Ethics in Security and Privacy 2

INFR11158/11230 Usable Security and Privacy

Dr. Jingjie Li

17/03/2026

THE UNIVERSITY *of* EDINBURGH

Stakeholder-based ethics analysis

- **Stakeholders:** You are expected to consider all possible stakeholders (people, including the research team and society at large, and entities including companies) that may be impacted by your research. You are expected to detail how each stakeholder may have been impacted by the research procedures you undertook and how those stakeholders may be impacted by the publication of your research now and in the future.
- For example, who are the stakeholders involved in a vulnerability disclosure?

https://www.usenix.org/conference/usenixsecurity26/call-for-papers#ethics

Stakeholder-based ethics analysis

- Impacts:
  - **Ethical principles:** You are expected to articulate the ethical principles you considered. A starting point is considering the principles in The Menlo Report in the context of each identified stakeholder: **"Beneficence", "Respect for Persons", "Justice", and "Respect for Law and Public Interest".**
  - **Harms:** There are at least two broad categories of potentially negative outcomes from the research and publication process: tangible harms (e.g., financial loss or exposure to psychologically disturbing content) and violations of human rights even if there are no directly tangible harms (e.g., the violation of a participants' right to informed consent or the violation of users' right to privacy via the study of data that users expect and desire to be private).
- Thinking about the harms and impact throughout the research / publication lifecycle

Stakeholder-based ethics analysis

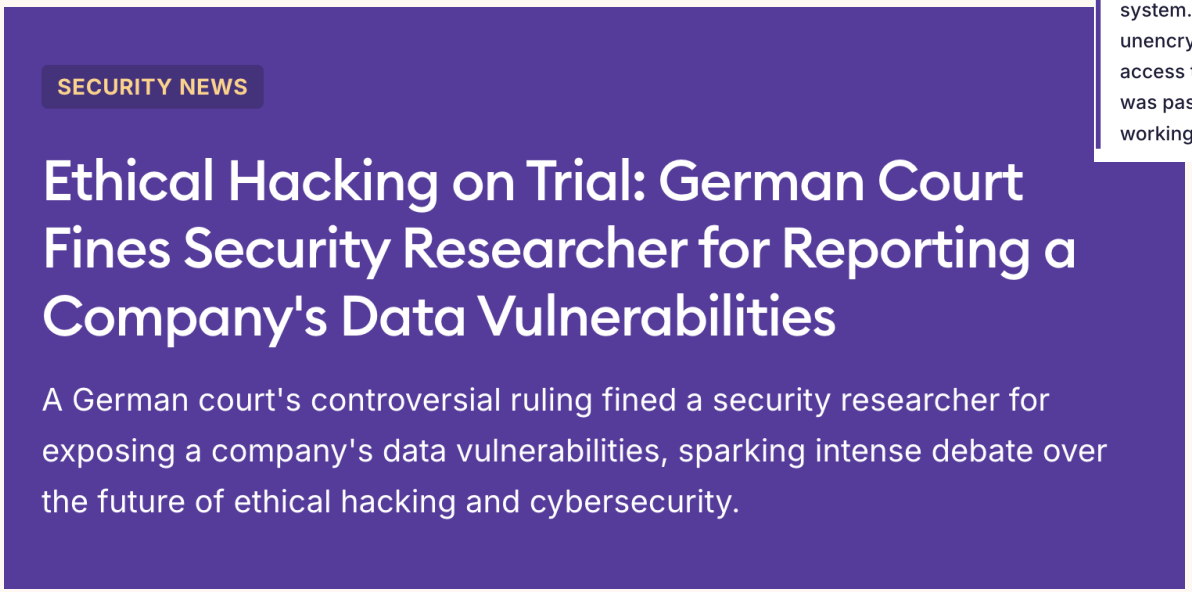
- **Mitigations:** You are expected to detail both mitigated and unmitigated (potential) harms of your work. You are expected to detail the steps taken to mitigate harms.
- Thinking about mitigations proactively, not reactively

Stakeholder-based ethics analysis

- **Decision:** You are expected to articulate why the decision to proceed with the research and the decision to publish the research was reached, respectively. One approach to reaching such a decision is to weigh ethical harms against ethical benefits; see the "Beneficence" principle in The Menlo Report. An alternative or additional approach is to focus on avoiding the violation of individuals' rights; see the "Respect for Persons" principle in The Menlo Report and the discussion of deontological ethics in the above-cited 2023 USENIX Security paper.
  - Analysis may lead to the same, or different outcomes. A risk assessment (likelihood / impact severity) would help

Example dimensions in computer security research ethics

> Today, June 23, 2021 at 8:09am, an 'ethical hacker' alerted us to a security vulnerability in our system. Due to this vulnerability, it was possible to access the password to our database and access unencrypted passwords and personal data. Using this database password, the hacker gained external access to our database and our ticketing system. We currently do not know to what extent this data was passed on or further used by the 'ethical hacker' and whether further access occurred. We are working intensively to investigate the incident.

