Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Digital Footprints and Open-Source Intelligence: Insights from Open Data Exploitation

Mohammed Suhail A
Cyber Forensics
Vivekananda Global University,
Jaipur, India

Krishna Pingale
Cyber Forensics
Vivekananda Global University,
Jaipur, India

Silmi Ali
Cyber Forensics
Vivekananda Global University,
Jaipur, India

Abstract: **The quantitative research looks at the use of digital footprints, such as mobile phone numbers, email addresses, social media, review sites, contact sharing sites and data breach sites to expose hidden network structures and allow open-source intelligence (OSINT) analysts to identify hidden structures through open-data mining. The digital traces resulting from our interactions in the digital world expose permanent and multi-platform connections that reveal identity, social network, platforms and exposure. This research fills the gap in empirical evidence of OSINT in user spaces, by conducting a survey via Google Forms to gather details on the profile and digital footprints of the participants in multiple platforms. Demographic details (age, occupation, location), social media platform use, public content, visibility, access to friends' content and public profile information were gathered to facilitate analysis of digital footprints. Further, the study includes OSINT-based verification of email identifiers using open-source breach intelligence databases to determine whether or not the respondents had been identified in data breaches, and to enhance the intelligence value of the data. This enables statistical analysis of the prevalence, cross-platform connectivity and intelligence potential of digital footprints through linking of individual identifiers to additional network connections. The findings reveal that breach data is important for exposure; and social and review sites are important for identity and profiling. Ultimately, this study provides empirical evidence of the assessment of intelligence value of open data, and OSINT for cybersecurity, privacy risk, digital forensics and ethical data analytics.**

Keywords: Digital footprints, OSINT, network mapping, open data exploitation, data breaches

INTRODUCTION

Digital footprints are the traces left by people as a result of their interactions with digital media platforms. These can include information such as telephone numbers, email addresses, social media profiles, reviews, location history, contacts and search history. These can be traces that are explicitly shared by users, as well as traces generated by platforms. As we build our digital presence, even basic actions, such as setting up an account, sharing on social media, leaving feedback, or giving an app permission to access our data, are all building our online footprint. Taken together, these snippets of data can come together to provide insights into our behaviours, interests, social networks and travel or communication patterns.

Open-source intelligence (OSINT) involves gathering and analysing publicly (or semi-publicly) accessible information. When applied to digital footprints, OSINT offers a methodical approach to linking information across networks and platforms, and to provide insight into patterns that may not be apparent from isolated data. Data that may not be notably relevant or suspicious when viewed in isolation can reveal a lot when combined with other available information. As a result, digital footprints are significant for research in cybersecurity and privacy, as well as in digital forensics, investigation and intelligence.

Increasing use of digital media has led to more information being generated. Social media apps, review sites, messaging apps and contact book sharing services now create an integrated environment in which information about identity can be shared across a range of services. But this also presents concerns regarding privacy, profiling and inadvertent disclosure. An email address or telephone number, or a public review, may seem an inconspicuous piece of information on its own, but when combined with other digital traces, it may contribute to the emergence of personal or social-level patterns.

Despite the growing importance of OSINT, most of the literature has focused on tools, methods and investigative uses of OSINT rather than on the digital activities of everyday users. Less emphasis has been placed on the digital traces that are formed through everyday online behaviour and which can be used for open-data processing. This research contributes to this body of knowledge by considering digital footprints from the user perspective, and through the use of open data to determine network patterns. The study adopts a quantitative questionnaire-based approach to examine social media platform usage, contact sharing, privacy, and vulnerability to breaches, including external verification via email identifiers

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

against databases of known breaches. This study aims to understand the role of digital routine in the OSINT-relevant visibility and how open data can be used for intelligence.

LITERATURE REVIEW

Open-source intelligence (OSINT) is a formalized approach to cybersecurity as it can infer exposure, patterns and relationships from publicly accessible data. The latest reviews demonstrate that OSINT helps with threat detection, attribution, and public exposure, so it can be used to infer covert identity connections from publicly available data.

Research on digital footprints demonstrates that online actions leave traces such as email addresses, phone numbers, social accounts, reviews, and metadata. Recent research stresses that these traces can be linked across platforms to infer identities and behaviour, aiding digital forensics and privacy data studies.

