PRIVY: OPERATIONALIZING PRIVACY POLICY WITH AI-ASSISTED PRIVACY IMPACT ASSESSMENT WORKFLOW

Hao-Ping (Hank) Lee[†], *Yu-Ju Yang*[‡], *Matthew Bilik*[§], *Isadora Krsek*[†], *Thomas Serban von Davier*[†], *Kyzyl Monteiro*[†], *Jason Lin*[†], *Shivani Agarwal*[†], *Jodi Forlizzi*[†], *Sauvik Das*[†]

Carnegie Mellon University[†], University of Illinois Urbana-Champaign[‡], University of Washington[§]

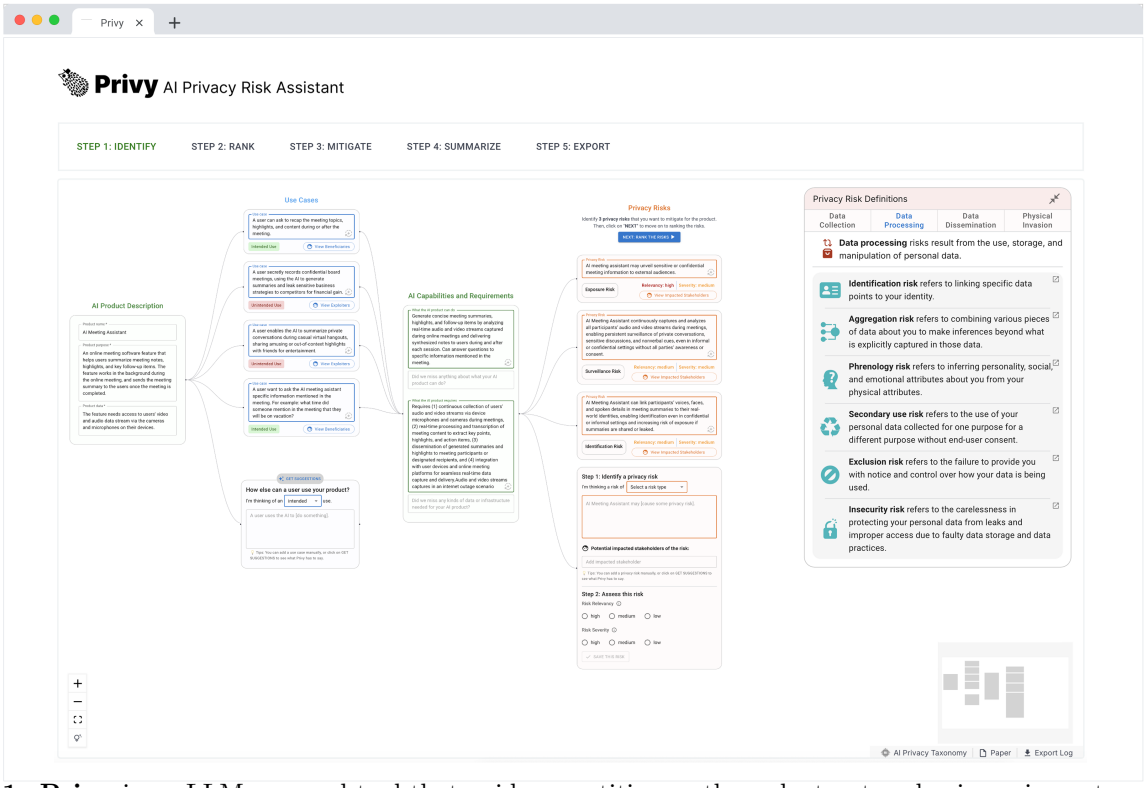

Figure 1. Privy is an LLM-powered tool that guides practitioners through structured privacy impact assessments to: (i) identify relevant risks in novel AI product concepts, and (ii) propose appropriate mitigations.

AI is reshaping the privacy landscape faster than public policy and organizational practice can adapt. Across government and industry, AI products increasingly rely on large-scale personal data, derive sensitive inferences from behavioral traces, and generate content that can be repurposed across contexts [4]. These capabilities challenge longstanding policy assumptions about notice-and-consent, purpose limitation, and data minimization: harms can arise even when teams do not intentionally collect new data, because models can infer it; risks can propagate downstream through model outputs; and privacy failures are often difficult to detect until systems are deployed.

Yet the common policy mechanism for organizations — the privacy impact assessment (PIA) — has not kept pace. PIAs are now required or strongly encouraged in AI product development, but existing frameworks and standards were largely designed for pre-AI systems (e.g., ISO's Information Technology Security Techniques[1]), and producing a high-quality PIA remains labor-intensive and dependent on scarce privacy expertise [1, 5, 3]. Organizations therefore face a policy-level "principle-practice gap": policymakers can mandate privacy assessments and principles, but product teams struggle to operationalize them early enough to meaningfully influence design decisions [7, 6].

