Privacy Through Design

A Design Methodology to Promote the Creation of Privacy-Conscious Consumer AI

Sauvik Das Georgia Institute of Technology Principal Investigator

Privacy through Design (PtD) is a novel research methodology to help creators of consumer-facing AI technologies: (i) model how acute, use-case specific privacy concerns among end-users among stakeholders trade-off against the envisioned utility or value of proposed AI concepts; and, (ii) understand how to (re-)design those concepts in a manner that respects the privacy-concerns of stakeholder groups while retaining the envisioned value or utility of their concept. To that end, this proposal makes three concrete contributions: (1) develop a taxonomy of algorithmic privacy intrusions to operationalize the unique privacy harms entailed by consumer AI and map those harms onto the unique capabilities and requirements of AI; (2) develop PtD using an iterative methodology incorporating experts and practitioners in industry and academia; and (3) comparatively evaluate of PtD and Google's PAIR guidebook with student teams to determine if and how PtD helps creators reduce privacy risk. This research will help bridge the principle-practice gap for privacy in human-centered AI (HAI).

The research is currently in the phases of **developing the taxonomy of algorithmic privacy intrusions**, and **understanding AI practitioners on privacy considerations** via interviews and surveys to investigate how designers and developers approach end-user privacy when creating AI technology.

Understanding AI Practitioners on Privacy Considerations

in preparation for CHI'23

We are conducting semi-structured interviews and planning to conduct an online survey study with experts in academia and industry who design and create consumer-facing AI products (i.e., systems that are trained on data from and/or make inferences about human users) to understand how preventing and/or mitigating potential end-user privacy infringements manifest in their existing design processes. The semi-structured interviews and online survey study are guided by the three research questions:

- *RQ1: What does privacy mean to AI practitioners?*
- *RQ2: How are AI privacy issues identified by AI practitioners?*
- *RQ3: What are the current processes for AI practitioners to improve privacy in their products?*

Specifically, we discuss topics with AI practitioners about 1) the meanings of "end-user privacy" for their products, 2) how AI is considered to be potentially compromising end-user privacy, 3) their motivations to consider end-user privacy in their products, 4) methods and tools used to ensure privacy in their product, 5) how privacy is discussed and communicated within the team, and 6) challenges when considering privacy during the development and design processes.

So far, we have conducted the interview study with ten AI practitioners working on different types of AI products and roles. While we have not yet conducted a formal data analysis, some preliminary observations emerge from our interview, such as how the discussion on end-user privacy is incorporated into practitioners' design processes. For example, some participants share a button-up approach where active privacy documents of the product are maintained, and end-user privacy discussion is a part of the routine in which other team members will review such privacy documents; on the other hand, some participants shared a top-down approach, in which end-user privacy is discussed and defined before the developing phase, and practitioners treated them as specifications of products to comply with.

We also see an emerging theme of the challenges in end-user privacy considerations in our interview data. For example, our participants shared that considerations about end-user privacy largely rely on practitioners' own self-awareness and knowledge about privacy. However, practitioners' levels of privacy literacy vary, and associated training in workplaces is normally deemed to be overly generic and unuseful for their products. Also, end-user privacy is often treated as compliance and/or afterthought in the product cycle, which could be seen as a factor to sabotage creativity and/or to slow down the development.

We aim to recruit 20-30 partitions for the interview study. Then, we will examine an online survey to validate and provide quantitative supplements to our findings found in the interview. The tentative plan for the submission of this work is the ACM CHI Conference on Human Factors in Computing Systems (CHI'23).

PID	APPLICATION DOMAIN	AI TECHNOLOGY	ROLE
P1	Process Mining	Predictive Analytics	Software Engineer
P2	Media and Entertainment	Search / Information Retrieval, Speech and Voice	Software Engineer
P3	Retail	Computer Vision / Image Analysis, Recommender Systems	Data Scientist, Software Engineer
P4	Holliday Rental Meta Search	Recommender Systems	Designer, Researcher
P5	Marketing	Predictive Analytics, Recommender Systems	Software Engineer
P6	Enterprise software	Conversational AI / Chatbots	PM, Researcher, Tech lead/manager
P7	Defense / Military, Education, General-purpose Machine Learning Tools (e.g., APIs), Healthcare, Government / Public Sector	Decision support, Predictive Analytics, Recommender Systems, Search / Information Retrieval	Designer
P8	General-purpose Machine Learning Tools (e.g., APIs), Financial Services: Lending / Mortgage, Financial Services: Other, Retail	Decision Suppor, Natural Language Processing, Predictive Analytics	Technical Lead / Manager
P9	Healthcare, Hiring / Recruiting, Media and Entertainment, Financial Services: Lending / Mortgage, Financial Services: Other, Public Transportation	Conversational AI / Chatbot, Natural Language Processing, Speech and Voice	Designer, Domain / Content Expert
P10	General-purpose Machine Learning Tools (e.g., APIs)	Computer Vision / Image Analysis, Recommender Systems	Software Engineer

Developing the Taxonomy of Algorithmic Privacy Intrusions

in preparation for CHI'23

We are constructing a taxonomy of privacy infringements and harms entailed by consumer AI technologies. In constructing such a taxonomy, our goal is to formally codify patterns of privacy infringement that we see across a corpus of examples and relate those infringements to the unique capabilities and requirements of AI.

We are working in a bottom-up fashion, using an iterative process of gathering examples, identifying common characteristics, defining meta-characteristics, and grouping the characteristics into dimensions. Specifically, we select criteria for identifying examples of algorithmic privacy infringements — e.g., attacks in which personal data is inferred from people's photos, the NYTimes expose on ClearView.ai. We are also curating more examples, including press articles mentioning data privacy and ethics and existing privacy consideration tools (e.g., Google's People+AI Guidebook), and will be performing the content analysis on those materials to form the said taxonomy. In the current formative phase of the taxonomy, we see emerging dimensions to potentially model our examples in terms of 1) where the data that the AI system trained on is collected, 2) who is using the AI systems, 3) who is benefiting from the AI systems.