A prominent focus of the literature is perceived privacy vs exposure. People don't appreciate the extent to which their browser-based activity can be inferred, in particular through processes such as indexing, aggregating or reusing data in OSINT workflows. Privacy-centered research shows that users adapt and take action to protect their digital footprint.

Breach-related studies indicate leaked identifiers, like email addresses, improve attribution and visibility when paired with open sources. OSINT research indicates that public breach data and leaked account credentials enhance risk assessment, exposure mapping and identify account and platform connections.

Network reconstruction is also a common OSINT topic. This work demonstrates that social media, contact-sharing and other public traces can serve as nodes in a network of visibility. This helps to uncover latent connections and multi-platform relationships that may not be evident in one data source.

While this work is emerging, less research includes questionnaire surveys of users with multi-platform exposure variables and breach verification through a quantitative framework. Much of this literature is tool-oriented or qualitative, so this study contributes by looking at digital footprint exposure, privacy choices and breach visibility in a single study.

METHODOLOGY

1. Research Design

This research, which sought to examine online traceability and exposure of online footprint from the point of view of open-source intelligence (OSINT) adopted a quantitative, descriptive research design. The purpose of the study was to collect data and interpret the prevalence of exposure related behaviours of social media, review site, privacy, contact sharing and awareness of the exposure of personal data but not to determine causality. This was suitable as the aim was to gather and explain the frequency of exposure related behaviours of a sample of participants, and link these behaviours with markers of public intelligence.

2. Study Instrument

The primary data was collected by designing a questionnaire (Google Forms). The questionnaire contained many multi-choice questions to gather data about: (1) demographic information, (2) online activities, (3) online privacy awareness, (4) users' experiences with reviews sites, (5) access to users' contacts, (6) data breaches and (7) the type of information users think is available to people they do not know personally. The questionnaire was organised into six sections to enable us to explore each of the characteristics of online footprint in a similar way across the respondents.

The first section was about the demographic information of the respondents - age, profession, university, organisation, city and state. Section 2 asked about the social media utilisation and sharing of posts, birthday, tagging of family and friends, sharing of login IDs and passwords of the social media, frequency of checking the privacy settings and if they would share unnecessary permission of the apps (microphone, camera and contact). The third part of the survey measured reviews, and asked the respondents where they review, and if they review food delivery apps such as Zomato or Swiggy.

The next section of the survey quantified sharing of contacts, by asking the respondents how many contacts they have in their phone and if apps are allowed to access contacts. The fifth part of the survey looked at data leakages, by asking if the respondent has checked if their data is leaked online and if they know of any tools/websites to do so. The last part of the survey measured the online presence of the individual, by asking what a stranger can see online such as name, pictures, phone number, email, location, workplace/school and friends.

3. Sample and Data

We were able to use survey data of 207 of the respondents in the data set. The sample consisted of students, professionals (including self-employed and homemakers) and retirees, which suggests that there would be a variety in the social media and digital use patterns of the sample. Data was self-reported and the data was used as the unit of analysis for descriptive analysis of privacy and presence.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

4. OSINT Verification Layer

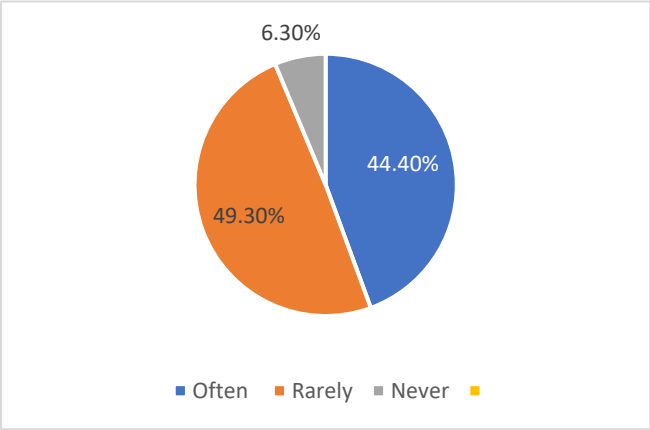
To add an additional layer to self-reported responses, an OSINT verification layer of some of the identifiers was added. Exposure of identifiers for emails was analysed using breach check tools (such as Have I Been Pwned) typically used in OSINT to check if an email identifier is exposed in a breach. This enabled us to understand if the respondents' perceptions of exposure of their personal information in breaches were linked to breach event.

