








# SoK: A Privacy Framework for Security Research Using Social Media Data

Kyle Beadle<sup>1</sup> , Kieron Ivy Turk<sup>2</sup> , Aliai Eusebi<sup>1</sup>, Mindy Tran<sup>3</sup> ,  
 Marilyne Ordekian<sup>1</sup> , Enrico Mariconti<sup>1</sup> , Yixin Zou<sup>3\*</sup> , and Marie Vasek<sup>1\*</sup>   
<sup>1</sup>University College London <sup>2</sup>University of Cambridge <sup>3</sup>Max Planck Institute for Security and Privacy

**Abstract**—The use of social media data in research is common, spanning fields from computer science to social science, from human-computer interaction to law and criminology. However, social media data often contains personal and sensitive information. While prior work discusses the ethics of research using social media data, focusing on ethics broadly can be insufficient to unravel granular privacy risks and possible mitigations. Focusing on research papers that use social media data to study security-related topics, we systematically analyze 601 papers across 16 years, covering a wide array of academic disciplines. Our findings highlight a lack of transparency in reporting — only 35% of papers mention any considerations of data anonymization, availability, and storage. Applying Solove’s taxonomy to classify the identified privacy risks in the social media setting, we observe that Solove’s taxonomy was prescient in capturing aggregation risk, but the volume, timeliness, and micro details of data, combined with modern data science, yield risks beyond what was considered 20 years ago. We present the implications of our findings for various stakeholders: researchers, ethics boards, and publishing venues. While there are already signs of improvement, we posit that some small behavioral changes from the academic community may make a big difference in user privacy.

## 1. Introduction

From measuring the spread of COVID-19 [40], predicting crime [2] to analyzing the impact of political activists [88], social media provides researchers with rich and expansive data to understand prominent events and human behaviors. Typically, researchers collect this data via API access, web scraping, or third parties which varies in both form and content. Platforms like Twitter and Reddit used to have open APIs, which further fostered easy data collection. Even for those using Reddit alone as a data source, Proferes et al. found 727 manuscripts from 2010 through 2020 [61].

Easy access to social media data begets access to sensitive information. While social media users often discuss mundane topics such as entertainment and travel, disclosures can delve into intimate personal stories related to sensitive topics such as sexual abuse [6], pregnancy loss [4], and gender transition [31] — both anonymously [48] and in identifying ways [67]. Disclosures are even more common among vulnerable and marginalized populations who

perceive social media as a safer space with a sense of community belonging [5], [6], [32], [58]. Even if the data disclosed is not seemingly sensitive, it is still possible for privacy attacks to infer attributes [28]. Social media users have diverse and nuanced views of researchers’ use of their data [38], with a majority unaware of such practices [22] yet advocating for the importance of consent [22].

In this SoK, we examine the usage of social media data within security research as a case study to discuss how researchers can better navigate the tensions between pursuing better science enabled by social media data and protecting social media users’ privacy. We focus on security research for two reasons. First, security research can involve sensitive topics such as misinformation, harassment, and abuse [78]. Second, security research reveals vulnerabilities that impact companies and users. The adversarial nature of security research means that security researchers should be held to higher standards when handling privacy issues.

Our SoK is guided by the following research questions:

- 1) How do security researchers handle privacy of social media data?
- 2) What privacy risks emerge from security research using social media data?
- 3) How do security researchers mitigate privacy risks?

Our research contributes to the security and privacy literature in three ways. (1) A cross-disciplinary comparison: Our SoK is grounded in 601 security research papers, spanning across six disciplines and 16 years, and screened out of an initial dataset of over 10k papers. (2) A framework of specific privacy risks, possible mitigations, and tradeoffs in light of Solove’s taxonomy of privacy [72]: We find that privacy considerations are insufficiently discussed in our analyzed papers; those that report privacy considerations focus more on data anonymization techniques rather than data availability or data storage. Moreover, Solove’s taxonomy requires adaptations to tackle specific privacy challenges that social media creates for data dissemination and intrusion. The mitigations for identified privacy risks are also not straightforward and often introduce nuanced tradeoffs. (3) A call to action for various stakeholders: We discuss our findings’ implications for researchers, institutions, and publication venues, highlighting possible ways forward for addressing data privacy issues — alongside broader research ethics issues — for research using social media data.

\*. Both authors advised this work equally.

## 2. Background

We define key terms then review related work on research ethics concerning public/social media data.

**Social media.** Social media, or social networking sites, is defined by boyd and Ellison [13] as “web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.” Our work refers to Wikipedia’s list [85] and ensures each platform matches boyd and Ellison’s definition [13].

**Sensitive data.** The EU’s GDPR is a framework for thinking rigorously about privacy and unintended harms. It is a canonically acceptable way to evaluate privacy harms in the literature; its ideals appear in laws worldwide. We refer to Article 9 of the GDPR’s definition of *sensitive* data, i.e., “personal data [as] revealing racial or ethnic origin, political opinions, religious or philosophical beliefs; health-related data; data concerning a person’s sex life or sexual orientation” [20]. This complements Article 4, which lays out *personal* data as any information relating to an identified or identifiable individual, or factors relating to physical, cultural, or economic identity. The GDPR typically necessitates clear consent for processing *sensitive* data [79].

**Ethical frameworks and practices.** Prior work provides various frameworks [93] and recommendations to guide the ethics of research using online data [14], [82], [90], [93], publicly available data [15], [63], or, specifically, social media data [42], [52], [91]. The recommended privacy-preserving practices include aggregating findings and anonymizing data to prevent reverse identification [8], [14], [49]; paraphrasing quotes [10]; receiving prior consent when spotlighting or spotlighting public figures if consent is impossible [42]; and keeping original links instead of downloading artifacts to preserve data subjects’ right to edit or remove their content [12].

On the ethical implications of collecting social media data, Fiesler et al. prompted researchers to make individual judgments based on specific circumstances rather than solely relying on platforms’ terms of services [21]. Drawing from interviews with social computing researchers, Vitak et al. recommend researchers be transparent with participants, have ethical deliberations with colleagues, and exercise caution in sharing results [82]. Other researchers have evaluated the ethical boundaries of using datasets of illicit origin, such as hacked data, arguing that while it is generally ethically problematic, exceptions can be made [35], [77].

In light of the ethical considerations, norms around academic publications and research infrastructure are also evolving. A growing number of research venues, such as NeurIPS, IEEE S&P, and USENIX Security, have required authors to include ethics statements and established dedicated research ethics committees [7], [34], [80]. In the United States, approvals from Institutional Review Boards (IRBs) are required for research with human subjects that receive federal funds. Nevertheless, research using social

media data often gets exempt from IRB review [81], and because of IRB’s focus on legal compliance, having IRB approval alone does not guarantee ethics [71].

**Public perceptions.** Social media users’ own opinions matter when it comes to research using their data. Prior work demonstrates knowledge gaps in social media users’ understanding of what is “public” [60] and awareness of their data being used for research purposes [22]. While users generally expect researchers to seek consent and anonymize their data [86], their acceptance and perception of data use are shaped by the study’s topic [22], purpose [11], analysis method [38], positionality [22], [38], and the platform [27]. Users express concerns about their data being used in unintended ways [11], [86], being distorted [11], or causing possible harm if not anonymized properly or presented to the wrong audience [19]. Users with marginalized identities, such as Black Twitter users [38] and LGBTQ+ communities [19], face more severe consequences from privacy violations (e.g. harassment or being accidentally outed) and are more likely to find researchers collecting their public data to be intrusive [38]. Some subreddit communities (e.g. r/gamergirls, r/IndianCountry, and r/Drugs) even have explicit rules addressing or opposing research requests.

**Related SoKs.** Similar SoKs analyze research using social media data, although with a different angle (e.g., ethics more broadly) and/or scope (e.g., focusing on one particular platform). Both Proferes et al. [61] and Fiesler et al. [23] analyze research ethics around research using Reddit data. While Reddit includes many small communities discussing sensitive topics, less than 14% of papers in Proferes et al.’s dataset ( $n=727$ ) mentioned IRB or ethics reviews [60]. Fiesler et al. uncover few ethics statements with deeper reflections; most offered justifications for the research, described related methodological decisions, or raised ethical concerns but without offering solutions [23]. Zimmer and Proferes create a typology of research using Twitter data between 2007 and 2012 ( $n=382$ ) across disciplines, methods, and ethics, finding that only 4% of the corpus mentioned any ethical considerations [92]. Focusing on empirical studies in human-computer interaction (not necessarily those relying on social media data), Niksirat et al. compared samples from CHI 2017 and 2022 and found marginal improvements in transparency, but not research ethics and openness [68].

Our work expands on these prior SoKs by providing a more recent, cross-platform analysis of security research using social media data, including platforms emerging after Twitter and Reddit, such as TikTok. Rather than focusing on ethics more broadly, we present a comprehensive framework of specific *privacy* risks in light of Solove’s taxonomy [72], outlining how the risks manifest as well as potential mitigations and tradeoffs. Our focus on security research and privacy risks makes this SoK particularly relevant to the S&P community. While we focus on research addressing topics related to security, our paper selection encompasses other subfields of computer science (e.g., data science and machine learning) as well as humanities and social science literature, enabling us to identify the prevalence of issues across a wide array of disciplines.

### 3. Methodology

We perform a systematic review of the security literature on social media using four coders. We follow the approach of Wolfswinkel et al. [87] to define our review’s thematic scope, conduct the literature search, and select papers. We lean on the work from Proferes et al. [61] for the specific information to extract from papers.

#### 3.1. Identifying Relevant Work

We seek to include any full-length academic research papers that (i) are written in English, (ii) use data collected from social media, and (iv) cover topics related to computer security. We search for potentially relevant papers across 120 venues. We curate our list of venues by selecting the top 20 venues ranked by h5-index in each of the following six categories as of August 2023, using Google Scholar’s *top publications* list<sup>2</sup>:

- Computer Security and Cryptography (CSC)
- Data Mining and Analysis (DMA)
- Human-Computer Interaction (HCI)
- Humanities, Literature & Arts, Communication (HLAC)
- Social Sciences, Criminology (SSC)
- Social Sciences, Forensic Science (SSFS)

Within the selected venues, we perform keyword searches using two sets of keywords, one on computer security — adapted from the 2024 IEEE S&P topics of interest — and the other on social media (see Table 1). Our search across 4 academic databases yields 10,363 publications.

#### 3.2. Filtering and Final Dataset

We check all initial papers against our inclusion criteria and scope of analysis (overview of our filtering process in Figure 1). We first remove 3,706 duplicates caused by overlaps in the four databases, yielding 6,657 papers.

For the remaining papers, we did a round of abstract-based filtering and excluded papers that match the following criteria: (1) empirical user studies<sup>3</sup> (e.g., surveys, interviews, and focus groups) without analyzing social media data; (2) literature reviews; (3) position papers; and (4) papers based on data from video games and other gaming platforms (e.g., in-game chats and messages, as the data is not supposed to be public). To ensure reliability in the filtering process, three researchers use Rayyan [57], a literature management platform, to collaboratively filter 6,657 papers.

