Consentful AI: Designing Consent Across Data, Models, and Human-AI Interaction

Jasmeet Kaur
CISPA - Helmholtz Center for Information Security
Saarbrücken, Germany

Sheza Naveed
University of Michigan
Michigan, USA

Vikram Mohanty
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Gabriel Grill
Interdisciplinary Transformation University
Linz, Upper Austria, Austria

Hana Habib
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Kentaro Toyama
University of Michigan
Michigan, USA

Lorrie Faith Cranor
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Jane Im
CISPA - Helmholtz Center for Information Security
Saarbrücken, Germany

Abstract

Generative AI systems raise pressing questions about consent. Many are built on data whose creators were never asked for permission. Consent is also often missing in human-AI interactions—generative AI rarely asks users how shared information should be remembered or used. The large-scale, ever-changing nature of these systems makes conventional notions of consent hard to apply, and the rise of agentic AI raises the stakes for how systems interpret users' consent. How, then, should we design consent mechanisms for both training and interacting with generative AI systems? This workshop investigates consentful AI as a human-centered approach to AI design, development, and governance. We connect consent for data used to develop AI with consent during human-AI interaction, treating consent as an ongoing, contextual relationship. Our aim is to build a community around consentful AI, and participants will develop design patterns and a research agenda for making consent meaningful in human-centered AI.

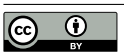
CCS Concepts

• **Security and privacy → Human and societal aspects of security and privacy**; • **Human-centered computing → Human computer interaction (HCI)**.

Keywords

Consentful AI, Human-AI Interaction, Generative AI, AI Agents, Training Data, Data Governance, User Agency

ACM Reference Format:
Jasmeet Kaur, Sheza Naveed, Vikram Mohanty, Gabriel Grill, Hana Habib, Kentaro Toyama, Lorrie Faith Cranor, and Jane Im. 2027. Consentful AI: Designing Consent Across Data, Models, and Human-AI Interaction. In *Proceedings of the Extended Abstracts of the 2027 CHI Conference on Human Factors in Computing Systems (CHI EA '27), May 10–14, 2027, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3856832.3871955



This work is licensed under a Creative Commons Attribution 4.0 International License.
CHI EA '27, Pittsburgh, PA, USA
© 2027 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-3073-3/2027/05
https://doi.org/10.1145/3856832.3871955

1 Motivation

Meaningful consent is difficult to obtain throughout the current generative AI lifecycle. A single system may depend on the contributions of millions of people - authors, artists, programmers, platform users, and others - often without those individuals having been asked whether their data or work could be used for AI development. Much of this data was created in different contexts and under different expectations about how they would be used [12]. Dataset documentation research has long called for greater transparency around collection processes, intended uses, distribution, and consent [2, 4], yet recent audits of AI datasets continue to identify substantial gaps in licensing and provenance information [14]. At the same time, data providers are increasingly attempting to communicate restrictions on AI use, but these preferences may be expressed inconsistently across mechanisms such as terms of service and machine-readable crawling protocols [15]. Together, these gaps raise a fundamental question for AI development – how can people meaningfully express whether, and under what conditions, their data may contribute to AI systems?

Consent challenges do not end once a model is trained. In interactions with deployed systems, users may disclose information, delegate tasks, connect external services, or allow systems to remember context and act on their behalf. Human–AI interaction research has emphasized communicating system capabilities, preserving user control, and adapting interactions over time [1]. Generative and increasingly agentic AI extends these concerns by introducing uncertainty not only about what a system can do, but what it has permission to do. As AI systems become capable of taking actions across applications and on behalf of users, consent increasingly becomes an interactional question - when should a system ask, what exactly should permission cover, and when should that permission expire or be renegotiated? These questions also raise

issues of enforcement and oversight, particularly, who should enforce consent-related rules and ensure that companies comply with them - companies themselves, independent regulators, or other institutions?

