

ANSWERING PEOPLE'S PRIVACY QUESTIONS: WEAVING TOGETHER USER-CENTERED DESIGN, NATURAL LANGUAGE PROCESSING AND LEGAL CONSIDERATIONS

Norman Sadeh (CMU), Shomir Wilson (Penn State), Tom Norton (Fordham),
and Joel Reidenberg (Fordham)

<https://usableprivacy.org>



Motivations

- ❖ In the United States, the dominant legal and public policy approach towards privacy concentrates on “Notice and Choice”.
- ❖ In practice, users seldom read privacy policies, and those who do often struggle to understand what the text really means.
- ❖ If users do not read policies or fail to understand key disclosures, how can they be assumed to have provided any sort of meaningful consent ?

Objective: privacy assistant functionality capable of answering those privacy questions that actually matter to users.

Challenges

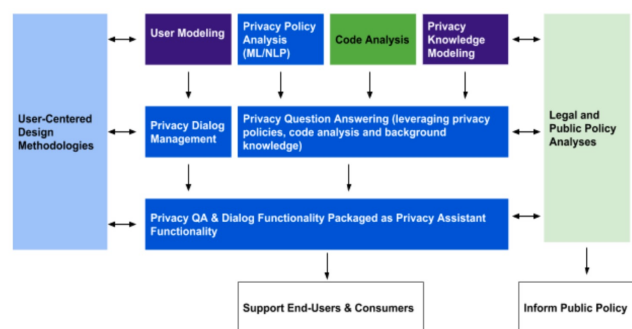
- ❖ Understanding **types of user privacy questions**;
 - ❖ What do they care about? Do their questions make sense?
- ❖ Understanding how they are likely to **phrase** their questions
 - ❖ Can they articulate their questions? Do they need help?
- ❖ Identifying to **what extent answers can be inferred** from the text of privacy policies;
- ❖ Determining **additional sources of information** that can be leveraged to answer user questions (e.g. other sources of text but also possibly code in the case of mobile apps)
- ❖ Identifying response formats that are **useful to the user but also satisfactory from a legal perspective**

Solution

- ❖ Collect large corpus of privacy questions (crowdsourcing), analyze these questions – including interviews with subjects but also legal analysis.
- ❖ Study to what extent the text of privacy policies is sufficient to answer questions from participants.
- ❖ Study to what extent lack of awareness or limited privacy literacy leads users to not ask or think about certain questions or issues or to struggle in articulating questions
- ❖ Explore different formats for interaction with users to address ambiguous questions, lack of knowledge, usability and legal considerations

Scientific Content

1. Privacy assistant functionality to automatically answer a rich collection of privacy questions.
 - Disambiguating questions
 - Dialogues that combine partial answers with requests for further clarification
2. Large corpus of privacy questions to be released to the research community.
3. Evaluation of privacy question answering functionality including evaluation with visually impaired users.
4. Public release of a privacy question answering prototype for mobile apps.
5. Detailed analysis of people's privacy questions.



Broader Impact: Society

- ❖ Release of a privacy assistant capable of answering privacy questions people have about mobile apps.
- ❖ App will add even more value to, and will also be evaluated with visually impaired users.
- ❖ Inform public policy discussions (e.g., to what extent the text of privacy policies is sufficient to answer those questions people care about).

Broader Impact: Education

- ❖ The project will directly train graduate students, including law students, undergrad students and post-doctoral fellows
- ❖ Will produce material used by the PIs in their courses and provided the basis for a number of class projects, effectively contributing to the training of many additional students

Conclusion

- ❖ “Privacy Notice and Choice” is broken.
- ❖ This state of affairs threatens our most basic right to operate as autonomous beings
- ❖ This work aims to improve this situation through fundamental research that spans machine learning, natural language processing, code analysis, human-computer interaction, user-centered design, artificial intelligence, law and public policy.

