

Singapore Management University

## Institutional Knowledge at Singapore Management University

---

Research Collection School Of Computing and Information Systems

School of Computing and Information Systems

---

5-2023

### She elicits requirements and he tests: Software engineering gender bias in large language models

Christoph TREUDE

Singapore Management University, [ctreude@smu.edu.sg](mailto:ctreude@smu.edu.sg)

Hideaki HATA

Follow this and additional works at: [https://ink.library.smu.edu.sg/sis\\_research](https://ink.library.smu.edu.sg/sis_research)



Part of the [Programming Languages and Compilers Commons](#), and the [Software Engineering Commons](#)

---

#### Citation

TREUDE, Christoph and HATA, Hideaki. She elicits requirements and he tests: Software engineering gender bias in large language models. (2023). *Proceedings of the 20th IEEE/ACM International Conference on Mining Software Repositories, Melbourne, Australia 2023 May 15-16*. 624-629.

Available at: [https://ink.library.smu.edu.sg/sis\\_research/8865](https://ink.library.smu.edu.sg/sis_research/8865)

This Conference Proceeding Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [cherylds@smu.edu.sg](mailto:cherylds@smu.edu.sg).

# She Elicits Requirements and He Tests: Software Engineering Gender Bias in Large Language Models

Christoph Treude  
The University of Melbourne, Australia  
christoph.treude@unimelb.edu.au

Hideaki Hata  
Shinshu University, Japan  
hata@shinshu-u.ac.jp

**Abstract**—Implicit gender bias in software development is a well-documented issue, such as the association of technical roles with men. To address this bias, it is important to understand it in more detail. This study uses data mining techniques to investigate the extent to which 56 tasks related to software development, such as assigning GitHub issues and testing, are affected by implicit gender bias embedded in large language models. We systematically translated each task from English into a genderless language and back, and investigated the pronouns associated with each task. Based on translating each task 100 times in different permutations, we identify a significant disparity in the gendered pronoun associations with different tasks. Specifically, requirements elicitation was associated with the pronoun “he” in only 6% of cases, while testing was associated with “he” in 100% of cases. Additionally, tasks related to helping others had a 91% association with “he” while the same association for tasks related to asking coworkers was only 52%. These findings reveal a clear pattern of gender bias related to software development tasks and have important implications for addressing this issue both in the training of large language models and in broader society.

## I. INTRODUCTION

Implicit gender bias is prevalent among professional software developers. For instance, a study of 142 professional software engineers undertaking an Implicit Association Test [1] found significant associations between men and technical leadership positions, general technical positions, and career advancement. Garcia et al. [2] discovered that *pink tasks*, which are tasks that require high standards and timeliness but offer little substantive development or visibility [3], were often assigned to female students in team projects. Terrell et al. [4] found that while pull requests submitted by women tend to be accepted more often than those submitted by men, this is only the case when the women’s identities are not immediately apparent.

Gender bias can lead to a lack of representation and opportunities for underrepresented groups, which can negatively impact innovation and productivity. For example, Vasilescu et al. [5] found gender diversity to be a significant positive predictor of productivity in GitHub teams. Gender bias can also perpetuate discrimination and create a hostile work environment, leading to high turnover rates with the associated knowledge loss [6] and a lack of diversity in the workforce.

To effectively address these biases, it is crucial to understand them in more detail. By identifying specific tasks and activities that are affected by gender bias, we can target our efforts to

eliminate bias more effectively. Understanding the nuances and complexities of bias, such as how it can manifest differently depending on the context, is crucial to creating an inclusive and equitable software development community.

The advent of large language models has given us a new powerful tool to study such gender bias based on data mining techniques rather than expensive surveys and experiments. Large language models process vast amounts of natural language data, making it possible to identify patterns of gendered language that may not be immediately apparent when looking at raw data, such as detecting problematic associations using Caliskan et al.’s Word Embedding Association Test [7] or investigating the use of gendered pronouns in association with certain tasks or positions [8].