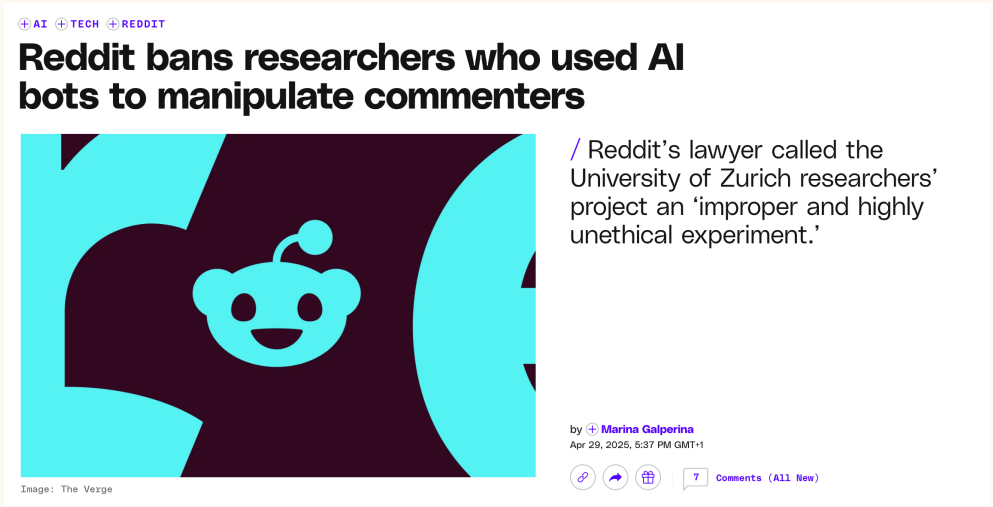


A German court handed down a chilling verdict in a recent case involving a security researcher who analyzed software on behalf of a client and found Modern Solution GmbH & Co.'s retail customer passwords stored in plain text. The court ruled that the programmer's actions constituted unauthorized access to external computer systems and spying on data, and issued a €3,000 ($3,265) fine.

https://socket.dev/blog/ethical-hacking-on-trial-german-court-fines-security-researcher?utm_source=chatgpt.com

- **Disclosures.** Vulnerabilities, if known to adversaries, can expose people to negative outcomes, such as harms or rights violations. Publicly disclosing vulnerabilities before they have been privately disclosed to the responsible parties, and hence before they have been mitigated, can therefore expose people to negative outcomes.
- In this case, why disclosing the password risk still leads to a legal consequence?
- How can we mitigate the ethical and legal risks?

https://socket.dev/blog/ethical-hacking-on-trial-german-court-fines-security-researcher?utm_source=chatgpt.com



AI TECH REDDIT

Reddit bans researchers who used AI bots to manipulate commenters

/ Reddit's lawyer called the University of Zurich researchers' project an 'improper and highly unethical experiment.'

by Marina Galperina
Apr 29, 2025, 5:37 PM GMT+1

7 Comments (All New)

Image: The Verge

- **Experiments with live systems without informed consent.** Researchers testing live services (e.g., for vulnerabilities) such as web services or APIs that give access to otherwise non-public algorithms or models must also consider ethics. Such experiments should only be performed after carefully analyzing the potential negative outcomes to the service provider, which may include cost (of CPU cycles or of human effort) or corrupting system state, and to end users who are using the same service provider for non-research purposes.
- What would you do instead if you want to study the persuasiveness of LLM in natural environments?