Overall, these developments raise consent questions at two ends of the AI lifecycle, which we propose to bridge in this workshop - **(1) What would meaningful consent look like for data owners whose data or work may be used to develop AI systems, and how could we design mechanisms to support it?** Publishing a photograph, writing a blog post, contributing source code, or participating in an online community does not necessarily mean that the owner anticipated the artifact becoming training data for a generative model. The technical availability of data and meaningful permission for a particular use are distinct questions; **(2) What would meaningful consent look like for users whose data an AI system may use in taking actions, and how could we design mechanisms to support it?** Conversational and agentic AI systems may remember information, combine information across contexts, infer information that users never explicitly disclosed, invoke external tools, communicate with third parties, or take actions with consequences outside the conversation itself.

HCI has extensive experience demonstrating that merely presenting people with choices does not guarantee meaningful consent [3, 17–19]. Effective privacy choices need to be relevant, understandable, actionable, and situated within the contexts in which decisions are made; they may also need to be revisited as practices continue over time [18]. Research on web consent interfaces further shows that seemingly small design decisions can substantially influence whether users accept data collection [16]. These findings caution against simply carrying existing notice-and-consent mechanisms into AI systems, where data practices, system capabilities, and user relationships with systems may evolve over time. We therefore propose ***consentful AI*** as an emerging research area concerned with how AI systems can respect, solicit, negotiate, and operationalize human consent throughout their lifecycle.

This workshop aims to build a community around consentful AI. Through interactive sessions, participants will map out research questions and themes around when consent is necessary, what makes it meaningful, whose consent matters, how it should be represented computationally, when it should be renewed, how refusal or withdrawal should work, and where consent itself may be an insufficient mechanism for governing AI.

2 Topics and Themes

The workshop is open to diverse interpretations of consent across generative AI, AI agents, and datasets. We invite completed works, works-in-progress, and provocations. Our discussions will be organized around questions including but not limited to:

- **Consent in Human-AI Interaction:** When and how should AI systems seek permission to act, remember, infer, or use information?
- **Consent and Agentic Action:** When AI agents autonomously pursue goals and take sequences of actions, what does meaningful user consent look like?
- **Consent for AI Training and Data:** What constitutes meaningful consent for collecting, using, transforming, and withdrawing data used to train AI?
- **Designing and Evaluating Consent:** How can consent mechanisms be informed, understandable, revocable, and effective without enabling dark patterns?
- **Beyond Consent:** Where are the limits of consent, and when are broader forms of governance, accountability, or collective decision-making needed?

3 Organizers

The organizers are an interdisciplinary group of researchers working on HCI, privacy, AI, and human-AI systems.

Jasmeet Kaur is a postdoctoral researcher at the CISPA Helmholtz Center for Information Security. Her research focuses on designing human-centered systems that support consent, privacy, and agency in AI. She is particularly interested in centering the consent and perspectives of people who are often positioned at the peripheries of technology design. Her work draws on expertise in the design and evaluation of technologies, including AI, in resource-constrained settings. [10].

Sheza Naveed is a University of Michigan graduate and researcher whose work focuses on consent in AI systems and the usable privacy and security of AI, particularly as AI becomes more integrated into health technologies. She brings expertise in user research and human-centered design in sensitive domains such as health, with a focus on communities often underserved by technology design.

Vikram Mohanty is a project scientist at Carnegie Mellon University's Human-Computer Interaction Institute, where he builds and studies large-scale human-AI collaborative systems for complex real-world problems. He developed Civil War Photo Sleuth, an AI-infused online platform used by more than 20,000 people to identify unknown historical portraits, through which he has studied the privacy risks of facial recognition [11]. His current research includes developing privacy-aware AI systems that help R&D enterprises solve their hardest challenges by finding novel analogical inspirations from unexpected domains.

Gabriel Grill is a postdoctoral researcher at the Human Rights and Technology Group part of the Interdisciplinary Transformation University Austria and a visiting researcher at the AI Futures Lab at TU Delft. His research is concerned with the socio-technical study of algorithmic infrastructures [5] like generative AI and their consequences to societies. He has critically studied surveillance technologies such as protest prediction and job seeker profiling and is currently examining the limits of current approaches to privacy [6].

Hana Habib is an Assistant Teaching Professor in the Software and Societal Systems Department (S3D) at Carnegie Mellon University and the Associate Director of CMU's Masters in Privacy Engineering and AI Governance program. She contributes expertise in usability evaluations of privacy choice and consent mechanisms [7]. Her past work has also explored how to support practitioners in their understanding of AI privacy risks [20].

