

Developers Need Protection, Too: Perspectives and Research Challenges for Privacy in Social Coding Platforms

Nicolás E. Díaz Ferreyra
Institute of Software Security
TU Hamburg
Hamburg, Germany
nicolas.diaz-ferreyra@tuhh.de

Abdessamad Imine
*Lorraine University and Loria-
Cnrs-Inria Nancy-Grand Est*
Nancy, France
abdessamad.imine@loria.fr

Melina Vidoni
School of Computing
Australian National University
Canberra, Australia
melina.vidoni@anu.edu.au

Riccardo Scandariato
Institute of Software Security
TU Hamburg
Hamburg, Germany
riccardo.scandariato@tuhh.de

Abstract—Social Coding Platforms (SCPs) like GitHub have become central to modern software engineering thanks to their collaborative and version-control features. Like in mainstream Online Social Networks (OSNs) such as Facebook, users of SCPs are subjected to privacy attacks and threats given the high amounts of personal and project-related data available in their profiles and software repositories. However, unlike in OSNs, the privacy concerns and practices of SCP users have not been extensively explored nor documented in the current literature. In this work, we present the preliminary results of an online survey (N=105) addressing developers' concerns and perceptions about privacy threats steaming from SCPs. Our results suggest that, although users express concern about social and organisational privacy threats, they often feel safe sharing personal and project-related information on these platforms. Moreover, attacks targeting the inference of sensitive attributes are considered more likely than those seeking to re-identify source-code contributors. Based on these findings, we propose a set of recommendations for future investigations addressing privacy and identity management in SCPs.

Index Terms—social coding platforms, privacy concerns, usable security, privacy engineering

I. INTRODUCTION

Over the last years, the landscape of modern software engineering has significantly evolved thanks to the emergence of Social Coding Platforms (SCPs) like GitHub and BitBucket [1, 2]. From software enthusiasts to more senior engineers, SCPs bring together individuals with diverse technical expertise and background beneath the premises of public and collaborative software development [3]. Furthermore, since their introduction in the late 2000s, SCPs have not only attracted new players into the Open Source Software (OSS) ecosystem, but also redefined the cooperative nature of OSS deployment by offering free hosting and version control capabilities enhanced with social features [4, 5].

At their core, SCPs resemble many characteristics of mainstream Online Social Networks (OSNs) like Facebook or Twitter. For instance, in order to network, collaborate, and disseminate their projects, SCP users often set up a profile like the one in Fig. 1. Such profiles condense plenty of information about developers' demographics and their behaviour within

the platform, including the projects and repositories they own, maintain, and contribute to. Moreover, some of the activity metrics one can obtain from a particular user include (i) the number of repositories they created, (ii) the number of issues they reported, (iii) the number of pull requests they submitted, and (iv) the number of commits they authored [6].

Developer-centred SCP metrics (and the information they aggregate) have helped researchers to characterise the GitHub community and unveil common software engineering practices in the wild. Still, such information can also be leveraged to carry on targeted cyberattacks on specific users and socio-technical groups. That is, on individuals of a certain gender, region, or with a particularly technical background (e.g., users prone towards insecure coding practices) [7]. Furthermore, with the current advances in Artificial Intelligence (AI) and Machine Learning (ML), even those keeping a more private, less visible profile could be re-identified thanks to the large amounts of data available on the platform.

Motivation: Privacy and identity management have been extensively investigated within the context of OSNs. Overall, the unintended disclosure of personal information, anonymity, and data misuse are some of the many issues addressed within the current literature [8]. Likewise, the inference of sensitive attributes from metadata and the re-identification of data subjects have received great attention among privacy researchers seeking to elaborate Privacy Enhancing Technologies (PETs) for online networking [9, 10]. Particularly, for shaping socio-technical solutions aiming to protect users' identity and personal attributes from re-identification, inference, and linkability attacks.

