Data-Driven Personas for Software Engineering Research

Jefferson Seide Molléri[a] and Bogdan Marculescu[b]
Kristiania University of Applied Sciences, Kirkegata 24-26, Oslo 0153, Norway
{jefferson.molleri, bogdan.marculescu}@kristiania.no

Keywords: Empirical Research, Persona, Survey Data, Demographics.

Abstract: This paper presents a proof-of-concept on creating data-driven personas for software engineering research using Stack Overflow survey data. We developed three archetypes to illustrate how quantitative data can inform research scenarios. The process involved addressing challenges such as interpreting quantitative data, balancing detail and applicability, ensuring realism, and iterative refinement. The work emphasizes personas as a flexible, human-centered tool that addresses methodological issues in SE research.

1 INTRODUCTION

Personas are a human-centric approach for understanding user behaviors. They offer archetypical representations that embody behaviors and motivations of real user groups (Junior and Filgueiras, 2005). In software engineering (SE) research, personas help conceptualize and conduct empirical studies by providing a deeper understanding of study contexts, allowing more tailored research designs.

This is particularly beneficial for large-scale case studies, action research, and design science projects, where in-depth familiarity with the field is needed. Moreover, in industry collaborations, personas can help simulate real-world scenarios, refining and validating study designs early in the process.

Online communities, such as Stack Overflow (SO), offers extensive and rich data on real-world developer practices, behaviors, skills, and demographics. This can serve as a resource for creating data-driven personas that accurately reflect how developers work and the problems they face.

The motivation of this work stems from the complexity of understanding developer and user behavior in SE research. Traditional surveys that rely on broad statistical summaries often lack the interpretative nuance needed to capture individual experiences and decision-making processes. This highlights a need for a more human-centered approach that integrates quantitative patterns with qualitative insights to create meaningful and realistic representations.

This paper proposes a methodological process to transform raw survey data into personas that are representative for the population being studied. We share practical insights and lessons learned from the process of developing data-driven personas for the purpose of SE research. Key challenges include selecting appropriate clustering techniques, interpreting behavioral patterns from raw survey data, and ensuring that the personas remain authentic representations of real user groups while being broadly applicable across diverse SE research contexts.

By sharing our approach, we aim to provide a foundational guide for researchers interested in leveraging personas in empirical SE studies, highlighting both the potential and the limitations of data-driven personas in this field.

2 RELATED WORK

2.1 Personas in SE Research

Personas have been used across various disciplines to create more user-centered products and services. In SE, researchers have adapted the method to gain a deeper understanding of developer and user behaviors. Ford et al. (2017) used personas to characterize different SE work styles, identifying variations in tasks, collaboration, and autonomy. Their work introduced personas such as debuggers, learners, and experienced advisors to capture the diversity of SE roles.

In requirements engineering (RE), personas are especially employed for modeling user needs. Researchers have explored integrating personas within

[a] https://orcid.org/0000-0001-5629-5256
[b] https://orcid.org/0000-0002-1393-4123

Molléri, J. S. and Marculescu, B.
Data-Driven Personas for Software Engineering Research.
DOI: 10.5220/0013475100003928

In *Proceedings of the 20th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2025)*, pages 798-805
ISBN: 978-989-758-742-9; ISSN: 2184-4895

the RE process to support specific activities (Schneidewind et al., 2012), helping requirements engineers capture varied user needs and roles. A systematic mapping study (Karolita et al., 2023) identified primarily qualitative methods for creating and validating personas in RE. These approaches promote a more human-centered process. The study also highlighted challenges in implementing personas and proposed future work.

Additionally, Ramos et al. (2021) developed five distinct personas based on user data, which were evaluated by users and RE professionals as being representative and of high quality. Dividing users into groups based on their behaviors helped identifying distinct patterns and tailor solutions accordingly. This behavior segmentation enables researchers to interpret user data through archetypes, enhancing understanding of the broader user landscape.

With advances in Machine Learning and Artificial Intelligence (ML/AI), personas can even extend beyond traditional use. For instance, they can simulate human participants in survey-based research (Steinmacher et al., 2024) or, potentially, in interviews. While not intended to replace human respondents, such AI-driven personas could provide insights for researchers and help validate design choices, such as data collection instruments.