We used phone intelligence (reverse phone) tools (such as IPQualityScore) to analyse phone numbers. IPQualityScore can be used to provide phone intelligence (carrier and phone line type) or phone activity/reputation/risk signals that can be used for OSINT phone-number tracing. Using these tools the study was able to explore the potential of typical personal identity information (email and phone) in online presence for public or semi-public intelligence applications.

5. Data Analysis

The data collected were analysed by descriptive statistics. They were broken down into frequencies and percentages and patterns were analysed between the variables to see if any patterns existed in sharing publicly, privacy settings, permission sharing, reviewing, exposure of contacts and the use of the breach-checking apps. Due to the nature of the research (descriptive, exploratory), the analysis looked at the frequency of behaviours, and patterns in exposure, as opposed to predicting and testing hypotheses.

The results of the OSINT verification exercise were not considered to be a forensic analysis in and of itself, but a further level of analysis of the survey results. This allowed us to demonstrate how open-source intelligence (OSINT) approaches could be combined with the survey reports of online behaviours to better understand other aspects of online exposure.



6. Ethical Considerations

We been mindful of academic and ethical procedures in this research. Only for academic research, the questionnaires data (personal details) were collected and kept anonymous as research data. The OSINT checks (if any) only used free software and open-source intelligence data, and only for the purpose of identifying patterns of exposure, rather than to expose, track or manipulate individuals.

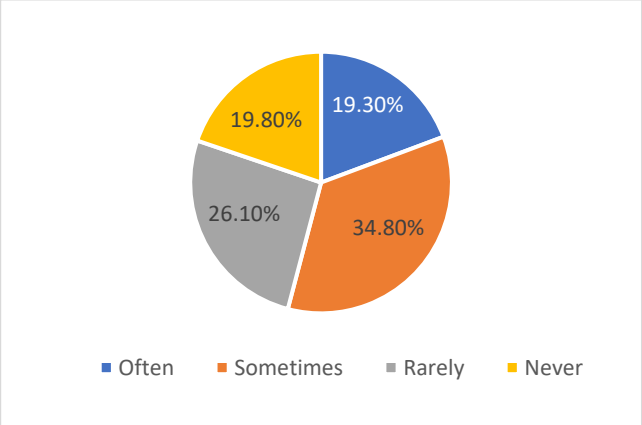
7. Methodological Rationale

The survey design and open-source intelligence (OSINT) checks were suitable for research on digital footprints because the traceability of digital footprints is not only determined by our online behaviours, but also the visibility of these behaviours. The survey data collected behavioural and awareness factors of online privacy, and OSINT enabled us to operationalise the measurement of the traceability of several personal identifiers (such as email addresses, phone numbers) through data breach websites and phone intelligence websites. This allowed us to investigate the visibility, privacy and impact of the open-source traceability of digital footprints in our everyday.

RESULT AND DISCUSSION

A total of **207 responses** were analyzed. The findings indicate moderate online activity but weak privacy practices and limited awareness of tools for checking data leaks. A substantial proportion of respondents also reported that identifiable personal information was visible online, suggesting meaningful exposure risk.

1. Public posting behavior



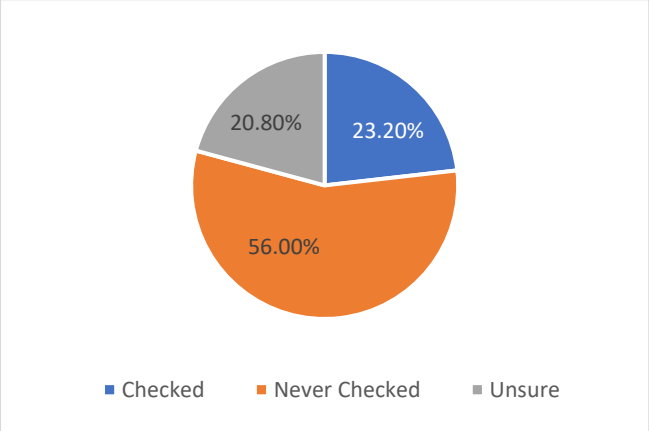
Public posting was common, though not extreme. The largest group reported posting **sometimes (72; 34.8%)**, followed by **rarely (54; 26.1%)**, **never (41; 19.8%)**, and **yes, often (40; 19.3%)**. This suggests that while frequent public posting was not dominant, many respondents still shared content publicly on at least an occasional basis.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

