Designing Support for Systematic Sociotechnical Risk Literacy

Ashley Marie Walker
Google
New York, New York, USA
amwalker@google.com

Renee Shelby
Google Research
San Francisco, California, USA
reneeshelby@google.com

Ari Schlesinger
School of Computing
University of Georgia
Athens, Georgia, USA
ari.schlesinger@uga.edu

Emily Tseng
Microsoft Research
New York, New York, USA
emilytseng@microsoft.com

Mark Diaz
Google Research
New York City, New York, USA
markdiaz@google.com

Andy Elliot Ricci
Digital and Computational Studies
Bates College
Lewiston, Maine, USA
aricci@bates.edu

Angela D. R. Smith
School of Information
University of Texas at Austin
Austin, Texas, USA
angela.smith@ischool.utexas.edu

Abstract

Risk literacy in technology contexts concerns the ability to understand, assess, and make informed decisions about risks and their implications. This requires understanding the landscape of sociotechnical harms, as well as cultivating skills for identifying social and technical sources of harms, interpreting probabilistic information, understanding uncertainty, evaluating risk-benefit tradeoffs, and making decisions with incomplete information. Despite efforts to educate diverse audiences on technological risks and harms, these efforts are often ad-hoc and issue-driven. To move beyond issue-based approaches, we need a systematic approach to understanding technology's role in harm and assessing risk across sociotechnical contexts - what we term "systematic sociotechnical risk literacy". In this workshop, we will experiment with existing harms frameworks to develop a toolkit for teaching systematic risks across different contexts. This will equip educators to teach digital safety and foster coherence through shared language and practices. Ultimately, systematic digital risk literacy will enable engineers, policymakers, and researchers to collaboratively shape computing's future and navigate future uncertainties.

CCS Concepts

• **Human-centered computing → HCI theory, concepts and models**.

Keywords

risk literacy, digital safety, risks, harms, pedagogy

ACM ISBN 979-8-4007-1968-4/25/08
https://doi.org/10.1145/3737609.3747101

ACM Reference Format:
Ashley Marie Walker, Renee Shelby, Ari Schlesinger, Emily Tseng, Mark Diaz, Andy Elliot Ricci, and Angela D. R. Smith. 2025. Designing Support for Systematic Sociotechnical Risk Literacy. In *Adjunct proceedings of the sixth decennial Aarhus conference: Computing X Crisis (AAR Adjunct 2025), August 18–22, 2025, Aarhus N, Denmark.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3737609.3747101

1 Motivation

Risk literacy in technology contexts concerns the ability to understand, assess, and make informed decisions about risks and their implications [10, 13]. This requires understanding the landscape of sociotechnical harms [22] as well as cultivating skills for identifying social and technical sources of harms, interpreting probabilistic information, understanding uncertainty, evaluating risk-benefit tradeoffs, and making decisions with incomplete information [7, 15, 19, 24]. Computing researchers have identified a wide range of challenges in technology risk literacy, including *mental model gaps* where people develop incomplete or inaccurate mental models of how technology works [9, 14, 28, 30], *knowledge-action divides*, where knowledge of risk does not translate into protective behaviors [25], *knowledge disparities* across background and standpoints [11, 18, 21], and *knowledge gaps* for new technologies such as AI given their complexity and novelty [17]. While researchers have developed measurement approaches (e.g., [1, 6, 8, 20]), the field still grapples with *measurement complexity* and lack of a standardized approach. Given the current moment of rapid social and technological change, acquiring the fundamental knowledge required to make well-informed decisions about technology use and development is essential for becoming risk literate—particularly for technology developers and policymakers.