2.2 Challenges in Implementing Personas

While personas are beneficial, their application is not without challenges. Chapman and Milham (2006) argue that methodological issues, such as difficulties in determining the representativeness of personas and threats to validity due to a lack of verifiability, may undermine their effectiveness.

Moreover, the adoption and efficacy of personas vary across projects. A recent study (Wang et al., 2024) found that human-centered aspects, which are assumed to be a core part of personas, are often overlooked. Similarly, Billestrup et al. (2014) report that practitioners often develop ad hoc persona usage practices, which do not always align with recommended best practices in the literature.

Despite these concerns, personas grounded in data from surveys and interviews can effectively represent user characteristics and support user-centered SE research (Ford et al., 2017). Guidelines for integrating personas into SE (Faily and Lyle, 2013) suggest that we should (1) offer rationale for persona traits to address skepticism about their 'fictional narratives,' (2) support the qualitative analysis processes that create personas rather than just storing persona data, and (3) facilitate the exchange of personas between projects and team members to encourage maintenance and adaptation.

3 METHODOLOGY

3.1 Research Goal

Our main objective is to present a methodology for creating data-driven personas in SE research. As an illustrative example, we will consider the exploration of mentorship dynamics within the StackOverflow community. The specific study would be focusing on the SO community, and using relevant data collected in that community to drive the development of personas. Interactions between archetypes could guide further studies into knowledge exchange in other communities.

Motivated by studies that investigated the interaction between users who ask and answer questions on SO, e.g. Vasilescu et al. (2013); Wang et al. (2013); Chua and Banerjee (2015), we aim to further explore how these archetypes reflect learning and mentorship dynamics. Specifically, two archetypes align closely to "the continuous learner" and "the experienced advisor" personas described in Ford et al. (2017). Beyond these two, we are also interested in the behaviors of individuals who use SO primarily as an information source, i.e. those who search and read the community posts without asking, answering or commenting.

This proof-of-concept seeks to create personas from the SO survey data. Those personas could then be used to accomplish a research goal: linking personas to specific behavioral patterns, offering insights into how learning and mentorship motivations may shape interactions and engagement within the SO community.

3.2 Context

Stack Overflow, as a large-scale developer-driven platform, provides a rich dataset that captures developer activities, preferences, and self-reported aspirations. Moreover, the platform conducts yearly surveys to capture the opinions and attitudes of their users on various relevant topics. In 2024, the survey addressed topics of interest to our research such as SO usage and community involvement, as well as AI and emerging technologies, and professional challenges (Stack Overflow, 2024a).

The survey was conducted between May 19 and June 20 2024, and collected 65,437 responses from 185 countries, with a median completion time of

21 minutes. Respondents were primarily recruited through SO's channels, which favored responses from highly engaged users. The survey collected a total of 114 variables over 68 questions, grouped into seven distinct categories: (1) basic information, (2) education, work, and career, (3) technology and tech culture, (4) Stack Overflow usage and community, (5) artificial intelligence, (6) professional developer series, and (7) final questions.

3.3 Data Collection and Preparation

To carry out this study, we use data from the SO Developer Survey (Stack Overflow, 2024a), which contains quantitative information on developer demographics, experience levels, behavioral tendencies, and learning aspirations. The dataset also includes responses related to how developers use the community (self-reported).

The scripts we developed to process the survey data and generate personas, along with the resulting artifacts, are available as supplementary material (Molléri, 2024). Additionally, we recommend that readers interested in replicating our results download the full survey dataset (Stack Overflow, 2024b).

3.4 Behavioral Segmentation

We employed a segmentation approach to group developers based on their engagement patterns captured in the survey data. Respondents were clustered into three behavioral segments, based on their answers to the question: *"How do you use Stack Overflow? Select all that apply."*

1. Quickly finding code solutions
2. Finding reliable guidance from community-vetted answers
3. Learning new-to-me technology/techniques
4. Learning new-to-everyone technology/techniques
5. Showcase expertise with code solutions
6. Engage with community by commenting on questions and answers or voting on questions and answers

These answers are ordered by engagement. Moreover, the answers are not orthogonal, they are not mutually exclusive, and they are not balanced. More users are using SO to find solutions, than are contributing. For example, users that contribute and showcase their expertise and engage with the community are often doing so in addition to finding existing solutions and learning new technologies.

