

This paper proposes a groundbreaking method to inspect crowds' privacy concerns in data practice. This method, called Lean Privacy Review (LPR), can make data practitioners collect privacy concerns (also their range and magnitude) from crowd workers in a quick, cheap, and easy-to-access way. It can serve as a low-cost and complementary method to ordinary privacy reviews, widely used in the industry to identify potential customer acceptance issues with their data usage.

The specific contributions of this research can be summarized as follows:

1. LPR is the first method that inspects privacy concerns of data practices using people without special knowledge. Existing solutions to review privacy rely on data privacy experts or practitioners, which requires a lot of time and cost to find a result. For example, to study users' privacy concerns through FVS, experimenters have to carefully design the controlled variables, collect many responses, and analyze the results to illustrate statistical significance. To solve this problem, LPR postulates that non-tech-savvy participants can contribute more than just answering multiple-choice questions. But there lies a potential problem: these participants may lack the motivation and expertise to engage in complex tasks and provide high-quality feedback. To address this challenge, LPR offers a specific workflow. It breaks the single privacy review task into two parts: expressing free-text privacy opinions and annotating free-text responses using a taxonomy of privacy concerns, inspired by the Find-Fix-Verify pattern (which uses crowd workers to help with document shortening and other text tasks).
2. LPR proposed a new storytelling technique - privacy storyboarding. Privacy storyboarding describes data practice in a way that is easy to understand for non-specialists and easy to generate for data practitioners. It also allows practitioners to iterate and modify the statement describing relevant data practice. LPR decomposes a data story into several data action units (each data action involves at least one data stakeholder), combines these units into data applications, and finally synthesizes all data applications into data practices in a tree structure. This method succinctly and expressively describes the data flow, making it easy for crowd workers to understand the case and think about privacy concerns in it.

3. LPR is inexpensive, fast, consistent, and can provide high-quality privacy review results. This paper introduced 12 real-world data usage cases into data ethics discussion and evaluated LPR using these practices with 240 crowd users and 24 data practitioners.

We can have much to do in future work. First, the current study focuses on the feasibility and benefits of crowd inspection, but we could explore the iterative process of data practice designs. Second, we may conduct privacy training through LPR for practitioners. Third, future work could research how to help novice LPR users create storyboards better. Fourth, future work would benefit from building a taxonomy for positive opinions and building interfaces to help users contrast positive views with negative ones on the same topics. The last one is creating new privacy indexes, which I think has a broad prospect.

I have some thoughts about future work. The first question is, in terms of users, I wonder if we can use LPR to design better privacy notices so that users can read them patiently and authorize their rights (this idea is from How Internet Users' Privacy Concerns Have Evolved Since 2002). Secondly, I think different ranges and magnitude of privacy concerns found by LPR could be used as the review standard of ethics boards. Third, building an LPR privacy index requires a large amount of data, I wonder if we can find a way to use the strength of the masses, like human-computer identity authentication. The LPR privacy index can be a reference for making a storyboard. In addition, we can build different privacy indexes for different groups to solve the target population problem.