In this study, we use back-translation [9] to mine gender bias associated with 56 tasks related to software development, taken from previous work [10]. We use the pronoun “she” to describe each task in English, translate the tasks into the genderless language of Finnish, and back into English using the DeepL translator.<sup>1</sup> We then analyse the pronouns in the result. To account for the impact of context on translation, we perform 100 permutations of the task list and aggregate the results.

We identify a significant disparity in the gendered pronoun associations with different tasks. For instance, the sentence “As a software engineer, she elicits requirements” was changed to “As a software engineer, he elicits requirements” in only 6% of all cases after round-trip translation (with “he or she” and “he/she” accounting for the remaining cases), while “As a software engineer, she tests” was changed to “As a software engineer, he tests” in 100% of the cases. Other tasks with a relatively weak association with “he” included task estimation, infrastructure setup, and support tasks, while providing comments and learning were exclusively associated with “he”.

Our findings reveal the specific software development tasks and activities that are most affected by gender bias, which can help everyone involved in software projects anticipate and proactively address potential issues. For example, role assignments should be made with conscious effort to avoid reinforcement of stereotypes. Additionally, mining techniques can be used to uncover patterns of gendered language in soft-

<sup>1</sup><https://www.deepl.com/translator>

ware repositories and internal communication, and necessary adjustments can be made to eliminate bias. Our work also highlights the importance of addressing bias in large language models at a fine-grained task level, in addition to efforts aimed at addressing more coarse-grained biases related to social [11], political [12], and sentiment [13] aspects.

## II. RELATED WORK

The biases that systems can contain have been studied in areas such as machine learning, natural language processing, and deep learning. Caliskan et al. [7] showed for the first time that applying standard machine learning to ordinary language results in human-like semantic bias. For biases in text corpora, Bordia and Bowman [8] proposed a metric to measure gender bias and a method to reduce gender bias. Mehrabi et al. [14] investigated real-world applications that demonstrated bias in different ways and listed different sources of bias that could affect artificial intelligence applications. They also examined different domains and sub-domains in AI and showed how researchers have observed and tried to address unfair results using state-of-the-art methods.

Various empirical studies have been conducted on gender bias in software development. Vasilescu et al. [15] assessed gender representation and social impact on Stack Overflow and reported that the majority of contributors to Stack Overflow are men, and that men have gained more reputation. Kuechler et al. [16] examined the mailing list subscription and posting statistics of FOSS participants in six FOSS projects and found that the participation rate of women is decreasing. Comparing the acceptance rates of contributions from men and women in the open source software community, Terrell et al. [4] reported that women’s contributions tend to be accepted more frequently than men’s, while for contributors who are outsiders to the project and whose gender is identifiable, the acceptance rate is higher for men. From an analysis of the impact of gender bias in the GitHub, Imtiaz et al. [17] reported that women provide less information, work on fewer projects and organisations, and are more restrictive in expressing their sentiments. Prana et al. [18] conducted a multi-regional geographic analysis of gender inclusion in GitHub and reported that gender diversity is low in all regions of the world, with no substantial differences by region. In an empirical study of newcomer women, Padala et al. [19] reported that female newcomers encountered gender bias in 63% of the barriers they faced. Rossi and Zacchiroli [20] analyzed the development histories of a large number of software projects and reported that while the percentage of female authors worldwide is low but steadily increasing, the percentage of female participation declined during the COVID-19 epidemic.

This study is unique in its subject and approach: it analyzes fine-grained tasks in software development from the perspective of current large language models.

## III. RESEARCH METHOD

In this section, we outline our research question and describe our methods for data collection and data analysis.

### A. Research Question

Our research question is:

**RQ** What implicit gender biases are embedded in large language models about software engineering tasks?

We aim to study software engineering gender bias in large language models and identify software engineering tasks that are typically associated with a specific gender.