We created three personas archetypes representing distinct behavior patterns: *Information seekers* are primarily focused on quickly find code solutions and guidance to technical challenges (i.e. answers 1 and 2). *Engaged Learners* use SO to learn new technologies and techniques (answers 3 and 4). Finally, *Knowledge Sharers* actively take a collaborative approach to enrich the knowledge base for others (answers 5 and 6).

For each behavioral segment, we analyzed response patterns to identify distinctions and similarities that allowed us to profile the three archetypes. In this proof-of-concept, we targeted survey questions that could provide us with a better understanding of our proposed research goal, focusing on education, work, and career. However, we also recognize the value of exploring different traits, particularly those not intuitively related to our main goal. To address this uncharted domain, we employed personas.

3.5 Persona Development Process

Based on the segments, we develop three persona archetypes, i.e., 'information seekers,' 'engaged learners,' and 'knowledge sharers'. Our process started with an automated data-driven sampling of the survey responses. For each behavioral segment, we carried out the following steps:

1. Iterated over each column in the filtered subset for a given behavioral segment
2. Depending on the column type:
   (a) For numerical data, we selected a random non-missing value from the sample
   (b) For single-choice categorical data, we randomly sampled a value from the sample
   (c) For multiple-selection columns, we randomly sampled a number of options representative of the typical number of responses

This process generated a unique persona profile for each behavioral segment, allowing us to capture a realistic set of attributes for each type. Once the individual personas were created for each behavioral pattern, we compiled them into a data frame, assigning a unique identifier based on the segment.

We then refined the persona profiles to transform the data-driven outputs into relatable archetypes suitable for research. This involved (1) ensuring traits such as coding experience aligned with plausible career paths; (2) synthesizing numerical and categorical data to infer behaviors like learning styles and motivations; and (3) balancing specificity and generality to keep the personas applicable across various contexts.

3.6 Methodological Challenges and Considerations

While data-driven personas provide a structured way to interpret behavioral patterns, several methodological challenges arise:

Handling Data Types Proved Challenging. We developed specific code to handle the various kinds of data—numerical, categorical, and open-ended text, but we could not satisfactorily visualize the results graphically. Each data type required unique handling, complicating both the analysis and the representation of personas.

In particular, regarding multiple-selection fields, the dataset included columns where respondents selected multiple options. To accurately reflect these choices in each persona, we developed a function to simulate realistic option selection: (1) split responses into individual options, (2) analyze the typical number of selections made by respondents, (3) use this distribution to sample a representative subset of options, then (4) aggregate them back into a single string for the persona.

Informed Segmentation vs. Organic Clustering. We opted for a segmentation approach based on specific usage behaviors, but other clustering methods, such as data-driven clustering algorithms, could have potentially revealed different groupings. This choice might limit the applicability of our findings and personas to scenarios framed by our specific segmentation.

Additionally, our behavioral segmentation was guided by self-reported data from survey responses. This introduces biases and potential inaccuracies, as respondents' self-assessments may not always reflect their actual behaviors or engagement levels. Further validation in real-world context is needed to ensure the realism and accuracy of personas.

Data Differences Among Segments Were Often Minimal. Descriptive statistics alone did not always highlight meaningful variations, making it challenging to distinguish unique traits for each segment. This required deeper analysis and more nuanced interpretation.

Balancing Objective Data with Interpretative Insights. While the segmentation process was quantitatively driven, each persona needed interpretive details to reflect plausible, realistic behaviors. Conflicting traits were sometimes drawn, requiring adjustments. For example, our information seeker persona initially had 10 years of coding experience, 20 of which professionally - a clear inconsistency. We reduced their professional experience to 5 years to align with a realistic career trajectory.

4 RESULTS

4.1 Profile of Behavioral Segments

We profiled the three archetypes based on their survey responses as follows:

Number of Respondents: Of the 65,437 responses, 39.9% are classified as information seekers, 16.3% as engaged learners, and 24.3% as knowledge sharers; 19.6% did not answer this question and were excluded from further analysis.