META

Research Cannot Be the Justification for Compromising People's Privacy

August 3, 2021

By Mike Clark, Product Management Director

LISTEN TO ARTICLE

For months, we've attempted to work with New York University to provide three of their researchers the precise access they've asked for in a privacy protected way. Today, we disabled the accounts, apps, Pages and platform access associated with NYU's Ad Observatory Project and its operators after our repeated attempts to bring their research into compliance with our Terms. NYU's Ad Observatory project studied political ads using unauthorized means to access and collect data from Facebook, in violation of our Terms of Service. We took these actions to stop unauthorized scraping and protect people's privacy in line with our privacy program under the FTC Order.

Earlier this year, we invited researchers, including the ones from NYU, to safely access US 2020 Elections ad targeting data through FORT's Researcher Platform. This offered the Ad Observatory researchers a more comprehensive data set than the one they created by scraping data on Facebook. The researchers had the opportunity to use the data set, which is designed to be privacy-protective, instead of relying on scraping, but they declined.

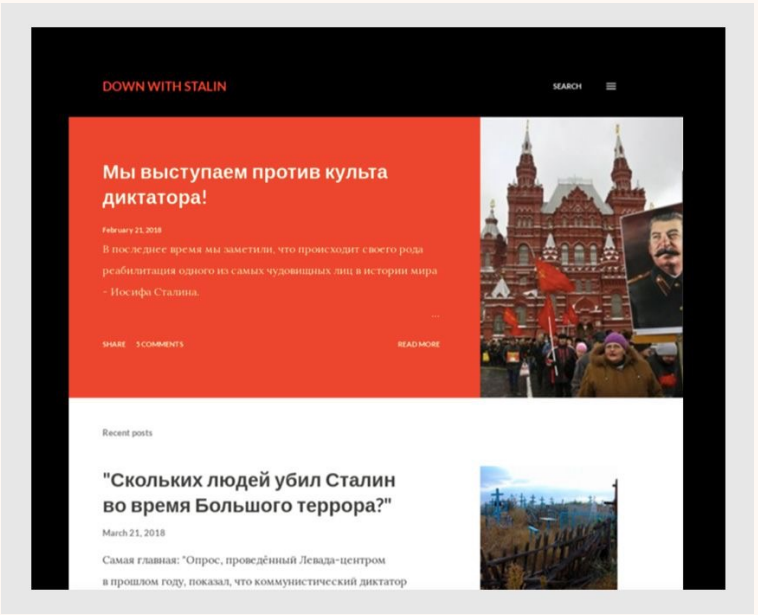
CCS News / Press Highlights August 21, 2021 - Lois Anne DeLong

Facebook Disables Ad Observatory; Academicians and Journalists Fire Back

Since September 15 of 2020, the Ad Observatory has been an effective resource for journalists and academicians seeking information about the placement and targeting of ads on social media. In doing so, the project increased transparency about what messages are being used, who is funding each ad, and how much is being spent to disseminate them.

On August 6, Facebook shutdown the NYU Ad Observatory by shutting down the Facebook accounts of Cybersecurity for Democracy team members Damon McCoy and Laura Edelson. Though the official reason given was that Facebook was complying with Federal Trade Commission rules because researchers did not have permission from users to scrape the information, the FTC denied the claim. According to an article in The Washington Post, "the Federal Trade Commission rejected Facebook's assertion in a letter sent to Facebook CEO Mark Zuckerberg on Thursday, and penned by Acting Director for the Bureau of Consumer Protection Samuel Levine."

https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy/?utm_source=chatgpt.com