### B. Data Collection

To answer our research question, we needed a list of the various activities that fit into a software developer’s job description. The most comprehensive list we identified was recently published by Masood et al. [10]. As part of their work, the authors surveyed literature on software development tasks and activities, and synthesized a list of example tasks and their categorization, primarily building on the work of Meyer et al. [21], Licorish and MacDonell [22], Graziotin et al. [23], Ford and Parnin [24], Milewski [25], Glass et al. [26], Murgia et al. [27], and Madampe et al. [28]. We manually analysed the examples (from Table I of Masood et al. [10]) and split them into separate activities where applicable, e.g., extracting the activities “assign GitHub issues” and “review pull requests” from the example “Assigning GitHub issue or reviewing pull request”, for a total of 56 software engineering tasks.

To identify bias, we relied on back-translation with a genderless language, i.e., a language that does not have grammatical gender, meaning that nouns, pronouns, and adjectives are not assigned a gender. We chose Finnish as the genderless language [29]. As an example, the translation of the phrases “she is a software engineer” and “he is a software engineer” both results in the Finnish phrase “hän on ohjelmistoinsinööri” since the Finnish word “hän” is the translation of both “she” and “he”. When translating the Finnish phrase back into English, the translator needs to decide whether to translate “hän” into “she” or “he”, and translators that rely on large language models tend to use their implicit gender biases to make this decision.

We chose the DeepL translator for our study. DeepL advertises itself as “the most accurate and nuanced machine translation”<sup>2</sup> and provides a free trial and an API. We were not able to use alternatives such as Google Translate or ChatGPT for our study since these tools have built-in heuristics to avoid biased output when translating from a genderless language to a gendered one, such as providing both translations (Google Translate) or providing disclaimers (ChatGPT).

Through our first experiments with DeepL, we noticed that the gendered pronoun in the final result depends on the surrounding sentences. For example, if only a single task was given, “she” was always back-translated into “he”. However, if multiple tasks were given in the same input, some tasks were assigned to “she”, some were assigned to “he/she”, some were assigned to “he or she”, some were assigned to “he”, and for a small minority, all pronouns were omitted. To systematically

<sup>2</sup><https://www.deepl.com/en/whydeepl>

TABLE I

RESULTS INDICATING THE FREQUENCY OF PRONOUN TRANSLATION FOR EACH SENTENCE, BROKEN DOWN BY “SHE”, “HE/SHE”, “HE OR SHE”, AND “HE”. THE PERCENTAGE OF INSTANCES IN WHICH THE PRONOUN WAS TRANSLATED AS “HE” IS PROVIDED IN THE FINAL COLUMN. NOTE THAT A SMALL NUMBER OF TRANSLATIONS DID NOT PRODUCE ANY PRONOUN, AND THAT ALL ORIGINAL SENTENCES BEGAN WITH “AS A SOFTWARE ENGINEER, ”.