Basic Information: The majority of respondents (36.8%) across all segments are aged between 25-34 years, with information seekers predominantly falling in this age range. Younger individuals (under 24 years old) are more likely to be engaged learners, whereas older respondents (35+ years) are often knowledge sharers.

Geographically, the largest group are from the United States (16.9%), Germany (7.5%), and India (6.5%). Knowledge sharers are more commonly based in India than in Germany, while engaged learners more frequently reside in the United Kingdom instead of India.

Education: Most participants across all segments have a bachelor or master degree (38.11% and 23.77%, respectively). The preferred method of learning coding is though 'resources like videos, blogs, forum, and online community.' Notably, knowledge sharers favor 'books and physical media' as their second choice, while other segments lean toward 'school (i.e., university, college, etc.).' For online learning platforms, information seekers prioritize 'technical documentation,' while SO is the top choice for engaged learners and knowledge sharers.

Work: The majority (58.12%) are employed full-time. For engaged learners, the second most common role is full-time student, whereas for other segments, it is full-time independent contractor, freelancer, or self-employed. The most common job roles across all segments are full-stack developer (27.9%), back-end developer (15.2%), student (7.8%), and front-end developer (5.12%).

The average annual compensation across all participants is $244,226 USD. As expected, knowledge sharers, being the most experienced group, reported the highest average salary at $284,578 USD per year. Interestingly, engaged learners had the lowest average compensation at $226,040 USD per year.

Career: Another key indicator of maturity is the number of years participants have been coding, both including and excluding formal education. Information seekers and engaged learners share a similar profile, with an average of over 13 years of coding experience, 9 of which are professional. Knowledge sharers, by contrast, have average of 16 years of coding experience, with more than 11 years spent professionally.

Outside of work, most respondents engage in coding as a hobby (31.8%), for professional development or self-paced learning (18.41%), or by contributing to open-source projects (11.74%). These trends are consistent across all three segments.

4.2 Resulting Personas

Here are the resulting three personas based on behavioral segments, focusing on their work style, experience, technologies, learning preferences, and other relevant traits[1].

Figure 1: Visual representation of Alex, Morgan and Jordan (generated by DALL-E).

4.2.1 Alex, the Information Seeker

“I’m always on the lookout for fixes to everyday problems. Stack Overflow might not be my first stop, but it’s a reliable repository of answers I can count on.”

Alex is a 28-year-old web applications developer from New Zealand, working as a freelancer. With 10 years of coding experience, including 5 years professionally, Alex is proficient in languages like JavaScript, PHP, and Java. They are now exploring tools like PostgreSQL, Firebase Realtime Database, and cloud platforms like AWS to expand their skill set.

Alex thrives in hybrid work environments and prefers structured learning resources like online courses, video tutorials, and AI-powered learning tools. They primarily seek solutions through technical documentation, blogs, and community contributions. Alex prefers ready-to-use solutions with minimal customization.

4.2.2 Morgan, the Engaged Learner

“For me, coding is all about leveling up. I love exploring fresh topics, bouncing around ideas, and testing out new tools.”

Morgan is a 25-year-old full-stack developer based in the United States, working in a large organization with over 1,000 employees. With 7 years of coding experience, 5 of which are professional, Morgan’s skills spans languages like C, C#, Python, and SQL, along with cloud platforms like AWS and Azure. Morgan is now experimenting with DevOps tools, automated testing frameworks, and microservices.

Morgan’s learning journey combines structured and hands-on methods, including tutorials, coding challenges, and certification videos. Morgan uses Stack Overflow daily to ask questions, debug issues, and engage with the developer community. Their preference for customizable technologies reflects a will to integrate innovative tools into their workflows.

4.2.3 Jordan, the Knowledge Sharer

“I get a kick out of helping others. It’s rewarding to share what I’ve learned and watch people pick it up.”

Jordan is a 44-year-old embedded applications developer and part-time student from Bangladesh, with 27 years of coding experience, including 16 years of professional expertise. Jordan is fluent in languages like C#, Go, SQL, and JavaScript and is currently exploring machine learning libraries such as TensorFlow and PyTorch. They are adept at using cloud services like AWS, Firebase, and Microsoft Azure.