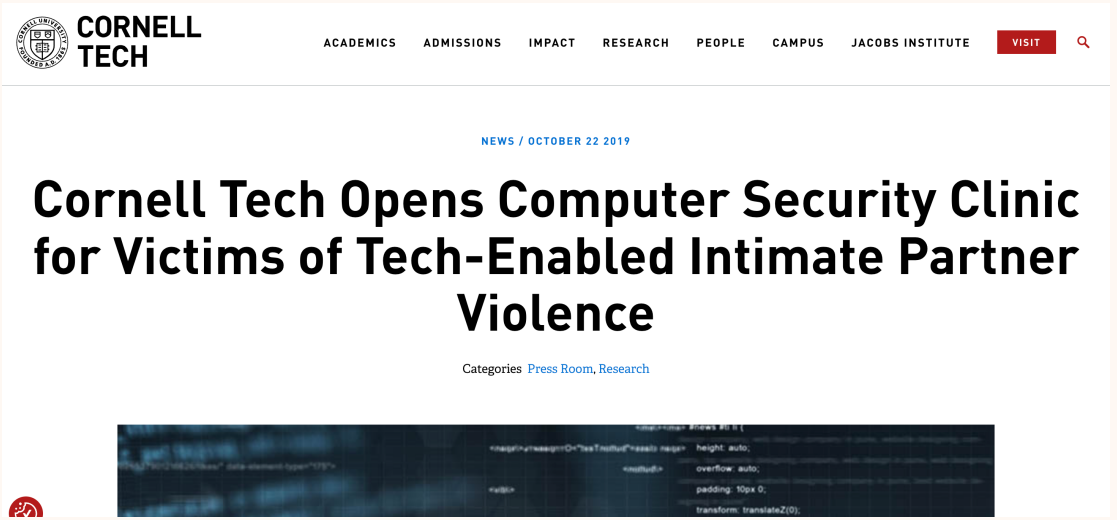


- **Deception.** In most cases, participants should be fully informed of the purposes and risks (among other things) of participating in experiments. If deception is used, the necessity of doing so should be carefully considered and the decision to use deception should be discussed in the ethics section. In general, participants in a deception study should be debriefed afterward to explain the necessity of the deception, even when the deception was mild.
- What are the (broader) risks in deception?

https://www.wired.com/story/jigsaw-russia-disinformation-social-media-stalin-alphabet/?utm_source=chatgpt.com

- **Innovations with both positive and negative potential outcomes.** Technologies that can positively impact one stakeholder group may negatively impact those same or other stakeholder groups. For example, advancements in anonymity systems could positively impact people that need anonymity under repressive regimes or excessive surveillance. At the same time, the mere use of those technologies could create negative impacts to those same people if the use of such technologies is detectable and hence subjects those individuals to additional scrutiny....Thus, researchers should think broadly about both the positive and negative potential impacts of their research throughout the research process, including during project selection and publication.
- Is there any good ways to mitigate risks from dual use?

https://www.reuters.com/technology/cybersecurity/britain-france-lead-35-nation-agreement-controlling-spyware-mercenary-hackers-2024-02-06/?utm_source=chatgpt.com



- **Wellbeing for team members.** In some cases, research activities have the potential to negatively impact team members. For example, research on hate speech could expose team members to disturbing content and negatively impact their psychological wellbeing. Or, crawling morally questionable websites from a home network could cause an ISP to (incorrectly) make inferences about the researcher that may not be true or that may be undesirable to the researcher.
  - Research teams are expected to articulate how they considered the wellbeing of their researchers and the steps taken to mitigate identified risks as well as what risks remained unmitigated.
  - It should be clear that the team's decision accounts for power dynamics, e.g., that the most junior people on the team were empowered to make decisions that were right for them

https://news.stanford.edu/stories/2025/08/ai-companions-chatbots-teens-young-people-risks-dangers-study
https://tech.cornell.edu/news/cornell-tech-opens-computer-security-clinic-for-victims-of-tech-enabled-intimate-partner-violence/

Think: we want to study how cybercriminals use generative AI to enable new crimes. What study methods we could use? What are the ethics risks and alternatives? How do we justify our decisions and tradeoffs?

https://www.youtube.com/watch?v=qX5hsuH2_QM

Discuss: Who are at-risk users?

2024 IEEE Symposium on Security and Privacy (SP)

SoK: Safer Digital-Safety Research Involving At-Risk Users

Rosanna Bellini* Emily Tseng* Noel Warford† Alaa Daffalla*
Tara Matthews‡ Sunny Consolvo‡ Jill Palzkill Woelfer§ Patrick Gage Kelley‡
Michelle L. Mazurek† Dana Cuomo¶ Nicola Dell* Thomas Ristenpart*

*Cornell Tech †University of Maryland ‡Google §JumpCloud ¶Lafayette College

***Abstract*—Research involving at-risk users—that is, users who are more likely to experience a digital attack or to be disproportionately affected when harm from such an attack occurs—can pose significant safety challenges to both users and researchers. Nevertheless, pursuing research in computer security & privacy (S&P) is crucial to understanding how to meet the digital-safety needs of at-risk users and to design safer**

In this paper, we systematize knowledge from the S&P and HCI research communities to develop pragmatic guidance about reducing risk of harm in the planning, execution, and sharing of digital-safety research involving at-risk users (i.e., *at-risk research* hereafter). Our guidance reflects a systemization of "good" practices based on an analysis of 196 academic works and oral histories from an expert panel

Some examples of at-risk groups

“We define a user(s) as being at-risk if they face an elevated likelihood of an attack to their digital safety, have factors that influence or exacerbate their chances of being targeted, and/or experience heightened harm as a result of a digitally-mediated attack”

- Survivors of intimate partner violence
- Political activist
- Identity based marginalization (e.g., queer, women, people of color....)

