details

Baby Panda Care

Metadata Policies Potential Compliance Issues Policy Analyses Static App Analysis

App Metadata from Google Play

Package ID	com.sinyee.babybus.care
Categories	Game_educational, Family_pretend
Description	For more videos, subscribe to BabyBus on YouTube! ▶▶▶ https://goo.gl/lll2fX Caring for a baby requires a good amount of patience and hard work, and the thought of changing shoes and switching roles must have crossed your mind. Well, here is the chance! We are giving your children the role of caring for another baby, a cute little baby panda. You might be surprised to find how much your children will enjoy this
Developer's Name	BabyBus Kids Games
Developer's Email	ser@babybus.com
Physical Address	Building 9#3F, Fuzhou Cross-Strait Creative Garden, Jingong Road 1, Cangshan District, Fuzhou City, Fujian, China
Editor's Choice	No
Rating	4.23 (101,181 ratings)
Interactive Elements	-
Content Rating	Everyone
Content Descriptors	-
Downloads	10,000,000+
Updated	January 30, 2018
Policy Link(s)	http://en.babybus.com/index/privacyPolicy.shtml

Potential Compliance Issues Analysis	Policy Analyses	Static App Analysis
Status Completed	Number of Policies 2	Status Completed
Created 4/7/2018, 8:29:41 AM		Number of Practices 33
Last Modified 7/22/2018, 1:30:07 PM		Created 7/22/2018, 1:29:58 PM
Number of Issues 40		Last Modified 7/22/2018, 1:29:58 PM
		App Checksum 2b3649a434aa7b78e9468cd348341ec5

Figure 36: MAPS: Screenshot of MAPS system showing different tabs that enable users to find details about individual apps, including detailed analysis results (“potential compliance issues”, “policy analysis”, and “static app analysis”).

App Analysis Details

Baby Panda Care

Metadata

Policies

Potential Compliance Issues

Policy Analyses

Static App Analysis

Policy 1 of 2

- (1) Toggle navigation Home Our Apps About Us English 简体中文 繁體中文 한국어 日本語 Français Deutsch Русский Português
- (2) User Agreement Privacy Policy Partenariat web PRIVACY POLICY This Privacy Policy ("Policy") provides important information regarding BabyBus (Fujian) Network Technology Co., Ltd ("BabyBus," "we," or "our") in respect to user information collected through our mobile applications and the use of the information. You provide consent to this Policy when your access and use BabyBus applications ("Products") and other related services ("Services") .
- (3) 1. Personal and non-personal information.
- (4) We understand the importance of our users' personal information, and more importantly the information pertaining to children under 13 years of age. We do not collect or require users to enter their personal information when using our Products. We do not collect any personal information from children with our Products.
- (5) When using our Products, we might read non-personal information (such as internet, Wi-Fi availability, network connection, wake lock, device status, internal and external storage availability and usability) for the use of Product development and Service improvement. The collection of information will be used for the development and functional improvement of our Products to ensure quality user experience. We only collect such information to the extent that allow us to conduct our normal business operation and Products' research and development.
- (6) Any information we received will be used internally for the purpose of product development and to third parties performing services on our behalf, who comply with our Privacy Policy. We will not disclose your information publicly unless we have received your consent or under government order. 2.COPPA
- (7) We comply with the Children's Online Privacy Protection Act. We do not knowingly collect personal information on children under the age of 13. When a user identifies himself or herself as a child under the age of 13 through a support request or through any feedback, we will not collect, store or use, and will delete in a secure manner, any personal information of such user. 3.Submitting personal information for prizes and customer support.
- (8) From time to time, we may host contests or surveys on our website at www.babybus.com. We may send out a letter to your email address after you have given us consent through your participation and agree to receive the prize for the event you have participated in. For contests and surveys, we typically require only the information necessary for a child to participate. We only contact the parent for more personalized information for prize-fulfillment purposes when the child wins the contest or completes a survey.

Completed (2 Sources)

- <http://en.babybus.com/index/privacyPolicy>. from Play Page retrieved on 4/7/2018, 8:59:52 AM.shtml
- <http://en.babybus.com/index/privacyPolicy>. from Play Page retrieved on 4/7/2018, 8:59:52 AM.shtml

Figure 37: MAPS: Looking at an app's privacy policy.

App Analysis Details

Baby Panda Care

[Metadata](#)
[Policies](#)
[Potential Compliance Issues](#)
[Policy Analyses](#)
[Static App Analysis](#)

Completed Analysis (2 Sources)

<http://en.babybus.com/index/privacyPolicy.shtml> from Play Page retrieved on 4/7/2018, 8:59:52 AM

<http://en.babybus.com/index/privacyPolicy.shtml> from Play Page retrieved on 4/7/2018, 8:59:52 AM

Columns
 Export Page
 Search:

Policy Analysis Practice ⓘ	App Analysis Practice ⓘ	Non-Compliance Type ⓘ	Specificity ⓘ	⚙
Identifier Cookie or similar Tech 3rd Party	Identifier Cookie or similar Tech 3rd Party	Omission	Specific	
Identifier Device ID 3rd Party	Identifier Device ID 3rd Party	Omission	Specific	
Identifier IMEI 3rd Party	Identifier IMEI 3rd Party	Omission	Specific	
Identifier IMSI 3rd Party	Identifier IMSI 3rd Party	Omission	Specific	
Identifier IP Address 1st Party	Identifier IP Address 1st Party	Omission	Specific	
Identifier IP Address 3rd Party	Identifier IP Address 3rd Party	Omission	Specific	
Identifier MAC 3rd Party	Identifier MAC 3rd Party	Omission	Specific	
Identifier Mobile Carrier 3rd Party	Identifier Mobile Carrier 3rd Party	Omission	Specific	
Identifier SIM Serial 3rd Party	Identifier SIM Serial 3rd Party	Omission	Specific	
Identifier SSID BSSID 3rd Party	Identifier SSID BSSID 3rd Party	Omission	Specific	
Identifier 1st Party	Identifier IP Address 1st Party	Omission	General	
Identifier 3rd Party	Identifier Cookie or similar Tech 3rd Party	Omission	General	
Identifier 3rd Party	Identifier Device ID 3rd Party	Omission	General	
Identifier 3rd Party	Identifier IMEI 3rd Party	Omission	General	
Identifier 3rd Party	Identifier IMSI 3rd Party	Omission	General	
Identifier 3rd Party	Identifier IP Address 3rd Party	Omission	General	

Figure 38: MAPS: Summary of potential privacy compliance issues for a given app.

App Analysis Details

Baby Panda Care

Metadata Policies Potential Compliance Issues Policy Analyses Static App Analysis

Complete Policy Analyses

Completed Analysis (2 Sources)
<http://en.babybus.com/index/privacyPolicy.shtml> from Play Page retrieved on 4/7/2018, 8:59:52 AM
<http://en.babybus.com/index/privacyPolicy.shtml> from Play Page retrieved on 4/7/2018, 8:59:52 AM

Supporting Policy Segments

Choose a practice to view policy segments.

Columns Export Page

Search:

Policy Analysis Practice	Performed	Not Performed
Contact E Mail Address 1st Party	Yes 2 Segments	No
Identifier Cookie or similar Tech 1st Party	Yes 2 Segments	No
Contact 1st Party	No	No
Contact 3rd Party	No	No
Contact E Mail Address 3rd Party	No	No
Contact Phone Number 1st Party	No	No
Contact Phone Number 3rd Party	No	No
Contact Postal Address 1st Party	No	No
Contact Postal Address 3rd Party	No	No
Contact ZIP 1st Party	No	No

Showing 1 to 10 of 50 entries

Show 10 entries

Previous 1 2 3 4 5 Next

Figure 39: MAPS: Automated analysis of a privacy policy.