Kentaro Toyama is the W.K. Kellogg Professor of Community Information at the University of Michigan School of Information

and a fellow of the Dalai Lama Center for Ethics and Transformative Values at MIT. His research examines how digital technologies interact with human agency and social and institutional contexts. His influential "Law of Amplification" shows how technology amplifies existing human intentions and capacities [21]. This perspective is particularly relevant to consentful AI, highlighting the need to consider the human and social conditions under which mechanisms for agency, participation, and control become meaningful.

Lorrie Faith Cranor is Director and Bosch Distinguished Professor of the CyLab Security and Privacy Institute and FORE Systems University Professor of Computer Science and of Engineering and Public Policy at Carnegie Mellon University where she directs the CyLab Usable Privacy and Security Laboratory (CUPS). She has published research on the usability of privacy choice and consent mechanisms [8] and risks of generative AI systems [13], and is currently conducting research on the consent mechanisms used in popular agentic AI systems.

Jane Im is a tenure-track faculty member at the CISPA Helmholtz Center for Information Security, where she directs the Real-world Interactions and Systems for Change (RISC) Group. In her research, she designs and builds systems to address consent, privacy, and safety in social media and AI. Her work on theorizing consentful systems [9] has been cited widely across HCI, privacy, and social computing communities.

4 Workshop Website

Our workshop website is: https://jasmeet23.github.io/consentful-ai/.

5 Workshop Structure

With the goal to build an interdisciplinary community, develop design resources, and establish a research agenda for consentful AI, we structure the workshop as below.

5.1 Pre-Workshop Plans

We will publicize the workshop through relevant HCI, CHI, CSCW, privacy, responsible AI, machine learning, data governance, design, and AI ethics communities along with social media. Participants will be invited to submit 2-4 page position papers, works-in-progress, design explorations, or provocations, excluding references. Submissions may focus primarily on consent in human-AI interaction, consent concerning data used for AI development, and conceptual or technical connections between the two. We particularly encourage submissions presenting concrete consent breakdowns - situations in which existing mechanisms for permission, refusal, or withdrawal fail to adequately represent stakeholder intentions. Each submission will be reviewed by members of the organizing team. Selection will consider relevance to the workshop, clarity of contribution, and potential to stimulate discussion. Selected papers will be made public on the workshop website.

5.2 Workshop Activities

5.2.1 ***Session Ia: Lightning Talks*** *(20 minutes).* We will begin with rapid lightning talks in which participants introduce their submissions in one minute each. Rather than providing comprehensive presentations of their work, participants will be encouraged to use their talks to surface provocative examples, unresolved questions, or assumptions about consent. Participants will note connections to others' work during the talks, providing an initial basis for forming interdisciplinary groups and identifying shared interests across the workshop community.

5.2.2 ***Session Ib: Consent Breakdown Scenarios*** *(30 minutes).* Participants will form interdisciplinary groups and analyze concrete scenarios. Example scenarios may include:

Data Creator: Public artwork is included in AI training data.

End User: Sensitive information may be retained for personalization.

AI Agent: Trip planning may require access to emails, calendars, finances, and third parties.

Bystander: Uploaded photos may contain other people.

Withdrawal: Permission is withdrawn after model training.

Each group will analyze a scenario from the perspectives of multiple stakeholders. Participants may take on different stakeholder role, as mentioned above, and negotiate what meaningful consent would require from each perspective. Groups will identify what requires permission, who can legitimately provide it, what information is needed for informed consent, when permission should be requested or renewed, how refusal should affect the system, and what revocation could realistically mean. Midway through the activity, groups will exchange scenarios. The receiving group will act as a team challenging the first group's assumptions by identifying overlooked stakeholders, conflicting permissions, unintended consequences, or ways in which the proposed consent mechanism could fail. This activity will surface both common principles and tensions that cannot be resolved through individual consent alone.

5.2.3 ***Session Ic: Envisioning Consentful AI: Designs and Anti-Designs*** *(30 minutes).* Groups will select one of the consent breakdowns identified earlier and turn it into a design challenge - *What would a system look like if meaningful consent were treated as a core design requirement?* Each group will generate at least five possible design responses. Participants will be encouraged to envision design beyond conventional consent mechanisms like pop-up dialogs, and to explore approaches such as contextual consent, machine-readable creator preferences, consent negotiation between stakeholders, mechanisms for inspecting and modifying permissions, and meaningful forms of refusal and revocation.