Original Sentence	“she”	“he/she”	“he or she”	“he”	% “he”
She elicits requirements.	0	51	43	6	6%
She estimates tasks/projects.	0	61	0	39	39%
She performs infrastructure setup.	0	39	14	47	47%
She performs support tasks.	0	44	6	49	49%
She archives code versions.	0	16	34	50	50%
She generates reports/documents.	0	47	0	51	52%
She submits changes.	0	23	26	51	51%
She asks coworkers.	0	46	2	52	52%
She performs administrative tasks.	0	35	7	57	58%
She performs personal debugging.	0	38	5	57	57%
She performs user training.	0	28	9	63	63%
She stores design versions.	1	11	21	67	67%
She assigns GitHub issues.	2	22	6	69	70%
She manages development branches.	1	17	8	74	74%
She mentors others.	0	13	13	74	74%
She browses FAQs.	0	24	0	76	76%
She browses documentation.	0	19	3	77	78%
She commits code.	0	14	7	78	79%
She reviews pull requests.	0	14	6	80	80%
She assesses potential problems.	0	18	1	81	81%
She fixes bugs.	0	16	2	82	82%
She reads/reviews code.	0	18	0	82	82%
She has meetings.	0	16	1	83	83%
She navigates code.	0	14	3	83	83%
She reads changes.	0	17	0	83	83%
She edits code.	0	13	3	84	84%
She edits artifacts.	0	12	3	85	85%
She writes documentation/wiki pages.	0	14	0	86	86%
She accepts changes.	0	10	2	87	88%
She produces on-line help.	0	9	2	87	89%
She submits pull requests.	0	11	2	87	87%
She classifies requirements.	0	12	0	88	88%
She inspects code.	0	9	2	88	89%
She networks.	0	11	1	88	88%
She provides comments on project milestones.	0	11	0	89	89%
She fixes defects.	0	8	2	90	90%
She provides comments on commits.	0	10	0	90	90%
She helps others.	0	9	0	91	91%
She produces user documentation.	0	0	1	91	99%
She provides enhancements.	0	8	0	92	92%
She releases code versions.	0	7	1	92	92%
She browses the web.	0	6	0	94	94%
She maintains changes.	0	2	0	94	98%
She reads artifacts.	0	5	0	95	95%
She produces tutorials.	0	1	0	96	99%
She browses articles.	0	3	0	97	97%
She identifies constraints.	0	1	0	97	99%
She codes.	0	2	0	98	98%
She plans.	0	2	0	98	98%
She writes emails.	0	2	0	98	98%
She removes dead code.	0	1	0	99	99%
She restructures code.	0	1	0	99	99%
She writes artifacts.	0	1	0	99	99%
She learns.	0	0	0	100	100%
She provides comments on issues.	0	0	0	100	100%
She tests.	0	0	0	100	100%

study this, we provided DeepL with the list of 56 tasks 100 times, randomly shuffling the order of tasks each time.

Several tasks, such as “accept changes”, would have a different meaning outside of a software engineering context. Therefore, we needed to ensure that the software engineering context was injected into the translation process. We experimented with different prefixes and found that the short phrase “As a Software Engineer” was not changed during back translation; it still said “As a Software Engineer” after translating in both directions. Since “As a Software Engineer” is not a task, we believe that this prefix would not introduce task-related bias. We also used “she” as the initial pronoun in all tasks, e.g., for the task “help others”, the complete phrase given to DeepL was “As a software engineer, she helps others.” This provided us with a total of 5,600 translation results, i.e., 100 back-translations for each of the 56 phrases.

### C. Data Analysis

To analyse the data, we aggregated for each of the 56 input sentences how often it had been back-translated using each pronoun variant (“she”, “he/she”, “he or she”, “he”) across the 100 runs. Since “he” was by far the most common pronoun in the back-translations (in  $4,490/5,600 = 80.2\%$  of all sentences), we then calculate for each input sentence how often it was back-translated using the pronoun “he”, compared to all other translations.

### D. Data Availability

We make our data and scripts available in our online appendix:<sup>3</sup> `tasks.txt` contains all sentences used as input for the back-translation, `llm-bias.py` contains the Python code that uses the DeepL API to back-translate each sentence 100 times, `output.txt` shows the raw output, and `llm-bias.csv` aggregates the results.

## IV. RESULTS

Table I shows the results. While the majority of tasks were translated using the pronoun “he”, several tasks were often translated using “he/she” or “he or she”, with only four out of 5,600 translations using the “she” pronoun that was originally given to the translator. The results indicate a significant disparity in the gendered pronoun associations with different tasks. Four tasks (elicit requirements, estimate tasks/projects, perform infrastructure setup, perform support tasks) were associated with “he” in the minority of cases, with “he/she” being prominent. On the other end of the spectrum, six tasks were associated with “he” in at least 99 out of 100 runs (remove dead code, restructure code, write artifacts, learn, provide comments on issues, test).