2. Privacy-setting management

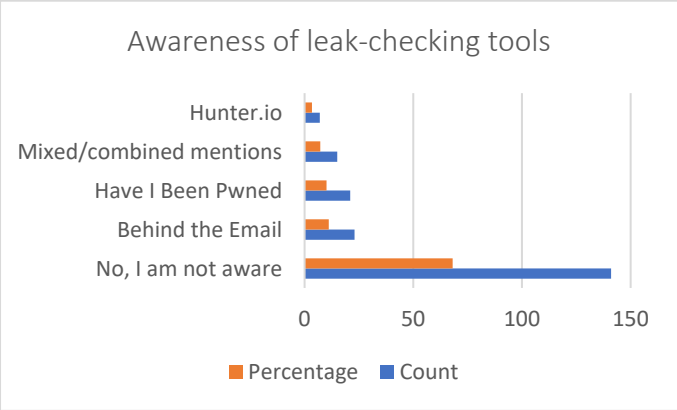
Privacy-setting management was inconsistent across the sample. Most respondents reported checking or managing privacy settings **rarely (102; 49.3%)**, followed by **often (92; 44.4%)**, and **never (13; 6.3%)**. This indicates that privacy awareness exists, but regular privacy maintenance is not fully established among respondents.

3. Data-breach checking



More than half of the respondents had **never checked** whether their data had leaked online. Specifically, **116 respondents (56.0%)** selected **never checked**, **48 (23.2%)** reported that they had **checked**, and **43 (20.8%)** were **unsure**. This shows a clear gap between online participation and active security monitoring.

4. Awareness of leak-checking tools



Awareness of websites and tools used to detect leaked personal data was low. The majority of respondents, **141 (68.1%)**, stated that they were **not aware** of such tools. Smaller groups mentioned **Behind the Email** (23), **Have I Been Pwned** (21), **Hunter.io** (7), and mixed or combined mentions (**15**). This reflects limited practical knowledge of available privacy-protection resources.

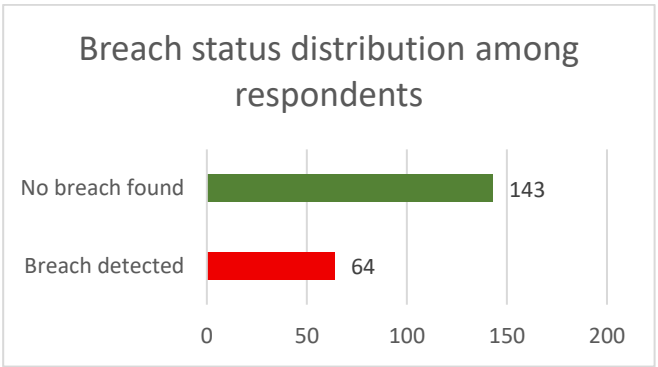
5. Visible online identifiers

Identifier visible online	Count	Percentage
Name	177	85.5%
Photos	90	43.5%
Email address	66	31.9%
Friends/family	65	31.4%
Workplace/school	59	28.5%
Phone number	52	25.1%
Location	37	17.9%

Note. Multiple responses were allowed.

When asked what strangers could see about them online, respondents most frequently reported their **name (177; 85.5%)**. Other visible identifiers included **photos (90; 43.5%)**, **email address (66; 31.9%)**, **friends/family (65; 31.4%)**, **workplace/school (59; 28.5%)**, **phone number (52; 25.1%)**, and **location (37; 17.9%)**. These findings show that even when respondents were cautious in some areas, multiple personal identifiers remained visible online.

6. Breach status distribution



Overall breach status showed that **64 respondents (30.9%)** had a **breach detected**, while **143 respondents (69.1%)** reported **no breach found**. This means nearly one-third of the sample had some form of breach exposure, which is significant for a privacy-focused study.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

7. Sources associated with respondent email breach exposure

Note. The email address was associated with 66 breach exposure entries across multiple distinguished categories of online services.