1. We did not search IEEE Access because it is indexed by both Scopus and Web of Science.

2. A complete list of included venues can be found at: [https://osf.io/v9ycf/?view\\_only=015aad64503e421b983c0fae2fa410c5](https://osf.io/v9ycf/?view_only=015aad64503e421b983c0fae2fa410c5)

3. While some similar SoKs such as Proferes et al. [61] included papers that involve user studies, we deliberately exclude papers from this category since informed consent is usually explicitly given. With our work’s focus on privacy (rather than ethics), we wanted to understand how prior work has tackled the difficulty of obtaining (informed) consent at scale as well as the risks of users not knowing they are being observed.

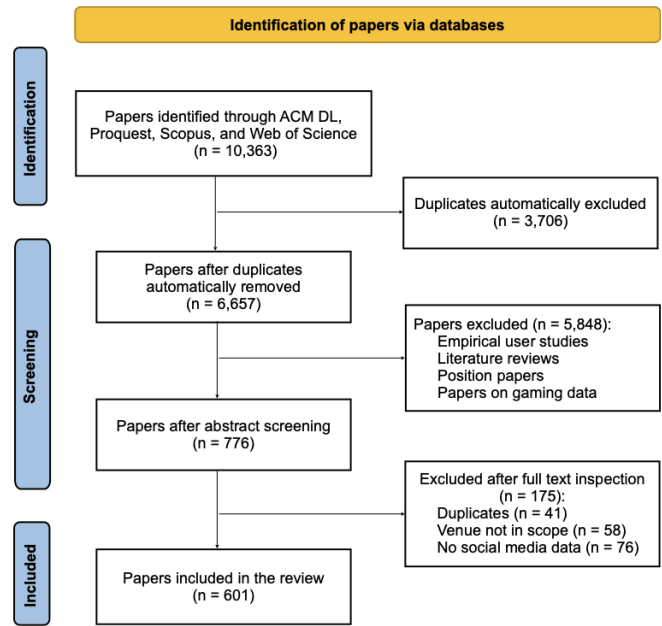


Figure 1. Flowchart for how we filtered out irrelevant publications and assembled the final corpus. Model adapted from Haddaway et al. [30].

Each researcher individually coded the same set of 100 papers, marking each paper as relevant or not after reading the abstract, and then compared their decisions until reaching a consensus (Fleiss kappa=0.84, representing almost perfect agreement [41]). The three researchers collectively made minor adjustments to the inclusion/exclusion criteria; each researcher then independently reviewed their own subsets of papers and discussed uncertain cases with the others. We exclude 5,848 papers after abstract review. We code each of the remaining 776 papers in-depth and exclude another 175 papers for violating our inclusion criteria upon reading the full text.

Our final dataset includes 601 papers.<sup>4</sup> Figure 2 shows the historical trends of these papers by venue categories. Most papers in our dataset come from DMA (327; 54%), followed by HLAC (113; 19%). Surprisingly, papers from CSC only make up for 16% of papers in our dataset, indicating the broad existence of security research leveraging social media data beyond traditional security venues. Figure 4 reports the platform distribution in our dataset, including all platforms appearing at least five times. Papers using Twitter data make up almost half of our dataset, making Twitter by far the most represented platform.

#### 3.3. Data Extraction and Analysis

Our goal is to identify privacy risks from research analyzing social media data. These risks can emerge from the topic and population selection, platform and dataset choice,

4. All included papers can be found at: [https://osf.io/v9ycf/?view\\_only=015aad64503e421b983c0fae2fa410c5](https://osf.io/v9ycf/?view_only=015aad64503e421b983c0fae2fa410c5)

TABLE 1. SUMMARY OF LITERATURE REVIEW SEARCH METHODS AND INITIAL RESULTS.

| Search Criteria   | Database <sup>1</sup> | # Results |
|---|-----------------------|-----------|
| In body text: cybersecurity OR information security OR cybercrime<br>OR cybersafety OR security OR network security OR privacy OR authentication<br>OR anonymity OR attack OR abuse OR illicit OR illegal OR fraud OR risk* OR harassment<br>OR hate OR trust OR safe spaces OR toxic OR sexism OR racism OR disinformation OR<br>manipulation (computer security search terms)<br>AND<br>In body text: social media or social network or social network site or online site or online<br>service or online platform or online site or online group or community or forum* (social<br>media search terms)<br>AND<br>(all included venues) | Proquest              | 2781      |
| same as above   | Web of Science        | 2493      |
| same as above   | Scopus                | 4691      |
| same as above   | ACM DL                | 398       |

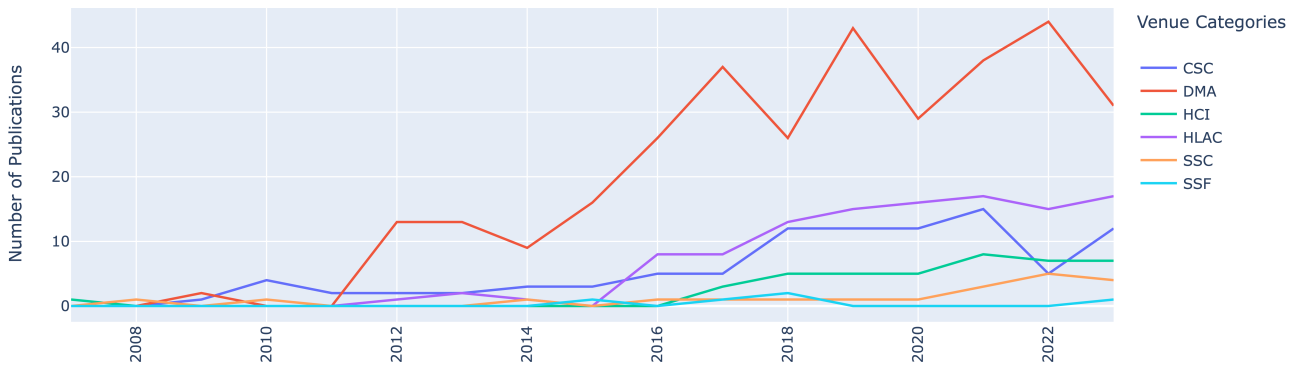


Figure 2. Publication trends by venue category over time.

granularity of the data, and how the data was analyzed and reported. Using these parameters as high-level categories, three researchers inductively build a codebook by analyzing a subset of 40 papers. The key parameters we extracted and coded from each paper include the studied topic, population, platform, data collection method, data analysis method, and description of ethics (in Appendix A). For parameter choice, we refer to prior similar SoKs (such as Proferes et al. [61] on ethics of research with Reddit data) and ensure that all parameters have privacy implications.

The three researchers discussed their initial coding results, reconciled disagreements, and iteratively refined the codebook. We then implemented a comprehensive template in Qualtrics to facilitate standardized extraction and analysis of data between multiple researchers. Subsequently, five authors tested inter-rater reliability on a new subset of 30 papers and reached substantial agreement [41] (Fleiss kappa=0.69 averaged across all questions). The researchers then split and coded the remaining papers.

**Mapping to Solove’s Taxonomy.** Inspired by Lee et al.’s taxonomy of AI privacy risks [43], we ground our secondary qualitative analysis and presentation of findings in Solove’s taxonomy of privacy [72]. Solove’s taxonomy is heavily based on legal scholarship, with the government as the key threat actor. The taxonomy was developed before the rise of social media. With Solove’s taxonomy, we aim

to identify which risks are fundamentally tied to the very notion of privacy risks, agnostic to any specific technological context, and which risks are uniquely enabled by research with social media data.

Similar to Lee et al. [43], we iteratively refine and adapt Solove’s taxonomy [72] to suit our research purpose. Table 2 provides the exact mapping. Specifically, “Interrogation,” “Breach of Confidentiality,” “Secondary Use,” “Exposure,” and “Appropriation” are the five risks included in Solove’s taxonomy [72] but not in our analysis. We exclude these risks based on their lack of applicability to the social media context and overlap with other existing risks.

“Interrogation” is excluded because it mostly applies to research that presents users with their social media data to gauge their reactions. Such research usually receives more ethical and privacy scrutiny from IRBs, following a clear procedure, compared to research using social media data, which is viewed as a gray area by IRB members [33].

“Secondary Use” is excluded because it is a generic consequence of increased accessibility of social media data and not a standalone risk that varies between different research. The accessibility of datasets, as well as the dissemination through publications, conferences, and press, all allow for secondary use of social media data.

“Appropriation” is excluded because it is a subjective call whether social media datasets or research results ul-

timately benefit those who are analyzed. Researchers are incentivized to argue that their research is for the public interest. Users are rarely able to provide opinions on social media research as they are rarely made aware of it. Risks that encapsulate the harm produced by “Appropriation” include “Disclosure,” “Distortion,” “Increased Accessibility,” “Intrusion,” and “Decisional Interference.” These risks similarly address the agency taken away from the exposed users or communities to control the observation, recording, and analysis of their data.

“Breach of Confidentiality” is excluded because the harms experienced are similar to those of “Disclosure” within a social media context. Prior work shows that users have different expectations of privacy depending on the location of disclosure [22], [50]. Users in public or semi-public online spaces may experience “Disclosure” that they were not expecting. Users in private online spaces may experience “Breach of confidentiality” when information from their private group is exposed. However, no papers in our dataset analyze data from private groups.

Our incorporation of Solove’s taxonomy is also reflected in the parameters of our data extraction (see Appendix A). This mapping is qualitative, so each parameter does not exclusively correspond to only one risk. Some parameters correspond to multiple risks, e.g., Q15 about the reporting of examples corresponds to both “Disclosure” and “Black-mail.” Other risks correspond to multiple parameters, e.g., “Surveillance” risk can manifest from whether marginalized populations are analyzed (Q7), use of existing datasets (Q10), types of data collected (Q13), and dataset size (Q14). We explain these mappings when applicable in §4.

### 3.4. Ethical Considerations

This study received a review and an exemption by the UCL Department of Security and Crime Science ethics committee. For papers, we collect the article URL, the year, the authors, the title, and the journal. This is made public in our online repository to ensure reproducibility. For our coding, to avoid aggregation risks, our data is stored on Qualtrics with its access control scheme. Only authors conducting the analysis have access to this data.

Our work does not aim to name and shame scholars in any of the disciplines mentioned. Rather, we review the literature to understand whether communities across disciplines follow precautions and consider the privacy risks of carrying out security research using social media data. We include both positive and negative examples in our findings. The positive examples demonstrate concrete ways forward for researchers to proactively consider privacy when using social media data. The negative examples are not to single out authors, but rather to identify how specific practices and communities may fall short.

## 4. Privacy Risks

Across all papers, a lack of reporting around privacy risks and mitigations is a prominent issue: only 35%

( $n=209$ ) discuss any considerations of data anonymization, availability, and storage issues. There is more transparency among the 65 papers that study marginalized and vulnerable groups, as 46% ( $n=30$ ) of them provide the considerations.

Table 3 further shows the breakdown of reporting by discipline. We find that reporting is uneven across disciplines: 78% ( $n=32$ ) of HCI papers, 69% ( $n=66$ ) of CSC papers, but only 20% of HLAC ( $n=23$ ) papers describe data privacy procedures. Figure 3 displays the lack of reporting increasing over time disproportionately to reported privacy or ethics considerations, without any improvement after the adoption or enforcement of GDPR.

