EPIC: Enhanced Privacy and Integrity Considerations for Research (Tutorial)

HANNA ALZUGHBI*, Clemson University, USA
KELLY CAINE, Clemson University, USA
MEHTAB IQBAL, Clemson University, USA
BART KNIJNENBURG, Clemson University, USA
HANSEN LEE, Clemson University, USA
SUSAN E. MCGREGOR, Columbia University, USA
EMILY SIDNAM-MAUCH, Clemson University, USA

This tutorial engages researchers in a series of collaborative activities towards Enhanced Privacy and Integrity Considerations (EPIC) for human subjects research in the artificial intelligence (AI) field. The tutorial aims to identify common challenges to study integrity, convey best practices for protecting participants at the point of study design, and discuss how to best design tools to support robust, privacy-enhancing human subjects research in AI. In particular, the tutorial provides hands-on training on how to determine sample size and collect participant demographics in a way that prioritizes data integrity, participant privacy, and sample representativeness. Tutorial participants discuss and troubleshoot the unique challenges to and opportunities for designing robust and ethical human-centered AI research.

CCS Concepts: • **Human-centered computing** → **User studies**; • **Security and privacy** → *Privacy protections.*

Additional Key Words and Phrases: participant recruitment, privacy, research integrity, human-centered AI

ACM Reference Format:
Hanna Alzughbi, Kelly Caine, Mehtab Iqbal, Bart Knijnenburg, Hansen Lee, Susan E. McGregor, and Emily Sidnam-Mauch. 2024. EPIC: Enhanced Privacy and Integrity Considerations for Research (Tutorial). In *29th International Conference on Intelligent User Interfaces - Companion (IUI Companion '24), March 18–21, 2024, Greenville, SC, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3640544.3645249

1 MOTIVATION AND RELEVANCE

Human-centered design and evaluation approaches are essential to creating effective and ethical artificial intelligence (AI) systems [21]. The AI research field's increasing focus on human-centered evaluations (e.g., controlled experiments with real human users) requires researchers to make difficult decisions regarding the *type* and *extent* of data that they will collect about study participants. Demographic data collection is necessary to ensure the representativeness of a study sample and to achieve fairness of resulting insights and systems; however, such data can easily put study participants at risk of re-identification—especially when combined with behavioral data collected to power the AI system itself [20]. Similarly, decisions about the number of participants to recruit must be made carefully to meet the

*Authors are listed in alphabetical order, as all will contribute equally to the tutorial development and implementation.

study’s needs for statistical power, methodological robustness, or generalizability, while also avoiding unnecessary data collection.

The EPIC tutorial gives HCI and AI researchers fundamental knowledge on designing robust, privacy-preserving human subjects studies. The interactive tutorial engages individuals with varying expertise and experiences, inviting diverse perspectives to identify privacy and integrity concerns for human-centered AI (HCAI) studies and discuss paths forward for robust and ethical human subjects studies in AI. Participants simultaneously gain and contribute to knowledge on research best practices for protecting research participants and improving study design for AI studies.

2 FORMAT, ACTIVITIES, AND OBJECTIVES

The full-day EPIC tutorial engages participants through guided discussions, interactive lectures, hands-on lessons, and group activities. The morning sessions cover topics pertaining to existing practices, challenges, and opportunities for human subjects research in HCAI. After reviewing novel data on the sample size and demographic data collection practices at IUI, attendees participate in an affinity diagramming exercise to map the challenges and opportunities HCAI researchers face regarding enhancing the privacy and integrity of research on AI systems. Next, subject matter experts present best practices for determining sufficient sample size for a study and designing privacy-preserving data collection tools in order to mitigate risks to participant privacy at the point of study design. Attendees then receive interactive, hands-on training to apply these best practices to their own planned studies. In the afternoon, attendees work together to apply their new knowledge to tackle the challenges to participant privacy and research integrity they identified earlier in the day in a participatory design session to inform solutions to support AI researchers in conducting robust, ethical HCAI research.

The tutorial’s objective is to instruct and explore the process of determining key aspects of study design (i.e., sample size and demographic data collection), aiming for a balance between data utility and privacy considerations. The planned outcomes include:

- Attendees develop the ability to identify and navigate trade-offs between data collection practices and privacy expectations. Participants gain an advanced understanding of the tensions arising from the interplay of sample size, privacy concerns, and representation.
- Attendees understand best practices for determining sample size and demographic data collection and become familiar with resources they can use to implement these practices in their research.
- Attendees’ discussion generates valuable insights to the HCAI community on enhancing the privacy and integrity of human subjects research. These insights are captured through focus groups, affinity diagramming, and participatory design sessions, and summaries of the takeaways are made available on the tutorial webpage.