If we consider Masood et al.’s grouping of tasks into 14 sub-categories [10], it is interesting to note that tasks related to requirements (elicit requirements 6%, estimate tasks/projects 39%, classify requirements 88%) have a weaker association with “he” on average than any other grouping of tasks, followed by tasks related to documents and their maintenance

(generate reports/documents 52%, store design versions 72%, maintain changes 98%). On the other hand, tasks related to providing comments are more strongly associated with “he” on average than any other grouping (provide comments on issues 100%, provide comments on commits 90%, provide comments on project milestones 89%).

For tasks that are directly related to each other, we observe several biases: asking coworkers has a relatively weak association with “he” (52%) while helping others has a much stronger association (91%). Browsing documentation and FAQs has a weaker association with “he” (78% and 76%, respectively) than browsing articles and the web (97% and 94%, respectively). Accepting changes has a stronger association with “he” (88%) than submitting them (51%). For artifacts, the strength of the association with “he” decreases from writing (99%) and reading (95%) to editing (85%).

## V. DISCUSSION

Our analysis reveals a significant disparity in the gendered pronoun associations with various software development tasks. Our findings indicate that the majority of tasks are translated using the pronoun “he” more often than “she”. However, the extent to which the large language model associates tasks with “he/she” or “he or she” instead of “he” varies. By being aware of these tendencies, individuals involved in software projects can take proactive steps to prevent and address potential issues. This includes being mindful of role assignments and making conscious efforts to avoid reinforcing stereotypes.

Addressing bias in language is a complex and ongoing process that requires a multifaceted approach. While some tools, such as Google Translate and ChatGPT, have implemented heuristics to suppress biased output in extreme cases, those do not address the underlying problem of bias in training data. Additionally, heuristics are unlikely to be complete since different grammatical forms could still expose biases.

Data mining techniques can be used to extract implicit biases from large language models to better understand what needs to be mitigated, such as by providing models with large amounts of unbiased text data and fine-tuning them to recognize and correct for gender bias. Our research highlights that even within the single domain of software engineering, different but related tasks (e.g., “elicit requirements” and “classify requirements”) have varying associations.

While we cannot say for certain what caused the differences reported here, we can speculate based on related work and the other results we found. For example, the difference between “elicit requirements” (6% “he”) and “classify requirements” (88% “he”) might be caused by the fact that the former may involve talking to customers while the latter can be done without customer involvement. The association of “she” with customer-facing tasks is consistent with related work indicating that “she” is more likely to be associated with roles that require communication [30]. This finding is also in line with our other results, which show that tasks that require communication are more closely related to “she” (e.g., mentor

<sup>3</sup><https://doi.org/10.5281/zenodo.7745436>

others, have meetings). More work is needed to mitigate such biases in large language models.

## VI. THREATS TO VALIDITY

The validity of our results is affected by several threats. Our findings are specific to back-translations with one specific genderless language and the large language model used by the DeepL translator. It is important to note that all large language models may contain biases, despite some user interfaces attempting to mask such biases. Other ways of identifying biases, such as using a different genderless language, might have revealed different results.

The identified biases are dependent on the exact phrasing of each task. To minimize the introduction of additional subjectivity, we reused the phrasing from related work [10] as much as possible. However, we observed that the exact translation of each sentence was influenced by the surrounding sentences, and while we attempted to mitigate this by running the translation 100 times in different permutations, we cannot guarantee that 100 runs were sufficient. The number 100 was chosen based on the API limits of the DeepL translator.

We did not delve into the role of the combined pronouns “he/she” and “he or she” which accounted for 19.3% of the output. Future research should investigate the intricacies of large language models that return combined pronouns and the reasons for the association of these forms with certain tasks. The same applies to other pronouns, such as “his”, “her”, and “they”, among others. We plan to investigate these in future work.