Groups will then deliberately design a contrasting "consent dark pattern," an interface or mechanism that appears to provide choice while making refusal, understanding, or withdrawal difficult. Comparing these designs will help participants articulate what distinguishes meaningful consent from merely procedural consent. Each group will document its proposed designs and anti-design to share in the following session.

5.2.4 ***Session IIa: Keynote Talk*** *(20 minutes).* Dr. Lorrie Faith Cranor[1] will be the keynote speaker. Her expertise in usable privacy and security, human-centered technology, and privacy decision-making closely aligns with the workshop themes and will provide valuable insights to stimulate discussion among participants.

[1] https://lorrie.cranor.org

5.2.5 ***Session IIb: Discussing and Prioritizing Consentful AI Designs*** *(30 minutes).* This activity will begin with a collaborative design discussion, where a volunteer from each group will briefly present their proposed designs and anti-design to the wider workshop. Participants from other groups will be invited to share their opinions, provide feedback, raise concerns, and suggest potential improvements. This discussion will encourage participants to critically reflect on different approaches to meaningful consent and explore how ideas from different groups might complement or strengthen one another.

After all groups have presented, participants will vote to identify the top five most promising designs. Each participant will receive five votes, which they can distribute across designs proposed by other groups. Participants will be encouraged to consider how effectively each design addresses consent challenges, its feasibility, and its potential applicability across different AI contexts. The five designs receiving the highest number of votes will be selected as the workshop's most promising design directions.

5.2.6 ***Session IIc: Research Agenda*** *(30 minutes).* The final activity will translate the day's discussions into a collective research agenda. Participants will first individually contribute open questions emerging from the workshop and place them under themes such as definitions of consent, dataset governance, technical enforcement, evaluation methods, and the limits of individual consent. The workshop will conclude with five prioritized design directions, a synthesis of key challenges, and a preliminary research agenda for future work on consentful AI.

5.3 Post-Workshop Plans

We intend the workshop to establish a sustained interdisciplinary research community around consentful AI. Following the workshop, organizers will synthesize the consent breakdown scenarios, designs and anti-designs, and research questions generated during the activities. This resource will be made publicly available through the workshop website and will support the broader CHI community, as well as industry practitioners, in designing and evaluating consent mechanisms across AI applications, user interfaces, and data governance practices. The proposed designs and anti-designs can inform the development of alternative consent mechanisms, and help identify opportunities and practical challenges for improving consent mechanisms in real-world AI deployments.

We will also explore developing the workshop outcomes into a collaborative research agenda for publication at an HCI or interdisciplinary venue and encourage continued collaborations across HCI, privacy, AI, and data governance. The longer-term objective is to move consent from the margins of AI development, where it is frequently treated as a legal notice, dataset metadata field, or interface checkbox, toward a substantive topic of human-centered AI research, while fostering practical approaches to meaningful consent in research and industry.

6 Call for Participation

Generative AI raises questions about consent across the AI lifecycle. Before deployment, models may learn from collections of human-created texts, images, code, music, conversations, and other data. What constitutes meaningful permission for these uses, and can that permission be refused, conditioned, or withdrawn? After deployment, AI systems can remember information, infer sensitive attributes, connect to external services, and act on people's behalf. When should these systems ask permission, and how should consent be negotiated over time? In this one-day workshop, we bring together researchers and practitioners from HCI, AI, privacy, design, data governance, machine learning, and related areas to investigate how to make AI consentful by design. We invite 2–4 page position papers, works-in-progress, design explorations, and provocations. Approximately 20 participants will engage in brainstorming and design sessions to develop consent scenarios, design insights, and a research agenda for consentful AI across data, models, and interactions. We plan to publish participants' submissions along with resources generated during the workshop on the workshop website. Please visit https://jasmeet23.github.io/consentful-ai/ for participation and contact details.

References

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

[2] Emily M Bender and Batya Friedman. 2018. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the association for computational linguistics* 6 (2018), 587–604.