App Analysis Details

Baby Panda Care

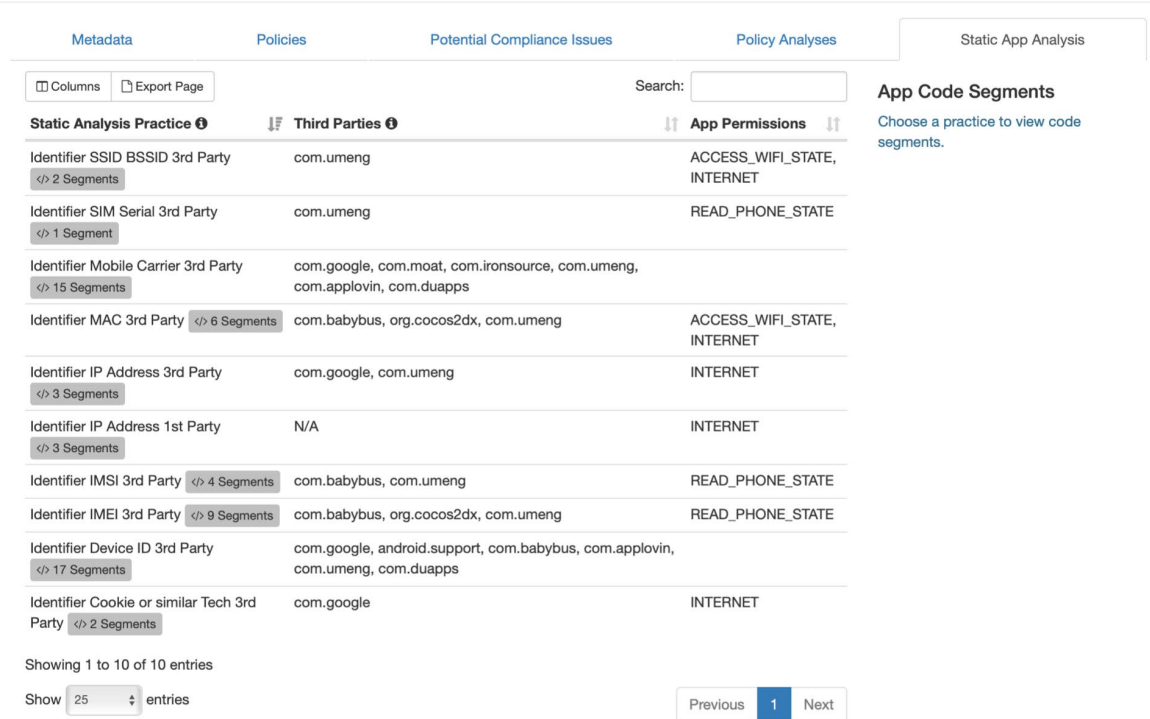


Figure 40: MAPS: Static analysis of a mobile app's code.

4.5.3 IFTTT

Our information-flow-based analysis of 19,323 unique IFTTT recipes found that around 50% of the 19,323 unique recipes we examined are potentially unsafe, as they contain a secrecy violation, an integrity violation, or both [SAB+17]. More specifically: 4,432 recipes (23%) contained only integrity violations, 3,220 recipes (17%) only secrecy violations, and 1,985 recipes (10%) both secrecy and integrity violations. These numbers include the recipes with both definite and possible ("maybe") violations. If we consider only definitely violating recipes, there are 7,150 (37%) unsafe recipes: 3,605 (19%) with only integrity violations, 1,927 (10%) with only secrecy, and 1,618 (8%) with both. Figure 41 illustrates these results.

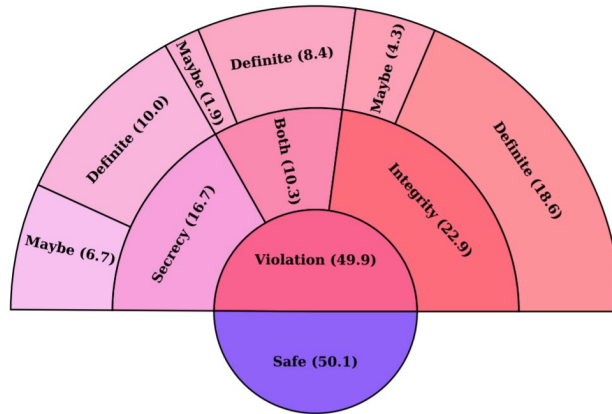


Figure 41: Violating recipes, broken down by violation type.

We next categorized the types of harm that these potentially unsafe recipes can cause to users. After manually examining a random selection of potentially unsafe recipes, we find that recipes cannot only lead to harms such as personal embarrassment but can also be exploited by an attacker, e.g., to distribute malware or carry out denial-of-service attacks.

We followed up this largely automated analysis with a more in-depth examination of the risks and harms of IFTTT recipes based on recipes collected from a user study of 28 active IFTTT users, through which we collected 732 applets that they installed as well as additional information that described their use [CSK+20]. Using the automated information-flow-based analysis we previously developed, updated for new devices that appeared in the rules we collected from users, we found that about 57% of participants' IFTTT rules had potential secrecy or integrity violations, which is roughly consistent with our prior findings. However, our more detailed manual analysis, which took advantage of the additional context collected from participants, revealed that although many applets might technically have secrecy or integrity violations, they are rarely harmful because of these violations. Only about 10% of the secrecy-violating rules (just over 3% of all rules) could lead to secrecy harms, and just under 20% of integrity-violating rules (8.6% of all rules) present serious integrity-related risks. Consistent with our manual evaluation, participants did not believe that their rules were likely to lead to secrecy- or integrity-related harms, though they did care about the security and privacy of their rules.

Our contextualized analysis of trigger-action rules and their security and privacy risks also led to unexpected findings, some which involved identifying risks to people that previously weren't being considered in automated analyses. For example, IFTTT rules can create surveillance risks to incidental users, i.e., people besides the IFTTT user who created the rule. More specifically, many rules cause data to be collected about

people other than the IFTTT user, possibly without their awareness. 22% of the rules we collected fell into this category, and about a third of them were not detected through automated analyses.

In summary, building on previous work on information flow, we developed one of the first automated analyses of IFTTT rules (and, more generally, trigger-action rules for home automation platforms). This analysis identified that almost half of publicly shared recipes were potentially unsafe. A detailed, systematic examination of individual users' specific recipes revealed that a smaller portion, roughly 10%, of recipes was unsafe and lead to substantial harms, while other unsafe recipes would likely cause minor harms at worst. This examination also revealed potential harms not accounted for by our and others' automated analyses and suggested features that future analyses should have to more accurately and more comprehensively capture the risks that users face in practice.

4.6 Privacy Risk in Machine Learning Pipelines

4.6.1 Proxy Use

We made precise the notions of second-order causation and correlation used by his notions of proxy. He has identified modal logics with multiple sorts of points as a possible framework for building second-order causal reasoning upon. One sort of point can represent different individuals in a world, to allow frequentist first-order causal reasoning. A second sort of point can represent different assignments of values to individuals. This second sort corresponds to data sets about the individuals, making it suitable for second-order causation. He also compared them to notions found in the scientific and legal prior work.

We also provided auditing methods. They rerun the ML algorithm on altered data sets to see how the classifier changes. To check whether the algorithm used the proxy, it removes or randomizes the proxy from the data set, retrains the classifier, and checks whether it has changed in a significant way (using multiple runs to find changes that are not merely due to randomization). In particular, it checks whether the correlation between the target and the decision remains. If so, it suggests that the use of the proxy causes this correlation. Similarly, it can check whether the correlation between the target and proxy causes the proxy's use by randomizing the target's value in the data set.