## VII. CONCLUSION AND FUTURE WORK

In this work, we used data mining techniques to examine the prevalence of implicit gender bias in software development tasks as manifested in large language models. Our analysis revealed a significant disparity in the gendered pronoun associations with various tasks, with a distinct pattern of bias observed in certain tasks, such as requirements elicitation. These findings have important implications for addressing gender bias in the training of large language models and in broader society, highlighting the importance of understanding and addressing this issue in more detail.

We will extend this work by considering other genderless languages, other translators, other pronouns, and other ways of phrasing software development tasks. In addition, we aim to conduct a detailed analysis of where these biases originate, such as project-specific documents or general-domain literature. We also plan to conduct a deeper analysis of the impact of the domain “software engineering” on these biases, focusing specifically on tasks that also exist in other domains, such as mentoring and planning.

## REFERENCES

[1] Y. Wang and D. Redmiles, “Implicit gender biases in professional software development: An empirical study,” in *Proceedings of the 41st ACM/IEEE International Conference on Software Engineering: Software Engineering in Society*. IEEE, 2019, pp. 1–10.

[2] R. Garcia, C.-J. Liao, A. Pearce, and C. Treude, “Gender influence on communication initiated within student teams,” in *Proceedings of the*

*53rd ACM Technical Symposium on Computer Science Education V. 1*, 2022, pp. 432–438.

[3] S. Brough, A. E. Bauer, K. Brooks, A. Hopkins, and S. Madison, “Women in astronomy workshop report,” *arXiv preprint arXiv:1106.6094*, 2011.

[4] J. Terrell, A. Kofink, J. Middleton, C. Rainear, E. Murphy-Hill, C. Parnin, and J. Stallings, “Gender differences and bias in open source: Pull request acceptance of women versus men,” *PeerJ Computer Science*, vol. 3, p. e111, 2017.

[5] B. Vasilescu, D. Posnett, B. Ray, M. G. van den Brand, A. Serebrenik, P. Devanbu, and V. Filkov, “Gender and tenure diversity in GitHub teams,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 3789–3798.

[6] M. P. Robillard, “Turnover-induced knowledge loss in practice,” in *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2021, pp. 1292–1302.

[7] A. Caliskan, J. J. Bryson, and A. Narayanan, “Semantics derived automatically from language corpora contain human-like biases,” *Science*, vol. 356, no. 6334, pp. 183–186, 2017.

[8] S. Bordia and S. Bowman, “Identifying and reducing gender bias in word-level language models,” in *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, 2019, pp. 7–15.

[9] S. Prabhunoye, Y. Tsvetkov, R. Salakhutdinov, and A. W. Black, “Style transfer through back-translation,” in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 866–876.

[10] Z. Masood, R. Hoda, K. Blincoe, and D. Damian, “Like, dislike, or just do it? how developers approach software development tasks,” *Information and Software Technology*, p. 106963, 2022.

[11] P. P. Liang, C. Wu, L.-P. Morency, and R. Salakhutdinov, “Towards understanding and mitigating social biases in language models,” in *Proceedings of the International Conference on Machine Learning*. PMLR, 2021, pp. 6565–6576.

[12] R. Liu, C. Jia, J. Wei, G. Xu, L. Wang, and S. Vosoughi, “Mitigating political bias in language models through reinforced calibration,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 17, 2021, pp. 14 857–14 866.

[13] P.-S. Huang, H. Zhang, R. Jiang, R. Stanforth, J. Welbl, J. Rae, V. Maini, D. Yogatama, and P. Kohli, “Reducing sentiment bias in language models via counterfactual evaluation,” in *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020, pp. 65–83.

[14] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, “A survey on bias and fairness in machine learning,” *ACM Computing Surveys*, vol. 54, no. 6, jul 2021. [Online]. Available: <https://doi.org/10.1145/3457607>

[15] B. Vasilescu, A. Capiluppi, and A. Serebrenik, “Gender, representation and online participation: A quantitative study of StackOverflow,” in *Proceedings of the International Conference on Social Informatics*, 2012, pp. 332–338.

