A Delicate Balance: The Promise of AI-driven Personalized Privacy Interventions

KATHARINA BARLAGE, LMU Munich, Germany

FLORIAN ALT, LMU Munich, Germany

Consent and control are central to usable privacy research, yet one-size-fits-all mechanisms have failed to resolve the privacy paradox in everyday technology use. Recent advances in AI enable highly personalized privacy nudges that may better align users' stated privacy attitudes with their actual behavior. At the same time, personalization introduces critical challenges, including blurred boundaries between supportive guidance and manipulative design, as well as the risk that privacy interventions themselves become privacy threats by design. We argue that contemporary AI systems allow for more complex forms of personalization by learning higher-level user patterns without explicit intermediate classifications, creating tensions between technical potential and methodological rigor in HCI research. We further discuss the use of large language models for adaptive privacy communication and the detection of manipulative design practices. We conclude that meaningful design and evaluation of AI-based privacy interventions require interdisciplinary dialogue across HCI, social sciences, ethics, and law.

CCS Concepts: • **Security and privacy → Privacy protections**; **Usability in security and privacy**.

Additional Key Words and Phrases: HCI, Consent, User Control, Privacy, Interventions, AI, LLMs

ACM Reference Format:

Katharina Barlage and Florian Alt. 2026. A Delicate Balance: The Promise of AI-driven Personalized Privacy Interventions. 1, 1 (February 2026), 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Consent and control are central mechanisms in privacy research and practice. Yet in everyday contexts, users often struggle to understand privacy policies and adopt privacy-protective behaviors, despite extensive research on the potential and limitations of these mechanisms [5]. The growing adoption of AI tools further complicates this landscape: while AI systems can threaten user privacy and control, they also enable new forms of consent and control, such as personalized nudges and interventions—though these may verge on manipulation.

Research on personalized privacy interventions exposes several gaps. Warberg et al. [7] found no significant effects of personality-based nudges tailored to Big Five traits, despite large-scale experiments. This contrasts with earlier findings suggesting associations between personality and privacy behavior; for example, Halevi et al. [3] report that individuals high in openness share more personal information and use less conservative Facebook privacy settings.

More promising approaches move beyond stable personality traits. Jackson and Wang [4] demonstrate that behavior- and attitude-based personalization, highlighting discrepancies between users' stated privacy attitudes and actual behavior, can yield more effective privacy interventions in mobile contexts. Similarly, Egelman and Peer [2] show that

Authors' Contact Information: Katharina Barlage, LMU Munich, Munich, Germany, katharina.barlage@ifi.lmu.de; Florian Alt, LMU Munich, Munich, Germany.

Manuscript submitted to ACM

tailored privacy mitigations grounded in cognitive and decision-making characteristics may outperform personality-based approaches. Supporting this shift, Amon et al. [1] find that users with Dark Triad traits exhibit heightened awareness of interdependent privacy while simultaneously engaging in increased sharing of others' photos.

2 Problem statement and core position

Traditional consent and control mechanisms, when applied in a one-size-fits-all manner, reach their limits in terms of user acceptance and adoption in everyday contexts[4]. As privacy researchers, we must navigate the tension between users' stated privacy attitudes and their often contradictory privacy-related behaviors. Users frequently act in ways that conflict with their expressed preferences, complicating the design of effective privacy interventions.

Advances in AI technologies enable new forms of nudging and manipulation, such as personalized privacy interventions that take personal privacy attitudes and prior knowledge into account. This enables, for example, privacy policies to be communicated in a way that reflects users' preferences and knowledge levels. However, the use of AI in this context introduces two key challenges. First, there is a fine, if not nonexistent, line between supportive guidance and manipulative design; in practice, all interventions, including those intended to be "supportive," exert some degree of influence. Second, the research community must ensure that AI-driven personalized interventions do not themselves become privacy threats by design, thereby undermining the very goals they aim to support. This can be achieved through techniques such as federated learning, differential privacy, on-device personalization, parameter-efficient fine-tuning (PEFT), contextual personalization using retrieval-augmented generation (RAG), and secure aggregation and encryption techniques, including secure multiparty computation (SMPC) and homomorphic encryption.

AI systems introduce new forms of personalized nudging that can account for more complex personalization factors than non-AI-based approaches (e.g., [1–4, 7]), enabling highly individualized interventions. These interventions could combine multiple personalization factors, such as personality traits, privacy attitudes, context, and prior behavior, to generate meaningful summaries of privacy policies, whereas traditional approaches had only limited means to translate personalization factors into user-facing communication. We argue that the primary distinction between supportive guidance and manipulative design lies in the underlying intention. However, as computer science and HCI researchers, it is not our role to determine where this normative boundary should be drawn. Instead, our responsibility lies in designing and evaluating interaction modalities and interfaces, while the ethical, social, and political distinctions must be negotiated collaboratively in multidisciplinary teams.