4.6.2 Whitebox Membership Inference in Non-parametric Models and Ensembles.

We proposed a class of white-box shadow model attacks on tree-based, non-parametric, models based on feature importance (see Figure 42). We evaluated these attacks on random forests (RF) and Gradient Boosting machines (GBMs) trained on four real world datasets and showed these as vulnerable to white-box attacks as comparable previously studied neural networks.

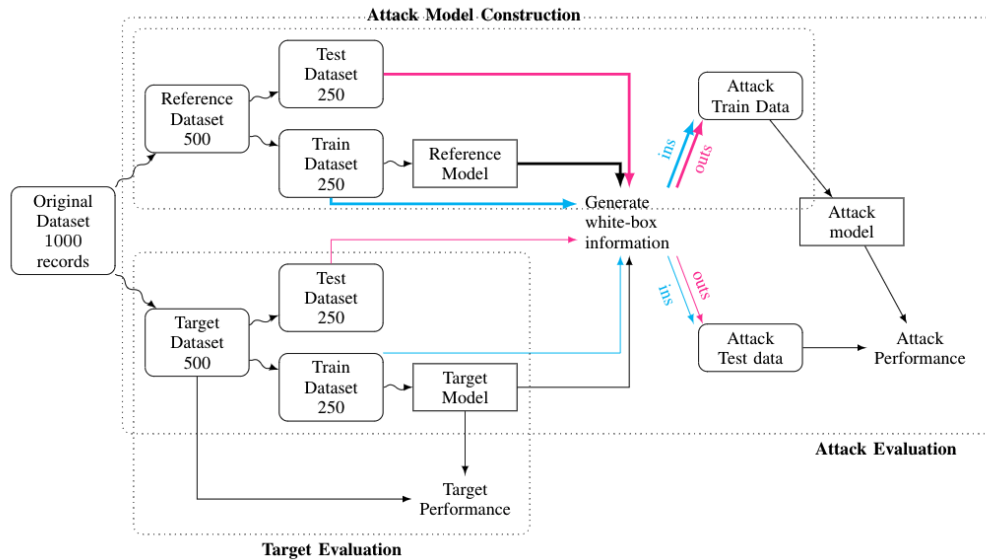


Figure 42: White-box Shadow Model Attack for Membership Inference.

We also evaluate the effects of the number of estimators on attack success and showed that with decreasing generalization error, attack accuracy increases, demonstrating a fundamental tradeoff between generalization typically induced by larger ensembles and membership privacy risk (see Figure 43).

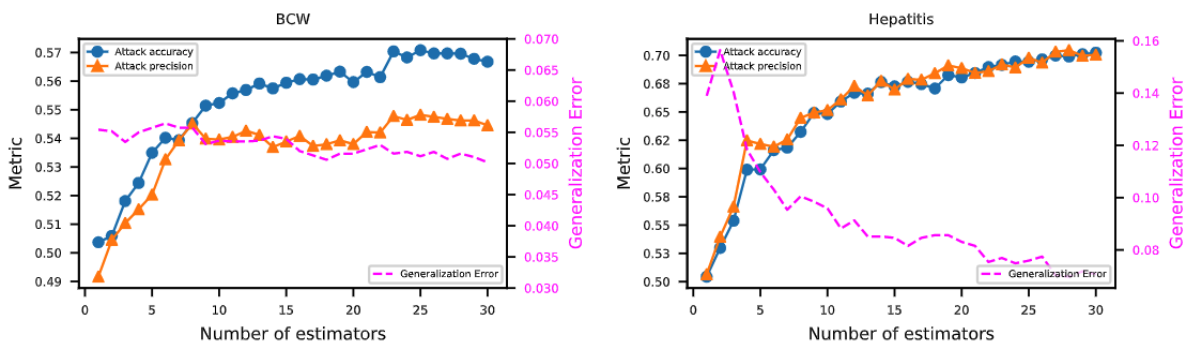


Figure 43: Membership inference attack success vs. number of estimators in ensembles.

4.7 Integration, Deployment, and Evaluation of Technologies and Tools

4.7.1 Integration of IoT Privacy Infrastructure in TA3 Systems

Under the IoT Collaboration Research Team (CRT) of the Brandeis program, our team worked closely with the team at University of California Irvine to integrate our IoT Privacy Infrastructure with their Testbed for IoT-based Privacy-Preserving PERvasive Spaces. The UCI team has published TIPPERS-based applications in IoT Privacy Infrastructure to provide UCI community privacy notice and choice (see Figure 14). Integration with TIPPERS and deployment at UCI including supporting opt out functionality required for UCI's Office of Information Technology. In December 2019, we presented IoT Privacy Infrastructure to a group of UCI students, who have developed and deployed "Zotbins", sensor-based smart trash cans around the UCI campus. The students used our IoT privacy infrastructure to publicize the locations of Zotbins on the UCI campus along with disclosure of their data practices.

4.7.2 Mobile app privacy assistant

An improved version of the mobile app privacy assistant piloted as part of the study discussed in [LAS+16] was released in the Google Play store for users of rooted Android phones. Details of improvements made to the privacy assistant prior to its release, including the development of enhanced privacy profiles are discussed in [Liu19]. Figure 44 shows a few screenshots of this app.

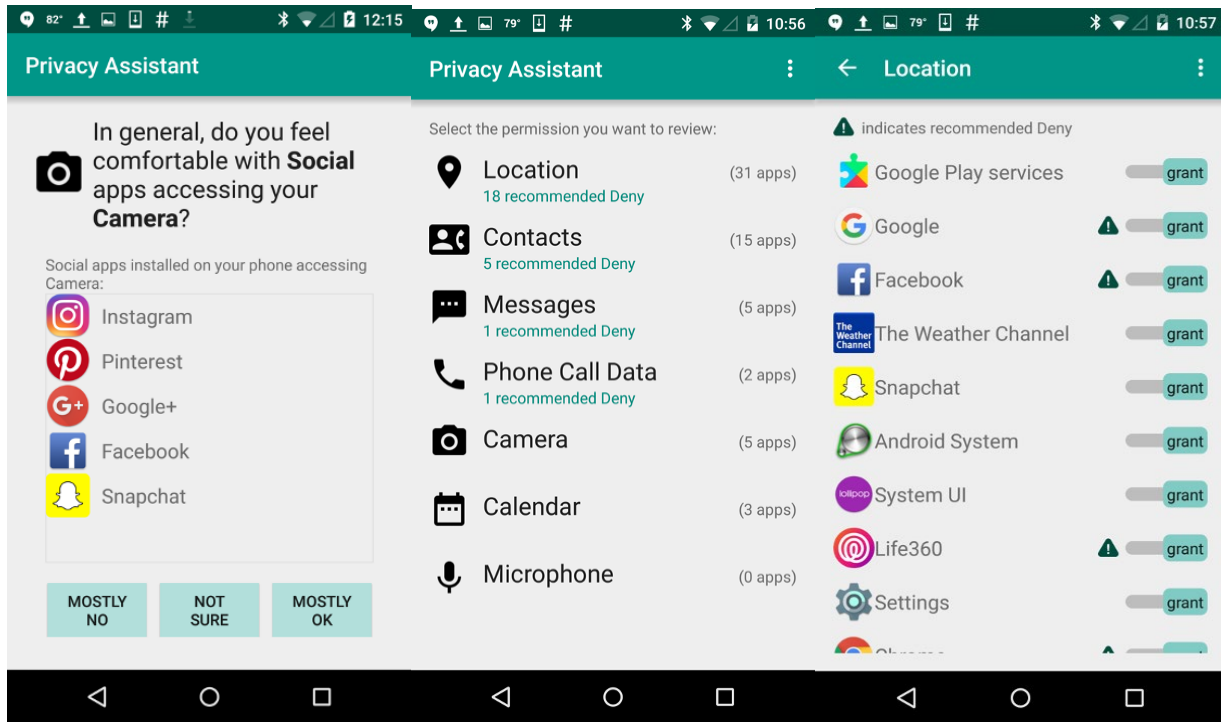


Figure 44: Screenshots of Mobile App Privacy Assistant deployed in Google Play Store for rooted Android phones.