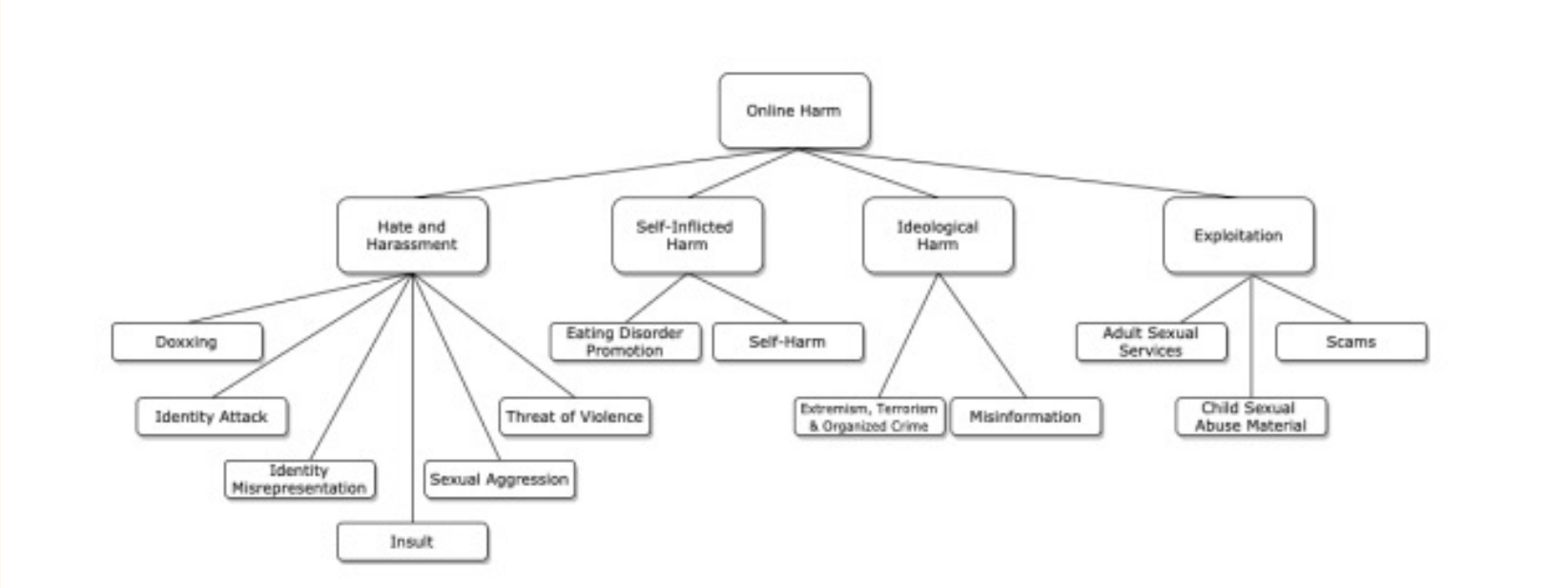
Research questions

- What digital-safety risks are associated with research involving at-risk users?
- What practices do researchers report employing to help mitigate digital-safety risk in at-risk research?
- What pragmatic guidance might researchers follow to reduce the risk of harm in their digital-safety research involving at-risk users?

Method

- Materials: 196 peer-reviewed papers in premier S&P and HCI venues after this initial dataset was collected - CCS, CHI, CSCW, IEEE S&P, NDSS, PETS, SOUPS, and USENIX Security
- Approach: qualitative coding and analysis

Agrafiotis, I., Nurse, J.R., Goldsmith, M., Creese, S. and Upton, D., 2018. A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate. *Journal of Cybersecurity*, *4*(1), p.tyy006.



A Unified Taxonomy of Harmful Content

What are the risks in research?