Jordan enjoys contributing to the developer community through open-source projects and mentoring peers. They actively engage on SO, using it to share expertise and provide guidance. Learning primarily through technical documentation and peer-reviewed resources, Jordan also participates in live coding sessions and AI-powered tools to validate ideas. With a collaborative spirit, Jordan serves as a bridge between academia and industry.

4.3 Mentorship Dynamics

At this stage, we have established distinct profiles for the three behavioral segments: Alex, Morgan, and Jordan. While we initially expected SO usage to correlate with overall maturity in education, work, and career, this was not always the case. For example, engaged learners are often younger students or early-career developers, whereas information seekers tend to be more established professionals, often with higher salaries.

[1] The source file resulting in the data-driven personas (‘refined_personas.csv’) is available in our Supplementary Material (Molléri, 2024)

Table 1: Top-10 technologies of interest relative to the total respondents within each behavioral segment.

	Information Seekers	Engaged Learners	Knowledge Sharers
Technologies	**Want to learn (in %)**		**Can mentor (in %)**
Visual Studio Code	54.3	57.8	70.3
ChatGPT	46.9	45.9	64.3
JavaScript	37.0	37.7	62.5
Docker	42.0	43.6	50.2
Python	37.4	44.3	50.0
SQL	34.1	36.7	53.8
PostgreSQL	39.1	38.2	43.6
HTML/CSS	32.6	33.7	53.0
Slack	31.4	32.1	42.7
TypeScript	33.1	32.7	37.4

To illustrate the interaction between the three segments, we propose a hypothetical mentorship scenario. *Morgan, an engaged learner, struggles to find specific knowledge about 'dynamically modifying object prototypes at runtime in JavaScript' through tutorials. Frustrated, they post a question on Stack Overflow, hoping for an answer. Jordan, a regular contributor with expertise in this area, spots the question and provides a clear, accurate response. While this exchange could end here, Alex stumbles upon the same thread while researching solutions to their own challenge of 'degrading performance in JavaScript code,' benefiting from the existing exchange.*

To further test this scenario, we identified technologies of interest for the three segments based on the data. The SO developer survey posed questions such as: *"Which programming, scripting, and markup languages have you done extensive development work in over the past year, and which do you want to work with over the next year?"* Similar questions were asked about other technologies, including database environments, cloud platforms, web frameworks, etc.

We assumed that Jordan as a knowledge sharer, is willing to mentor others in technologies they have worked with, while Alex and Morgan are more interested in learning technologies they want to work with. By examining connections between these interests, we identified technologies that are most likely to benefit from interactions among these segments, as outlined in Table 1.

5 DISCUSSIONS

5.1 Insights and Lessons Learned

The primary contribution of this work lies in demonstrating data-driven personas as a methodological tool in SE research. We (1) introduced a process for developing personas through behavioral segmentation and (2) provided an example of characterizing developer engagement in online communities. Our proof-of-concept personas illustrate the potential of data-driven personas to uncover insights into complex phenomena, such as learning and mentorship dynamics.

By systematically clustering respondent profiles by behavioral patterns, we created a meaningful structure to capture distinct traits, similar to Ramos et al. (2021). The combination of data-driven methods and reflexive process allowed us to create relatable and realistic personas. The process has broad applicability across SE research, enabling researchers to simulate realistic scenarios, validate study designs, and interpret respondent behaviors.

Nonetheless, several challenges arose during the process (see Section 3.5). Key lessons for SE researchers include the importance of a refinement process and the need of meaningful segmentation criteria to create plausible archetypes. Combining automated techniques with researcher reflection ensured the coherence of the personas, while iterative validation enhanced their credibility. It is also important to note that this process was resource-intensive, requiring significant effort to align persona details with realistic scenarios.

5.1.1 Methodological Reflection

Personas can play a role in designing and validating research instruments, such as interview guides, diaries, and questionnaires. They help identify key themes and questions tailored to specific archetypes. By reflecting on personas' behaviors and preferences, researchers can design data collection strategies that align with real-world challenges and workflows.