4.7.3 IoT Privacy Infrastructure

We successfully released the IoT Privacy Infrastructure to the public. One project goal for the IoT Privacy Infrastructure is to evaluate its scalability and usability through a real-world deployment. By the end of 2020, the IoT Assistant app has been downloaded by over 17,000 users and the IoT privacy infrastructure hosts descriptions of over 150,000 IoT resources (e.g., see Figure 15 and Figure 17).

4.7.4 The IoT Assistant App and Video Obfuscation Application

In collaboration with CMU's Eberly Center for Teaching Excellence and Education Innovation, we demonstrated a working prototype of our IoT Assistant infrastructure in support of smart classroom functionality, where participants in the classroom can use their IoT Assistant app to individually opt in or out of being recorded in footage being captured during a lesson. IoT Assistant users can discover and access descriptions of individual classroom sessions along with privacy settings, which enable them to opt in or out of video capture. If they elect to not be capture, their face is obfuscated as part of the footage being captured during the classroom session. To opt into being captured,

users access an opt-in setting associated with the classroom session from their IoT Assistant app. Once they have selected this setting, they are required to train the system to recognize their face as part of a short training session also mediated by the IoT Assistant app. Obfuscation is enforced in real-time using the OpenFace system, as detailed in [WAD+17,WAD+18,DDW+17]. This prototype could be extended to a variety of other contexts to support compliance with regulations such as CCPA and GDPR (e.g., video analytics in malls, amusement parks, classrooms, gyms, places of worship and much more).

4.7.5 MAPS

As already discussed, the MAPS system was used to assist a large European electronic device manufacturer analyze several legacy mobile apps for potential compliance violations just prior to the EU GDPR regulation taking effect in Europe. The system was also used to analyze a number of automotive apps for potential compliance issues as part of a collaboration with the Center for Democracy and Technology. Early versions of the system were also used as part of a collaboration with the California Office of the Attorney General. And finally, the system was used to systematically scan and analyze over 1 million Android apps on the Google Play store with results of this analysis used to compile reports shared with personnel at the Federal Trade Commission [ZPS+19]

4.7.6 Opt-Out Easy

We released our Opt-Out Easy browser plugin on the Google Chrome and Mozilla Firefox extension stores [OptOutChrome, OptOutFirefox]. Opt-Out Easy uses machine learning to identify privacy choices in the text of websites' privacy policies, and presents these choices to users as they browse the web. Before launching, we analyzed many popular websites for opt-outs, and Opt-Out Easy allows users to request analyses of additional websites. We regularly re-analyze websites to ensure our results are up-to-date. Currently, Opt-Out Easy contains opt-out choices for more than 4000 websites. Figure 45 shows the main screen of the Opt-Out Easy user interface.

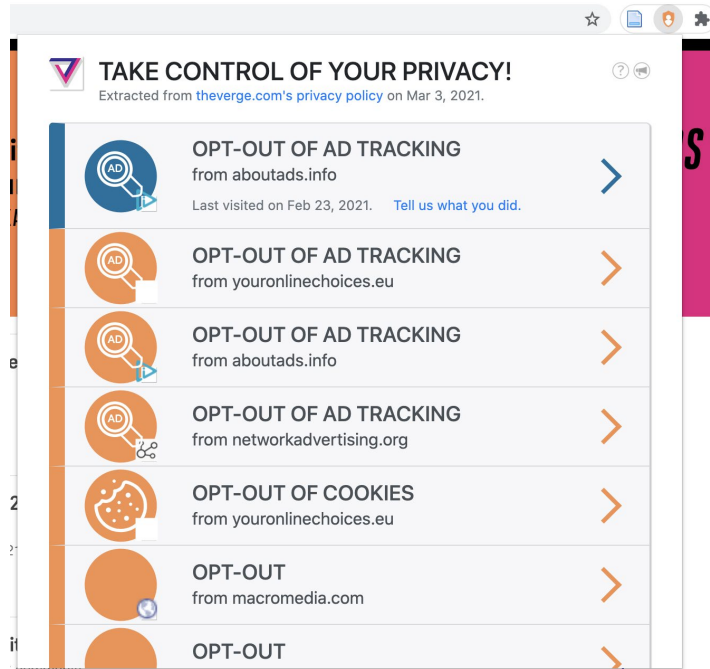


Figure 45: The User Interface of Opt-Out Easy.


4.7.7 IoT Security and Privacy Nutrition Labels






We have documented our label design in a complete label specification on our website [IoTLabel]. Figure 46 and Figure 47 below show the primary layer and the secondary layer of the IoT security and privacy nutrition label, respectively. In addition, we have implemented an interactive wizard for generating labels according to our specification.

Security & Privacy Overview



Smart Device Co.

Smart Video Doorbell NS200
 Firmware version: 2.5.1 - updated on: 11/12/2020
 The device was manufactured in: China

 Security Mechanisms	Security updates Automatic - Available until at least 1/1/2022
	Access control Password - Factory default - User changeable, Multi-factor authentication, Multiple user accounts are allowed

 Data Practices	Sensor data collection	 Visual	 Audio	 Physiological	 Location
	Sensor type	Camera	Microphone		
	Purpose	Providing device functions	Providing device functions, Research		
	Data stored on device	Identified	No device storage		
	Data stored on cloud	Identified	Identified - Option to delete		
	Shared with	Manufacturer, Government	Manufacturer		
	Sold to	Not disclosed	Not sold		
	Other collected data	Motion, Account info, Payment info, Contact info, Device setup info, Device tech info, Device usage info			

Privacy policy www.NS200.smartdeviceco.com/policy

 More Information	Detailed Security & Privacy Label: www.iotsecurityprivacy.org/labels	
--	---	---

CMU IoT Security and Privacy Label **CISPL 1.0** iotsecurityprivacy.org







Figure 46: Primary layer of IoT Security and Privacy Nutrition Label.

Security & Privacy Details

Smart Device Co.