Despite the similarities between SCPs and OSNs, the privacy issues of the former have not received as much attention as the ones of the latter. However, information available in SCP platforms can be leveraged to perform social engineering attacks and, in turn, compromise the security of companies and public agencies at large. For instance, in early 2022, a group of cyberattackers gained access to 190 Gb of Samsung's data using secret keys exposed in public repositories [11]. Moreover, a recent report by GitGuardian [12] shows that the

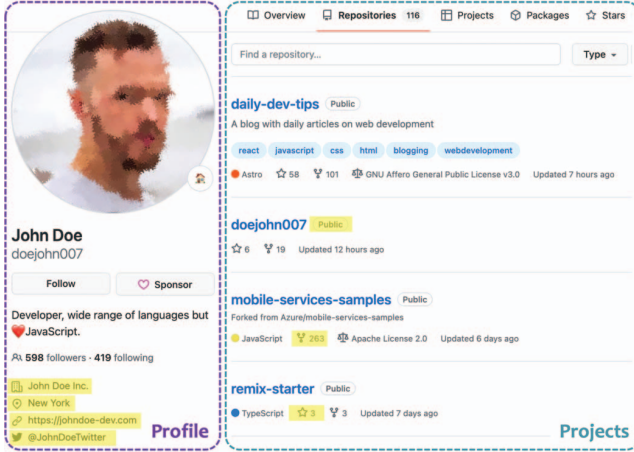


Fig. 1. An example of a developer page on GitHub.

extent of publicly-exposed secrets on GitHub has more than doubled since 2020. Though alarming, such numbers should not be surprising as it just takes a developer inadvertently pushing code to a public repository to compromise the security of a whole system [13]. Hence, there is a call for further investigations seeking to characterise the privacy threats and concerns emerging within collaborative coding environments.

Contribution and Research Questions: In this work, we present and discuss the preliminary results of an empirical study about privacy concerns in SCPs. Such a study consisted of an online survey (N=105) addressing the following Research Questions (RQs):

- **RQ1:** *How concerned are users of SCPs about their privacy?* To answer this RQ, we designed a questionnaire based on privacy-related constructs available in the current literature. Particularly, we assessed developers' perceived control over their personal and project-related information, their concerns about social/organisational threats, and their prior experiences of privacy invasion in SCPs, among others.
- **RQ2:** *What is their perceived likelihood of certain privacy threats emerging within SCPs?* The purpose of this RQ is to determine whether SCP users underestimate the chances of experiencing privacy attacks. For this, participants were asked to assess the likelihood of two hypothetical privacy attack scenarios describing (i) the inference of sensitive attributes based on profile data and (ii) the re-identification of code authorship from information available in SCP repositories.

All in all, our findings suggest the presence of concerns related to social and organisational privacy threats among SCP users. Still, many participants reported (i) feeling safe publishing personal and project-related data in these environments and (ii) having a sense of control over such data. We discuss the implications of these insights and elaborate on some perspectives for further research endeavours.

II. BACKGROUND AND RELATED WORK

A. Privacy Research in Online Social Networks

Many investigations have addressed users' concerns and attitudes towards service providers, information access, and personal data sharing in OSNs. These efforts have yielded significant findings about the role such concerns play on people's self-disclosure behaviour, access-control preferences, and overall perception of online privacy risks [8]. For instance, it has been repeatedly shown that users of OSN often fail to preserve the contextual integrity of their personal data despite their high privacy concerns [14]. This phenomenon is usually associated with a low usability of security controls and the lack of transparency with which platforms collect, process and store users' data [8]. Recent publications have also drawn attention to the cognitive biases affecting people's privacy decision-making processes [15]. Particularly, they suggest that users of OSNs are prone to replace rational risk judgements with "rules of thumb" or heuristics to reduce the complexity of their cybersecurity choices [16]. In turn, they often adopt sub-optimal privacy configurations and expose their data to untrusted audiences or malicious users.

Prior work has also sought to characterise potential privacy attacks in OSNs. Such attacks typically use information that users disclose in their profiles and posts to feed ML models capable of inferring sensitive attributes and behavioural patterns [17]. For instance, Gu et al. [18] developed a probabilistic model that employs user-centric and social-relationship data to infer individuals' home location. Such a model can achieve more than 90% accuracy if the targeted user had disclosed some location data in the past (e.g., check-in at a bar or a restaurant) and around 60% when not. Likewise, Comito [19] introduced a supervised learning framework for predicting the next location of an individual based on her prior mobility behaviour. Recent publications also give an account of the effectiveness of Deep Neural Networks (DNNs) for detecting sensitive information on images, including peoples' age [20], gender [20], and race [21]. Overall, DNN-based privacy attacks outperform other state-of-the-art approaches due to their high performance and efficiency over large datasets [17].