Our resulting personas can be applied in the following research contexts: (1) to guide a **case study** into how different archetypes interact with tools, communities, and learning resources; (2) in a **design science research**, to simulate and evaluate the design of tools aimed at specific behavioral groups; (3) in an

action research to test the effectiveness of interventions or workflows informed by personas in practical settings; (4) enhancing an **ethnography** by providing a structured framework to interpret observed behaviors; or (5) as a reference point for triangulating findings of qualitative interviews and quantitative surveys in a **mixed-methods research**.

By integrating these personas into the research process, we demonstrate their value as a methodological tool for empirical SE beyond UX/UI and requirements engineering. This enables more contextualized studies that provide insights into developer and user behavior in real-world settings while bridging the gap between theoretical and practical insights.

5.1.2 Limitations

While the persona development process provided valuable insights, limitations should be acknowledged. First, the reliance on survey data introduces inherent biases, as respondents self-select to participate, often skewing the sample toward more engaged users. This may limit the representativeness of personas for less active or non-contributing users. Although our process reflects distributions within the dataset, it does not fully resolve questions about representativeness. This highlights the need for further evaluation, such as involving professionals in assessing their alignment with real-world scenarios.

Additionally, behavioral segmentation based on survey responses may oversimplify human behavior. For example, close-ended fields can be challenging to interpret, as they capture preferences without reflecting the intensity or context. Moreover, segmentation criteria are sensitive to the chosen thresholds, affecting the applicability of personas.

The development and usage of personas in SE research have historically been ad hoc, lacking systematic approaches. While this work introduces a structured, data-driven methodology, the relevance of the personas may still vary across contexts. Researchers applying this method should account for differences between their target populations and the SO respondent base to ensure appropriate adaptation.

Finally, the personas are influenced by subjective interpretation during the refinement process, which could introduce researcher bias. While efforts were made to ensure reality and balance, the subjective nature of this step would benefit from iterative validation and stakeholder involvement. Moreover, it is essential to recognize that personas, even when based on real-world data, are not exhaustive representations of the population. Instead, they serve as tools for communication, providing practical insights and facilitating understanding of specific user groups, while complementing broader analyses of user behaviors.

5.2 Implications for SE Research

By using real-world data, segmented by behavioral patterns, we can create personas that are representative of community being studied, and that can provide a grounded understanding of its behavior. Thus, these personas can be a useful tool for studying motivations and learning preferences within the SO community.

Personas can be used to frame a realistic research scenario that align with actual behaviors. For example, the knowledge sharer, characterized by high engagement and expertise, can be pivotal in exploring how experienced developers influence others on online communities. Similarly, the information seeker can help us investigate how new developers navigate and adopt knowledge within these communities.

Additionally, our process can be used to investigate hidden populations within the dataset. For example, 19.6% of respondents did not answer how they use SO, which could indicate less engaged behaviors or barriers to participation. Similarly, individuals who chose not to disclose their salaries or those identifying as members of minority groups point out to populations with limited data visibility. Personas can help form hypotheses about these hidden groups. However, it is important to acknowledge that with fewer participants, the resulting personas may be more susceptible to biases or reflect characteristics that are overly specific to a single individual.

5.2.1 Future Work

Future research aims to build upon our proof-of-concept by validating the personas in real scenarios. This could involve a case study to test their utility in studying mentorship dynamics in online communities. We also intend to explore how these personas could guide study designs, such as creating tailored interview guides or targeted surveys.

Additionally, expanding our data sources beyond SO's survey responses could strengthen the personas' representativeness. Advanced segmentation techniques, like clustering algorithms (Ford et al., 2017), may further refine behavioral segments. Collaboration with practitioners will also be essential to ensure that the personas address practical needs and remain applicable across various SE contexts.

Future research aims to build upon our proof-of-concept by validating the personas in real scenarios. To assess representativeness, we propose evaluation by Stack Overflow users, similar to the approach used by Ramos et al. (2021), who evaluated personas for alignment with RE professionals. Our proposed evaluation could involve a case study to test the personas' utility in studying mentorship dynamics in online communities. We also intend to explore how these personas could guide study designs, such as creating tailored interview guides or targeted surveys.