Despite efforts to educate diverse audiences on technological risks and harms, these efforts are often ad-hoc and issue-driven [26]. Existing frameworks for analyzing technology's consequences are abundant but siloed, varying across digital literacy levels, spheres of influence, and scopes of concern (e.g., [2–4, 27, 29]). This fragmentation is evident in educational settings: some computer science students may gain expertise in accessibility requirements due to a professor's specialization, but lack understanding of privacy theory and practice. Similarly, industry engineers may become well-versed in privacy compliance through internal processes, yet remain unaware of broader sociotechnical risks unless these are central to their work culture, such as concern for algorithmic fairness or issues like tech-facilitated violence. Policymakers also learn about risks and harms in fragmented ways, through siloed briefings or serendipitously through social and professional networks [12]. Consequently, translating mitigation approaches across these fragmented frameworks becomes a challenge, particularly for those new to digital safety, and it unnecessarily divides efforts among coalitions that could otherwise be aligned.

To move beyond issue-based approaches, we need a systematic approach to understanding technology's role in harm and assessing risk across sociotechnical contexts - what we term ***systematic sociotechnical risk literacy***. This will equip educators to teach digital safety across various contexts and foster coherence through shared language and practices. Ultimately, systematic digital risk literacy will enable engineers, policymakers, and researchers to collaboratively shape computing's future and navigate future uncertainties.

In this workshop, we will experiment with existing systematic harms frameworks to develop a toolkit for teaching systematic risks across different contexts. We will particularly focus on how audience characteristics necessitate adaptations for effective learning. While teaching contexts vary (e.g, contact time and desired outcomes from awareness to better design to targeted advocacy), understanding audience-specific adaptations is vital to identifying core framework elements that must remain consistent and those that can be tailored to individual circumstances. Understanding the design space for what adaptations are useful for different audiences will help to clarify what aspects are necessary to hold consistent across contexts to provide a workable framework and what can be meaningfully adapted for individual circumstances. In collaboration with workshop attendees, we will work to react to and build off preliminary prototypes of what this kind of teaching and resource support might look like. This will include activities such as:

- Brainstorming situations where this systematic kind of approach might be useful and where it might not
- Developing strategies for addressing common concerns that arise when teaching these concepts
- Identifying limitations and potential sticking points for different audiences and circumstances

Aarhus2025 provides a unique opportunity to engage different stakeholders in a generative environment, fostering relationships and resources for effective communication about the full spectrum of digital safety risks and harms. Our organizing team, with expertise across harm domains and institutional contexts, has a track record of developing harm and safety frameworks [5, 16, 22, 23] and conducting successful workshops at ACM venues on topics such as synthesizing across digital safety topics [26]. This workshop addresses the pressing need for tools and methodologies that enable stakeholder groups to systematically understand the full spectrum of digital risks and harms. Coming together as a community to build relationships and resources to help stakeholders across academia, industry, and policy understand how to impart an understanding of how sociotechnical risks are interlinked with one another will help improve digital literacies with those best positioned to proactively prevent harm.

2 Intended Audience & Expected Outcomes

The intended audience of this workshop is people who educate others across a range of different contexts. This includes: faculty, graduate students, and industry practitioners. Educators may teach in traditional classroom settings, facilitate community programs, or work with engineers and designers who are unfamiliar with research on online harms. We expect that participants will have a variety of use cases and backgrounds. For example, a professor who teaches graduate students over the course of a whole semester as well as an industry practitioner that leads 2-hour training sessions for designers and engineers. We have intentionally designed our day-of programming to include collective sensemaking time upfront so that we can create a flexible repository of tools to support digital safety education in many contexts.