Smart Video Doorbell NS200
 Firmware version: 2.5.1 - updated on: 11/12/2020
 The device was manufactured in: China

 Security Mechanisms	Security updates	Automatic - Available until at least 1/1/2022		
	Access control	Password - Factory default - User changeable, Multi-factor authentication, Multiple user accounts are allowed		
	Security oversight	No security audits		
	Ports and protocols	www.NS200.smartdeviceco.com/ports		
	Hardware safety	Not disclosed		
	Software safety	www.NS200.smartdeviceco.com/sw_safety		
	Personal safety	www.NS200.smartdeviceco.com/user_safety		
	Vulnerability disclosure and management	www.NS200.smartdeviceco.com/vul_report		
	Software and hardware composition list	www.NS200.smartdeviceco.com/BOM		
	Encryption and key management	www.NS200.smartdeviceco.com/encryption		
 Data Practices	Sensor data collection	Visual	Audio	Motion
	Sensor type	Camera	Microphone	Motion sensor
	Collection frequency	Continuous - Option to opt out	Continuous - Option to opt in	Continuous - Option to opt out
	Purpose	Providing device functions	Providing device functions, Research	Providing device functions, Research
	Data stored on the device	Identified	No device storage	Pseudonymized
	Local data retention time	Up to a year	No retention	Up to a month
	Data stored in the cloud	Identified - Data subject access request	Identified - Option to delete	No cloud storage
	Cloud data retention time	Up to 10 years	Up to two months	No cloud storage
	Data shared with	Manufacturer, Government	Manufacturer	Manufacturer, Third parties
	Data sharing frequency	Periodic	Periodic - Adjustable	Periodic - Adjustable
	Data sold to	Not disclosed	Not sold	Third parties
	Other collected data	Account info, Payment info, Contact info, Device setup info, Device tech info, Device usage info		
	Data linkage	Data will not be linked with other data sources		
	What will be inferred from user's data	Not disclosed		
	Special data handling practices for children	No		
In compliance with	GDPR			
Privacy policy	www.NS200.smartdeviceco.com/policy			
 More Information	Call Smart Device Co. with your questions at	1 000-000-0000		
	Email Smart Device Co. with your questions at	info@smartdeviceco.com		
	Functionality when offline	Limited functionality on offline mode		
	Functionality with no data processing	Limited functionality on dumb mode		
	Physical actuations and triggers	Device blinks when motion is detected		
	Compatible platforms	Amazon Alexa		


CMU IoT Security and Privacy Label **CISPL 1.0** iotsecurityprivacy.org 

Figure 47: Secondary layer of IoT Security and Privacy Nutrition Label.

5. Conclusions

Managing one's privacy is becoming increasingly complex in today's data centric economy. The expectations placed on users when it comes to learning about data practices and making privacy decisions associated with mobile technologies, the Internet of Things and Big Data are simply unrealistic. We need new technologies that can help users regain control over the collection and use of their data, as well as technologies to assist technology providers and regulators ensure that deployed technologies comply with relevant privacy regulations.

Work in this project produced a number of new technologies, tools, models, and findings designed to address this challenge. These results have already influenced and are expected to continue influencing privacy practices. This includes the deployment of a publicly accessible privacy infrastructure for the Internet of Things to help publicize the presence of IoT systems and their data collection and use practices, including any available privacy choices. This includes the availability of an "IoT Assistant" app for people to discover these IoT systems and their data practices, including accessing any privacy choices they make available - the IoT Assistant app is available in both the iOS and Android app stores. This also includes the release of a personalized privacy assistant app for rooted Android phones to help people manage their mobile app permissions, the development of a mobile app privacy compliance tool ("MAPS") that has been used to automatically analyze over 1 million Android apps for potential privacy compliance issues, and the launch of the "Opt-Out Easy" browser extension to help people discover opt-out choices buried deep in the text of privacy policies (tool available in the Chrome and Firefox stores). In addition to these tools, research conducted in the project helped develop rich models of people's privacy preferences and expectations across a variety of mobile app and IoT contexts, including the development of predictive models using machine learning techniques. These techniques have been shown to have the potential of drastically reducing user burden when it comes to helping users manage the myriad of privacy choices they need to make in mobile and IoT environments. This research has included exploring tradeoffs between user burden and control when it comes to configuring these technologies, including identifying differences in the way different people feel about relying on automated recommendations and even delegating some of their privacy decisions to predictive models. Research on privacy nudges has demonstrated that nudging can be configured to be particularly effective at helping people make privacy decisions they are less likely to regret. Privacy nutrition labels have been developed to help simplify the presentation of key data practices to help users when it comes to evaluating different IoT devices. Work on transparency has also yielded techniques to identify and analyze data flows and to help evaluate

privacy risk in machine learning pipelines. In addition to the above, research conducted in this project has influenced ongoing privacy discussions with findings being presented at high profile events such as the FTC Privacy Conference, the International Conference of Data Protection and Privacy Commissioners, events organized by the International Association of Privacy Professionals and many other venues, while also receiving significant exposure in the press. Tools such as our Mobile App Privacy Compliance system have been used to automatically analyze large numbers of mobile apps for potential privacy compliance issues and generate reports shared with regulatory agencies.

Most recently, the project also designed and evaluated a number of possible icons and textual descriptions, and its recommended design was eventually adopted as the official recommendation for the new California Consumer Protection Act [HZY+21].

Finally, work conducted in this project is influencing public policy discussions, with investigators invited to attend a number of high-profile public policy events (see Subsection 7.2 below) and this project's research also receiving a fair bit of coverage in the press (see news section at [PPAsite]).

6. Recommendations

With the widespread adoption of Artificial Intelligence and the accelerating deployment of IoT technologies, it is simply impossible for people to keep up with the number of ways in which their data is collected and used. While the past few years have seen the emergence of new privacy regulations around the world, the impact of these regulations on people's ability to manage their privacy will remain limited unless people can be provided with technologies that empower them to take advantage of the more detailed privacy disclosures and richer sets of privacy settings required by these regulations. Without such technologies the main limitation is usability: people are simply unable to deal with the complexity of the dataflows, the amount of information they are expected to read and the number of privacy decisions they are expected to make.

Industry by default has no incentive to go beyond "best practices" when it comes to showing that it complies with applicable regulations. What is urgently needed is a significant increase in government funding for the development of new user-oriented technologies that help people manage their privacy, technologies such as those developed as part of this project. Results from our project have shown that these technologies can have a major impact on restoring people's control over their data. Yet significantly more research is required to further develop, test and fine-tune these technologies. Only government can provide the necessary funding. Given DARPA's role in supporting the development of cutting-edge AI technologies, the agency also has a responsibility to continue funding privacy research and to contribute to the development of privacy enhancing technologies that counter balance the privacy challenges arising from the broad deployment of AI. Incidentally a good part of the research conducted in this project also involved the use of AI to help people regain control over their data.

7. References

7.1 Conference Papers and Journal Articles

[AMH17] Alvim, M., Mardziel P., Hicks, M., "Quantifying Vulnerability of Secret Generation Using Hyper-Distributions." *International Conference on Principles of Security and Trust*, Springer, Berlin, Heidelberg, 2017.

<https://arxiv.org/pdf/1701.04174.pdf>

[ASS+15] H. Almuhiemedi, F. Schaub, N. Sadeh, Y. Agarwal, A. Acquisti, I. Adjerid, J. Gluck, L. Cranor, "Your Location Has Been Shared 5398 Times! A Field Study on Mobile Privacy Nudges", in Proc. CHI 2015, Jul 2015 [pdf]

[CSK+20] Cobb, C., Surbatovich, A., Kawakami, A., Sharif, M., Bauer, L., Das, A., Jia, L., "How Risky are Real Users' IFTTT Applets?" *SOUPS*, 2020.

https://camillec.com/SOUPS_2020_IFTTT.pdf

[CFP+20] Colnago, J., Feng, Y., Palanivel, T., Pearman, S., Ung, M., Acquisti, A., Cranor, L., Sadeh, N., "Informing the Design of a Personalized Privacy Assistant for the Internet of Things," *ACM CHI*, 2020.

<https://dl.acm.org/doi/abs/10.1145/3313831.3376389>

[DDS+18] Das, A., Degeling, M., Smullen, D., Sadeh, N., "Personalized Privacy Assistants for the Internet of Things: Providing Users with Notice and Choice," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 35-46, 2018.

https://www.privacyassistant.org/media/publications/IEEE_magazine_2018.pdf

[DDW+17] Das, A., Degeling, M., Wang, X., Wang, J., Sadeh, N., Satyanarayanan, M., "Assisting Users in a World Full of Cameras: A Privacy-Aware Infrastructure for Computer Vision Applications," *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1387-1396, 2017.