B. Privacy Research in Social Coding Platforms

Unlike in OSNs, research addressing SCPs has not placed much attention on users' concerns about privacy and data protection. Moreover, at its best, current investigations can be seen as attack models aiming to unveil correlations between developers' demographic attributes and behavioural patterns emerging within these platforms. For instance, Wu et al. [6] observed that the number of pull requests and commits generated by GitHub users are linked to specific categories of profile pictures (e.g., human, logo, or animal) and demographic characteristics. Research addressing gender biases has also contributed to characterising developers' behaviour in SCPs. A study by Imtiaz et al. [22] showed that female users in GitHub are more restrained in communication than men and tend to concentrate their work across fewer projects and organisations.

Likewise, Newton et al. [23] found that GitHub contributors identified as women have a shorter tenure in open source projects compared to men.

Research addressing code attribution has also helped, to a large extent, depict the landscape of privacy attacks in SCPs. Particularly, it has been shown that code available across public software repositories can be exploited to re-identify its contributors through stylistic patterns reflecting individual programming styles. These patterns range from simple artefacts in comments and code layouts to subtle habits in syntax usage and control flow. For instance, Caliskan-Islam et al. [24] applied ML methods to de-anonymize the authors of C/C++ source code using a stylometry feature set extracted from Abstract Syntax Trees (ASTs). Alongside, Alsulami et al. [25] introduced an LSTM model for the automatic extraction of such features, whereas Abuhamad et al. [26] proposed a similar technique targeting large-scale, language-oblivious, and obfuscation-resilient code authorship identification. This last one is a language-sensitive approach capable of identifying authors across four programming languages (e.g., C, C++, Java, and Python), even when using a combination of them (e.g., Java/C++, Python/C++).

III. METHODOLOGY

A. Survey Design

We conducted an online survey to answer the research questions presented in Section I. Such a survey had two main sections, the first addressing privacy concerns in SCPs (RQ1) and the second focusing on privacy attacks stemming from information available in profiles and code repositories (RQ2). For the elicitation of privacy concerns, we referred to the following psychological constructs and scales available in the current literature on OSNs:

- **Perceived Control (CTRL)** over the personal and project-related information disclosed in SCPs [27].
- **Perceived Privacy Risks (RSK)** in SCPs [27].
- **Prior Invasion of Internal Privacy (PIIP)** in SCPs, such as privacy violations experienced in the past [28].
- **Privacy Concerns on Organisational Threats (PCOT)**, including the collection and secondary usage of data by service providers and third parties [29].
- **External Informational Privacy Concerns (EIPC)** on information disclosure practices that may affect others' privacy in SCPs (e.g., colleagues and collaborators) [28].
- **Privacy Concerns on Social Threats (PCST)**, such as embarrassment and potentially malicious actions performed by other SCP users [29].

These scales have been elaborated and validated by their corresponding authors: CTRL and RSK by Krasnova et al. [27]; EIPC and PIIP by Morlok [28]; PCOT and PCST by Krasnova et al. [29]. We have slightly adapted them for the purpose of this study by replacing the term “OSNs” with “SCPs” on each scale item. All constructs were measured using a 6-point Likert scale ranging from *completely disagree* to *completely agree*. We have assessed their reliability by

SCENARIO 1: *Your activity in SCPs (e.g., commits, comments, and reactions) can be used to infer your gender and other socio-demographic information.*

SCENARIO 2: *Someone (e.g., a colleague) pushes code you wrote into a public repository using his/her account. Then the authorship of this code can be traced back to you using information available in SCPs (e.g., other code snippets you wrote and your profile information).*

Fig. 2. Hypothetical privacy attack scenarios in a SCP.

computing the corresponding Cronbach’s Alpha coefficients, which resulted higher than 0.7 in all cases. This suggests that the items of each scale have relatively high internal consistency since values above 0.7 are considered “acceptable”.