Risks posed			Description	Example papers
to participants	...from data collection	Breach of confidentiality	Researchers may be compelled to disclose participant data to an authority without participants' consent, due to subpoena, duties to law enforcement, or parental rights.	[26, 56, 58, 152]
		Unauthorized access	Even when using best-practice data-security tools, adversaries may gain unauthorized access to sensitive participant data.	[83, 85]
	...from direct research, including primary interviews or when researchers offer digital-safety advice	Coercion of contributions	Adversaries may accompany participants to studies and provide or discourage responses, especially when the adversary is an intimate (e.g., a partner, family member, or caregiver).	[44, 56, 88, 90]
		Disruption to support	Researchers may disrupt the normal functioning of digital-safety services and place a participant's security in jeopardy. Participants may also conflate research activities with service provision and feel compelled to participate in research to receive support.	[23, 43]
		Distress and re-traumatization	At-risk participants may be prompted to recount moments where they experienced digital-safety harms, which may cause distress. This can extend to viewing the researcher as a physical threat to a participant's wellbeing.	[12, 31, 44, 56, 137]
		Escalation of abuse	Research activities may require or encourage participants to break routines or take protective actions like removing spyware, which may incite adversaries to escalate their abuse or retaliate against the participant.	[56, 80, 85, 140]
		Withhold benefit	If researchers do not inform participants about the viability of reported threats or available protective practices, participants may be at greater risk.	[73, 113]
	...from the publication of research products	Adversarial feedback	Research may publicize protective strategies in ways that inform adversaries, who then correspondingly adapt or escalate their attacks.	[21, 26, 40, 44, 82, 138]
		Deanonymization	Unsuccessfully paraphrased quotes or poor redaction of participant information might reveal the identities of at-risk participants, particularly those who are public figures.	[34, 44, 45]
		Misrepresentation	Research may inadvertently mischaracterize participants' digital-safety needs, which may disrupt their safety strategies or encourage risky or ineffective interventions.	[83, 90, 118]

to researchers	Burnout and vicarious trauma	Immersion in stories of hate, harassment, and abuse may incur vicarious trauma or secondhand traumatic stress, which may result in burnout or exhaustion.	[11, 31, 43, 91, 100, 139]
	Harassment and intimidation	Researchers may themselves experience hate and harassment due to public statements about their research. Scholars with marginalized identities are particularly susceptible.	[12, 40]
	Liability exposure	Researchers may be subject to criminal prosecution or civil litigation for failing to disclose observed vulnerabilities (of at-risk groups or technical systems) uncovered during their research.	[26, 88, 144]
	Surveillance	Adversaries who have strategies for digitally tracking and monitoring at-risk groups may extend these tactics to researchers.	[104, 114, 121]

What are the practices?

What are the practices?