TABLE 3. OCCURRENCE OF DATA PRIVACY REPORTING BY VENUE.

|                    | DMA | HLAC | CSC | HCI | SSC | SSFS |
|--------------------|-----|------|-----|-----|-----|------|
| Data Availability  | 66  | 4    | 26  | 6   | 1   | 0    |
| Data Anonymization | 83  | 18   | 32  | 19  | 6   | 1    |
| Data Storage       | 17  | 1    | 8   | 7   | 1   | 0    |
| <b>Total</b>       | 166 | 23   | 66  | 32  | 8   | 1    |

Adapting Solove’s taxonomy to our findings, we observe that while Solove’s taxonomy is still useful for understanding and mitigating risks related to information collection and processing, research using social media data requires context-specific considerations for risks during information dissemination and invasion. Information dissemination risks are exacerbated by datasets of social media data becoming accessible and being published at large volumes. Moreover, researchers may be viewed as threat actors and may disrupt dynamics in online communities from which they collect data. We next elaborate on the risk manifestation, mitigation, and tradeoffs for each of the 11 privacy risks.

### 4.1. Surveillance

We define surveillance as *collecting, aggregating, and analyzing social media data that users are not aware of*. What makes surveillance uniquely interesting in the social media context is the uncertainty users face, not knowing whether their online activity is monitored without their knowledge. We can also quantify surveillance by the size of datasets (Table 5). Another metric relating to surveillance is the types of data collected, with text data (76%;  $n=457$ ), profile data (35%;  $n=210$ ), and metadata (31%;  $n=187$ ) being the three most common types in our dataset.<sup>5</sup>

**Risk Manifestation.** Surveillance can happen even through seemingly mundane activities. In the example of WeChat, one can use networks of red packets (a traditional form of monetary gift in many east Asian cultures) to identify the relationships between users, which increases the risk of users being surveilled through their WeChat activity [84]. APIs, third-party data sources, and scrapers also enable the surveillance of social media users. In terms of data collection methods, 36% ( $n=215$ ) of papers use existing datasets, 35%

5. The aggregated percentages sometimes go beyond 100% when the question allows the selection of multiple options (in this case, one paper could collect multiple data types).

TABLE 2. MAPPING OF PRIVACY RISKS TO SOLOVE'S TAXONOMY

| Solove's Taxonomy [72]  | Social Media Data Risk   | Explanation   |
|---|--|---|
| <b>Information Collection</b>   |  |   |
| Surveillance<br><i>"the watching, listening to, or recording of an individual's activities"</i>   | Collecting, aggregating, and analyzing social media data which users are not aware of.       | Even though social media data is public, data collection methods, such as web crawlers and APIs, enable the recording, studying, and storage of user's data beyond their intended purpose.  |
| Interrogation<br><i>"various forms of questioning or probing for information"</i>   | N/A  | People may be subject to interrogation on social media, such as having to answer for a post others disagree with, or someone may have an image or video of them posted online without their consent. However, researchers indiscriminately accumulate information.  |
| <b>Information Processing</b>   |  |   |
| Aggregation<br><i>"the combination of various pieces of data about a person"</i>  | Combining social media data to make inferences about individuals and groups.                 | Different social media platforms have different contexts. What is expected in one may be prohibited in another. Aggregation demolishes the boundaries between different platforms and enables more sensitive inferences to be made.   |
| Identification<br><i>"linking information to particular individuals"</i>  | Linking social media activity to a specific user account or offline identity.                | Social media sites implement identification in a variety of ways: real-names, anonymity, and pseudonymity. Collecting data across platforms can lead to the identification of a user through methods via machine learning and stylometry.   |
| Insecurity<br><i>"carelessness in protecting stored information from leaks and improper access"</i>   | Storing social media data irresponsibly.   | Despite its public nature, researchers should handle social media data as other private information due to other privacy risks such as surveillance and identification.   |
| Secondary Use<br><i>"the use of information collected for one purpose for a different purpose without the data subject's consent"</i>                 | N/A  | While collected social media data exposes users to risk through the ability to cross-reference multiple datasets, social media collapses information processing and information dissemination. Secondary use becomes a consequence of the increased accessibility of social media data and not a standalone risk. |
| Exclusion<br><i>"the failure to allow the data subject to know about the data that others have about her and participate in its handling and use"</i> | Failing to notify users or include their feedback into the processing of social media data.  | When researchers only treat social media as bits, they strip users of autonomy over their own data.   |
| <b>Information Dissemination</b>  |  |   |
| Breach of Confidentiality<br><i>"breaking a promise to keep a person's information confidential"</i>  | N/A  | Most social media data isn't private; it is public or semi-public. As a result, researchers often cannot breach user's confidentiality.   |
| Disclosure<br><i>"the revelation of truthful information about a person that impacts the way others judge her character"</i>                          | Revealing the contents of collected social media data.                                       | Sharing social media data in a publication may give increased attention to a private subject matter and may release sensitive information that embarrasses.   |
| Exposure<br><i>"revealing another's nudity, grief, or bodily functions"</i>   | N/A  | Social media flattens the risk of exposure and disclosure. On social media, user posts are not categorized by how sensitive they are considered. Every data point can be treated equally during research. Exposure becomes a sub-risk of disclosure.  |
| Increased Accessibility<br><i>"amplifying the accessibility of information"</i>   | Providing easy access to social media data and datasets.                                     | Collected social media data may remain publicly available for years after collection, exposing users in that data to "potential future risk."   |
| Blackmail<br><i>"the threat to disclose personal information"</i>   | Using social media as a means or motivation to threaten or damage a user or researcher.      | Social media data may often include embarrassing or sensitive information. Researchers risk being blackmailed or enabling blackmail when they disseminate their outputs. This risk is exacerbated when combined with other risks, such as aggregation and identification.   |
| Appropriation<br><i>"the use of the data subject's identity to serve the aims and interests of another"</i>   | N/A  | While social media data is being used for the aims and interests of researchers, these interests are largely subjective. The disruption that dissemination of social media research causes is closer to disclosure and distortion risks.  |
| Distortion<br><i>"the dissemination of false or misleading information about individuals"</i>   | Intentionally or unintentionally misrepresenting a phenomenon observed in social media data. | The way researchers analyze or disseminate their research may impact perceptions of users or communities. It is important for researchers to reflect how their biases and decisions impact their study.   |
| <b>Invasion</b>   |  |   |
| Intrusion<br><i>"invasive acts that disturb one's tranquility or solitude"</i>  | Impacting the interpersonal relationships of users and how platforms interact with them.     | The Internet allows researchers to easily find any online community or groups. However, entering these digital spaces without proper training or experience increases the risk that research disturb users.   |
| Decisional Interference<br><i>"the government's incursion into the data subject's decisions regarding her private affairs"</i>                        | Impacting the interpersonal relationships of users and how platforms interact with them.     | Beyond disturbing online communities or groups, researchers may directly affect the relationships people have with one another and the social media platform. Researchers may create or exacerbate existing tensions within groups or highlight ways that platforms can further marginalize vulnerable voices.    |



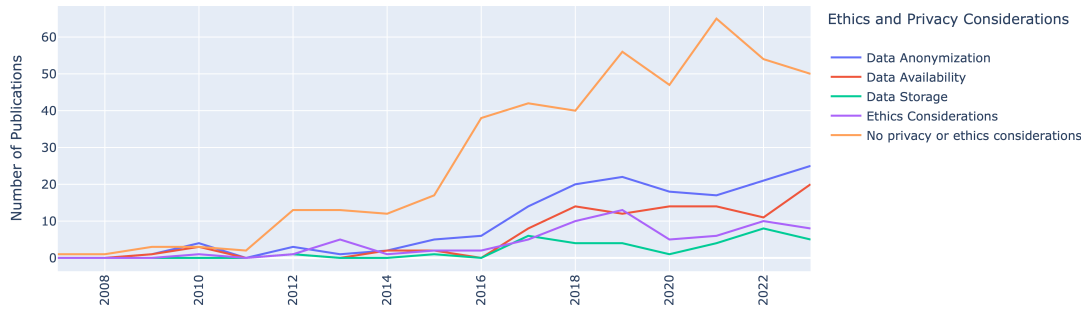


Figure 3. Reporting of ethical and privacy considerations over time. GDPR was adopted in 2016 and enforced starting in 2018.



Figure 4. Distribution of ethics and privacy considerations by venue. The size of each bubble represents the absolute frequency of papers reporting a specific ethics/privacy consideration within each venue category. The color intensity indicates the percentage of papers within a venue category that reported each consideration.

TABLE 4. TOP 18 PLATFORMS STUDIED IN OUR DATASET.

| Platform    | Count (%)   |
|-------------|-------------|
| Twitter     | 296 (49.2%) |
| Facebook    | 108 (18.0%) |
| Forums      | 73 (12.1%)  |
| Reddit      | 47 (7.8%)   |
| YouTube     | 37 (6.2%)   |
| Instagram   | 26 (4.3%)   |
| Flickr      | 20 (3.3%)   |
| Weibo       | 19 (3.2%)   |
| Epinions    | 19 (3.2%)   |
| Google+     | 15 (2.5%)   |
| Foursquare  | 15 (2.5%)   |
| LiveJournal | 13 (2.2%)   |
| Slashdot    | 6 (1.0%)    |
| TikTok      | 6 (1.0%)    |
| Gowalla     | 6 (1.0%)    |
| Orkut       | 5 (1.0%)    |
| LinkedIn    | 5 (1.0%)    |
| Advogato    | 5 (1.0%)    |

TABLE 5. FREQUENCY OF DATASET SIZES

| Dataset Size      | Count (%)   |
|-------------------|-------------|
| 5,000,001+        | 139 (23.1%) |
| 500,001-5,000,000 | 112 (18.6%) |
| 50,001-500,000    | 118 (19.6%) |
| 5,001-50,000      | 95 (15.8%)  |
| 501-5,000         | 63 (10.5%)  |
| 51-500            | 25 (4.2%)   |
| 1-50              | 5 (1.0%)    |
| Not Reported      | 44 (7.3%)   |

( $n=211$ ) use APIs, and 21% ( $n=127$ ) use scraping/crawling to collect social media data. Surveillance can amplify risks of experiencing coercion, discrimination, and chilling effects for social media users who are vulnerable and marginalized, 11% ( $n=65$ ) of papers in our dataset.

**Risk Prevention.** Kim et al. demonstrate how to mitigate surveillance risks when studying the social networks of migrants on Twitter; they conducted a privacy risk analysis to validate that their Twitter data can be safely used without exposing sensitive information of the users and minimize risks of re-identification [37].

**Trade-offs.** One challenge to preventing surveillance is researchers’ need for data granularity. Removing usernames, timestamps, location, and profile data does enhance privacy, but it limits the types of analysis researchers can conduct. Using synthetic data may mitigate surveillance, but it is not always appropriate. User trust is necessary to mitigate surveillance; users must feel confident that researchers are not violating their privacy. However, building trust will challenge researchers’ time constraints.

## 4.2. Aggregation

We define aggregation as *combining social media data to make inferences about individuals and groups*. Users have bounded rationality [1], and their decisions on what to share are usually bounded by the thinking process toward the particular platform. When users post on separate platforms, they may share enough information to link the online profiles (e.g. by using the same username) and researchers can aggregate this data to create a larger profile of the user.