The primary goal of this workshop is for participants to be better equipped to teach other people how to systematically think about harms. This may take the form of a support network for people engaging in teaching about sociotechnical risk in a systematic way, the development of a library of resources and pedagogical approaches to be used as reference, and facilitating new collaborations to develop more evidence-based best practices for design and evaluation of teaching. One way that people might be better equipped to teach audiences how to systematically think about risk is through a set of resources that help be adapted to different institutional contexts and audiences. Potential resource that could emerge from day-of discussions include:

- Activities to scaffold on existing experience and interest
- Approaches for counteracting participant resistance and encouraging engagement
- Best practices for how to adapt terminology to audiences from different contexts
- Strategies for speculative practices to help build competencies for avoiding potential harms from technology

3 Workshop Structure

3.1 Participant Submission

Prospective participants will apply by submitting an overview of their interest and/or experience with educating on digital risk and harm: addressing the context(s) they teach/work in, the audiences they teach/work with, the response to digital safety from those they teach, the specific sub-areas of digital safety they have experience with, and the goals they have for designing and utilizing a toolbox to support systemic risk literacy. Our aim is to create an environment that engenders more people to feel ownership of the digital safety process, thus we will encourage participation from individuals who are seeking to become more involved with digital safety, people working on crises in computing that impact the digital safety landscape, and experts across varied digital safety domains.

3.2 Day of Workshop Plan

We want the format of this workshop to be responsive to the specific group of participants attending. To ensure that we can make the most of our time together, we will spend the morning building a shared understanding of the problem space, developing connections among participants, and surfacing the desired discussion space for the afternoon sessions. In the afternoon, we will focus on rotating small group discussions to iterate on ideas quickly before sharing out to the whole group at the end of the day, highlighting next steps and collaboration opportunities. Overall, the goal is to include multiple modalities for interaction to ensure that we encourage new connections between participants, cover a range of preferred engagement styles, and foster an engaging and collaborative environment.

Morning: Exploring the Systemic Risk Literacy Problem Space

- **Session 1:** (30 minutes) Introductions and Agenda Setting
  This session will establish the desired outcomes for the day's work and outline the workshop logistics.
- **Session 2:** (45 minutes) 1:1 Relationship Building
  Over the course of running workshops, we find that having multiple modes of interaction throughout the day helps ensure participants stay engaged. Starting with a session that prioritizes 1:1 interactions, where participants rotate through multiple opportunities to connect with each other encourages links with new potential collaborators and leads to more variation in group composition for later small group discussions.
- **Break** (15 minutes)
- **Session 3:** (60 min) Possible Futures Brainstorm
  This session will focus on a brainstorming activity to elicit possible outcomes of increased systemic risk literacy. The goal here is to encourage participants to construct a concrete and creative list of steps and outcomes that will promote increased systemic risk literacy and mitigate computing crises.
- **Session 4:** (30 min) Discussion Elicitation
  During this session participants will write down questions that they would like to focus on in the afternoon sessions and then collaboratively develop thematic affinity groupings that will anchor each of the discussion groups. We anticipate that these discussions will focus on 3 main aspects of adaptation: different audiences (e.g., programmers, designers, undergraduate students, policymakers, community members), different levels of granularity (e.g., what level of detail is useful for expert audiences, end users, or classroom settings), and different roadblocks (e.g., what kinds of language adaptions are useful for different audiences, what artifacts are useful to support learning goals, and how to make the most of short contact times for teaching).

Lunch Break

Afternoon: Designing Ways to Advance Systematic Risk Literacy

- **Session 5:** (45 min) Discussion Section 1
  The small group discussions are intended to give participants time to dive deeply into the practicalities of teaching a systematic approach to risk literacy. We have 2 rounds of discussion planned as, based on previous experience with this format of workshop, it is common for participants to be excited about more than one discussion topic. Having multiple rounds of discussion means that participants can either choose to go deeper on topics they feel strongly about or move to another discussion topic to explore a different angle.
- **Break** (15 min)
- **Session 6:** (45 min) Discussion Section 2
  Participants will either continue discussions or rotate to another discussion topic they would like to explore.
- **Session 7:** (60 min) Report Out and Next Steps
  In our closing session, we will come back together as a large group to share out high level themes from the small discussion groups and start drafting concrete next steps we can take as a group to advance (teaching) systemic risk literacy. We will anticipate that next steps will include resource sharing about best practices for different audiences and constraints, suggestions and references for teaching tools that would support, etc.