In foundational work, we introduced a taxonomy of privacy risks that emerge from modern AI systems [4]. In this talk, we extend that agenda by presenting Privy (Fig. 1), a privacy risk-envisioning tool that translates AI privacy risk concepts into product-specific PIA reports during early-stage product development. Our goal is to build critical infrastructure for privacy policy implementation: a practical workflow that (i) makes AI-specific privacy risks legible

[1] `https://www.iso.org/standard/86012.html`

to non-privacy-experts while design choices are still flexible, (ii) supports the creation of mitigation plans that can be evaluated and negotiated, and (iii) produces structured documentation that can travel across organizational and policy interfaces (e.g., between product teams, privacy counsel, external auditors, and regulators).

In short, PRIVY guides practitioners to produce high-quality PIA reports through an AI-assisted workflow. First, the tool helps practitioners articulate the capabilities and requirements of their AI product concepts grounded in their envisioned use cases. Second, it supports risk identification by mapping product specifics to our AI privacy risk taxonomy, prompting teams to consider risks such as sensitive inference, unintended disclosure through generated outputs, covert data collection, and cross-context data repurposing [4]. Third, it foregrounds risk mitigation by facilitating user-centered privacy design iterations — an element central to end-user privacy empowerment but often absent from early design discussions [2]. PRIVY uses large language models (LLMs) to assist with risk envisioning and mitigation brainstorming, while intentionally preserving practitioner control through thoughtful friction: users steer system outputs and generate PIA reports tailored to their privacy work. The tool is designed to empower non-privacy-experts rather than replace expert privacy review.

We evaluated PRIVY with 24 AI practitioners, whose resulting PIA reports were assessed by 13 privacy experts. Our findings show that PRIVY effectively scaffolds practitioners in creating high-quality PIAs. Using PRIVY, practitioners identified privacy risks that experts judged to be relevant and severe, and they developed mitigation strategies that were both effective and appropriate. Participants also found PRIVY to be useful and usable, and the tool helped overcome long-standing barriers to privacy practice by improving awareness (through structured risk identification), fostering motivation (by encouraging reflection and engagement), and supporting ability (by increasing self-efficacy in privacy work).

We conclude by discussing PRIVY's implications for public privacy policy. First, AI-assisted PIAs can serve as a practical bridge between abstract regulatory requirements and concrete product decisions, enabling earlier and more reviewable evidence of "privacy *by* and *through* design," rather than post hoc documentation. Second, tools like PRIVY, which empower non-privacy-experts to examine the privacy risks of novel AI product concepts, can reduce informational asymmetries between vendors, deployers, and impacted stakeholders — supporting more informed product decisions and negotiations around data use, retention, and downstream sharing. Third, because PRIVY produces structured, taxonomy-aligned descriptions of risks and mitigations, its outputs can be aggregated across products and organizations to surface recurring patterns of risk and mitigation. Such evidence can help policymakers calibrate requirements, identify gaps in existing standards, and target guidance or enforcement where it is most needed. PRIVY demonstrates how privacy-enhancing tooling for practitioners can operationalize emerging AI privacy risk frameworks and strengthen the feedback loop between policy objectives and product development practice.

References

[1] Roger Clarke. "Privacy impact assessment: Its origins and development". In: *Computer Law & Security Review* 25.2 (Jan. 2009), pp. 123–135.

[2] Sauvik Das et al. "The Security & Privacy Acceptance Framework (SPAF)". In: *Foundations and Trends® in Privacy and Security* 5.1-2 (Dec. 2022). Publisher: Now Publishers, Inc., pp. 1–143.

[3] Hao-Ping (Hank) Lee et al. ""I Don't Know If We're Doing Good. I Don't Know If We're Doing Bad": Investigating How Practitioners Scope, Motivate, and Conduct Privacy Work When Developing AI Products". In: *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association, Aug. 2024, pp. 4873–4890.

[4] Hao-Ping (Hank) Lee et al. "Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks". In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. CHI '24. Association for Computing Machinery, May 2024, pp. 1–19.

[5] Jeroen van Puijenbroek and Jaap-Henk Hoepman. "Privacy Impact Assessment in Practice". en. In: *International Workshop on Privacy Engineering: Proceedings of the 3rd International Workshop on Privacy Engineering, co-located with 38th IEEE Symposium on Security and Privacy (S&P 2017)*. May 2017.

[6] Ben Shneiderman. "Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems". In: *ACM Transactions on Interactive Intelligent Systems* 10.4 (2020), pp. 1–31.

[7] Alan FT Winfield and Marina Jirotka. "Ethical governance is essential to building trust in robotics and artificial intelligence systems". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2133 (2018), p. 20180085.