<http://www.cs.cmu.edu/~anupamd/paper/CV-COPS-2017.pdf>

[DFK+17] Datta, A., Fredrikson, M., Ko, G., Mardziel, P., Sen, S., "Use Privacy in Data-Driven Systems Theory and Experiments with Machine Learnt Programs," *ACM Conference on Computer and Communications Security (CCS)*, 2017.

<https://dl.acm.org/doi/pdf/10.1145/3133956.3134097>

Datta, A., Sen, S., Zick, Y., "Algorithmic Transparency via Quantitative Input Influence," in *Proceedings of the 37th IEEE Symposium on Security and Privacy*, May 2016.

<http://www.andrew.cmu.edu/user/danupam/datta-sen-zick-oakland16.pdf>

[EDB+18] Emami-Naeini, P., Degeling, M., Bauer, L., Chow, R., Cranor, L., Haghghat, M.R., Patterson, H., "The Influence of Friends and Experts on Privacy Decision

Making in IoT Scenarios,” *21st ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*, pp. 1-26, 2018.

<https://users.ece.cmu.edu/~lbauer/papers/2018/cscw2018-social-influence.pdf>

[EDA+21] Emami-Naeini, P., Dheenadhayalan, J., Agarwal, Y., Cranor, L., “Which Privacy and Security Attributes Most Impact Consumers’ Risk Perception and Willingness to Purchase IoT Devices?” *42nd IEEE Symposium on Security and Privacy (S&P)*, 2021.

https://www.synergylabs.org/yuvraj/docs/Emami-Naeini_Oakland21_IoTLabelWhichAttributes.pdf

[EAC+20] Emami-Naeini, P., Agarwal, Y., Cranor, L., Hibshi, H., “Ask the Experts: What Should Be on an IoT Privacy and Security Label?” *IEEE S&P*, pp. 447-464, 2020. <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9152770>

[EBH+17] Emami-Naeini, P., Bhagavatula, S., Habib, H., Degeling, M., Bauer, L., Cranor, L., Sadeh, N., “Privacy Expectations and Preferences in an IoT World,” *13th Symposium on Usable Privacy and Security (SOUPS)*, pp. 399-412. 2017.

<https://www.usenix.org/system/files/conference/soups2017/soups2017-naeini.pdf>

[EDA+19] Emami-Naeini, P., Dixon, H., Agarwal, Y., Cranor, L., “Exploring How Privacy and Security Factor into IoT Device Purchase Behavior,” *CHI*, 2019.

<https://dl.acm.org/doi/pdf/10.1145/3290605.3300764>

[FYN21] Feng, Y., Yao, Y., Sadeh, N., “A Design Space for Privacy Choices: Towards Meaningful Privacy Control in the Internet of Things,” *CHI*, 2021.

<https://www.privacyassistant.org/media/publications/chi21-design-space.pdf>

Gluck, J., Schaub, F., Friedman, A., Habib, H., Sadeh, N., Cranor, L., Agarwal, Y., “How Short is Too Short? Implications of Length and Framing on the Effectiveness of Privacy Notices,” *Symposium on Usable Privacy and Security (SOUPS)*, June 2016.

<https://www.usenix.org/system/files/conference/soups2016/soups2016-paper-gluck.pdf>

[HZY+21] Habib, H., Zou, Y., Yao, Y., Acquisti, A., Cranor, L., Reidenberg, J., Sadeh, N., Schaub, F., “Toggles, Dollar Signs, and Triangles: How to (In)Effectively Convey Privacy Choices,” *ACM Conference on Human Factors in Computing Systems (CHI)*, 2021. <https://yixinzou.github.io/publications/pdf/chi2021-habib-preprint.pdf>

[KIN+20] Kumar, V.B., Iyengar, R., Nisal, N., Feng, Y., Habib, H., Story, P., Cherivirala, S., Hagan, M., Cranor, L., Wilson, S., Schaub, F., Sadeh, N., “Finding a Choice in a Haystack: Automatic Extraction of Opt-Out Statements from Privacy Policy Text,” *Web Conference*, April 2020.

https://usableprivacy.org/static/files/kumar_iyengar_www_2020.pdf

[LBF+2018] Leino, K., Black, E., Fredrikson, M., Sen S., Datta, A., "Feature-wise Bias Amplification," *International Conference on Learning Representations (ICLR)*, 2018. <https://openreview.net/pdf?id=S1ecm2C9K7>

[LAS+16] Liu, B., Andersen, M.S., Schaub, F., Almuhiemedi, H., Zhang, S., Sadeh, N., Acquisti, A., Agarwal, Y., "Follow My Recommendations: A Personalized Assistant for Mobile App Permissions," *Symposium on Usable Privacy and Security (SOUPS '16)*, Jun 2016. <https://www.usenix.org/system/files/conference/soups2016/soups2016-paper-liu.pdf>

[LML+20] Lu, K., Mardziel, P., Leino, K., Fredrikson, M., Datta, A., "Influence Paths for Characterizing Subject-Verb Number Agreement in LSTM Language Models," *Annual Meeting of the Association for Computational Linguistics*, May 2020. <https://arxiv.org/pdf/2005.01190.pdf>

[PDY+17] Pappachan, P., Degeling, M., Yus, R., Das, A., Bhagavatula, S., Melicher, W., Emami Naeini, P., Zhang, S., Bauer, L., Kobsa, A., Mehrota, N., Sadeh, N., Venkatasubramanian, N., "Towards Privacy-Aware Smart Buildings: Capturing, Communicating, and Enforcing Privacy Policies and Preferences," *International Workshop on the Internet of Things Computing and Applications (IoTCA)*, June 2017. 2017. <https://www.cs.cmu.edu/~anupamd/paper/IoTCA2017.pdf>

Rao, A., Schaub, F., Sadeh, N., Acquisti, A., Kang, R., "Expecting the Unexpected: Understanding Mismatched Privacy Expectations Online," *Symposium on Usable Privacy and Security (SOUPS)*, June 2016. <https://www.privacyassistant.org/media/publications/rao.pdf>

[SAB+17] Surbatovich, M., Aljuraidan, J., Bauer, L., Das, A., Jia, L., "Some recipes can do more than spoil your appetite: Analyzing the security and privacy risks of IFTTT recipes," *In Proceedings of the 26th International World Wide Web Conference*, April 2017. <https://users.ece.cmu.edu/~lbauer/papers/2017/www2017-ifttt-info-flows.pdf>

[SBS+18] Schiffner, S., Berendt, B., Siil, T., Degeling, M., Riemann, R., Schaub, F., Wuyts, K., Attores, M., Guerses, S., Klabunde, A., Polonetsky, J., Sadeh, N., Zanfir-Fortuna, G., "Towards a Roadmap for Privacy Technologies and the General Data Protection Regulation: A Transatlantic Initiative," *Annual Privacy Forum*, Barcelona, 2018. <https://www.esat.kuleuven.be/cosic/publications/article-2909.pdf>

[SFZ+20] Smullen, D., Feng, Y., Zhang, S., Sadeh, N., "The Best of Both Worlds: Mitigating Trade-offs Between Accuracy and User Burden in Capturing Mobile App Privacy Preferences," *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2020. <https://privacyassistant.org/media/publications/pets-paper52-2020.pdf>

[SYS+21] Smullen, D., Yao, Y., Sadeh, N. "Managing Intrusive Practices In The Browser: A User Centered Perspective," *Proceedings of Privacy Enhancing Technologies (PoPETs)*, 2021. To appear.