**Risk Manifestation.** The reuse of existing datasets increases the risk of aggregation. Aggregated Netflix user data with a reused IMDB movie ratings dataset aided researchers to uncover sensitive information about Netflix subscribers [53]. In our dataset, 36% ( $n=215$ ) of the papers rely on existing datasets. Notably, researchers can extract personal information even from anonymized datasets. The aggregation of four real-world datasets led to de-anonymization and privacy leakages in heterogeneous social networks [45]. Moreover, the risk is unevenly distributed due to disparities in data availability; when popular datasets go offline, researchers may selectively share data within

familiar groups. Aggregation-related risks are also particularly relevant to studies focusing on detection, prediction, or classification, as these often involve making inferences about users. Such analyses can reveal sensitive topics or private attributes about individuals, such as mental health status [16].

**Risk Prevention.** While anonymization of data is important, particularly when examining sensitive topics, more work needs to be done to minimize the risks of aggregation. For example, aiming to detect suicide risk, researchers [16] built a detection model using anonymized datasets from Weibo, then applied the approach to another collected Reddit dataset [24]; while the authors anonymized the Weibo dataset before labeling, re-identification is still possible from the cross-platform analysis. It is vital to consider the implications before conducting research. Researchers must also consider whether their use of data is something that would be expected by that person or group, specifically when investigating marginalized or vulnerable populations.

**Trade-offs.** One of the main challenges in preventing the risks of aggregation is that IRBs often do not review the use of third-party datasets. This oversight gap means that researchers may assume their use of the data is automatically ethical if it comes from public or previously studied sources. We note that IRBs require approval for data reuse through the same board approving the initial study, though this is often ignored. Researchers must go beyond institutional ethics review toward reflecting on the potential for harm in re-using or combining datasets.

### 4.3. Identification

We define identification as *linking social media activity to a specific user account or offline identity*. While some social media encourage anonymity and others maintain a real-name policy, having accounts across different platforms means these accounts can be related to the same user.

**Risk Manifestation.** The primary risk with identification is that pseudo-anonymous online users can be linked to offline identities. Information about a user, while not directly linked to their person, may be enough to identify them—for example, a user’s first name, gender, date of birth, and possibly nearest city are frequently shared on platforms such as Facebook, and are enough information to identify a single individual. Identification is further possible when researchers include exact quotes from Reddit posts [26] which can be easily linked to usernames. Researchers further heighten identification risk further when they aggregate data (§4.2). Distinct online accounts can be linked to a single user, e.g., by combining Facebook and Twitter to re-identify users based on locations and organizations from profiles and user-generated content [89].

Identification risk also depends on the study’s aims and analysis methods. Network analysis, constituting 22% ( $n=134$ ) of our sample, is prone to identification risks as networks have complex structures—often unique and can be identified on non-anonymized websites. Machine learning methods, 55% ( $n=332$ ) of papers in our dataset, are also

open to identification risks as researchers often deploy ML to make inferences across large datasets.

**Risk Prevention.** Identification is most commonly mitigated through data anonymization: removing personally-identifying metadata to ensure data cannot be linked to an online or offline user. Nevertheless, only 26% ( $n=158$ ) of papers reference data anonymization—either in the text of the paper or in the dataset used. Of those, 60% ( $n=95$ ) use anonymized data, 36% ( $n=57$ ) use non-anonymized data, and 7% ( $n=11$ ) use both anonymized and non-anonymized data. Disconcertingly, 11% ( $n=17$ ) used data where individuals can be re-identified. Data anonymization prevents trivial identification but does not prevent more complex identification techniques. Modern researchers can use data modification or perturbation, which maintains the main concepts or results of the data without providing an exact copy by modifying the data.

**Trade-offs.** Being able to link users creates an invasion of privacy. This is a drawback, as users have distinct online accounts to prevent the inferences we discuss. On the other hand, certain public figures will always be identifiable, or may be intrinsically identifiable based on study aims—for example, discussing a prominent politician’s posts. In these cases, avoiding identification provides little to no benefit while requiring more work from researchers.

### 4.4. Insecurity

We define insecurity as *storing social media data irresponsibly*. While researchers collect social media data under the justification of it already being public, scraped data should still be stored securely to prevent aggregated data leaks.

**Risk Manifestation.** We find only 6% ( $n=34$ ) of papers mention data storage. Of those papers, 50% ( $n=17$ ) use cloud storage while 38% ( $n=13$ ) use local storage; 12% ( $n=4$ ) of papers discuss storage without disclosing how their data was stored. Data stored in the cloud, e.g., using Amazon Web Services [69], increases the risk of data being unintentionally accessible over the Internet. Moreover, storing data on an exposed server without proper authentication may allow remote access to unauthorized users. Storing data locally may lead to unintentional modifications by other local users. A malicious user with access to local files may view or intentionally tamper with improperly stored data if the data is not encrypted properly. Among the papers that mention data storage, only three further mention their data is encrypted.

**Risk Prevention.** Standard storage practices should be adhered to: storing data locally where possible, encrypting all stored data, and limiting access to the minimal set of users who require access. If researchers require external access (e.g. for data sharing with an external organization), secure remote access must be configured and shared servers restricted behind firewalls to avoid unauthorized access to exposed services. Additionally, researchers should interact with the proper stakeholders to prevent governmental data seizures when appropriate. For example, when studying



Instagram direct messages from adolescents, Razi et al. locally stored the data on a secure server and obtained a “Certificate of Confidentiality” issued by the National Institute of Health, which would protect participant privacy and prohibit the data from being subpoenaed during the legal discovery process [64].

**Trade-offs.** Allowing Internet access to data is a modern necessity, and data sharing agreements with external researchers further prevents secure storage methods from being used. Researchers must consider the trade-offs between placing their data on physical servers or external hard drives. Using physical servers for data storage offers stronger security measures and controlled access, but may require significant resources and increase vulnerability to internal threats. Alternatively, storing data on external hard drives might reduce costs and enhance portability, yet increases the risk of data loss or theft due to mishandling. Including an audit process can help monitor and enforce secure data handling, but may slow down the research process and introduce additional administrative overhead.

#### 4.5. Exclusion

We define exclusion as *failing to notify users or include their feedback in the processing of social media data*. While users make their data public for anyone on the internet to observe, they cannot explicitly consent to this data being collected and used for research purposes. In this way, users are *excluded* from the process of data collection.

**Risk Manifestation.** Researchers may cause harm to vulnerable online communities if they do not have sufficient prior experience or skip consultation with these communities for the engagement. For instance, researchers studied online sex work using covert online ethnography and semi-supervised ML, without consulting with the community first [39]. Data subjects have the right to be forgotten under the GDPR. When researchers do not inform users that their data is being scraped and stored, users are unable to exercise their right to be forgotten.

Exclusion risk also comes from the method of data collection. Research using external datasets amplifies exclusion risk, as the researchers would still retain a copy of the data even if the data subject requests removal from the original version. When gathering data, there may be clauses in the terms of service (ToS) for the social media platform that state scraping their site is not allowed and/or constraints on the usage of data collected via the API. Researchers may not check for these clauses before collecting data and may violate the ToS. Only 2% ( $n=11$ ) of papers in our dataset discussed compliance with the relevant platform’s ToS.

**Risk Prevention.** One way to mitigate or prevent exclusion is to coordinate with social media platforms and users before collection. For example, Gong et al. consulted with GitHub before conducting data collection for their research on detecting malicious accounts among developer communities [29]. Another possible prevention is to publicize the option to be excluded from studies and be able to prove to users that their data has been removed.

**Trade-offs.** Researchers can avoid exclusion by ensuring all participants explicitly opt in to the study. However, it is incredibly difficult to achieve consent when scraping data from millions of users. It also makes publication of shared data an ethical concern, as future studies using shared data would need to contact all data subjects again to be able to maintain this consent. Alternatively, careful precautions can be taken to allow for a non-exclusionary opt-out model of data collection. These would require contacting data subjects to ensure they are aware that their data has been collected, which creates issues similar to the opt-in approach.

#### 4.6. Disclosure

We define disclosure as *revealing the contents of collected social media data*. In the context of social media research, disclosures happen when researchers share data online and in papers, such as through examples. We find that 39% ( $n=237$ ) of papers in our dataset include examples; among them, 68% ( $n=162$ ) use plain text examples and only 48% ( $n=113$ ) anonymize or paraphrase examples. A further 23% ( $n=55$ ) of those papers include images and 5% ( $n=11$ ) include censored images. The disclosure of sensitive information through revealing examples, such as plain text, is even more prevalent among research studying marginalized populations, as shown in Figure 5.

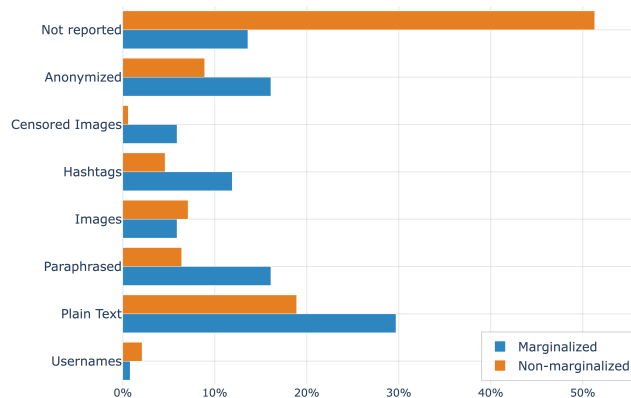


Figure 5. Examples used by publications studying marginalized and vulnerable populations

**Risk Manifestation.** Disclosure risk varies with different data types—disclosing images, location, and video data is riskier than direct quotes because it may lead to the user’s being identified. However, users behind exact quotes in papers can still be identified through a quick online search. Researchers amplify disclosure risk when researchers include direct usernames in publications. For instance, researchers published examples of usernames along with corresponding location data when researching query processing and authentication in social networks [83]. It requires more caution to deal with disclosure risk from research publications when studying vulnerable groups. When inferring eating disorder networks on Twitter, researchers exposed and drew attention

to sensitive users when publishing direct quotes from pro-eating disorder accounts and sharing the usernames of the most popular users in the network [55].

**Risk Prevention.** In light of disclosure risks from sharing specific details, an alternative could be focusing on the broader phenomenon. For example, Sun et al. examined concession-abuse-as-a-service by focusing on the actors and supporting services involved rather than describing how to commit the scam [73]. Another risk mitigation strategy is to refrain from disclosing in publications or paraphrasing quotes, particularly for sensitive topics and high-risk populations. In detecting suicidal ideation, Sawhney et al. paraphrased all posts they included in the paper, anonymized their Reddit data, and separately stored their annotation data from their raw data [70].

**Trade-offs.** When paraphrasing examples, it is important not to distort the example or lose its meaning. This can often be resource-intensive as it may require an additional auditor to assess the paraphrasing. Increasing barriers to conducting research may further the study of vulnerable populations, as they are already understudied.

#### 4.7. Increased Accessibility

We define increased accessibility as *providing easy access to social media data and datasets*. The risks of increased accessibility are associated with “potential future risk” that comes from the availability of existing datasets. We find that only 16% ( $n=98$ ) of all papers discuss data availability. Among them, only 39% ( $n=38$ ) use data that is publicly available and anonymized while 60% use data that is publicly available and revealing. Furthermore, 12 papers claim that their data is available upon request.

**Risk Manifestation.** Increased accessibility may lead to the recontextualization of data beyond the specific platform. As such, surveillance, aggregation, identification, insecurity, and exclusion are all exacerbated by increased accessibility. Researchers may also draw attention to privacy-violating datasets using existing datasets that were collected when norms around social media data were different. Researchers still use large dataset repositories [44], [51] despite data collection occurring while social media was still growing in popularity and user privacy was not yet regulated. Researchers may not update existing datasets after some users delete their data online. This intertwines with disclosure and blackmail risk as deleted data can still be linked to a live social media account.