4 Post-Workshop Plan

Following the workshop, we will create a collaborative document for collecting and reporting key outcomes. This document will highlight the collective knowledge of participants and disseminate workshop outcomes to a broader audience. The document will be a repository of experiences and strategies for different contexts to support educators in scaffolding knowledge construction, identifying misconceptions, responding to resistance, and empowering people across organizational contexts to advance systematic risk literacy.Moreover, the post-workshop report will identify future strategic research goals and collaborative actions the computing community can take to understand, identify, and mitigate digital harm in systemic ways that help address crises that are amplified and enabled by computing. This document will enable stakeholders working across the range of industry, academic, community and civil society groups to design context-specific teaching tools and to advance systematic risk literacy across contexts.

References

[1] Farman Afzal, Shao Yunfei, Mubasher Nazir, and Saad Mahmood Bhatti. 2021. A Review of Artificial Intelligence Based Risk Assessment Methods for Capturing Complexity-risk Interdependencies: Cost Overrun in Construction Projects. *International Journal of Managing Projects in Business* 14, 2 (2021), 300–328.

[2] Travis Breaux and David Gordon. 2013. What Engineers Should Know about US Security and Privacy Law. *IEEE Security and Privacy* 11, 3 (2013), 72–76. https://doi.org/10.1109/MSP.2013.74

[3] Ann Cavoukian. 2011. Privacy by Design: The 7 Foundational Principles.

[4] Andy Coverdale, Sarah Lewthwaite, and Sarah Horton. 2022. Teaching Accessibility as a Shared Endeavour: Building Capacity across Academic and Workplace Contexts. In *Proceedings of the 19th International Web for All Conference (W4A'22)*. ACM, 1–5. https://doi.org/10.1145/3493612.3520451

[5] Mark Díaz, Sunipa Dev, Emily Reif, Emily Denton, and Vinodkumar Prabhakaran. 2024. SoUnD Framework: Analyzing (So)cial Representation in (Un)structured (D)ata. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7. 371–383.

[6] Roel Dobbe. 2022. System Safety and Artificial Intelligence. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1584–1584.

[7] Roel Dobbe, Thomas Krendl Gilbert, and Yonatan Mintz. 2021. Hard Choices in Artificial Intelligence. *Artificial Intelligence* 300 (2021), 103555.

[8] Serge Egelman and Eyal Peer. 2015. Scaling the Security Wall: Developing a Security Behavior Intentions Scale (sebis). In *Proceedings of the 33rd annual ACM*