[SSA+20] Story, P., Smullen, D., Acquisti, A., Cranor, L., Sadeh, N., Schaub, F., “From Intent to Action: Nudging Users Towards Secure Mobile Payments,” *Sixteenth Symposium on Usable Privacy and Security (SOUPS)*, 2020.

https://usableprivacy.org/static/files/story_soups_2020.pdf

[SZR+19] Story, P., Zimmeck, S., Ravichander, A., Smullen, D., Wang, Z., Reidenberg, J., Russell, N.C., Sadeh, N., “Natural Language Processing for Mobile App Privacy Compliance,” *AAAI Spring Symposium on Privacy Enhancing AI and Language Technologies (PAL)*, May 2019.

https://usableprivacy.org/static/files/story_pal_2019.pdf

[SZS+18] Story, P., Zimmeck, S., Sadeh, N., “Which Apps Have Privacy Policies? An Analysis of Over One Million Google Play Store Apps,” *Annual Privacy Forum*, Barcelona, 2018. https://usableprivacy.org/static/files/Story_APF_2018.pdf

[SAB+17] Surbatovich, M., Aljuraidan, J., Bauer, L., Das, A., Jia, L., “Some recipes can do more than spoil your appetite: Analyzing the security and privacy risks of IFTTT recipes,” *In Proceedings of the 26th International World Wide Web Conference*, April 2017. <https://dl.acm.org/doi/pdf/10.1145/3038912.3052709>

[UPB+16] Ur, B., Pak Yong Ho, M., Brawner, S., Lee, J., Mennicken, S., Picard, N., Schulze, D., Littman, M. L. “Trigger-action programming in the wild: An analysis of 200,000 IFTTT recipes.” *In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, 2016.

[WWD+20] Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., Hu, X., “Score CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks,” *IEEE/CVPR Workshop on Fair, Data Efficient and Trusted Computer vision*, 2020.

https://openaccess.thecvf.com/content_CVPRW_2020/papers/w1/Wang_Score-CAM_Score-Weighted_Visual_Explanations_for_Convolutional_Neural_Networks_CVPRW_2020_paper.pdf

[WAD+18] Wang, J., Amos, B., Das, A., Pillai, P., Sadeh, N., Satyanarayanan, M., “Enabling Live Video Analytics with a Scalable and Privacy-Aware Framework,” *ACM Computing, Communications and Applications (TOMM)*, June 2018.

<https://www.privacyassistant.org/media/publications/tomm2018.pdf>

[WAD+17] Wang, J., Amos, B., Das, A., Pillai, P., Sadeh, N., Satyanarayanan, M., “A Scalable and Privacy –Aware IoT Service for Live Video Analytics,” *8th ACM Multimedia Systems Conference (MMSys)*, June 2017.

<https://dl.acm.org/doi/pdf/10.1145/3083187.3083192>

[WMD+20] Wang, Z., Mardziel, P., Datta, A., Fredrikson, M., “Interpreting Interpretations: Organizing Attribution Methods by Criteria,” *IEEE/CVPR Workshop on Fair, Data Efficient and Trusted Computer Vision*, 2020.

https://openaccess.thecvf.com/content_CVPRW_2020/papers/w1/Wang_Interpreting_Interpretations_Organizing_Attribution_Methods_by_Criteria_CVPRW_2020_paper.pdf

[WWR+20] Wang, Z., Wang, H., Ramkumar, S., Fredrikson, M., Mardziel, P., Datta A., “Smoothed Geometry for Robust Attribution,” *Advances in Neural Information Processing Systems*, Oct 2020
<https://arxiv.org/pdf/2006.06643.pdf>

[WAS+19] Warberg, L., Acquisti, A., Sicker, D., “Can Privacy Nudges be Tailored to Individuals’ Decision Making and Personality Traits?” *ACM Workshop on Privacy in the Electronic Society (WPES)*, November, 2019.
<https://dl.acm.org/doi/pdf/10.1145/3338498.3358656>

[ZFB+21] Zhang, S., Feng, Y., Bauer, L., Cranor, L., Das, A., Sadeh, N., “Did You Know This Camera Tracks Your Mood? Modeling People’s Privacy Expectations and Preferences in the Age of Video Analytics,” *PoPETS*, 2021.
<https://www.petsymposium.org/2021/files/papers/issue2/popets-2021-0028.pdf>

[ZFS21] Zhang, S., Feng, Y., Sadeh, N., “Facial Recognition: Understanding Concerns and Privacy Attitudes Across Increasingly Diverse Deployment Scenarios,” *USENIX Symposium on Usable Privacy and Security (SOUPS)*, Aug. 2021. To appear.

[ZFD+20a] Zhang, S., Feng, Y., Das, A., Bauer, L., Cranor, L., Sadeh, N., “Understanding People’s Privacy Attitudes Towards Video Analytics Technologies,” *Federal Trade Commission PrivacyCon*, 2020.
https://www.ftc.gov/system/files/documents/public_events/1548288/privacycon-2020-shikun_zhang.pdf

[ZSS+19] Zimmeck, S., Story, P., Smullen, D., Ravichander, A., Wang, Z., Reidenberg, J., Russell, N.C., Sadeh, N., “MAPS: Scaling Privacy Compliance Analysis to a Million Apps,” *PoPETS*. Stockholm, 2019.
<https://usableprivacy.org/static/files/popets-2019-maps.pdf>

[ZWZ+16] Zimmeck, S., Wang, J., Zou, Y., Iyengar, R., Liu, Schaub, F., Wilson, S., Sadeh, N., Bellovin, S., Reidenberg, J., “Automated Analysis of Privacy Requirements for Mobile Apps,” *AAI Fall Symposium on Privacy and Language Technologies*, Nov 2016. https://shomir.net/pdf/publications/plt_2016s.pdf

7.2 Short Selection of Conference Presentations, Tutorials, Panels and other Dissemination Events

- Acquisti gave a Keynote Talk at the IAPP Europe Data Protection Congress, November 2019, including references to our work.

- Acquisti gave the Annual Distinguished Lecture on Big Data Law and Policy at the Ohio State Moritz College of Law, November 2019, including references to our work.
- Acquisti gave a Keynote Talk at the Purdue 2050: Conference of the Future, Purdue University, 2019, including references to our work.
- Acquisti gave a Keynote Talk at the University of Bologna Business School, Marketing Effectiveness through Customer Journeys and Multichannel Management Conference, June 2019, including references to our work.
- Bauer presented our work on IFTTT at the CERT Data Science in Cybersecurity Symposium, Arlington, VA, Aug. 29, 2018 (talk title: “IoT, Privacy, and Machine Learning: (Avoiding) Some Challenges and Pitfalls”).
- Bauer presented our work on IFTTT at University College London, London, UK, Jul. 9, 2018 (talk title: “Back to the Future: From IFTTT to XSS, it's all about the information-flow lattice”).
- Cranor presented our work on PPA’s at Davos conference in late January 2016.
- Sadeh presented our work on PPA’s at FTC’s 1st Privacy Conference (Priv Con) in January 2016.
- Cranor and Sadeh were panelists at the 2018 IAPP Global Summit’s Inaugural Privacy Engineering Section Forum, which featured representatives from both government (e.g., FTC, NIST, EU EDPS, NSF), industry (Google, Microsoft, Facebook, and many more) and academia.
- Cranor gave a keynote talk at NSF CyberCorps in January 2020 on “Security and Privacy for Humans.”
- Cranor gave a closing keynote at the RSA Conference in February 2020 that focused on the human element in security, including a discussion of research on IoT security and privacy nutrition labels.
- Datta presented our work on transparency and explanations in machine learning at the March 2018 Stanford Fairness Workshop and the June 2018 NYU workshop on Economics and Data Science in Conversations about Algorithms.
- Datta presented our work on explanations and privacy for machine learning systems at the Machine Learning and Formal Methods Summit, Oxford University, July 2018, at A personal journey: Symposium in applications of contextual integrity, Princeton University, September 2018, and at Aspen Institute/EPIC round table on Disrupting the Consumer Debt Market: The Role of Technology and Innovation, September 2018.
- Sadeh participated in a Panel at Privacy + Security Forum in early May 2020 along with Achim Klabunde (advisor to the European Data Protection Supervisor) and Gabriela Zanfir-Fortuna (Future of Privacy Forum).