The second part of the survey consisted of two hypothetical privacy threat scenarios that participants had to assess. As shown in Fig. 3, such scenarios represent privacy attacks documented in the current literature that can be performed with information available in SCPs (Section II-B). That is, (i) finding sensitive correlations between developers’ activities and their demographic characteristics and (ii) re-identify their source code contributions through stylistic coding patterns. Particularly, we asked participants how likely they think these scenarios could occur in real life using a scale with the values *very likely*, *likely*, *somehow likely*, *somehow unlikely*, *unlikely*, and *very unlikely*.

B. Population and Sampling

Participants were recruited via Prolific¹ and pre-screened based on their (i) self-reported knowledge of software development techniques and (ii) self-reported computer programming skills. They had to be at least 18 years old to join the study and were rewarded with GBP 2.00 (paid through Prolific) for a completed survey of 15 minutes duration on average. As a standard quality control, we targeted users who already took part in at least 10 other studies and had a minimum approval rate of 98% [30]. Two attention questions were also included in the survey to identify and discard answers from unengaged participants. As a result, we collected 105 valid responses from 110 full submissions.

C. Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of German Association for Experimental Economic Research. All participants received information about the study procedure (including data privacy statements) and were asked to give their informed consent before joining the experiment. They were also given the possibility to withdraw at any time without their answers being recorded. Survey instruments, consent forms, and study results are available in the paper’s **Replication Package**².

¹<https://prolific.co/>

²<https://doi.org/10.5281/zenodo.7692654>

TABLE I
SURVEY ITEMS EXCERPT.

Construct	Analysed construct item	1	2	3	4	5	6
CTRL	I feel in control over the personal and technical information I provide on SCPs	0	9	8	33	43	12
RSK	I feel safe publishing my personal information and project-related information on SCPs	4	5	22	37	29	8
PIIP	I have already experienced a violation of my privacy on SCPs	44	42	9	5	4	1
PCOT	I am often concerned other parties could actually collect my publicly available information on SCPs	6	20	20	32	21	6
EIPC	I am concerned that unauthorised people may access the code of my colleagues I shared on SCPs	9	17	23	37	14	5
PCST	I am concerned that other users may take advantage of me based on the information they learned about me through SCPs	14	24	23	26	16	2

Note: 1= Completely disagree, 2= Disagree, 3= Somewhat disagree, 4= Somewhat agree, 5= Agree, 6= Completely agree.

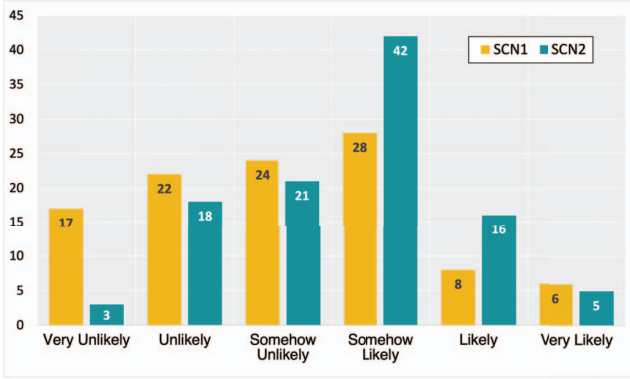


Fig. 3. Assessment of privacy attack scenarios.

IV. RESULTS AND DISCUSSION

Of 105 study participants, 94 were male, 10 were female, and 1 was non-binary. Around 48.6% reported having 2 to 5 years of experience using SCPs, 14.3% between 6 to 10, 13.3% less than 2, and 3.8% more than 10 years. Regarding software development, 35.2% had less than 2 years of experience, another 35.2% between 2 and 5, 16.2% between 6 and 10, and 13.3% more than 10 years. A complete overview of the sample's demographics is shown in Table II.

Each of the constructs elicited in the first part of the survey (i.e., CTRL, RSK, PIIP, PCOT, EIPC, and PCST) contains between 3 and 7 statements or items. We have selected one representative item per construct to conduct a preliminary analysis of developers' privacy concerns in SCPs. Though exploratory, such an approach meets the purpose of this work, which is to provide some initial insights and motivate further investigations in this regard. A complete overview of the results obtained for each scale item is available in the paper's Replication Package.