conference on human factors in computing systems. 2873–2882.
[9] Andrew J Flanagin and Miriam J Metzger. 2000. Perceptions of Internet Information Credibility. *Journalism & Mass Communication Quarterly* 77, 3 (2000), 515–540.
[10] Paul Gilster. 1997. Digital Literacy.
[11] Eszter Hargittai and Kerry Dobransky. 2017. Old Dogs, New Clicks: Digital Inequality in Skills and Uses Among Older Adults. *Canadian Journal of Communication* 42, 2 (2017), 195–212.
[12] Adnan A Hyder, Adrijana Corluka, Peter J Winch, Azza El-Shinnawy, Harith Ghassany, Hossein Malekafzali, Meng-Kin Lim, Joseph Mfutso-Bengo, Elsa Segura, and Abdul Ghaffar. 2011. National Policy-makers Speak Out: Are Researchers Giving Them What They Need? *Health Policy and Planning* 26, 1 (2011), 73–82.
[13] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. 2021. Conceptualizing AI Literacy: An Exploratory Review. *Computers and Education: Artificial Intelligence* 2 (2021), 100041.
[14] Emilee Rader and Janine Slaker. 2017. The Importance of Visibility for Folk Theories of Sensor Data. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. 257–270.
[15] Inioluwa Deborah Raji and Roel Dobbe. 2023. Concrete Problems in AI Safety, Revisited. arXiv:2401.10899 [cs.CY] https://arxiv.org/abs/2401.10899
[16] Bogdana Rakova, Renee Shelby, and Megan Ma. 2023. Terms-we-serve-with: Five Dimensions for Anticipating and Repairing Algorithmic Harm. *Big Data & Society* 10, 2 (2023), 20539517231211553.
[17] Maribeth Rauh, Nahema Marchal, Arianna Manzini, Lisa Anne Hendricks, Ramona Comanescu, Canfer Akbulut, Tom Stepleton, Juan Mateos-Garcia, Stevie Bergman, Jackie Kay, et al. 2024. Gaps in the Safety Evaluation of Generative AI. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7. 1200–1217.
[18] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. 2017. Where is the Digital Divide? A Survey of Security, Privacy, and Socioeconomics. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 931–936.
[19] Shalaleh Rismani, Roel Dobbe, and AJung Moon. 2024. From Silos to Systems: Process-oriented Hazard Analysis for AI Systems. *arXiv preprint arXiv:2410.22526* (2024).
[20] Shalaleh Rismani, Renee Shelby, Andrew Smart, Renelito Delos Santos, AJung Moon, and Negar Rostamzadeh. 2023. Beyond the ML Model: Applying Safety Engineering Frameworks to Text-to-Image Development. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. 70–83.
[21] Yukiko Sawaya, Mahmood Sharif, Nicolas Christin, Ayumu Kubota, Akihiro Nakarai, and Akira Yamada. 2017. Self-confidence Trumps Knowledge: A Cross-cultural Study of Security Behavior. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2202–2214.
[22] Renee Shelby, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla-Akbari, Jess Gallegos, Andrew Smart, Emilio Garcia, and Gurleen Virk. 2023. Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES '23)*. ACM. https://doi.org/10.1145/3600211.3604673
[23] Renee Shelby, Shalaleh Rismani, and Negar Rostamzadeh. 2024. Generative AI in Creative Practice: ML-Artist Folk Theories of T2I Use, Harm, and Harm-Reduction. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.
[24] André Steimers and Moritz Schneider. 2022. Sources of Risk of AI Systems. *International Journal of Environmental Research and Public Health* 19, 6 (2022), 3641.
[25] Monika Taddicken. 2014. The 'Privacy Paradox' in the Social Web: The Impact of Privacy Concerns, Individual Characteristics, and the Perceived Social Relevance on Different Forms of Self-disclosure. *Journal of Computer-mediated Communication* 19, 2 (2014), 248–273.
[26] Ashley Marie Walker, Michael Ann DeVito, Karla Badillo-Urquiola, Rosanna Bellini, Stevie Chancellor, Jessica L. Feuston, Kathryn Henne, Patrick Gage Kelley, Shalaleh Rismani, Renee Shelby, and Renwen Zhang. 2024. "What Is Safety?": Building Bridges Across Approaches to Digital Risks and Harms. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing (CSCW '24)*. ACM, 736–739. https://doi.org/10.1145/3678884.3681824
[27] Noel Warford, Tara Matthews, Kaitlyn Yang, Omer Akgul, Sunny Consolvo, Patrick Gage Kelley, Nathan Malkin, Michelle L Mazurek, Manya Sleeper, and Kurt Thomas. 2022. Sok: A framework for unifying at-risk user research. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2344–2360.
[28] Rick Wash and Emilee Rader. 2015. Too Much Knowledge? Security Beliefs and Protective Behaviors Among United States Internet Users. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. 309–325.
[29] Richmond Y. Wong, Michael A. Madaio, and Nick Merrill. 2023. Seeing like a Toolkit: How Toolkits Envision the Work of AI Ethics. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 1–27. https://doi.org/10.1145/3579621
[30] Yaxing Yao, Justin Reed Basdeo, Oriana Rosata Mcdonough, and Yang Wang. 2019. Privacy Perceptions and Designs of Bystanders in Smart Homes. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–24.