CMU/PwC DTIC - Privacy Engineering Capstone Project [Fall 2023]

Privacy Threat Modeling Framework for 'User Notice & Choice'

Team members: Seo Young Ko Geetika Gopi Alexandra Li Ziping Song Yanhao Qiu

Capstone Sponsor: PwC

For the capstone project, as a team, we worked on drafting a privacy threat modeling framework for User Notice and Choice. The framework is designed to be practical and user-oriented, addressing the limitations of existing high-level guidance. It builds on the concept of Privacy-by-Design and aims to provide a systematic approach to identifying and mitigating risks associated with user-oriented privacy notice and choice, particularly in the context of AI. This is a critical need in today's rapidly evolving technological landscape, where new technologies like AI are outpacing the ability of regulation to address privacy risks directly. The framework aims to move beyond a compliance-by-design approach to a more proactive and practical approach to privacy.

Our framework includes four steps. Here is a layout of the four steps.

1. **Data flow diagram(DFD) creation:** The user of this framework will first create a data flow diagram with regard to the targeted system. Organizations should examine the interactions that take place in the system for each data flow, such as data collection/usage purposes and inherent risks.
2. **Notice and Choice Requirements:** The second step requires organizations to conduct compliance analysis and identification of required notifications and choices for the data flow. They should also incorporate existing organizational policies into this analysis to see if any additional notices and choices are needed in order to meet the demands of said policies. Contextual integrity analysis in the form of user studies may also come into place for a better understanding of stakeholders' expectations.
3. **DFD Consolidation:** The third step includes the consolidation of data flows into a limited set of 'touch points'', ideally in the context of "user journeys."
4. **Usability threat identification:** The last step will help take care of potential usability issues, such as awareness, language/content-related, delivery, and presentation design that may hinder the effectiveness of privacy notice and choice.

While we have achieved a significant milestone in this capstone project, we acknowledge that this marks only the initial phase of a long-term endeavor. We anticipate expanding and enhancing this project in the following ways.

1. **AI threats:** This work provides partial coverage of threats related to user notice and choice in the context of artificial intelligence and machine learning solutions. In our future

research, we aim to extend the coverage of AI threats to offer a more comprehensive understanding.

2. **Mitigation Strategies:** Mitigation strategies for usability threats will be incorporated in future iterations of this work. This enhancement will make the framework more comprehensive, enabling not only the identification of threats but also helping users to implement effective mitigation strategies.

3. **Requirement identification using 3 levels:** The current framework only introduces the initial idea of three-level consideration factors to decide whether each system interaction requires notice or choice. Eventually, we need to further expand this framework to ask specific/accurate questions to verify whether the system interaction satisfies regulatory compliance, organization policies, and contextual needs. Currently, we do not have a systematic approach in terms of determining whether notice and choice should be required for specific data flows according to the 3 levels. For instance, how should we go from GDPR to deciding if a privacy notice and choice is needed for a data flow that happens at the order placement stage? What does satisfying users' contextual needs entail about requiring a notice and choice? We hope to dig deeper into these questions in future research.

4. **Inherent risk calculation:** Our framework introduces the inherent risk associated with each system interaction, reflecting the sensitivity of individual data flows. While we propose several factors for consideration in eliciting this inherent risk, such as data type and the purpose of data processing, we anticipate developing a more systematic and objective method for calculating the inherent risk in the future.

5. **Taxonomy of touchpoints:** This work demonstrates the methodology for eliciting touchpoints in each system interaction and aggregating diverse data flows through common touchpoints. Our next step involves expanding on this concept by creating a taxonomy of touchpoints, and providing a reference framework for individuals when eliciting touchpoints.

6. **Validation:** This work introduces the preliminary concept of our threat modeling framework for user notice and choice. We aim to validate our framework through a comprehensive user study, simulating various scenarios such as IoT and AI systems. This validation process will be beneficial in refining and enhancing our work to ensure an optimal user experience and comprehensiveness.