As shown in Table I, most participants agreed with the CTRL statement about feeling in control over the personal and technical information they disclose in SCPs. Likewise, about 71% agreed with the corresponding RSK statement (scores 4-6), whereas 29% reported not feeling safe publishing personal and project-related information (scores 1-3). Regarding prior

experiences, only 9.5% of respondents had suffered some privacy violation in the past, while the rest tended to disagree with this construct item. On the other hand, about 59% of the study participants reported concerns related to data collection from third parties (PCOT), and 56% were worried about unauthorised access to others' code they had shared (EIPC). Finally, when it came to social threats, 42% tended to agree with the PCST statement referring to the exploitation of publicly-available information by other SCP users.

Fig. 3 illustrates participants' assessment of the proposed privacy threat scenarios. We observe that 60% placed Scenario 1 (SCN1) in the "unlikely" side of the scale (i.e., *very unlikely*, *unlikely*, or *somewhat unlikely*), whereas 40% deemed it "likely" to a certain degree (i.e., *somewhat likely*, *likely*, or *very likely*). Surprisingly, these values get reversed when it comes to Scenario 2 (SCN2), where 60% of participants considered it to some extent "likely" and the remaining 40% "unlikely". When analysing each likelihood value separately, we can also spot some considerable differences between SCN1 and SCN2. Particularly, a larger number of participants considered SCN1 to be *very unlikely* as opposed to SCN2, though fewer times *likely* and *somewhat likely*.

A. Implications for Privacy Research

(i) **Privacy Decision-Making:** Our results suggest that, although users of SCPs usually feel in control of their information, they are also concerned about privacy threats stemming from (i) data collection by service providers and third parties, (ii) unauthorised access to code authored by other users, and (iii) the exploitation of publicly-available information. Prior investigations in OSNs have also reported a similar offset phenomenon often referred to as the "privacy paradox" [31]. That is, a discrepancy between users' privacy concerns and actual privacy-related behaviour [32]. Nevertheless, a feeling of control normally contributes to mitigating these concerns and ultimately to higher levels of information disclosure [32]. Hence, this observation should be further investigated and discussed with the help of additional empirical evidence.

(ii) **Usability and Transparency:** Participants' answers to the PCOT, EIPC, and PCST items also raise questions about the usability of current privacy-enhancing mechanisms

TABLE II
SURVEY SELF-REPORTED DEMOGRAPHIC DATA.

Demographic	Ranges	Freq.	%
Gender	Male	94	89.5%
	Female	10	9.5%
	Non-Binary	1	1.0%
Educational level	High School or Less	14	13.3%
	Some College	19	18.1%
	Undergraduate (BSc, BA)	48	45.7%
	Graduate (MSc, PhD)	24	22.9%
Employment status	Student	26	24.8%
	Unemployed, looking for work	10	9.5%
	Unemployed, not looking for work	2	1.9%
	Working full-time	13	12.4%
	Working part-time	54	51.4%
Software development experience	<2 years	37	35.2%
	2-5 years	37	35.2%
	6-10 years	17	16.2%
	>10 years	14	13.3%
Experience using SCPs	<2 years	35	13.3%
	2-5 years	51	48.6%
	6-10 years	15	14.3%
	>10 years	4	3.8%

in SCPs. The perception of organisational threats is often a product of a lack of transparency about how (personal) information is collected, processed, stored, and shared by the platform [33]. In this regard, it would be beneficial to analyse the extent to which SCP users have a fair understanding of common data-processing practices (e.g., from SCPs and third parties) and how their knowledge is translated into more (or less) restrictive privacy configurations (e.g., visibility of software repositories). To some degree, similar concerns have also been reported by practitioners within the Mining Software Repositories (MSR) community [34, 35]. Particularly on the ethical implications of retrieving information from SCP repositories without obtaining full consent from their users. It is thus critical not only to review the potentially invasive nature of these practices, but also to assess the limitations and drawbacks of current access-control mechanisms in SCPs.