3 TUTORIAL MATERIALS

Tutorial materials such as lecture slides and summaries of group insights generated through the activities will be available to the audience and general public on the tutorial webpage (http://hatlab.org/epic/).

4 TUTORIAL ORGANIZERS

The tutorial organizers are an interdisciplinary group of faculty and student researchers currently researching how to secure research workflows by developing usable tools for designing methodologically robust, data-minimizing human subjects data collection processes for AI research. This team brings a wide range of experience and expertise in research methods, usable privacy and security, and human-centered AI.

Kelly Caine is an expert on research methods, usable privacy and security (e.g.,[15],[19]), human factors, human-computer interaction, AI [9], and engineering psychology. Dr. Caine co-authored the book *Understanding Your Users* [1], a comprehensive text on conducting studies with human participants, and her paper investigating sample sizes of human subjects studies published at CHI [2] guides researchers on choosing appropriate sample sizes. **Bart Knijnenburg** is a central figure in the human-centered AI research community. He is an expert in scale development and statistical evaluation [14] and has taught extensively on these topics at human-centered AI conferences and summer schools (e.g., [7]). He is also an expert in usable privacy [4, 10] and user-tailored solutions to online privacy issues and privacy measurement [3, 5, 6, 8, 11–13, 16–18, 26]. **Emily Sidnam-Mauch**, *Clemson University, esidnam@clemson.edu.* Dr. Sidnam-Mauch is an expert in computer-mediated communication and social media [24, 25, 28], quantitative methods, and survey research [22, 23, 27]. She has developed curricula on determining target sample size and composition, including calculating, interpreting, and reporting power analysis and effect size. **Hanna Alzughbi, Mehtab Iqbal, and Hansen Lee** are Ph.D. students in Human-Centered Computing at Clemson University, studying usable privacy and security, the design of transparent AI tools, and human-AI interaction, respectively.

ACKNOWLEDGMENTS

This tutorial is based upon work supported by the National Science Foundation under Grant No. 2232690.

REFERENCES

[1] Kathy Baxter, Catherine Courage, and Kelly Caine. 2015. *Understanding your users: a practical guide to user research methods.* Morgan Kaufmann.

[2] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (Santa Clara, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 981–992. https://doi.org/10.1145/2858036.2858498

[3] Hichang Cho, Bart Knijnenburg, Alfred Kobsa, and Yao Li. 2018. Collective Privacy Management in Social Media: A Cross-Cultural Validation. *ACM Transactions on Computer-Human Interaction* 25, 3 (June 2018), 17:1–17:33. https://doi.org/10.1145/3193120

[4] Arik Friedman, Bart P. Knijnenburg, Kris Vanhecke, Luc Martens, and Shlomo Berkovsky. 2015. Privacy Aspects of Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, Boston, MA, 649–688. https://doi.org/10.1007/978-1-4899-7637-6_19

[5] Reza Ghaiumy Anaraky, Kaileigh Angela Byrne, Pamela J. Wisniewski, Xinru Page, and Bart Knijnenburg. 2021. To Disclose or Not to Disclose: Examining the Privacy Decision-Making Processes of Older vs. Younger Adults. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. https://doi.org/10.1145/3411764.3445204

[6] Yangyang He, Paritosh Bahirat, Bart P. Knijnenburg, and Abhilash Menon. 2019. A Data-Driven Approach to Designing for Privacy in Household IoT. *ACM Trans. Interact. Intell. Syst.* 10, 1 (Sept. 2019), 10:1–10:47. https://doi.org/10.1145/3241378

[7] Bart P. Knijnenburg. 2012. Conducting user experiments in recommender systems. In *Proceedings of the sixth ACM conference on Recommender systems - RecSys '12*. ACM Press, Dublin, Ireland, 3. https://doi.org/10.1145/2365952.2365956

[8] Bart P. Knijnenburg. 2017. Privacy? I Can't Even! Making a Case for User-Tailored Privacy. *IEEE Security & Privacy* 15, 4 (2017), 62–67.

[9] Bart P Knijnenburg, Nicole Bannister, and Kelly Caine. 2021. Using Mathematically-Grounded Metaphors to Teach AI-Related Cybersecurity. In *IJCAI-21 Workshop on Adverse Impacts and Collateral Effects of Artificial Intelligence Technologies (AIofAI), Montréal, Canada.* http://ceur-ws.org/Vol-2942/paper2.pdf

[10] Bart P. Knijnenburg and Shlomo Berkovsky. 2017. Privacy for Recommender Systems: Tutorial Abstract. In *Proceedings of the Eleventh ACM Conference on Recommender Systems.* ACM, Como Italy, 394–395. https://doi.org/10.1145/3109859.3109935

[11] Bart P Knijnenburg and Hongxia Jin. 2013. The Persuasive Effect of Privacy Recommendations. In *Twelfth Annual Workshop on HCI Research in MIS.* Milan, Italy. http://aisel.aisnet.org/sighci2013/16

[12] Bart P. Knijnenburg and Alfred Kobsa. 2013. Helping users with information disclosure decisions: potential for adaptation. In *Proceedings of the 2013 ACM international conference on Intelligent User Interfaces.* ACM Press, Santa Monica, CA, 407–416. https://doi.org/10.1145/2449396.2449448 Forthcoming.