- Sadeh participated in a plenary panel on the first day of the 40th International Conference of Data Protection and Privacy Commissioners in Brussels on October 24, 2018.
- Sadeh gave a keynote at the 2018 Security Assured Cyberinfrastructure workshop (SAC-PA2), focusing on our IoT Privacy Infrastructure, etc.
- Sadeh and Cranor organized a high profile privacy day event at CMU, featuring Ed Felten, US Deputy CTO at the White House (<http://cups.cs.cmu.edu/privacy-day/2016/>). The event also included a panel and poster session, which featured our new DARPA Brandeis project and included a demo of a first Personalized Privacy Assistant to help users manage Android app permissions.
- Sadeh presented the project at the IAPP Global Summit conference in Washington DC on May 2, 2019.
- Sadeh and Cranor each presented results of DARPA Brandeis project at the International Association of Privacy Professionals (IAPP) Global Privacy Summit in Washington DC in early May 2019.
- Sadeh presented a SOUPS 2019 tutorial on “Contextual Integrity: From Theory to Practice” covering techniques and results of this DARPA Brandeis project
- Sadeh participated in a panel at Online Privacy+Security Forum on “IoT Privacy in the Age of CCPA and GDPR” in May 2020.
- Sadeh gave a presentation on our IoT Privacy Infrastructure to members of the Future of Privacy Forum in May 2020.
- Sadeh presented “Design of a Privacy Infrastructure for the Internet of Things” at 2020 USENIX Conference on Privacy Engineering Practice and Respect on October 16, 2020.
- Sadeh and post-docs Anupam Das, Martin Degeling and Sebastian Zimmeck gave a tutorial on results coming out of our privacy assistant research at the 2017 Symposium on Usable Privacy and Security on July 12, 2017.

7.3 URLs

[IoTAGoogle] <https://play.google.com/store/apps/details?id=io.iotprivacy.iotassistant> - IoT Assistant Mobile App in the Google Play Store (live)

[IoTaiOS] <https://apps.apple.com/us/app/iot-assistant/id1491361441#?platform=iphone> - IoT Assistant Mobile App in the Apple iOS App Store (live)

[IoTAvideo] <https://youtu.be/ar4HIY0FePc> Video about the IoT Privacy Assistant app

[IoTlabel]<https://iotsecurityprivacy.org> - Website for IoT security and privacy label

[IoTPIPEPR] <https://www.usenix.org/conference/pepr20/presentation/sadeh> - video of Norman Sadeh's presentation on the [Design of a Privacy Infrastructure for the Internet of Things](#) at 2020 USENIX Privacy Engineering Practice and Respect Conference (PEPR'20).

[IoTPIvideo] <https://youtu.be/haiuzC4kfC4> - Video about the IoT Privacy Infrastructure

[IoTPortal] <https://www.iotprivacy.io/login> - IoT Privacy Infrastructure portal (live)

[OptOutChrome] <https://chrome.google.com/webstore/detail/opt-out-easy/hikefgklfabiiecechanbafeficfojik> - Opt-Out Easy Google Chrome Extension (live)

[OptOutFireFox] <https://addons.mozilla.org/en-US/firefox/addon/opt-out-easy/> - Opt-Out Easy Firefox browser add-on (live)

[OptOutvideo] https://www.youtube.com/watch?v=_2eWrPHyGJM - Opt-Out Easy youtube video

[PPAsite] <https://www.privacyassistant.org/> - Overall Privacy Assistant project website

[PPAvideos] <https://www.privacyassistant.org/iot/> - Early project videos

[ProjectNews] <https://www.privacyassistant.org/news/>

7.4 Technical Reports

[Alm17] Almuhimedi, H., "Helping Smartphone Users Manage their Privacy through Nudges," PhD Dissertation, School of Computer Science, Institute for Software Research, CMU Technical Report CMU-ISR-17-111, December 2017. <http://reports-archive.adm.cs.cmu.edu/anon/anon/usr/ftp/usr0/ftp/isr2017/CMU-ISR-17-111.pdf>

[Liu19] Liu, B. "Can Machine Learning Help People Configure Their Mobile App Privacy Settings?" PhD Dissertation, School of Computer Science, Institute for Software Research, CMU Technical Report CMU-ISR-19-105, December 2019. <http://reports-archive.adm.cs.cmu.edu/anon/isr2019/CMU-ISR-19-105.pdf>

[ZFD+20b] Zhang, S., Feng, Y., Das, A, Bauer, L., Cranor, L., Das, A., Sadeh, N., Understanding people's privacy attitudes towards video analytics technologies. Technical Report CMU-ISR20-114, Carnegie Mellon University, School of Computer Science, December 2020. <http://reports-archive.adm.cs.cmu.edu/anon/isr2020/CMU-ISR-20-114.pdf>

7.5 Issued Patents

[SLD+21] N. Sadeh, B. Liu, A. Das, M. Degeling, F. Schaub, "**Personalized Privacy Assistant**," US Patent 10956586 issued March 23, 2021. Assigned to Carnegie Mellon University. <http://patentsgazette.uspto.gov/week12/OG/html/1484-4/US10956586-20210323.html>

8. List of Symbols, Abbreviations, and Acronyms

ACM	Association for Computing Machinery
AFRL	Air Force Research Laboratory
AI	Artificial Intelligence
API	Application Programming Interface
Brandeis	Name of the DARPA privacy research program under which this project was conducted
CCPA	California Consumer Privacy Act
CMU	Carnegie Mellon University
CPRA	California Privacy Rights Act
COPPA	Children's Online Privacy Protection Act
CRT	Collaborative Research Theme bringing together multiple Brandeis projects.
DARPA	Defense Advanced Research Projects Agency
EU	European Union
FTC	Federal Trade Commission
GBM	Gradient Boosting Machines
GDPR	General Data Privacy Regulations
GLMM	Generalized Linear Mixed Model
HCI	Human-Computer Interaction
IEEE	Institute of Electrical and Electronics Engineers
IFTTT	If-this-then-that
IoT	The Internet of Things
IoT Assistant	(or IoT Assistant): The Internet of Things Assistant
IoT PI	(or IoT Privacy Infrastructure): The Internet of Things Privacy Infrastructure
IoT Portal	(or IoT Web Portal): The Internet of Things Web Portal
IUIPC	Internet Users' Information Privacy Concerns
LSTM	Long Short Term Memory
MAPS	Mobile App Privacy System
MTurk	Mechanical Turk
NGO	non-governmental organization
NLP	Natural Language Processing
POM	Privacy Option Management
PPA	Personalized Privacy Assistant
QII	Quantitative Input Influence
QR	Quick Response
RF	Random Forests
TIPPERS	Testbed for IoT-based Privacy-Preserving PERvasive Spaces
UCI	University of California Irvine