(iii) Multi-Party Privacy: Users’ social and external privacy concerns have motivated recent investigations addressing the resolution of multi-party privacy conflicts in OSNs [36]. Such conflicts often arise when users share information that can also compromise the privacy of others, such as a group picture or post mentions. All in all, multi-party privacy takes a collective view on the norms and boundaries of information disclosure, putting special emphasis on the conflicting privacy preferences among the co-owners of particular data items [37]. Secrets sprawl across software repositories can be considered as an instance of such conflicts in SCPs. Furthermore, pull request-based workflows could help (in principle) to prevent secret sharing as they promote the monitoring and reviewing of code being pushed to a particular repository [38]. Still, specific sources of multi-party privacy conflicts, concerns and countermeasures should be systematically investigated.

(iv) Risk Assessment: Prior research has shown that past experiences can shape, to a great extent, people’s privacy

perception and behaviour [39]. Therefore, it is not surprising that a similar number of participants tended to disagree with the PIIP statement but also agree with the RSK statement (90% and 70%, respectively). Furthermore, participants’ assessment of the proposed threat scenarios is also aligned with their overall perception of privacy control and safeness, as many believed SCN1 and SCN2 to be *very unlikely*, *unlikely*, or *somehow unlikely*. Given the central role of risk perception in privacy decision-making [15], it is critical to raise awareness among developers about the cybersecurity threats emerging within SCPs. Equally important is to endow them with methods and tools that help preserve the contextual integrity of the information they share while assessing the potential risks.

B. Limitations and Threats to Validity

The present work is subjected to limitations related to the size and composition of the studied sample. On the one hand, its relatively small size does not allow us to generalise the study results to the whole SCP community. Hence, the insights gained from it should be seen as preliminary instead of conclusive and motivate further investigations in this area. On the other hand, participants recruited via crowd sourcing platforms may incur in dishonest practices such as falsely claiming to meet the experiment’s eligibility criteria or provide random answers. As mentioned in Section III-B, we have sought to minimise these threats by employing Prolific’s built-in qualification features to recruit suitable participants who may actively engage in our study. In the future we plan enhance this step by applying expertise questions like the ones proposed by Danilova et al. [40].

V. CONCLUSION AND FUTURE WORK

SCPs are rich sources of information about developers’ sociotechnical skills and have helped, to a great extent, characterise current trends in software engineering. However, despite their importance, little attention has been placed on the privacy-related behaviour of SCP users and the threats steaming from unsavvy information disclosure practices. In this work, we provided some empirical insights in this regard with the aim of paving the road for future investigations. Overall, our results suggest the need for further research assessing the usability of the current privacy-enhancing mechanism available in SCPs. That is, to determine whether such mechanisms meet users’ individual and collective goals of transparency, access control, and anonymity.

Developers’ knowledge and perception of privacy threats also call for additional research efforts, as it is critical to assess their (potential) lack of awareness in order to outline adequate cybersecurity training programs. The role of risk awareness in the adoption of privacy-enhancing technologies has been thoroughly studied and documented across the OSN literature [8]. In future work, we plan to delve into the interplay between risk awareness and privacy-related behaviour in SCPs. Particularly, on developers’ knowledge about potential privacy threats (e.g., secrets sprawl or code re-attribution) and their adoption of cybersecurity best practices (e.g., see [38]).