**Risk Prevention.** It is essential to consider the original purpose for which the data was collected. Anonymized data collected may not be ethically appropriate for other uses. Researchers in the EU must also comply with GDPR, which mandates that data be stored only for as long as necessary for the intended research purposes. Storing data permanently may pose a significant privacy risk, as it increases the potential for unauthorized access and misuse over time. Conversely, in studying parenting of children with developmental disabilities using YouTube videos, Borgos-Rodriguez et al. demonstrate that an alternative could be only keeping

original links to videos rather than downloading the videos to preserve the uploader’s right to be removed [12].

**Trade-offs.** Restricting access to user data can impede the reproducibility of studies and counter open science. Without consistent access to the social media datasets, researchers may struggle to validate or replicate results. Additionally, ensuring user data privacy becomes challenging, especially when monitoring for deletions and take-down requests. While researchers might be able to respond to deletion requests (despite scalability issues), it is difficult to inform individuals once their data has been externally archived, potentially violating their expectations of control over personal information shared online.

#### 4.8. Distortion

We define distortion as *researchers’ intentional or unintentional misrepresentation of a phenomenon observed in social media data*. User-generated content usually comes with a purpose or motivation. An individual’s posting behaviors are influenced by advertisements, algorithms, and others in their social network. It is crucial to consider how social media data is generated and the ecosystem in which it is created. Distortion risk has privacy implications as it changes how people view an individual, a community, or a phenomenon, and distortion may also lead to other risks such as blackmail.

**Risk Manifestation.** Distortion risk can happen in a few ways, usually tied to the research methodology: self-selection bias, platform choice, and sampling. Self-selection bias refers to researchers choosing a digital platform or population that they believe best supports their hypothesis or pre-conceived notion of a phenomenon. The views of specific social media users may also not be representative of a larger population. Similarly, choosing a platform whose demographics do not match the demographics of the studied phenomenon will problematize the validity of results. For example, discrepancies in using social media data to model the risk of the Zika virus in Florida [25] may be a result of the average age of a Twitter user in 2019 [17] being lower than the average age of people in several Florida counties [56]. Non-random sampling will also affect the validity of results. For example, taking a convenience sample of social media data will lead to a phenomenon seeming more prevalent than it is.

**Risk Prevention.** One way to prevent distortion risks is to provide more context about the platform choice and limitations in the paper. Informing the reader of this context makes it clear whether the proposed social media platform of study is appropriate and how alternatives may not be better suited to attain the aims of a study. In justifying why studying a particular platform, 64% ( $n=384$ ) of papers noted the importance of platform features and demographics, 31% ( $n=187$ ) noted the platform’s popularity, 23% ( $n=141$ ) noted the dataset’s accessibility, 22% ( $n=133$ ) referred to prior literature, and 10% ( $n=63$ ) stated the platform was understudied. Notably, 11% ( $n=66$ ) papers did not state any reason.

Researchers can also present more context about a chosen platform by sharing its demographics and explaining how it is situated within the broader social media ecosystem. Becoming familiar with the data being analyzed, such as through manual review or exploratory data analysis is another means of preventing distortion risk. Transparently reporting data sampling and data cleaning methods can also mitigate the impact of distortion. These practices motivate a study along with more information for a reviewer to scrutinize potential biases in the results.

**Trade-offs.** Major obstacles to preventing distortion include access to data and the size of datasets. In our sample, the most studied platforms (see Figure 4) are usually the ones with more accessible and stable data collection tools or APIs. Even though the demographics of Twitter/X might not always be appropriate for a study's objectives, easy access to data incentivizes researchers to study the site. Furthermore, being familiar with the data is challenging and time-consuming when datasets are large.

#### 4.9. Blackmail

We define blackmail as *using social media data as a means or motivation to threaten or damage a user or researcher*. People change and regret the ideas that they once openly shared. Marginalized and vulnerable people use social media to discuss what is not safe for them to say offline. Yet, digital footprints are forever growing. Researchers risk enabling blackmail when disseminating their findings.

**Risk Manifestation.** Disseminating social media data through publications and public datasets exacerbates the risk of blackmail to both users and researchers. The risk of blackmail affects individual users when authors disclose their sensitive and potentially stigmatizing information such as medication usage [65]. Researchers' analysis and dissemination of traumatizing experiences may also take away a survivor's autonomy over their story and allow perpetrators to discover their online accounts, such as in the case of studying rape-related discussions on Reddit [36]. Conversely, researchers themselves may be physically or digitally threatened when they study sensitive topics such as online extremism [18].

**Risk Prevention.** Similar to increased accessibility and disclosure, preventing blackmail involves limiting access to data and ensuring that social media data cannot be traced back to an individual username or person. Withholding public access to data prevents data from resurfacing in the future. Moreover, Saha et al. protect LGBTQ+ people from being blackmailed for their sexuality by paraphrasing quotations [66]. Another mitigation can be using censored images, as done in Niu et al.'s study of drug-addiction videos on YouTube [54]. To protect researchers, venues could begin allowing researchers to publish under pseudonyms.

**Trade-offs.** Data sharing is important for open science and reproducibility. Providing data access upon request requires time-consuming management of databases, and data requests often go unanswered across disciplines [74]. Researcher pseudonymity would present problems for attribut-

ing credit during grant and job applications. Pseudonymity also increases administrative overhead for venues as they seek to verify authors' identities. Attributing quotes to people and groups may be unavoidable and may even empower those attributed.

#### 4.10. Intrusion

We define intrusion as *forcibly entering a cultural, experiential, and perspective-specific digital space*. Online communities allow people with similar interests and experiences to form bonds with one another and create tight-knit, niche groups. We find that 20% ( $n=120$ ) of all papers study some form of online community platforms, such as Discord and Reddit. The communities organize around topics such as politics, mental health, and underground markets. These communities maintain specific norms which may not be understood or recognized by outsiders. The public nature of the internet still allows anyone to find these spaces, and researchers risk intruding into these spaces when conducting research with social media data.

**Risk Manifestation.** Collecting social media data increases the risk of intrusion into a community that does not wish to be studied. Researchers lacking training or experience in entering into existing relationships may misunderstand them and cause harm with false inferences and faulty generalizations. For example, one study interviewed users of a dating forum to detect sex workers, yet then applied their own categorization of how likely a user is to become a sex worker [39]. Researchers may also disrupt these digital spaces by extracting information without repaying the favor. These intrusions may lead to online communities' lack of trust in researchers and impede future research.

**Risk Prevention.** One way to prevent intrusion is to include community/group members and insiders directly in the research. Ali et al. include young people in their study detection of unsafe private messages by allowing them to donate their data rather than it unknowingly being scraped [3]. Consulting with area experts and researchers with prior experience with the target groups is another prevention method. Positionality statements provide transparency and enable reflection of risk, even though they may not prevent intrusion.

**Trade-Offs.** A challenge to preventing intrusion is ensuring subjects are fairly evaluated. With insider access comes the risk of being pressured to reach a favorable outcome for the community. Similarly, being too familiar with a particular community may lead to biased results. Building long-standing relationships with communities is also challenging when research timelines and funding are finite. Positionality statements are not always necessary or appropriate, as they may even force researchers into publicly disclosing sensitive information [46].

#### 4.11. Decisional Interference

We define decisional interference as *impacting the interpersonal relationships of users and how platforms interact*

with them. Digital spaces are already vulnerable to trolling, disinformation, and platform changes. Researchers working with social media data may exacerbate these disruptions when they enter online communities and publish their work.

**Risk Manifestation.** Decision inference can come from how researchers' work is disseminated and used. Sex workers rely on dating apps and forums to find clientele safely, and work detecting sex work may be used by platforms to affect their livelihood [39]. Once a community knows that they are being watched or insiders have given researchers access, community members may change how they interact with the community. For example, measuring how users talk about specific drugs or their sobriety journey on Reddit may lead users to change their language to evade detection [47]. Additionally, once the research is made public online platforms may investigate the issue and enact changes that could negatively impact the studied issue and group.

**Risk Prevention.** To prevent decisional interference, it is important to disclose research goals early and often. These aims do not always need to be shared with the community being researched. However, frequently returning to research goals may help the team reflect on the possible impact of their work. For example, recognizing the vulnerability of gig workers, Ramesh et al. study them on Reddit rather than asking them to participate in a research study that could potentially endanger them during the COVID-19 pandemic [62]. Doing "member-checking" with research participants and communities can also mitigate decisional interference, as it allows communities and users to provide feedback. Researchers may also directly engage with platform stakeholders to shape how their research directly impacts users.

**Trade-offs.** Similar to preventing intrusion, researchers must consider the trade-offs between active and passive observation. Active participation with users or a community may better orient researchers, but researchers must be reflective of the impact of their presence. Meanwhile, passive participation may not directly impact relationships, but prior work shows that users do not want to be unknowing research participants [22]. Interacting with platform stakeholders may not always be feasible or possible. Social media platforms do not often cooperate with researchers, and the collaboration may also lead to biases or self-censorship.

## 5. Implications for Stakeholders

Regarding **RQ1**, our findings suggest a **lack of transparency around how security researchers handle the privacy of social media data**. Only 35% ( $n=209$ ) of examined papers report considerations of data anonymization, availability, and storage issues. This lack of transparency might result from the absence of clear guidelines for considering and reporting privacy implications, akin to how the Menlo report guides researchers to reason about ethics. There are further gaps here — 60% of analyzed papers use data that is both publicly available and revealing.

Regarding **RQ2**, we define **11 privacy risks emerging from security research using social media data under the**

**lens of Solove's taxonomy** (Table 2) and how they manifest throughout §4. For instance, using data from *public* social media platforms, such as X/Twitter, Bluesky, and Mastodon, may disproportionately expose users to "Identification" and "Aggregation" risks. Non-anonymized datasets can expose users, while anonymized datasets can be combined to re-identify users based on similar profiles. Using data from *pseudonymous*, *semi-private* platforms, such as Reddit and Discord, may disproportionately expose users to "Disclosure" and "Intrusion" risks, especially when researchers present exact quotes without or conduct research without prior engagement with the online community to understand their norms. As such, **the choice of social media platforms changes the scope and prevalence of privacy risks**.

Regarding **RQ3**, **common privacy-preserving practices researchers employ include aggregating findings, anonymizing data, and paraphrasing quotes**. While prior work has identified these practices more broadly (§2), we find specific implementations in our dataset, such as analyzing scam service as a broader phenomenon instead of focusing on individual users [73], using anonymized SNAP datasets [44], and paraphrasing quotes when studying disclosures of suicide ideation [70] and life events [67].

We also identify **three novel privacy-preserving practices not captured by prior work: data donations, obtaining certificates of confidentiality, and developing large-scale legal data sharing agreements**. Data donations [3] minimize "Surveillance," "Exclusion," "Disclosure," "Distortion," and "Intrusion" risks by actively involving users in the data collection, obtaining informed consent while still ensuring ecological validity. Certificates of confidentiality [64] minimize "Disclosure," "Increased Accessibility," and "Blackmail" risks by protecting users against government subpoenas. Large-scale legal data sharing agreements [59] minimize "Insecurity" and "Increased Accessibility" risks by tightly controlling data access to data.