Category	ID	Digital-safety practices	Example papers
Professional partnerships & Ethical review	SP1	Elicit expert (academic) opinion on topic area	[17, 31, 67, 70, 82, 83, 112, 132, 136]
	SP2	Form professional partnerships (e.g., support services for at-risk users)	[44, 52, 72, 80, 82, 99, 105, 124, 134, 145]
	SP3	Invite and include an at-risk user to join research team	[17, 83, 97, 112]
	SP4	Seek external (non-institutional) ethical review approval or monitoring	[30, 43, 44, 78]
Positionality & Participant engagement	SP5	Build rapport with participants for understanding digital-safety needs	[1, 33, 34, 38, 73, 91, 97, 113, 137]
	SP6	Conduct pilot studies with general (non-at-risk) users	[5, 30, 33, 64, 67, 95, 101]
	SP7	Conduct studies with proxies for at-risk users (e.g., advocacy groups)	[2, 24, 33, 70, 74, 104, 132]
	SP8	Include researchers whose identities affirm participants'	[2, 6, 38, 64, 97, 110, 112, 113, 132, 134]
	SP9	Practice responsiveness in data collection sessions to potential threats	[3, 38, 49, 89, 100, 101, 124, 127, 128, 132]
	SP10	Provide professional therapeutic support for emotive topics	[7, 11, 30, 48, 95, 100, 101, 115, 144]
	SP11	Train team members in working with digital-safety risks	[7, 38, 115, 121]
Privacy-preserving data collection	SP12	Discourage participant self-disclosure (e.g., personal histories)	[1, 7, 25, 52, 70, 75, 118, 123, 137, 144]
	SP13	Focus data collection on supporting participant safety needs	[24, 34, 38, 66, 81, 97, 120, 121, 123, 129]
	SP14	Do not collect or ask for participant demographic data	[17, 26, 64, 83, 84, 104, 120, 124, 136, 145]
	SP15	Do not collect personally identifiable information on participants	[30, 43, 44, 52, 54, 58, 73, 85, 95, 143]
	SP16	Implement protocols for researchers to prevent stalking by adversaries	[30, 60, 80]
	SP17	Separate potential threats from at-risk users during data collection	[6, 72, 88, 96, 97, 100, 110, 115]
	SP18	Permit participants to contribute false information (e.g., pseudonyms)	[17, 54, 58, 78, 83, 100]
	SP19	Offer participants many modalities to contribute (e.g., audio, notes)	[4, 7, 24, 34, 57, 67, 90, 107, 117, 130]
	SP20	Secure confidentiality and privacy of online and in-person research sites	[6, 24, 30, 43, 44, 77, 100, 113, 134, 139]
Secure data storage & processing	SP21	Implement strict data access control measures for research data	[1, 7, 34, 51, 80, 112, 134, 136, 139, 147]
	SP22	Redact participant information prior to analysis by research team	[59, 86, 95, 107, 114, 128, 130, 140, 143, 156]
	SP23	Use encryption for research data in-transit and at-rest	[52, 60, 75, 85, 86, 87, 101]
	SP24	Use non-encrypted safe storage for research data in-transit and at-rest	[7, 30, 34, 90, 97, 114, 130, 132]
Researcher accountability	SP25	Conduct data collection sessions around participant schedules	[1, 35, 54, 65, 97, 111, 120, 128, 139]
	SP26	Offer formal proof of identity as professional researchers	[70, 82, 97, 112, 114, 115]
	SP27	Only use data from publicly accessible sites (e.g., no authorization)	[11, 32, 40, 97, 103, 138, 147, 155]
	SP28	Provide proportional incentives to participants for contributions	[54, 64, 72, 73, 82, 110, 134, 139, 145, 151]
	SP29	Be transparent with participants about risks incurred by research	[24, 26, 38, 54, 57, 69, 95, 110, 113, 128]
Sharing & evaluating deliverables	SP30	Do not attribute reported data contributions with participant identifiers	[7, 8, 9, 34, 55, 84, 114, 117, 134]
	SP31	Do not report participant demographics in research deliverables	[17, 24, 43, 77, 78, 83, 117, 120, 144, 145]
	SP32	Do not report participant names, pseudonyms, or identifiers	[9, 48, 71, 78, 101, 114, 121, 143, 145, 155]
	SP33	Paraphrase or withhold sources of data (e.g., websites they use)	[2, 9, 17, 40, 59, 69, 78, 123, 136, 155]
	SP34	Evaluate research deliverables for adversarial feedback or education	[34, 38, 44, 59, 82, 113]
	SP35	Selectively edit participant data in research deliverables	[7, 9, 11, 40, 55, 124, 139, 140, 150, 151]
	SP36	Provide participants control of their contributions (e.g., permit redaction)	[7, 47, 54, 75, 91, 113, 114, 117, 136]

Better practices?

Safer practices

ID	Strategy title	Description	Example digital-safety practices
S1	Engage experts early	Consult or partner with domain experts from the beginning to inform and help facilitate safe research plans.	SP1, SP2, SP3, SP4, SP10
S2	Assess and mitigate risks by threat modeling	Apply the S&P practice of threat modeling to research protocols, and continuously update threat models to guide ongoing safety mitigations.	SP11, SP16, SP17, SP20
S3	Select the lowest risk method that addresses the research goals	Before soliciting at-risk users for high-touch methods like interviews, consider proxies (e.g., advocates), or indirect methods (e.g., online measurement).	SP6, SP7, SP12, SP14, SP15, SP27
S4	Respect that at-risk users self-manage risk	At-risk users are often experts in managing their safety risks. Give them choice in how they engage with research safety protocols, and respect the choices they make.	SP9, SP18, SP19, SP25, SP26, SP29
S5	Be an advocate for at-risk users' needs	Research, by its nature, can be extractive. Build reciprocity with at-risk users, and work to help them achieve their goals.	SP5, SP8, SP13, SP28, SP36
S6	Handle data and publications carefully	Data collection and analysis should follow security best-practice, and publications should avoid revealing identities or informing adversaries.	SP21, SP22, SP23, SP24, SP30, SP31, SP32, SP33, SP34, SP35