REFERENCES

- [1] T. Kinsman, M. Wessel, M. A. Gerosa, and C. Treude, "How Do Software Developers Use GitHub Actions to Automate Their Workflows?" in *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*, 2021, pp. 420–431.
- [2] T. Mens, M. Cataldo, and D. Damian, "The Social Developer: The Future of Software Development [Guest Editors' Introduction]," *IEEE Software*, vol. 36, no. 1, pp. 11–14, 2019.
- [3] J. E. Montandon, M. T. Valente, and L. L. Silva, "Mining the Technical Roles of GitHub Users," *Information and Software Technology*, vol. 131, p. 106485, 2021.
- [4] V. Cosentino, J. L. C. Izquierdo, and J. Cabot, "A systematic mapping study of software development with GitHub," *IEEE Access*, vol. 5, pp. 7173–7192, 2017.
- [5] M. Meli, M. R. McNiece, and B. Reaves, "How Bad Can It Get? Characterizing Secret Leakage in Public GitHub Repositories," *Proceedings 2019 Network and Distributed System Security Symposium*, 2019.
- [6] Y. Wu, Y. Zhang, T. Wang, and H. Wang, "Exploring the Relationship Between Developer Activities and Profile Images on GitHub," in *Proceedings of the 11th Asia-Pacific Symposium on Internetwork*, 2019, pp. 1–10.
- [7] B. Lazarine, S. Samtani, M. Patton, H. Zhu, S. Ullman, B. Ampel, and H. Chen, "Identifying vulnerable GitHub repositories and users in scientific cyberinfrastructure: An unsupervised graph embedding approach," in *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*. IEEE, 2020, pp. 1–6.
- [8] S. Oukemeni, H. Rifà-Pous, and J. M. M. Puig, "Privacy analysis on microblogging online social networks: a survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, 2019.
- [9] J. Henriksen-Bulmer and S. Jeary, "Re-identification attacks—A systematic literature review," *International Journal of Information Management*, vol. 36, no. 6, pp. 1184–1192, 2016.
- [10] G. Beigi and H. Liu, "A survey on privacy in social media: Identification, mitigation, and applications," *ACM Transactions on Data Science*, vol. 1, no. 1, pp. 1–38, 2020.
- [11] SecurityWeek. (2022) Thousands of Secret Keys Found in Leaked Samsung Source Code. Accessed: 22.12.2022. [Online]. Available: <https://www.securityweek.com/thousands-secret-keys-found-leaked-samsung-source-code>
- [12] GitGuardian. (2022) The State of Secrets Sprawl 2022. Accessed: 22.12.2022. [Online]. Available: https://res.cloudinary.com/da8kiytlc/image/upload/v1646148528/GitGuardian_StateOfSecretsSprawl2022.pdf
- [13] T. Segura. (2022, April) Has Your Code Leaked? Accessed: 22.12.2022. [Online]. Available: <https://www.infosecurity-magazine.com/blogs/has-your-code-leaked/>
- [14] H. Schwartz-Chassidim, O. Ayalon, T. Mendel, R. Hirschprung, and E. Toch, "Selectivity in posting on social networks: The role of privacy concerns, social capital, and technical literacy," *Heliyon*, vol. 6, no. 2, p. e03298, 2020.
- [15] S. Ostendorf and M. Brand, "Theoretical conceptualization of online privacy-related decision making—Introducing the tripartite self-disclosure decision model," *Frontiers in Psychology*, vol. 13, 2022.
- [16] N. E. Díaz Ferreyra, S. Ostendorf, E. Aïmeur, M. Heisel, and M. Brand, "ENAGRAM: An App to Evaluate Preventative Nudges for Instagram," in *Proceedings of the 2022 European Symposium on Usable Security (EuroUSEC '22)*. ACM, 2022, p. 53–63.
- [17] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin, "When Machine Learning Meets Privacy: A Survey and Outlook," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1–36, 2021.
- [18] Y. Gu, Y. Yao, W. Liu, and J. Song, "We Know Where You Are: Home Location Identification in Location-Based Social Networks," in *25th International Conference on Computer Communication and Networks (ICCCN '16)*. IEEE, 2016, pp. 1–9.
- [19] C. Comito, "NexT: A framework for next-place prediction on location based social networks," *Knowledge-Based Systems*, vol. 204, p. 106205, 2020.
- [20] A. Garain, B. Ray, P. K. Singh, A. Ahmadian, N. Senu, and R. Sarkar, "GRA_Net: A Deep Learning Model for Classification of Age and Gender From Facial Images," *IEEE Access*, vol. 9, pp. 85 672–85 689, 2021, publisher: IEEE.