6 CONCLUSION

Our work demonstrates the potential of data-driven personas as a methodological tool for SE research. The main contributions of this paper are (1) the methodology for developing data-driven personas, applied in the context of SE research, and (2) the accompanying practical insights and lessons learned from this application.

By segmenting behavioral patterns from survey data, we created personas that capture distinct motivations, preferences, and expertise levels within the StackOverflow community. Key methodological insights include refining data for realism, balancing detail with generality, and integrating quantitative findings with practical applications. Limitations still exist, for example the reliance on the availability of sufficient good quality data.

Personas provide a human-centered approach to SE research, guiding the design of tailored instruments like interview guides and validating study scenarios. Ultimately, data-driven personas can help bridge theoretical research with real-world behavior, providing a structured framework for exploring the human aspects of software development.

REFERENCES

Billestrup, J., Stage, J., Bruun, A., Nielsen, L., and Nielsen, K. S. (2014). Creating and using personas in software development: experiences from practice. In *Human-Centered Software Engineering: 5th IFIP WG 13.2 International Conference, HCSE 2014, Paderborn, Germany, September 16-18, 2014. Proceedings 5*, pages 251–258. Springer.

Chapman, C. N. and Milham, R. P. (2006). The personas' new clothes: methodological and practical arguments against a popular method. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 634–636. SAGE Publications Sage CA: Los Angeles, CA.

Chua, A. Y. and Banerjee, S. (2015). Answers or no answers: Studying question answerability in stack overflow. *Journal of Information Science*, 41(5):720–731.

Faily, S. and Lyle, J. (2013). Guidelines for integrating personas into software engineering tools. In *Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems*, pages 69–74.

Ford, D., Zimmermann, T., Bird, C., and Nagappan, N. (2017). Characterizing software engineering work with personas based on knowledge worker actions. In *2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 394–403. IEEE.

Junior, P. T. A. and Filgueiras, L. V. L. (2005). User modeling with personas. In *Proceedings of the 2005 Latin American conference on Human-computer interaction*, pages 277–282.

Karolita, D., McIntosh, J., Kanij, T., Grundy, J., and Obie, H. O. (2023). Use of personas in requirements engineering: A systematic mapping study. *Information and Software Technology*, 162:107264.

Molléri, J. S. (2024). Supplementary material for creating data-driven personas for software engineering research. Available at: https://doi.org/10.5281/zenodo.14182731.

Ramos, H., Fonseca, M., and Ponciano, L. (2021). Modeling and evaluating personas with software explainability requirements. In *Human-Computer Interaction: 7th Iberoamerican Workshop, HCI-COLLAB 2021, Sao Paulo, Brazil, September 8–10, 2021, Proceedings 7*, pages 136–149. Springer.

Schneidewind, L., Hörold, S., Mayas, C., Krömker, H., Falke, S., and Pucklitsch, T. (2012). How personas support requirements engineering. In *2012 First International Workshop on Usability and Accessibility Focused Requirements Engineering (UsARE)*, pages 1–5. IEEE.

Stack Overflow (2024a). Stack Overflow Developer Survey 2024. Available at: https://survey.stackoverflow.co/2024/.

Stack Overflow (2024b). Stack Overflow Insights - Developer Hiring, Marketing, and User Research. Available at: https://survey.stackoverflow.co/.

Steinmacher, I., Penney, J. M., Felizardo, K. R., Garcia, A. F., and Gerosa, M. A. (2024). Can chatgpt emulate humans in software engineering surveys? In *Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, pages 414–419.

Vasilescu, B., Filkov, V., and Serebrenik, A. (2013). Stackoverflow and github: Associations between software development and crowdsourced knowledge. In *2013 International conference on social computing*, pages 188–195. IEEE.

Wang, S., Lo, D., and Jiang, L. (2013). An empirical study on developer interactions in stackoverflow. In *Proceedings of the 28th annual ACM symposium on applied computing*, pages 1019–1024.

Wang, Y., Arora, C., Liu, X., Hoang, T., Malhotra, V., Cheng, B., and Grundy, J. (2024). Who uses personas in requirements engineering: The practitioners' perspective. *arXiv preprint arXiv:2403.15917*.