[16] V. Kuechler, C. Gilbertson, and C. Jensen, “Gender differences in early free and open source software joining process,” in *Proceedings of Open Source Systems: Long-Term Sustainability*, I. Hammouda, B. Lundell, T. Mikkonen, and W. Scacchi, Eds., 2012, pp. 78–93.

[17] N. Intiaz, J. Middleton, J. Chakraborty, N. Robson, G. Bai, and E. Murphy-Hill, “Investigating the effects of gender bias on GitHub,” in *Proceedings of the 41st International Conference on Software Engineering*, ser. ICSE ’19. IEEE Press, 2019, p. 700–711. [Online]. Available: <https://doi.org/10.1109/ICSE.2019.00079>

[18] G. A. A. Prana, D. Ford, A. Rastogi, D. Lo, R. Purandare, and N. Nagappan, “Including everyone, everywhere: Understanding opportunities and challenges of geographic gender-inclusion in OSS,” *IEEE Transactions on Software Engineering*, vol. 48, no. 9, pp. 3394–3409, 2022.

[19] H. S. Padala, C. Mendez, F. Fronchetti, I. Steinmacher, Z. Steine-Hanson, C. Hilderbrand, A. Horvath, C. Hill, L. Simpson, M. Burnett, M. Gerosa, and A. Sarma, “How gender-biased tools shape newcomer experiences in OSS projects,” *IEEE Transactions on Software Engineering*, vol. 48, no. 1, pp. 241–259, 2022.

[20] D. Rossi and S. Zacchiroli, “Worldwide gender differences in public code contributions: And how they have been affected by the COVID-19 pandemic,” in *Proceedings of the 44th ACM/IEEE International Conference on Software Engineering: Software Engineering in Society*, 2022, p. 172–183.

- [21] A. N. Meyer, E. T. Barr, C. Bird, and T. Zimmermann, "Today was a good day: The daily life of software developers," *IEEE Transactions on Software Engineering*, vol. 47, no. 5, pp. 863–880, 2019.
- [22] S. A. Licorish and S. G. MacDonell, "Exploring software developers' work practices: Task differences, participation, engagement, and speed of task resolution," *Information & Management*, vol. 54, no. 3, pp. 364–382, 2017.
- [23] D. Graziotin, F. Fagerholm, X. Wang, and P. Abrahamsson, "On the unhappiness of software developers," in *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering*, 2017, pp. 324–333.
- [24] D. Ford and C. Parnin, "Exploring causes of frustration for software developers," in *Proceedings of the 8th IEEE/ACM International Workshop on Cooperative and Human Aspects of Software Engineering*. IEEE, 2015, pp. 115–116.
- [25] A. E. Milewski, "Global and task effects in information-seeking among software engineers," *Empirical Software Engineering*, vol. 12, no. 3, pp. 311–326, 2007.
- [26] R. L. Glass, I. Vessey, and S. A. Conger, "Software tasks: Intellectual or clerical?" *Information & Management*, vol. 23, no. 4, pp. 183–191, 1992.
- [27] A. Murgia, P. Tourani, B. Adams, and M. Ortu, "Do developers feel emotions? An exploratory analysis of emotions in software artifacts," in *Proceedings of the 11th Working Conference on Mining Software Repositories*, 2014, pp. 262–271.
- [28] K. Madampe, R. Hoda, and P. Singh, "Towards understanding emotional response to requirements changes in Agile teams," in *Proceedings of the 42nd ACM/IEEE International Conference on Software Engineering: New Ideas and Emerging Results*, 2020, pp. 37–40.
- [29] E. A. Renström, A. Lindqvist, G. Akbas, L. Hekanaho, and M. G. Sendén, "Are gender-neutral pronouns really neutral? Testing a male bias in the grammatical genderless languages Turkish and Finnish," *Journal of Language and Social Psychology*, p. 0261927X221146229, 2022.
- [30] P. M. Fishman, *Interaction: The work women do*. Springer, 1997.