Distinguished category	Breach / exposure source
Social media and communication platforms	LinkedIn, Twitter, Dubsmash, MemeChat, Gravatar
E-commerce and retail platforms	BigBasket, IndiaMART, Paytm, Boat, Domino's India, Bookchor
Travel and mobility platforms	RailYatri, Ixigo, Zoomcar, Gaadi, ParkMobile, CarGurus
Survey, marketing, and analytics platforms	SurveyLama, ClearVoice Surveys, Quinstreet, Appen, Adda, Nitro
Cloud, productivity, and content tools	Canva, Dropbox, Trello, Dailymotion, Cutout.pro, Wattpad
Education and knowledge platforms	Vedantu, Internet Archive, Z-Lib, Raaga
Finance, business, and service platforms	Moneycontrol, Robinhood, Under Armour, Modern Business Solutions, Adobe, Covve
Entertainment and gaming platforms	Zynga, MGM Resorts, 1Win, Eatigo
Government and public services	TRAI
Breach repositories and aggregated exposure sources	Stealer logs, combolists, breached repositories, aggregated leak datasets,
	Exploit.in, Trik Spam Botnet
Other exposed services	Hathway, ClickASnap, PropTiger, Dunzo

Breach-check analysis identified 66 email breach exposure entries associated with the respondent record, spanning social media and communication platforms, e-commerce and retail platforms, travel and mobility platforms, survey and marketing platforms, cloud and productivity tools, education and knowledge platforms, finance and business services, entertainment and gaming platforms, government and public services, breach repositories, and other exposed services.

Overall, the findings indicate that although respondents showed some privacy awareness, many personal identifiers remained visible online and breach exposure was substantial.

DISCUSSION

Our findings indicate that the online footprint of respondents who are somewhat privacy aware is still very exposed. While many respondents reported that they have average control over their public and privacy settings, this study shows that it may not necessarily lead to good privacy behaviours. This is significant as it suggests that while users may be aware of privacy online, they may not be taking action to safeguard their data in their behaviour online.

A key finding is the public nature of the simple identifiers (name, photos, email and school or work) of the respondent. This behaviour is consistent with the aggregative nature of online footprints as seemingly less significant information becomes more significant when taken in aggregate. From an OSINT perspective, these identifiers are not only stand-alone data, but also linkable data which can be used to help identify, map and further analyse exposure. The results, therefore, confirm the notion of sharing as not being the outcome of a single sharing act, but multiple low-risk sharing acts on multiple platforms.

The findings on data-breach checking and knowledge of tools to check for data breaches is also important. Over 50% of our sample had not checked if their data had been breached and most were not aware of tools to check for breaches. This indicates users are more likely to be victims of data breaches, rather than taking steps to protect their data. This is worrying from a practical standpoint as users may not be aware if their

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

data has been breached, even if their data has been found in breach data sources.

This is supported by the analysis of the breach. The respondent's email address was linked to different sources of breach and exposure including social media, e-commerce, travel, survey, cloud, entertainment and finance. This demonstrates exposure is not service or function specific. Rather, the online identity is spread across multiple services and the potential for an identifier to be associated with multiple contexts and used for profiling, account matching or intelligence gathering is higher. This multi-source exposure in an OSINT world enhances the value and exposure of information.

The other key consideration is that privacy is reactive. Our findings from the survey indicate that users might not be considering privacy settings and exposure all the time, but reactively. This finding is in line with privacy fatigue, where users are conscious of the risks, but don't always think about them. To some degree, the study shows a disconnect between online skills and self-protective behaviours.

The research is also interesting in terms of the approach taken. The combination of self-reported survey data and breach verification shows that OSINT data can be used to complement the questionnaire data. This provides an external validity check on the questionnaire and that perceptions of exposure and exposure can be combined. This is valuable for the digital footprint research as it's not just perceptions, but can be used to measure potential exposure of users in various systems.

Lastly, the study shows that digital footprints are not only a consequence of digital interactions, but also pose a privacy and intelligence risk. The study participants were not prolific posters to the public, but had multiple identifiers and risk of exposure to breaches. This suggests the risks of privacy in the digital world are aggregative, diffuse and uncontrollable. The journal article discussion should focus on the fact that knowledge is only as good as the privacy management, breach monitoring and knowledge of how online traces can be combined that underpins it.