Our research focuses on personalized privacy interventions based on diverse personalization factors, designed for different interface types and use cases. In particular, we see strong potential in using large language models (LLMs) to adapt privacy policy explanations to users' prior knowledge, contextual situations, behaviors, and privacy attitudes. The resulting tools and models are intended to remain neutral in their technical design, without embedding an explicit intention to manipulate or guide users. Formats such as the CHI '26 Beyond Clicks Workshop are therefore essential for situating our findings within a broader context and for critically reflecting on their implications for users and society.

Compared to traditional personalization approaches, which rely on intermediate classification steps (e.g., inferring personality traits[1–3], privacy attitudes[4, 6], or behavioral profiles), contemporary AI systems show potential to bypass such explicit categorizations and instead learn higher-level user patterns directly. While this may increase the effectiveness of privacy interventions, it also poses a challenge for the HCI community: robust evaluation requires controlled experimental settings to ensure meaningful and interpretable results. This creates a tension between rapidly advancing technical capabilities and the methodological rigor expected in HCI research.

Beyond personalization, we also see potential for AI, and LLMs in particular, in detecting manipulative design practices. AI-enhanced tools could support the identification of dark patterns that actively undermine user consent and control in digital systems, thereby contributing to more transparent and accountable interface design.

To date, conventional consent and control mechanisms have failed to resolve the privacy paradox in real-world settings. We therefore argue for the exploration of AI-enhanced approaches that enable highly personalized nudges and privacy communication. By increasing the accessibility of information and control mechanisms, users gain greater agency and are better able to provide informed consent, which can lead to behavioral changes that better align with their privacy attitudes. We identify an emerging research space at the intersection of HCI and computer science, where usable consent and control mechanisms are designed alongside privacy-preserving system architectures and privacy-aware AI to enable personalized privacy interventions for informed consent without introducing new privacy risks.

3 Contribution to the workshop

By combining HCI research on personalized privacy interventions with a background in applied cryptography, we can contribute meaningful insights to the workshop from both HCI and technical perspectives. Our background does not allow us to decide whether a nudge is friendly guidance or a manipulation; we need to start a dialogue with professionals from other disciplines. We can provide insights into technical solutions for novel consent and control mechanisms and have a good understanding of how humans interact with these, but we need valuable insights from social scientists, ethicists, and policymakers to create human-centered, ethical consent and control mechanisms in the age of AI.

4 The Author

Katharina Barlage is a first-year PhD student at LMU Munich in Munich, Germany. Her research focuses on personalized privacy and security interventions enhanced by large language models (LLMs). During her Master's degree in computer science at HU Berlin, Germany, she worked in a smart cities research group on privacy-preserving technologies, with a focus on homomorphic encryption. Bridging user-centered privacy research and privacy-preserving system design, she explores how to develop human-centered personalized privacy solutions that are not inherently privacy-invasive.

References

[1] Mary Jean Amon, Aaron Necaise, Nika Kartvelishvili, Aneka Williams, Yan Solihin, and Apu Kapadia. 2023. Modeling User Characteristics Associated with Interdependent Privacy Perceptions on Social Media. *ACM Trans. Comput.-Hum. Interact.* 30, 3 (June 2023), 40:1–40:32. doi:10.1145/3577014

[2] Serge Egelman and Eyal Peer. 2015. The Myth of the Average User: Improving Privacy and Security Systems through Individualization. In *Proceedings of the 2015 New Security Paradigms Workshop (NSPW '15)*. Association for Computing Machinery, New York, NY, USA, 16–28. doi:10.1145/2841113.2841115

[3] Tzipora Halevi, James Lewis, and Nasir Memon. 2013. A pilot study of cyber security and privacy related behavior and personality traits. In *Proceedings of the 22nd International Conference on World Wide Web (WWW '13 Companion)*. Association for Computing Machinery, New York, NY, USA, 737–744. doi:10.1145/2487788.2488034

[4] Corey Brian Jackson and Yang Wang. 2018. Addressing The Privacy Paradox through Personalized Privacy Notifications. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2 (July 2018), 68:1–68:25. doi:10.1145/3214271

[5] Valentin Schwind, Netsanet Zelalem Tadesse, Estefania Silva da Cunha, Yeganeh Hamidi, Soltan Sanjar Sultani, and Jessica Sehrt. 2025. A Scoping Review of Informed Consent Practices in Human–Computer Interaction Research. *ACM Trans. Comput.-Hum. Interact.* 32, 4 (Aug. 2025), 35:1–35:60. doi:10.1145/3721284

[6] Hervais Simo and Michael Kreutzer. 2022. Towards Automated Detection and Prevention of Regrettable (Self-) Disclosures on Social Media. In *2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. 638–645. doi:10.1109/TrustCom56396.2022.00092 ISSN: 2324-9013.

[7] Logan Warberg, Alessandro Acquisti, and Douglas Sicker. 2019. Can Privacy Nudges be Tailored to Individuals' Decision Making and Personality Traits?. In *Proceedings of the 18th ACM Workshop on Privacy in the Electronic Society (WPES'19)*. Association for Computing Machinery, New York, NY, USA, 175–197. doi:10.1145/3338498.3358656