Drawing from our findings, we discuss implications for researchers, institutions (particularly ethics review boards), and publication venues. While researchers have an obligation to preserve user privacy when scoping, performing, and writing up their research, regardless of any oversight, ethics review boards and publication venues have the duty to verify that work under their purview is done sufficiently.

### 5.1. Implications for Researchers

Researchers need to consider privacy in all stages of their research. While it is crucial to support researchers in making discoveries about the world, new insights, however novel, are not always justified by the degree of privacy risks taken.

**The importance and tradeoffs of community engagement.** Community engagement to gauge the community's consent and develop the data collection practices in a ground-up approach is a clear way to mitigate exclusion risk. However, doing this in the early stages of research, such as when scoping out a new project, can also subvert the initial

hypotheses that the researchers wanted to investigate and leave the research questions unanswered.

Community engagement can also be tricky if the researchers are outsiders to the communities that they're studying, posing an intrusion risk. Given the Western world's history of colonialism, some see this work as inherently colonialist, especially if the researchers are studying marginalized communities. Insiders are not necessarily more advantageous. Their position as part of the community might allow them to participate more fully and might allow them more access to the community. But their preconceived notions from the community could lead to missing important mechanisms in understanding behavior.

There is no scientific consensus about the most ethical way to conduct research here. Some great research has been done so in coordination with platforms; other great research does not allow professional distance, particularly when measuring highly politicized communities via a critical lens. Alternatives need to be considered at each stage of involvement or not and the potential risks/harms need to be quantified.

**Risk disclosure.** Revealing privacy risks depends on the researcher's thoughts on who the relevant stakeholders are, which isn't straightforward. The platform itself? The people on the platform? The general group behind the platform that might also exist in other spaces? There's a concern of Exclusion when not notifying the right people, but also in Surveillance by the platform if informing them.

When designing the study, researchers need to consider the stakeholders and their possible reactions to notices with appropriate levels of information. Research in communicating privacy risks has drastically improved in the past two decades, and leveraging this to communicate appropriately with end users as well as broader community groups as relevant is vital. Many researchers here assume public social media data can be freely used due to its public nature, or that platform terms of service cover consent requirements already. We note that GDPR asks for clear consent when processing sensitive data.

**Nuances around third-party data collection.** Using third-party tools to collect social media data could pose an aggregation risk. Nonetheless, a large number of researchers repeatedly collecting the same data from the source can pose a financial cost to the platforms or risk bad science from improper data collection. Sometimes, third-party data platforms more tightly control access, limiting surveillance risks. Data from third-party data platforms could also be easier for researchers with little oversight to obtain, creating surveillance risks. The tradeoff here is nuanced based on the underlying sensitivity of the data, ease of collection, and barriers to entry for the data platform. We generally encourage third party data providers and ask them to control access to allow researcher access without external surveillance.

**More attention is needed for risks from data storage.** Data storage is an under-reported area of consideration, only mentioned by 6% ( $n=36$ ) of papers in our dataset. We posit that the underlying problem is resources. Researchers tackling security problems with social media data need

extensive physical resources, from computers to network connections, and from secure data storage to API/data access as applicable. Secure data storage needed to ensure against insecurity risks can be of issue for those with lower research budgets, calling for efforts towards communal provisioning.

Altogether, in order to do privacy-preserving research using social media data, researchers need an external review of their study design and safe communities to bring up potential privacy issues without blame. With the increasingly fast pace of research, even diligent researchers sometimes can skip these steps, not fully thinking through the privacy ramifications of their work. Communal norm setting among researchers goes a long way toward ensuring the adoption of best practices.

## 5.2. Implications for Institutions

Academic institutions in the US often leave institutional ethics approval to mixed-discipline IRBs. Under the Common Rule, IRBs require a panel containing at least one scientist, one non-scientist, and one person external to the institution to review research proposals [76]. Outside the US, ethics boards are often loosely based on the US IRB model but less standardized.

**We note a broader need for IRBs to understand the implications of research using social media data,** as well as how focusing on ethics broadly is insufficient to tackle the granular privacy risks. Huh-Yoo and Rader's interview study with IRB members shows that their risk perceptions increase when the research data is "digital" [33]. However, this awareness does not necessarily extend to research using social media data, which carries additional complexities due to the data's semi-public nature and lack of direct interactions with human subjects. In our dataset, only 7% ( $n=40$ ) of papers explicitly mention seeking ethics or IRB approval for their research, and most of them were found to be exempt from IRB oversight. We posit that there is a greater need for IRB reviewers to understand the privacy risks inherent to research using social media data and conduct thorough reviews, as doing this might also prompt them to reflect on exempt decisions. Formal exceptions require the research to "not reasonably place the subjects at risk" or rely on public information where "the identity of the human subjects cannot readily be ascertained." As shown by our SoK, both conditions could be violated in research using social media when the privacy risks are real and data subjects can be (re)identified.

IRB members also need to consider additional **privacy challenges from research concerning EU data under the GDPR.** These principles, while only necessary to apply to EU data, are a useful framework to consider data privacy protection for all; this is why others, like the US State of California, have rolled out similar legislation. Art. 17 grants users the right to demand the deletion of their personal data if it is no longer needed or the consent is withdrawn. IRB members should hold researchers accountable for best practices in honoring this right, such as by keeping links to artifacts rather than downloading the artifacts directly.

Meanwhile, GDPR Art. 89(1) introduces some exemptions for scientific research, provided that the erasure would impede achieving the research purposes. This particular provision allows researchers to potentially deny users' requests to exercise the right to be forgotten if doing so would compromise the integrity of the research. The legal community has been debating whether the right to be forgotten applies to academic publishing [9], [75].

### 5.3. Implications for Venues

While researchers can set norms, external enforcement of these is crucial to see them implemented. **Publication venues need to set and enforce expectations of social media data privacy.** However, there needs to be careful considerations made to not restrict research into privacy-sensitive topics. Some institutions have better IRB processes than others. Some individual researchers have better privacy considerations than others. Our results (§4) show that some disciplines are better than others at reporting privacy considerations—a majority of HCI and CSC papers report on data availability, anonymization, or storage. Given that the most common finding was a lack of mentioning privacy considerations, it is hard to automatically find potential privacy violations before the work is published. Venues such as IEEE S&P and USENIX Security have ethics committees that reviewers can flag papers to. However, this is necessarily ad hoc and may not always lead to increased levels of privacy-minded study design.

**Individual reviewers can and should consider privacy when reviewing papers.** Considering the trade-offs here is important, as reviewers could use data privacy as an excuse to sink an otherwise fine paper. This could discourage research into sensitive areas. Reviewers need to handle data privacy reviews with care to ensure consideration of sensitive research without requiring perfection.

Some reviewers encourage authors to remove sensitive comments in their work. On one hand, this can cause work to be more palatable to the masses. On the other, this can water down research in areas that are stigmatized or otherwise blunt the effects of words used in certain communities. There is a broad cultural context around what is “offensive,” “taboo,” or “non-academic,” which is not the same in every location. Editors need to handle this issue with care to ensure that reviewer bias is balanced with general stewardship when considering sensitive or potentially offensive topics, words, or studied participant disclosures.

## 6. Conclusion

Social media data is increasingly being used in security research; however, a lack of privacy considerations is exposing millions of people's data. In this work, we investigate these threats by proposing a taxonomy of privacy risks, based on Solove's taxonomy of privacy and a systematic analysis of 601 research papers. We find that while Solove's taxonomy is suitable for understanding some aspects of

social media research, such as risks during information processing, it fails to capture the severity of privacy risks during research dissemination, such as the accessibility, speed, and volume of existing datasets and research outputs. We also find a lack of reporting around data privacy, with only 35% (209/601) of papers discussing data anonymization, availability, and storage issues. The reporting is slightly better among papers that study marginalized and vulnerable populations, with 46% (30/65) reporting data privacy considerations. Our findings indicate that authors, academic institutions, and venues must do better in considering and reporting the privacy considerations of their social media research. By engaging with our framework, future researchers can ensure that social media research is conducted in a privacy-conscious way.

As security researchers, we must hold ourselves to a higher standard when handling social media data. We can use our experiences researching sensitive topics to apply data protection practices that respect user privacy and minimize data retention. We can question our use of large, open data repositories and tools that put users' privacy at risk. By embracing privacy-conscious social media research practices, we can ensure that users continue to feel comfortable online without research negatively impacts their lives.

## Acknowledgments

The authors would like to thank Rainer Böhme for his helpful early feedback. This project was funded by the UK EPSRC grant EP/S022503/1 that supports the Centre for Doctoral Training in Cybersecurity delivered by UCL's Departments of Computer Science, Security & Crime Science, and Science, Technology, Engineering & Public Policy; UK EPSRC grant EP/T517847/1; the Max Planck Society; and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the Germany's Excellence Strategy—EXC 2092 CASA—390781972.

## References

- [1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. Privacy and human behavior in the age of information. *Science*, 347(6221):509–514, 2015.
- [2] Somayyeh Aghababaei and Masoud Makrehchi. Mining social media content for crime prediction. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 2016.
- [3] Shiza Ali, Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Chen Ling, Munmun De Choudhury, Pamela J Wisniewski, and Gianluca Stringhini. Getting meta: A multimodal approach for detecting unsafe conversations within instagram direct messages of youth. *ACM on Human-Computer Interaction*, 7(CSCW1):1–30, 2023.
- [4] Nazanin Andalibi and Andrea Forte. Announcing pregnancy loss on facebook: A decision-making framework for stigmatized disclosures on identified social network sites. In *CHI conference on Human Factors in Computing Systems*, pages 1–14, 2018.
- [5] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. Social support, reciprocity, and anonymity in responses to sexual abuse disclosures on social media. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 25(5):1–35, 2018.