- [21] T. Vo, T. Nguyen, and C. T. Le, "Race Recognition Using Deep Convolutional Neural Networks," *Symmetry*, vol. 10, no. 11, p. 564, 2018.
- [22] N. Imtiaz, J. Middleton, J. Chakraborty, N. Robson, G. Bai, and E. Murphy-Hill, "Investigating the Effects of Gender Bias on GitHub," in *2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*. IEEE, 2019, pp. 700–711.
- [23] O. B. Newton and J. Song, "Modeling Gender Differences in Membership Change in Open Source Software Projects," *arXiv preprint arXiv:2206.08485*, 2022.
- [24] A. Caliskan-Islam, R. Harang, A. Liu, A. Narayanan, C. Voss, F. Yamaguchi, and R. Greenstadt, "De-Anonymizing Programmers Via Code Stylometry," in *24th USENIX Security Symposium (USENIX Security '15)*, 2015, pp. 255–270.
- [25] B. Alsulami, E. Dauber, R. Harang, S. Mancoridis, and R. Greenstadt, "Source code authorship attribution using long short-term memory based networks," in *European Symposium on Research in Computer Security*. Springer, 2017, pp. 65–82.
- [26] M. Abuhamad, T. AbuHmed, A. Mohaisen, and D. Nyang, "Large-Scale and Language-Oblivious Code Authorship Identification," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18)*, 2018, pp. 101–114.
- [27] H. Krasnova, S. Spiekermann, K. Koroleva, and T. Hildebrand, "Online Social Networks: Why We Disclose," *Journal of Information Technology*, vol. 25, no. 2, pp. 109–125, 2010.
- [28] T. Morlok, "Sharing is not Caring: The Role of External Privacy in Users' Information Disclosure Behaviors on Social Network Sites," in *Proceedings of the 20th Pacific Asia Conference on Information Systems (PACIS 2016)*. AIESeL, 2016.
- [29] H. Krasnova, O. Günther, S. Spiekermann, and K. Koroleva, "Privacy concerns and identity in online social networks," *Identity in the Information Society*, vol. 2, no. 1, pp. 39–63, Dec 2009.
- [30] J. Salminen, S.-g. Jung, and B. J. Jansen, "Suggestions for online user studies," in *International Conference on Human-Computer Interaction*. Springer, 2021, pp. 127–146.
- [31] S. Barth, M. D. de Jong, M. Junger, P. H. Hartel, and J. C. Roppelt, "Putting the privacy paradox to the test: Online privacy and security behaviors among users with technical knowledge, privacy awareness, and financial resources," *Telematics and Informatics*, vol. 41, pp. 55–69, 2019.
- [32] B. Kim and D. Kim, "Understanding the Key Antecedents of Users' Disclosing Behaviors on Social Networking Sites: The Privacy Paradox," *Sustainability*, vol. 12, no. 12, p. 5163, 2020.
- [33] D. Wilkinson, M. Namara, K. Patil, L. Guo, A. Manda, and B. Knijnenburg, "The Pursuit of Transparency and Control: A Classification of Ad Explanations in Social Media," in *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS)*, 2021, p. 763.
- [34] N. E. Gold and J. Krinke, "Ethics in the mining of software repositories," *Empirical Software Engineering*, vol. 27, no. 1, pp. 1–49, 2022.
- [35] M. Vidoni and N. E. Diaz Ferreyra, "Should I Get Involved? On the Privacy Perils of Mining Software Repositories for Research Participants," in *1st International Workshop on Recruiting Participants for Empirical Software Engineering (RoPES'22)*, 2022.
- [36] M. Humbert, B. Trubert, and K. Huguenin, "A Survey on Interdependent Privacy," *ACM Computing Surveys (CSUR)*, vol. 52, no. 6, oct 2019.
- [37] J. M. Such and N. Criado, "Multiparty Privacy in Social Media," *Communications of the ACM*, vol. 61, no. 8, pp. 74–81, 2018.
- [38] A. Krause, J. H. Klemmer, N. Huaman, D. Wermke, Y. Acar, and S. Fahl, "Poster: Committed by Accident – Prevention and Remediation Strategies Against Secret Leakage," In *43rd IEEE Symposium on Security and Privacy (IEEE S&P)*, May 2022.
- [39] K. Renaud, M. Dupuis, and R. Searle, "Cybersecurity regrets: I've had a few je ne regrette," in *New Security Paradigms Workshop (NSPW '22)*, New Hampshire, 2022.
- [40] A. Danilova, A. Naiakshina, S. Horstmann, and M. Smith, "Do You Really Code? Designing and Evaluating Screening Questions for Online Surveys With Programmers," in *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*. IEEE, 2021, pp. 537–548.