[13] Bart P. Knijnenburg, Alfred Kobsa, and Hongxia Jin. 2013. Dimensionality of information disclosure behavior. *International Journal of Human-Computer Studies* 71, 12 (2013), 1144–1162. https://doi.org/10.1016/j.ijhcs.2013.06.003

[14] Bart P. Knijnenburg and Martijn C. Willemsen. 2015. Evaluating Recommender Systems with User Experiments. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, Boston, MA, 309–352. https://doi.org/10.1007/978-1-4899-7637-6_9

[15] Yifang Li and Kelly Caine. 2022. Obfuscation Remedies Harms Arising from Content Flagging of Photos. In *CHI Conference on Human Factors in Computing Systems*. 1–25.

[16] Yao Li, Hichang Cho, Reza Ghaiumy Anaraky, Bart Knijnenburg, and Alfred Kobsa. 2022. Antecedents of collective privacy management in social network sites: a cross-country analysis. *CCF Transactions on Pervasive Computing and Interaction* 4, 2 (June 2022), 106–123. https://doi.org/10.1007/s42486-022-00092-8

[17] Yao Li, Reza Ghaiumy Anaraky, and Bart Knijnenburg. 2021. How Not to Measure Social Network Privacy: A Cross-Country Investigation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–32. https://doi.org/10.1145/3449218

[18] Yao Li, Alfred Kobsa, Bart P. Knijnenburg, and M.-H. Carolyn Nguyen. 2017. Cross-Cultural Privacy Prediction. *Proceedings on Privacy Enhancing Technologies* 2017, 2 (April 2017), 113–132. https://doi.org/10.1515/popets-2017-0019 Publisher: Sciendo Section: Proceedings on Privacy Enhancing Technologies.

[19] Heather Richter Lipford, Pamela J. Wisniewski, Cliff Lampe, Lorraine Kisselburgh, and Kelly Caine. 2012. Reconciling Privacy with Social Media. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion* (Seattle, Washington, USA) *(CSCW '12)*. ACM, New York, NY, USA, 19–20. https://doi.org/10.1145/2141512.2141523

[20] Arvind Narayanan and Vitaly Shmatikov. 2010. Myths and Fallacies of "Personally Identifiable Information". *Commun. ACM* 53, 6 (jun 2010), 24–26. https://doi.org/10.1145/1743546.1743558

[21] Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (2020), 495–504. https://doi.org/10.1080/10447318.2020.1741118

[22] Emily Sidnam. 2017. Annual Convention of the International Communication Association. In *Mobile Communication Division Top Papers*.

[23] Emily A Sidnam. 2015. *Accessing information and social capital on Facebook: A theoretical and empirical investigation of an accelerated knowledge gap model*. Ph. D. Dissertation. Purdue University.

[24] Emily Sidnam-Mauch. 2019. Counting the costs and projecting the future of numbering technologies.

[25] Emily Sidnam-Mauch and Leila Bighash. 2021. How controversy leads to commitment: Predecisional distortion in reactions to premarket products through online review systems. *Computers in Human Behavior* 124 (2021), 106902.

[26] Pamela J. Wisniewski, Bart P. Knijnenburg, and Heather Richter Lipford. 2017. Making privacy personal: Profiling social network users to inform privacy education and nudging. *International Journal of Human-Computer Studies* 98 (Feb. 2017), 95–108. https://doi.org/10.1016/j.ijhcs.2016.09.006

[27] Bei Yan, Lian Jian, Ruqin Ren, Janet Fulk, Emily Sidnam-Mauch, and Peter Monge. 2021. The paradox of interaction: Communication network centralization, shared task experience, and the wisdom of crowds in online crowdsourcing communities. *Communication Research* 48, 6 (2021), 796–818.

[28] Lindsay E Young, Emily Sidnam-Mauch, Marlon Twyman, Liyuan Wang, Jackie Jingyi Xu, Matthew Sargent, Thomas W Valente, Emilio Ferrara, Janet Fulk, and Peter Monge. 2021. Disrupting the COVID-19 misinfodemic with network interventions: network solutions for network problems. *American journal of public health* 111, 3 (2021), 514–519.