- [6] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *CHI conference on Human Factors in Computing Systems*, pages 3906–3918, 2016.
- [7] Carolyn Ashurst, Emmie Hine, Paul Sedille, and Alexis Carlier. AI ethics statements: analysis and lessons learnt from neurips broader impact statements. In *ACM conference on Fairness, Accountability, and Transparency*, pages 2047–2056, 2022.
- [8] John W Ayers, Theodore L Caputi, Camille Nebeker, and Mark Dredze. Don’t quote me: reverse identification of research participants in social media studies. *NPJ digital medicine*, 1(1):30, 2018.
- [9] Basak Bak. Reviving a european idea: Author’s right of withdrawal and the right to be forgotten under the EU’s general data protection regulation (GDPR). *SCRIPTed*, 19:120, 2022.
- [10] Rosanna Bellini, Emily Tseng, Noel Warford, Alaa Daffalla, Tara Matthews, Sunny Consolvo, Jill Palzkill Woelfer, Patrick Gage Kelley, Michelle L Mazurek, Dana Cuomo, et al. Sok: Safer digital-safety research involving at-risk users. In *IEEE Symposium on Security and Privacy*, pages 71–71, 2023.
- [11] Kelsey Beninger. Social media users’ views on the ethics of social media research. *The SAGE handbook of social media research methods*, 1, 2017.
- [12] Katya Borgos-Rodriguez, Kathryn E. Ringland, and Anne Marie Piper. Myautsomefamilylife: Analyzing parents of children with developmental disabilities on youtube. *ACM on Human-Computer Interaction*, 3(CSCW), November 2019.
- [13] danah boyd and Nicole Ellison. Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, 13(1):210–230, 2007.
- [14] Amy Bruckman. Studying the amateur artist: A perspective on disguising data collected in human subjects research on the internet. *Ethics and Information Technology*, 4:217–231, 2002.
- [15] Amber M Buck and Devon F Ralston. I didn’t sign up for your research study: The ethics of using “public” data. *Computers and Composition*, 61:102655, 2021.
- [16] Lei Cao, Huijun Zhang, Xin Wang, and Ling Feng. Learning users inner thoughts and emotion changes for social media based suicide risk detection. *IEEE Transactions on Affective Computing*, 2021.
- [17] Jessica Clement. Distribution of Twitter users worldwide as of october 2019, by age group, 2019. <https://web.archive.org/web/20200107132905/https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/>.
- [18] Periwinkle Doerfler, Andrea Forte, Emiliano De Cristofaro, Gianluca Stringhini, Jeremy Blackburn, and Damon McCoy. “I’m a professor, which isn’t usually a dangerous job”: Internet-facilitated harassment and its impact on researchers. *ACM on Human-Computer Interaction*, 5(CSCW2):1–32, 2021.
- [19] Brianna Dym and Casey Fiesler. Ethical and privacy considerations for research using online fandom data. *Transformative works and cultures*, 33, 2020.
- [20] European Parliament and Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council, 2016. <https://data.europa.eu/eli/reg/2016/679/oj>.
- [21] Casey Fiesler, Nathan Beard, and Brian C. Keegan. No robots, spiders, or scrapers: Legal and ethical regulation of data collection methods in social media terms of service. *International AAAI Conference on Web and Social Media*, 14(1):187–196, May 2020.
- [22] Casey Fiesler and Nicholas Proferes. “Participant” perceptions of Twitter research ethics. *Social Media + Society*, 4(1):2056305118763366, 2018.
- [23] Casey Fiesler, Michael Zimmer, Nicholas Proferes, Sarah Gilbert, and Naiyan Jones. Remember the human: A systematic review of ethical considerations in reddit research. *ACM on Human-Computer Interaction*, 8(GROUP):1–33, 2024.
- [24] Manas Gaur, Amanuel Alambo, Joy Prakash Sain, Ugur Kursuncu, Krishnaprasad Thirunarayan, Ramakanth Kavuluru, Amit Sheth, Randy Welton, and Jyotishman Pathak. Knowledge-aware assessment of severity of suicide risk for early intervention. In *The World Wide Web Conference*, pages 514–525, 2019.
- [25] Meysam Ghaffari, Ashok Srinivasan, Anuj Mubayi, Xiuwen Liu, and Krishnan Viswanathan. Next-generation high-resolution vector-borne disease risk assessment. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 621–624, 2019.
- [26] Shalmoli Ghosh, Janardan Misra, Saptarshi Ghosh, and Sanjay Podder. Utilizing social media for identifying drug addiction and recovery intervention. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 3413–3422. IEEE, 2020.
- [27] Sarah Gilbert, Katie Shilton, and Jessica Vitak. When research is the context: Cross-platform user expectations for social media data reuse. *Big Data & Society*, 10(1):20539517231164108, 2023.
- [28] Neil Zhenqiang Gong and Bin Liu. Attribute inference attacks in online social networks. *ACM Transactions on Privacy and Security (TOPS)*, 21(1):1–30, 2018.
- [29] Qingyuan Gong, Yushan Liu, Jiayun Zhang, Yang Chen, Qi Li, Yu Xiao, Xin Wang, and Pan Hui. Detecting malicious accounts in online developer communities using deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(10):10633–10649, 2023.
- [30] Neal R Haddaway, Biljana Macura, Paul Whaley, and Andrew S Pullin. Roses reporting standards for systematic evidence syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environmental Evidence*, 7:1–8, 2018.
- [31] Oliver Haimson. Social media as social transition machinery. *ACM on Human-Computer Interaction*, 2(CSCW):1–21, 2018.
- [32] Oliver Haimson and Gillian Hayes. Changes in social media affect, disclosure, and sociality for a sample of transgender americans in 2016’s political climate. In *International AAAI Conference on Web and Social Media*, 2017.
- [33] Jina Huh-Yoo and Emilee Rader. It’s the wild, wild west: Lessons learned from irb members’ risk perceptions toward digital research data. *ACM on Human-Computer Interaction*, 4(CSCW1):1–22, 2020.
- [34] IEEE S&P 2023. Research ethics committee, 2023. <https://www.ieee-security.org/TC/SP2023/cfpapers.html>.
- [35] Marcello Ienca and Effy Vayena. Ethical requirements for responsible research with hacked data. *Nature Machine Intelligence*, 3(9):744–748, 2021.
- [36] Nur Shazwani Kamarudin, Vineeth Rakesh, Ghazaleh Beigi, Lydia Manikouda, and Huan Liu. A study of reddit-user’s response to rape. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 591–592, 2018.
- [37] Jisu Kim, Francesca Pratesi, Giulio Rossetti, Alina Sirbu, and Fosca Giannotti. Where do migrants and natives belong in a community: a twitter case study and privacy risk analysis. *Social Network Analysis and Mining*, 13(1):15, 2022.
- [38] Shamika Klassen and Casey Fiesler. “This isn’t your data, friend”: Black twitter as a case study on research ethics for public data. *Social Media + Society*, 8(4):20563051221144317, 2022.
- [39] Panos Kostakos, Lucie Špráchalová, Abhinay Pandya, Mohamed Aboeleinen, and Mourad Oussalah. Covert online ethnography and machine learning for detecting individuals at risk of being drawn into online sex work. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1096–1099, 2018.
- [40] Theresa Kuchler, Dominic Russel, and Johannes Stroebel. Jue insight: The geographic spread of covid-19 correlates with the structure of social networks as measured by facebook. *Journal of Urban Economics*, 2022.

- [41] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.
- [42] Steve Lauterwasser and Nataliya Nedzhvetskaya. Privacy in public?: The ethics of academic research with publicly available social media data. *Berkeley Journal of Sociology*, 64:126–144, 2023.
- [43] Hao-Ping Lee, Yu-Ju Yang, Thomas Serban Von Davier, Jodi Forlizzi, and Sauvik Das. Deepfakes, phrenology, surveillance, and more! a taxonomy of ai privacy risks. In *CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2024.
- [44] Jure Leskovec and Julian McAuley. Learning to discover social circles in ego networks. *Advances in Neural Information Processing Systems*, 25, 2012.
- [45] Huaxin Li, Qingrong Chen, Haojin Zhu, Di Ma, Hong Wen, and Xuemin Sherman Shen. Privacy leakage via de-anonymization and aggregation in heterogeneous social networks. *IEEE Transactions on Dependable and Secure Computing*, 17(2):350–362, 2017.
- [46] Calvin A Liang, Sean A Munson, and Julie A Kientz. Embracing four tensions in human-computer interaction research with marginalized people. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 28(2):1–47, 2021.
- [47] John Lu, Sumati Sridhar, Ritika Pandey, Mohammad Al Hasan, and George Mohler. Investigate transitions into drug addiction through text mining of reddit data. *25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, page 2367–2375, 2019.
- [48] Xiao Ma, Jeff Hancock, and Mor Naaman. Anonymity, intimacy and self-disclosure in social media. In *CHI conference on Human Factors in Computing Systems*, pages 3857–3869, 2016.
- [49] Annette Markham. Fabrication as ethical practice. *Information, Communication & Society*, 15(3):334–353, 2012.
- [50] Alice E Marwick and danah boyd. I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society*, 13(1):114–133, 2011.
- [51] Alan Mislove, Massimiliano Marcon, Krishna P Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. In *7th ACM SIGCOMM conference on Internet measurement*, pages 29–42, 2007.
- [52] Megan A. Moreno, Natalie Goniou, Peter S. Moreno, and Douglas Diekema. Ethics of social media research: Common concerns and practical considerations. *Cyberpsychology, Behavior, and Social Networking*, 16(9):708–713, 2013. PMID: 23679571.
- [53] Arvind Narayanan and Vitaly Shmatikov. How to break anonymity of the netflix prize dataset. *arXiv preprint cs/0610105*, 2006.
- [54] Shuo Niu, Katherine G McKim, and Kathleen Palm Reed. Education, personal experiences, and advocacy: Examining drug-addiction videos on youtube. *ACM on Human-Computer Interaction*, 6(CSCW2):1–28, 2022.
- [55] Fayika Farhat Nova, Amanda Coupe, Elizabeth D Mynatt, Shion Guha, and Jessica A Pater. Cultivating the community: inferring influence within eating disorder networks on twitter. *ACM on Human-Computer Interaction*, 6(GROUP):1–33, 2022.
- [56] Florida Department of Health. Florida population median age. <https://web.archive.org/web/20250207012049/https://www.flhealthcharts.gov/ChartsDashboards/rdPage.aspx?rdReport=NonVitalIndRateOnly.Dataviewer>. Accessed: 2024-11-09.
- [57] Mourad Ouzzani, Hossam Hammady, Zbys Fedorowicz, and Ahmed Elmagarmid. Rayyan—a web and mobile app for systematic reviews. *Systematic reviews*, 5:1–10, 2016.
- [58] Erin O’Callaghan and Hannah M Douglas. # metoo online disclosures: A survivor-informed approach to open science practices and ethical use of social media data. *Psychology of Women Quarterly*, 45(4):505–525, 2021.
- [59] Sergio Pastrana, Daniel R Thomas, Alice Hutchings, and Richard Clayton. Crimebb: Enabling cybercrime research on underground forums at scale. In *World Wide Web Conference*, pages 1845–1854, 2018.
- [60] Nicholas Proferes. Information flow solipsism in an exploratory study of beliefs about twitter. *Social Media + Society*, 3(1):2056305117698493, 2017.
- [61] Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media + Society*, 7(2):20563051211019004, 2021.
- [62] Divya Ramesh, Caitlin Henning, Nel Escher, Haiyi Zhu, Min Kyung Lee, and Nikola Banovic. Ludification as a lens for algorithmic management: A case study of gig-workers’ experiences of ambiguity in instacart work. In *ACM Designing Interactive Systems Conference*, pages 638–651, 2023.
- [63] Signe Ravn, Ashley Barnwell, and Barbara Barbosa Neves. What is “publicly available data”? exploring blurred public–private boundaries and ethical practices through a case study on instagram. *Journal of empirical research on human research ethics*, 15(1-2):40–45, 2020.
- [64] Afsaneh Razi, Ashwaq Alsoubai, Seunghyun Kim, Shiza Ali, Gianluca Stringhini, Munmun De Choudhury, and Pamela J Wisniewski. Sliding into my dms: Detecting uncomfortable or unsafe sexual risk experiences within instagram direct messages grounded in the perspective of youth. *ACM on Human-Computer Interaction*, 7(CSCW1):1–29, 2023.
- [65] Boshu Ru, Kimberly Harris, and Lixia Yao. A content analysis of patient-reported medication outcomes on social media. In *IEEE International Conference on Data Mining Workshops*, pages 472–479, 2015.
- [66] Koustuv Saha, Sang Chan Kim, Manikanta D Reddy, Albert J Carter, Eva Sharma, Oliver L Haimson, and Munmun De Choudhury. The language of lgbtq+ minority stress experiences on social media. *ACM on Human-Computer Interaction*, 3(CSCW):1–22, 2019.
- [67] Koustuv Saha, Jordyn Seybolt, Stephen M Mattingly, Talayah Al-davood, Chaitanya Konjeti, Gonzalo J Martinez, Ted Grover, Gloria Mark, and Munmun De Choudhury. What life events are disclosed on social media, how, when, and by whom? In *CHI Conference on Human Factors in Computing Systems*, pages 1–22, 2021.
- [68] Kavous Salehzadeh Niksirat, Lahari Goswami, Pooja SB Rao, James Tyler, Alessandro Silacci, Sadiq Aliyu, Annika Aebli, Chat Wacharamanatham, and Mauro Cherubini. Changes in research ethics, openness, and transparency in empirical studies between CHI 2017 and CHI 2022. In *CHI Conference on Human Factors in Computing Systems*, pages 1–23, 2023.
- [69] Anna Sapienza, Alessandro Bessi, Saranya Damodaran, Paulo Shakarian, Kristina Lerman, and Emilio Ferrara. Early warnings of cyber threats in online discussions. In *IEEE International Conference on Data Mining Workshops*, pages 667–674, 2017.
- [70] Ramit Sawhney, Harshit Joshi, Saumya Gandhi, and Rajiv Ratn Shah. Towards ordinal suicide ideation detection on social media. In *14th ACM International Conference on Web Search and Data Mining*, pages 22–30, 2021.
- [71] Katie Shilton and Sheridan Sayles. “We aren’t all going to be on the same page about ethics”: Ethical practices and challenges in research on digital and social media. In *49th Hawaii International Conference on System Sciences*, pages 1909–1918. IEEE, 2016.
- [72] Daniel J Solove. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154:477, 2005.
- [73] Zhibo Sun et al. Having your cake and eating it: An analysis of Concession-Abuse-as-a-Service. In *30th USENIX Security Symposium*, pages 4169–4186, 2021.
- [74] Leho Tedersoo et al. Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific data*, 8(1):192, 2021.