CONCLUSION

This research shows that online traces are persistent, and potentially verifiable, exposure. This study showed that while the participants were concerned about privacy, they were not concerned about identifiers. Identifiers, photographs, email addresses, employment information and other incidental information were persistent across media and suggested a persistence of online traces.

The breach-verification aspect also implied that email identifiers were linked to a variety of known sources of exposure, and implied the possibility of republished data in other media. This implies for OSINT that user actions not only result in social exposure, but intelligence. The multi-platform identifier tracing suggests the use of open-data traces for privacy, exposure and risk.

In general, the research demonstrates that online privacy is not about actions, but traces. Users should be more conscious, check breaches and privacy to prevent exposure in the hyper-connected world. This insight adds to the emerging field of online traces, traceability and OSINT exposure studies.

FUTURE DIRECTIONS

The study should be repeated with a larger and more diverse sample to see if there are any variations in the visibility of online footprints by age group, occupation, geographical location and online literacy. Cross-sectional studies with different online platforms and users would also indicate the impact of online platforms on visibility and privacy breach.

Longitudinal studies will be valuable as it can help us evaluate changes in privacy behaviours. This will help us to assess the impact of various education programs, data breaches and other incidents.

Future research can expand the OSINT verification process to obtain other information such as usernames, mobile numbers, profile URLs and other email addresses. This will give useful information of data and online footprints.

Secondly, the other focus should be on psychological and behavioural issues. Privacy fear, trust, risk perception and self-protecting online may be useful to understand why people sometimes know the risks of online activities, but don't change their behaviours.

Intervention studies would be useful. For example, this may include investigating the impact of privacy education, data breach monitoring and online literacy to enhance self-protecting online behaviour and manage risks. This will fill the research gap from descriptive to prescriptive.

REFERENCES

[1] Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science, 347*(6221), 509–514.

[2] Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature, 54*(2), 442–492.

[3] boyd, d., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210–230.

[4] Debatin, B., Lovejoy, J. P., Horn, A.-K., & Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication, 15*(1), 83–108.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

[5] Dwyer, C., Hiltz, S. R., & Passerini, K. (2007). Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. *AMCIS 2007 Proceedings*.

[6] Hargittai, E., & Litt, E. (2013). New strategies for employment? Internet skills and online privacy behaviors. *Digital Culture & Education, 5*(1), 16–31.

[7] Kang, R., Dabbish, L., Fruchter, N., & Kiesler, S. (2015). "My data just goes everywhere": User mental models of the internet and implications for privacy and security. *Proceedings of the Eleventh Symposium on Usable Privacy and Security*, 39–52.

[8] Koohikamali, M., French, A. M., & Kim, D. J. (2017). A systematic review of privacy research in social networking sites from 2005 to 2015. *Computers in Human Behavior, 67*, 251–263.

[9] Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security, 64*, 122–134.

[10] Marwick, A. E., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society, 13*(1), 114–133.

[11] Mayer-Schönberger, V. (2009). *Delete: The Virtue of Forgetting in the Digital Age*. Princeton University Press.

[12] Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.

[13] Omand, D., Bartlett, J., & Miller, C. (2012). Introducing social media intelligence (SOCMINT). *Intelligence and National Security, 27*(6), 801–823.

[14] Pötzsch, S. (2015). The emergence of iPrivacy: Investigating the self, surveillance and social networking. *Surveillance & Society, 13*(2), 221–236.

[15] Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review, 154*(3), 477–560.

[16] Stoyanovich, J., Abiteboul, S., & Jagadish, H. V. (2019). Responsible data management. *Communications of the ACM, 62*(6), 40–43.

[17] Trottier, D. (2012). *Social Media as Surveillance: Rethinking Visibility in a Converging World*. Ashgate.

[18] Zheleva, E., & Getoor, L. (2009). To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles. *Proceedings of the 18th International Conference on World Wide Web*, 531–540.

[19] Have I Been Pwned. (2026). *Pwned websites and data breach notifications*. https://haveibeenpwned.com

[20] IPQualityScore. (2026). *Phone intelligence and reputation lookup tools*. https://www.ipqualityscore.com