[3] Adrienne Porter Felt, Serge Egelman, Matthew Finifter, Devdatta Akhawe, and David Wagner. 2012. How to ask for permission. *HotSec* (2012), 0.

[4] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Commun. ACM* 64, 12 (2021), 86–92.

[5] Gabriel Grill. 2025. *Algorithms as Forward Salients: Examining the Politics of Algorithmic Infrastructure.* Ph. D. Dissertation. University of Michigan, Ann Arbor, MI. hdl:2027.42/199116 doi:10.7302/26854

[6] Gabriel Grill. 2026. Surveillance in the Name of Security: A Tension at the Heart of SOUPS?. In *1st Workshop on Surveillance and Usable Privacy (SUP)*.

[7] Hana Habib. 2021. *Evaluating the Usability of Privacy Choice Mechanisms.* Ph. D. Dissertation. Carnegie Mellon University.

[8] Hana Habib and Lorrie Faith Cranor. 2022. Evaluating the Usability of Privacy Choice Mechanisms. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, 273–289. https://www.usenix.org/conference/soups2022/presentation/habib

[9] Jane Im, Jill Dimond, Melody Berton, Una Lee, Katherine Mustelier, Mark S Ackerman, and Eric Gilbert. 2021. Yes: Affirmative consent as a theoretical framework for understanding and imagining social platforms. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–18.

[10] Jasmeet Kaur and Pushpendra Singh. 2025. Exploring the potential of chatgpt for supporting family planning needs of postpartum women in resource-constrained areas of india. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–26.

[11] Manisha Kusuma, Vikram Mohanty, Marx Wang, and Kurt Luther. 2022. Civil war twin: Exploring ethical challenges in designing an educational face recognition application. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. 369–384.

[12] Chung Peng Lee, Rachel Hong, Harry H Jiang, Aster Plotnik, William Agnew, and Jamie Heather Morgenstern. 2026. How do data owners say no? A case study of data consent mechanisms in web-scraped vision-language AI training datasets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 40. 38808–38816.

[13] Megan Li, Wendy Bickersteth, Ningjing Tang, Lorrie Cranor, Jason Hong, Hong Shen, and Hoda Heidari. 2025. A Closer Look at the Existing Risks of Generative AI: Mapping the Who, What, and How of Real-World Incidents. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 8, 2 (Oct. 2025), 1561–1573. doi:10.1609/aies.v8i2.36655

[14] Shayne Longpre, Robert Mahari, Anthony Chen, Naana Obeng-Marnu, Damien Sileo, William Brannon, Niklas Muennighoff, Nathan Khazam, Jad Kabbara, Kartik Perisetla, et al. 2024. A large-scale audit of dataset licensing and attribution in AI. *Nature Machine Intelligence* 6, 8 (2024), 975–987.

[15] Shayne Longpre, Robert Mahari, Ariel Lee, Campbell Lund, Hamidah Oderinwale, William Brannon, Nayan Saxena, Naana Obeng-Marnu, Tobin South, Cole Hunter, et al. 2024. Consent in Crisis: The Rapid Decline of the AI Data Commons.(2024). *URL https://arxiv. org/abs/2407.14933* (2024).

[16] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[17] Ashwini Rao, Florian Schaub, Norman Sadeh, Alessandro Acquisti, and Ruogu Kang. 2016. Expecting the unexpected: Understanding mismatched privacy expectations online. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. 77–96.

[18] Florian Schaub, Rebecca Balebako, and Lorrie Faith Cranor. 2017. Designing effective privacy notices and controls. *IEEE Internet Computing* 21, 3 (2017), 70–77.

[19] Florian Schaub, Rebecca Balebako, Adam L Durity, and Lorrie Faith Cranor. 2015. A design space for effective privacy notices. In *Eleventh symposium on usable privacy and security (SOUPS 2015)*. 1–17.

[20] Aseem Shrey, Weiqian Zhang, Ece Gumusel, and Hana Habib. 2026. Risk vs. Reward: Supporting Practitioners' Understanding of LLM Privacy Risks. In *Proceedings of the 2026 Workshop on Human-Centered AI Privacy and Security (HAIPS)*.

[21] Kentaro Toyama. 2011. Technology as amplifier in international development. In *Proceedings of the 2011 iConference*. 75–82.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009