- [75] Jaime A Teixeira da Silva and Serhii Nazarovets. Can the principle of the 'right to be forgotten' be applied to academic publishing? probe from the perspective of personal rights, archival science, open science and post-publication peer review. *Learned Publishing*, 36(4):651–666, 2023.
- [76] The Office of Research Integrity. ORI introduction to RCR: Chapter 3. the protection of human subjects, 2024.
- [77] Daniel R. Thomas, Sergio Pastrana, Alice Hutchings, Richard Clayton, and Alastair R. Beresford. Ethical issues in research using datasets of illicit origin. In *Internet Measurement Conference, IMC '17*, page 445–462, New York, NY, USA, 2017.
- [78] Kurt Thomas et al. Sok: Hate, harassment, and the changing landscape of online abuse. In *IEEE Symposium on Security and Privacy*, pages 247–267, 2021.
- [79] Christina Tikkinen-Piri, Anna Rohunen, and Jouni Markkula. EU general data protection regulation: Changes and implications for personal data collecting companies. *Computer Law & Security Review*, 34(1):134–153, 2018.
- [80] USENIX Security 2025. Usenix Security '25 ethics guidelines, 2025. <https://www.usenix.org/conference/usenixsecurity25/ethics-guidelines>.
- [81] Jessica Vitak, Nicholas Proferes, Katie Shilton, and Zahra Ashktorab. Ethics regulation in social computing research: Examining the role of institutional review boards. *Journal of Empirical Research on Human Research Ethics*, 12(5):372–382, 2017.
- [82] Jessica Vitak, Katie Shilton, and Zahra Ashktorab. Beyond the belmont principles: Ethical challenges, practices, and beliefs in the online data research community. In *19th ACM conference on Computer-Supported Cooperative Work & Social Computing*, pages 941–953, 2016.
- [83] Yong Wang, Abdelrhman Hassan, Xiaoran Duan, and Xiaosong Zhang. An efficient multiple-user location-based query authentication approach for social networking. *Journal of Information Security and Applications*, 47:284–294, 2019.
- [84] Yunpeng Weng, Liang Chen, and Xu Chen. Identifying user relationship on wechat money-gifting network. *IEEE Transactions on Knowledge and Data Engineering*, 34(8):3814–3825, 2020.
- [85] Wikipedia. List of social networking services, 2023. [https://en.wikipedia.org/wiki/List\\_of\\_social\\_networking\\_services](https://en.wikipedia.org/wiki/List_of_social_networking_services).
- [86] Matthew L Williams, Pete Burnap, and Luke Sloan. Towards an ethical framework for publishing twitter data in social research: Taking into account users' views, online context and algorithmic estimation. *Sociology*, 51(6):1149–1168, 2017. PMID: 29276313.
- [87] Joost F Wolfswinkel, Elfi Furtmueller, and Celeste PM Wilderom. Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, 22(1):45–55, 2013.
- [88] Volker Wulf, Konstantin Aal, Ibrahim Abu Kteish, Meryem Atam, Kai Schubert, Markus Rohde, George P Yerosusis, and David Randall. Fighting against the wall: Social media use by political activists in a Palestinian village. In *SIGCHI conference on Human Factors in Computing Systems*, 2013.
- [89] Wenjing Zeng, Rui Tang, Haizhou Wang, Xingshu Chen, and Wenxian Wang. User identification based on integrating multiple user information across online social networks. *Security and Communication Networks*, 2021(1):5533417, 2021.
- [90] Michael Zimmer. Addressing conceptual gaps in big data research ethics: An application of contextual integrity. *Social Media + Society*, 4(2):2056305118768300, 2018.
- [91] Michael Zimmer. "But the data is already public": on the ethics of research in Facebook. In *The ethics of information technologies*, pages 229–241. Routledge, 2020.
- [92] Michael Zimmer and Nicholas John Proferes. A topology of twitter research: disciplines, methods, and ethics. *Aslib Journal of Information Management*, 66(3):250–261, 2014.
- [93] Matthew Zook, Solon Barocas, danah boyd, Kate Crawford, Emily Keller, Seeta Peña Gangadharan, Alyssa Goodman, Rachelle Hollander, Barbara A Koenig, Jacob Metcalf, et al. Ten simple rules for responsible big data research. *PLoS computational biology*, 13(3):e1005399, 2017.

## Appendix A. Analysis Template

- 1) Who is viewing this paper? [answer options anonymized]
- 2) What is the ID of the paper? [free text]
- 3) What is the population being studied in the paper? ☐ General user ☐ Other [free text]
- 4) (Optional) Do you have any additional thoughts on the population being studied in the paper?
- 5) Is the topic of the paper sensitive (e.g., elections, mental health, and hate speech)? ☐ Yes ☐ No ☐ Maybe [free text]
- 6) (Optional) Do you have any additional thoughts on the topic being studied in the paper?
- 7) Is the population vulnerable and/or marginalized (e.g. minors, migrants, people with disabilities, and racial/gender/sexual minorities)? ☐ Yes ☐ No ☐ Maybe [free text]
- 8) (Optional) Do you have any thoughts on the population's vulnerable and/or marginalized status in the paper?
- 9) What are the platforms being studied? Choose all that apply. ☐ Advogato ☐ Discord ☐ Douban ☐ Facebook ☐ Flickr ☐ Forums ☐ Foursquare ☐ Gowalla ☐ Instagram ☐ LiveJournal ☐ Quora ☐ Reddit ☐ Shareteches ☐ Tumblr ☐ Twitter ☐ VKontakte ☐ Weibo ☐ Yelp ☐ YouTube ☐ Other [free text]
- 10) How is the data collected? Choose all that apply. ☐ API ☐ Data breach ☐ Existing datasets ☐ Scraping ☐ Third-party tool ☐ Not reported ☐ Other [free text]
- 11) (Optional) Do you have any thoughts on the method of data collection?
- 12) What type of data is collected? Choose all that apply. ☐ Text ☐ Video ☐ Images ☐ Profile data ☐ Network data ☐ Location data ☐ Interaction data ☐ Metadata ☐ Other [free text]
- 13) (Optional) Do you have any thoughts on the type of data collected?
- 14) What is the size of the dataset collected? ☐ 1-50 ☐ 51-500 ☐ 501-5,000 ☐ 5,001-50,000 ☐ 50,001-500,000 ☐ 500,001-5,000,000 ☐ 5,000,001+ ☐ Not reported
- 15) What examples are used in the paper? Choose all that apply. ☐ Plain text ☐ Anonymized ☐ Paraphrased ☐ Images ☐ Censored images ☐ Hashtags ☐ None ☐ Other [free text]
- 16) (Optional) Do you have any thoughts on the examples used in the paper?
- 17) What analysis methods do the authors use? Choose all that apply. ☐ Model trained on data ☐ Model evaluated with data ☐ Statistical analysis ☐ Network analysis ☐ Thematic analysis ☐ Content analysis ☐ Discourse analysis ☐ Privacy risk analysis ☐ Sentiment analysis ☐ Topic modeling ☐ Conjunctive analysis of case configurations ☐ Other quantitative [free text] ☐ Other qualitative [free text] ☐ Other NLP [free text] ☐ Other [free text]
- 18) (Optional) Do you have any thoughts on the analysis method(s) of the study?
- 19) What is the author(s) explanation of platform choice? Choose all that apply. ☐ Dataset accessibility, availability, and verifiability ☐ Filling a research gap (platform is understudied) ☐ Importance of platform to study objectives (features, demographics) ☐ Platform popularity ☐ Reference to prior literature ☐ Not reported ☐ Other [free text]
- 20) (Optional) Do you have any thoughts on the author(s) explanation of platform choice?
- 21) Are there any data security or ethics considerations presented? ☐ Data availability ☐ Data storage ☐ Ethics process ☐ No ethics considerations ☐ Other [free text]
- 22) Please copy and paste the paper's ethics consideration section, if any.
- 23) How do the author(s) handle data anonymization? ☐ Anonymized data ☐ Non-anonymized data ☐ Individuals can be re-identified from the dataset ☐ Other [free text]
- 24) (If "data availability" selected in Q21) How available is the dataset? ☐ Data is publicly available (anonymized) ☐ Data is publicly available (revealing) ☐ Other [free text]
- 25) (If "data storage" selected in Q21) How is the data stored? ☐ Use of cloud services ☐ Locally stored data ☐ Other [free text]
- 26) (If "ethics process" selected in Q21) What ethical process did the author(s) undergo? ☐ Ethics or IRB approval ☐ Follow third-party ethical guidelines ☐ Terms of service compliance ☐ Legal compliance ☐ Other [free text]
- 27) (Optional) Do you have any thoughts on the author(s) explanation of data security considerations?
- 28) What are the aim(s) of the study? ☐ Detection model ☐ Prediction model ☐ Characterization of dataset ☐ Classification model ☐ Estimation model ☐ Understanding an event ☐ Other (free text)
- 29) (Optional) Do you have any thoughts on the aim(s) of the study?

## **Appendix B. Meta-Review**

The following meta-review was prepared by the program committee for the 2025 IEEE Symposium on Security and Privacy (S&P) as part of the review process as detailed in the call for papers.

### **B.1. Summary**

This paper is a systematization of knowledge of research from the past 16 years that employs data from social media in its publications. The work highlights potential harms to users whose data is collected and how researchers mitigate these risks.

### **B.2. Scientific Contributions**

- Provides a Valuable Step Forward in an Established Field

### **B.3. Reasons for Acceptance**

- 1) It is a helpful synthesis for those working in the field of research over social media to ensure they are aware of both the risks and potential mitigation efforts they can make to protect those whose social media data is included in research.